Outeurskapidentifikasie en mikroblogs: ’n eksploratiewe forensiese analise in ’n digitale era

Mariska Nel

orcid.org 0000-0001-7130-5803

BA, BA Hons, MA

Proefskrif voorgelê ter nakoming vir die graad Philosophiae Doctor in Afrikaans en Nederlands aan die Noordwes-Universiteit

Promotor: Prof JAK Olivier

Gradeplegtigheid: Junie 2018

Studentenommer: 20728395

hoor die taal van my hart
en al klink ek soms gebroke
gebroke en verward
dit is die taal van my hart

Taal van my hart 1995 Stef Bos

VOORWOORD

Ek sweer hierdie was een van die moeilikste goed wat ek tot dusver aangepak het, en daar is nie ’n manier waarop ek dit op my eie sou kon doen nie. Ek onthou aan die begin van my studie het iemand eendag gesê dat ’n PhD die alleenste ding is wat jy ooit kan aanpak, en dit is so. Tog moet ek sê dat ek besef het, dat self wanneer ek op my alleenste gevoel het, was daar altyd mense wat daar was vir my. Hierdie sou nie moontlik gewees het sonder die ondersteuning van die mense in my lewe nie.

Ek wil graag die volgende persone en instansies bedank. Sonder hulle sou die uitvoering en afhandeling van my PhD-studie nie moontlik kon wees nie:

My ouers, Ria en Jannie Nel. Mamma, Pappa, dankie vir julle onwrikbare geloof in my, julle liefde en hoe julle my grootgemaak het. As dit nie vir julle was nie sou ek nie oor die deursettingsvermoë beskik het nie. Sonder julle voorbeeld van hoe mens werk om jou drome te laat realiseer, van hoe mens nie fokus op die slegte tye nie, julle geloof, julle menswees, die manier waarop julle julle hande deurwerk en net nooit ophou nie! Dankie is nie genoeg nie! Lief vir Mamma en Pappa.

My boetie en sussie, Jannie en Leandrie (en natuurlik Finley – die klein rampokker het op 7 maande byna my dataverwerkings verwoes). Dankie dat ek julle deur hierdie tyd as die mense wat julle is kon leer ken. Dankie vir julle ondersteuning en gesels. Lief vir julle!

Prof Jako Olivier, ek kon nie vir ’n beter promotor gevra het nie en daar is soveel goed waarvoor ek moet dankie sê. Prof het my nie net deur my studie gehelp nie, maar ook my gevorm tot ’n navorser en my liefde vir die akademie versterk. Dankie Prof, hierdie is nie sonder Prof moontlik gewees nie!

Prof Attie de Lange, dankie vir Prof se ondersteuning en geloof in my. Dankie dat Prof se deur altyd oop was vir my, en dat Prof my laat besef het ek kan dit doen!

Die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, sonder wie se finansiële ondersteuning hierdie studie sonder twyfel nie voltooi sou word nie.

Die Noordwes-Universiteit vir finansiële ondersteuning. Alle opinies, bevindings, gevolgtrekkings en aanbevelings is dié van die navorser en nie van die Noordwes-Universiteit nie.
Mari Grobler, vir die redigering van die teks.

Peter Mekgwe vir die vertaling van die opsomming na Setswana.

Charika Swanepoel. Ek weet jy wil nie hê dat ek dankie sê nie, maar jy het immers die vertaling van die opsomming na Engels behartig. So dankie daarvoor, ook vir elke koffie en alles anders waarmee jy gehelp het, jy het meer beteken as wat jy ooit sal besef.

Steph en Neil. Ek is nie seker hoe ek vir julle kan dankie sê nie. Die daar wees, die verstaan as ek nie enige iets anders kon doen as *gat-op-stoel* nie. Julle het ʼn deel van my lewe geword terwyl ek op my moeilikste was, en het my aanvaar. Dankie dat julle in my glo! Ek waardeer julle.

Caro en Annie, nie eers duisende kilometers kon julle keer om my te ondersteun nie. Al het ons nie altyd tyd gekry om lukrake oppervlakkige geselsie te maak nie was julle slegs ʼn oproep/boodskap ver. Julle ondersteuning beteken die wêreld vir my. Anja, Shade, Nàdine en Sheugne, dankie dat julle daar was, dat julle gehelp het waar julle kon en ook vir al die grappies, koffies, eet en spanpogings.

Jody, jy het my keer op keer daaraan herinner dat ek net moet positief bly, jy het my aangemoedig, self wanneer ek nie meer wou nie. Jy is ʼn voorbeeld van hoe mens jou kop moet oplig en aangaan, al is dit nie lekker nie. Dankie vir die “trots wees” op elke klein mylpaal in hierdie laaste jaar – al was dit net ʼn afdeling wat ek voltooi het. Dankie dat ek geweet het dat jy onvoorwaardelik daar is, en dat jy bereid was om deel van die “D” te wees! Dankie vir jou Jody, en vir dit wat jy my van menswees geleer het.

Ons het baie oor die manier gepraat waarop woorde soos “dankie” en “jammer” rondgegooi word, waarop sekere woorde hul ware betekenis verloor. Nou moet ek vir jou probeer dankie sê vir die rol wat jy die afgelope paar jaar in my lewe gespeel het, vir jou bydra om hierdie D te voltooi. Jy het geklankbord, geluister as ek wou huil en opskop (en my NOOIT geglo nie), jy het saam met my gestem oor lukrake goed – wat natuurlik dan verander moes word. Vinnige koffies en gesprekke waar die D nie genoem is nie. Musiek (gestoor as onvanpaste name) en spesiale motiverings wat my laat sterf het in my siel (wie dink daardie nonsens uit). Rooiwynsondae om

my siel skoon te maak. Dankie Charonike, dankie vir die besef dat dit "*okei* is om nie *okei* te wees nie." Dankie vir vriendskap, en dankie vir jou geloof in my vermoë om hierdie te kan doen. O-ja, Team Tamsin!

Vir elke ander persoon wat op een of ander wyse ʼn bydrae gelewer het gedurende hierdie deel van my lewe. Almal by SAPT, wie ek kon pla wanneer my brein nie meer lekker wou dink nie, of wat bereid was om ʼn oomblik van hul besige dag aan my af te staan, te luister, of net te vra hoe ek vorder.

Laastens, Dankie aan God die Vader wat my die vermoë gegee het om hierdie te kon doen.

ABSTRACT

KEY TERMS: Forensic linguistics, authorship identification, microblogs, Twitter, forensic stylistics, corpus linguistics, communication, digital humanities, Internet Afrikaans.

Currently, we find ourselves in the era of Web 3.0 (Semantic Web) and Web 4.0 (Symbiotic Web). Not only are people now able to share content with each other, but they can also create the content themselves. The mobile nature of Web 3.0 and 4.0, that is, the manner in which users have access to the internet via various devices (users are no longer solely dependent on a computer to use the internet), has resulted in a change in social interaction and communication.

People are moving away from the more traditional communication mediums and they are finding news, information, and companionship online. It, therefore, appears to be evident that the role of the user has changed and should now be viewed as an online profile through which newsworthy occurrences can be shared in no time. Communication now also has to be short, the reader wants to know what is happening at a glance.

The development of these new technologies is consequently not only changing the ways in which communication takes place or the ways in which language is used to convey the message, but also creates an ideal opportunity for negative communication (slander and bullying or trolling) and fake profiles. In view of the increasing amount of anonymous profiles, one has to ask who is truly speaking and can the author of a short electronic text (in Afrikaans) be identified?

Against the broad background of the digital era, this explorative forensic analysis investigated the possibility of identifying the authors of microblog entries (on Twitter) while also attempting to lay bare the characteristics of the Afrikaans language found on these social media platforms.

In order to conduct the above-mentioned research, a theoretical overview was conducted in which key concepts such as language, forensic linguistics and corpus linguistics were investigated. Also presented, were the existing methods used to identify the authors of short texts from the internet (in other languages such as English) and the changes that occur in language for it to successfully communicate on these social media platforms. Seeing that authorship identification is carried out by means of a corpus analysis, reference and specialised corpora were compiled. Furthermore, three authors from the specialised corpora were chosen

and an extra corpus was compiled from each. The largest “suspicious text” consisted of 91 Tweets, or 1 409 words, while the smallest “suspicious text” consisted of 32 Tweets or 412 words. The size of these corpora is therefore considerably smaller than the corpora used in authorship identification in Afrikaans thus far. These suspicious texts were used to test the presented method of authorship identification regarding the authors of Afrikaans microblog entries. After the theoretical overview as well as the compiling and processing of the data, the empirical analysis was done.

The method presented for the identification of the author of an Afrikaans microblog entry includes stylometric, stylistic, and text analyses. Thirteen aspects that can be used in the process of authorship identification were identified. These aspects include determining keyness, ratio analyses along with the analysis of sentences, words, and characters per Tweet, language relationships, n-gram analyses, readability tests, common features (Twitter-specific features – hashtags, mentions of users and hyperlinks – punctuation and capital letters), syntactic features, morphological features, semantic features, interjections, curse words, emoticons, and an error analysis.

The stylometric, stylistic, and text analyses indicated that similar traits between authors can still be identified despite the limited size of the suspicious texts. It has been determined that the smallest amount of similarities occur between the third suspicious text and the real author of the text. In this case, in 9 of the 13 aspects analysed, the real author could be identified without reasonable doubt, a result that may be considered successful.

It has further been determined that Afrikaans is indeed adjusted by the various users to reach their communicative goals and that certain distinctive language features are identifiable in the Afrikaans used in microblogs. Finally, this study proposes a process that can be used to identify the authors of microblogs.

TSHOBOKANYO

DIKGOPOLO TSA KONOKONO: Bomankge ba puo ba diforensiki, go bona gore mokwadi ke mang, disaete tsa inthanete tse go kwalwang mo go tsone, Twitter, mekgwa ya diforensiki, badira ka puo ba ba dirisang corpus , tlhaeletsano, bokgoni jwa go kwala, Seaforikanse sa Inthenete.

Ga jaanong jaana re iphitlhela re le mo motlheng wa Web 3.0 (Webo ya Mafokoi) le wa Web 4.0 (Webo e Botlhe mo go yone ba Ikaegang ka ba Bangwe). Batho ba kgona, e seng fela go abelana diteng, mme gape ba kgona go tlhama diteng tseo ka bobone. Mokgwa wa didirisiwa tsa mmobaele tsa Web 3.0 le 4.0 ke gore, mokgwa o badirisi ba kgonang go tsena mo inthaneteng ka one ka tiriso ya difouno tsa mmobaele, (badirisi ga ba tlhole ba ikaegile fela ka khomphiutha gore ba dirise inthanete), o fetotse mokgwa wa tirisano mmogo ya bone le tlhaeletsano.

Batho ga ba tlhole ba dirisa mekgwa e e tlwaelegileng thata ya tlhaeletsano mme ba batlana le dikgang, tshedimosetso, le go tsalana le ba bangwe mo inthaneteng. Ka jalo, go bonala sentle gore seabe sa modirisi se fetogile mme jaanong se tshwanetse go lejwa jaaka porofaele ya inthanete e ba ka abelanang dikgang ka yone ka ponyo ya leitlho. Gape tlhaeletsano e tshwanetse go nna e khutshwane, mmadi o batla go itse ka ponyo ya leitlho fela gore go diregang.

Go tlhamiwa ga dithekenoloji tse di ntšha tseno ga go fetole fela tsela e tlhaeletsano e dirwang ka yone kgotsa tsela e puo e dirisiwang ka yone go fetisa molaetsa, mme gape e bula ditšhono tsa puisano e e seng molemo (go senya ba bangwe maina le go kgerisa ba bangwe le go dirisa patlisiso) le go tlhama diporofaele tse e seng tsa boammaaruri. Ka ntlha ya go oketsega ga diporofaele tse di senang maina a beng ba tsone, motho o ipotsa gore tota ke mang yo o buang le gore a go a kgonega go tlhaola mokwadi wa mokwalo o mokhutshwane wa eleketeroniki (ka Seaforikanse)?

Mo motlheng ono wa dijethale e e boitshegang e re nang le yone, ditshekatsheko tsa seforensiki di ne tsa dira patlisiso ka kgonego ya go batlisiso go bona gore bakwadi ke bomang ba ba kwadileng mo disaeteng tsa inthanete tse go kwalwang mo go tsone(tsa Twitter) le go leka gape go senola mokgwa o puo ya Seaforikanse e fitlhelelwang mo teng ga dipolatefomo tseno tsa mmedia wa botsalano e agegileng ka one.

Gore go kgonege go dira patlisiso e e umakilweng fa godimo, go ne ga dirwa ditshekatsheko tsa ka kakaretso tse mo go tsone go neng go dirwa patlisiso ka dikgopolo tsa konokono tse di jaaka puo, bomankge ba puo ba diforensiki le bomankge ba puo ba corpus. Gape go ne ga tlotlwa ka mekgwa e e leng teng e e dirisediwang go tlhaola bakwadi ba mekwalo e mekhutshwane ya inthanete (ka dipuo tse dingwe tse di jaaka ya Seesemane) le diphetogo tse di nnang gone mo puong gore e kgone go utlwala sentle mo dipolatefomong tseno tsa mmedia wa botsalano. E re ka tselaya go batlisiso go bona gore bakwadi ke bomang e dirwa ka ditshekatsheko tsa corpus, go ne ga kwala ditshupiso mmogo le di-corpus tse di kgethegileng. Mo godimo ga moo, go ne ga tlhophiwa bakwadi ba le bararo ba di-corpus tse di kgethegileng mme ga dirwa corpus e nngwe go tswa mo go nngwe le nngwe ya tsone. "Mokwalo o o belaetsang" o mogolo go e feta yotlhe e ne e le wa di *Tweet* di le 91, kgotsa mafoko a le 1 409, mme "mokwalo o o belaetsang" o monnye go e feta yotlhe e ne e le wa di *Tweet* di le 32 kgotsa mafoko a le 412. Ka jalo, go tla go fitlha jaanong jaana, bontsi jwa mekwalo eno ya corpus bo bonnye thata go feta corpus e e dirisediwang go tlhaola mokwadi wa Seaforikanse. Mekwalo e e belaetsang eno e ne e dirisediwa go lekeletsa mokgwa o o tlhamilweng wa go tlhaola bakwadi ba mekwalo ya Seaforikanse mo disaeteng tsa inthanete tse go kwalwang mo go tsone. Morago ga dikgopolo-kakaretso mmogo le go kwala deitha le go e tlhotlha, re ne ra dira ditshekatsheko go ya ka maitemogelo a re nang le one.

Mokgwa o o neng wa tlhamiwa wa go bona gore mokwadi wa mafoko a Seaforikanse ke mang mo saeteng ya inthanete e go kwalwang mo go yone o akaretsa setaele sa mokwadi, mokgwa wa gagwe wa go kgabisa puo, le go sekaseka mokwalo. Go ne ga umakiwa dikarolo di le sometharo tse di ka dirisiwang mo thulaganyong eno tsa go bona gore ke mang a kwadileng molaetsa. Dikarolo tseno di akaretsa dintlha tsa konokono, tshekatsheko ya rešio mmogo le go sekaseka diele, mafoko, le tlhaka ka nngwe ya *Tweet*, dikamano fa gare ga puo, ditshekatsheko tsa dikarolopuo, diteko tsa go buisega ga mokwalo, diponalo tsa ka gale tsa mokwalo (Dikarolo tse di kgethegileng tsa *Twitter* – di-*hashtag*, go umakiwa ga badirisi le dihaephalinki – matshwao a thutapuo le ditlhakakgolo), dikarolo tsa go aga diele, dinoko, bokao jwa mafoko, matshwao-tlhaeletso, matlhapa, difatlhegonyana tse di tsenngwang mo mokwalong, le phoso ya bosekaseki.

Setaele sa mokwadi, mokgwa wa go kgabisa puo, le tshekatsheko ya mokwalo di ne di bontsha gore bakwadi ba ba farologaneng ba sa ntse ba ka kgona go nna le setaele se se tshwanang sa go kwala go sa kgathalasege gore mokwalo o o belaelwang o ke wa bogolo jo bo lekanyeditsweng go le go kae fela. Go lemogilwe gore ditshwani di le mokawanyana tse dinnye di nna teng fa gare ga mokwalo wa boraro o o belaelwang le yo e leng mokwadi wa mokwalo

oo. Mo kgannyeng eno, di le 9 tsa dikarolo tse di sekasekilweng, mokwadi wa mmatota o ne a ka lemogwa kwantle ga pelaelo epe, e leng dipholo tse di ka lejwang e le tse di atlegileng.

Gape go ne ga elwa tlhoko gore batho ba ba farologaneng ga ba bue Seaforikanse ka tsela e e tshwanang fa b a bua se ba batlang go se bua le gore mokgwa o ba buang ka one o ka bonwa mo disaeteng tsa inthanete tse go kwalwang mo go tsone. Kwa bofelong, patlisiso e tshitshinya gore go nne le thulaganyo e e ka dirisiwang ya go bona gore bakwadi ke bomang ba disaete tsa inthanete tse go kwalwang mo go tsone.

OPSOMMING

SLEUTELTERME: Forensiese linguistiek, outeurskapidentifikasie, mikroblogs, Twitter, korpuslinguistiek, kommunikasie, digitale humaniora, Internetafrikaans.

Ons bevind onsself tans in die tydperk van web 3.0 (semantiese web) en web 4.0 (simbiotiese web). Individue is nou nie net in staat om inhoud met mekaar te deel nie, maar ook om self die inhoud te skep. Die mobiele aard van web 3.0 en 4.0 – dit wil sê die wyse waarop gebruikers deur middel van verskeie toestelle internettoegang het (die gebruiker is nie meer afhanklik van slegs ŉ rekenaar om die internet te kan gebruik nie) – gee aanleiding daartoe dat sosiale interaksie en kommunikasie verander het.

Individue beweeg weg van die meer tradisionele kommunikasiemediums, en vind nuus, inligting, sowel as gespreksgenote aanlyn. Die norm blyk dus nou te wees dat die gebruiker se rol verander het, en dat die rol van hierdie aanlyn profiel as ŉ profiel waardeur nuuswaardige gebeure "vinnig" gedeel kan word, beskou moet word. Verder moet kommunikasie nou kort wees, die leser wil in ŉ enkele oogopslag weet wat is besig om te gebeur.

Die ontwikkeling van hierdie nuwe tegnologieë verander gevolglik nie net die wyse waarop kommunikasie plaasvind, of die wyse waarop taal gebruik word ten einde die boodskap oor te kan dra nie, maar hierdie nuwe ontwikkelings skep ook die ideale geleentheid vir negatiewe kommunikasie (laster en afknouery – oftewel *trolling)*, sowel as nagemaakte profiele. In die lig van die toenemende aantal anonieme profiele, moet die vraag gestel word, wie is werklik aan die woord, en kan die outeur van ŉ kort elektroniese teks geïdentifiseer word?

In hierdie eksploratiewe forensiese analise in ŉ digitale era is daar ondersoek ingestel na die moontlikheid om die outeurs van mikrobloginskrywings (op Twitter) te identifiseer, en terselfdertyd is daar ondersoek ingestel na kenmerke van Afrikaans wat op hierdie sosialemediaplatform aangetref word.

Ten einde die bogenoemde ondersoek uit te voer is ŉ literatuurstudie onderneem, waarin sleutelbegrippe soos taal, forensiese en korpuslinguistiek ondersoek is. Die bestaande metodes (in ander tale) om die outeurs van kort tekste vanaf die internet te identifiseer, sowel as die veranderinge wat in taal plaasvind ten einde suksesvol op hierdie sosialemediaplatforms te kommunikeer is ook voorgehou. Aangesien outeurskapidentifikasie deur middel van ŉ korpusanalise uitgevoer word, is daar ŉ verwysings- en gespesialiseerde korpora saamgestel.

Verder is daar drie outeurs vanuit die gespesialiseerde korpora gekies van wie daar ’n ekstra korpus elk saamgestel is. Die grootste “verdagte teks” het uit 91 twiets, oftewel 1 409 woorde bestaan, terwyl die kleinste “verdagte teks” uit 32 twiets of 412 woorde bestaan het. Die grootte van hierdie korpora is dus aansienlik kleiner as die korpora wat tot dusver in outeurskapidentifikasie in Afrikaans gebruik is. Hierdie verdagte tekste is gebruik om die voorgestelde metode van outeurskapidentifikasie vir die outeurs van Afrikaanse mikrobloginskrywings te toets. Na afhandeling van die literatuurstudie, en die samestelling en verwerking van die data kon die empiriese analises aanvang neem.

Die metode wat voorgestel is vir die identifisering van die outeur van ’n Afrikaanse mikrobloginskrywing behels beide ’n stilometriese, stilistiese en teksanalise. Daar is dertien aspekte geïdentifiseer wat in die proses van outeurskapidentifikasie gebruik kan word. Die dertien aspekte sluit in sleutelwaarde-analises, ratio-analises asook die analise van sinne, woorde en karakters per twiet, taalverhoudings, n-gramanalises, leesbaarheidstoetse, algemene kenmerke (Twitter-spesifieke kenmerke – hutswoorde, verwysings na gebruikers en hiperskakels – leestekengebruik en hoofletters), sintaktiese kenmerke, morfologiese kenmerke, semantiese kenmerke, tussenwerpsels, kraswoorde, emotikons en ’n foute-analise.

Beide die stilometriese, stilistiese en teksanalise het getoon dat, ten spyte van die beperkte grootte van die verdagte tekste, daar steeds ooreenstemmende kenmerke tussen outeurs geïdentifiseer kan word. Daar is bevind dat die kleinste aantal ooreenkomste tussen die derde verdagte teks en die werklike outeur van die teks aangetref word. In die geval kon daar in nege van die 13 aspekte wat geanaliseer is, sonder redelike twyfel aangevoer word wie die werklike outeur van die teks is, ’n bevinding wat as suksesvol beskou kan word.

Daar is verder bevind dat Afrikaans wel, deur die onderskeie gebruikers aangepas word, ten einde aan hul kommunikatiewe doelwitte te voldoen en dat bepaalde eiesoortige taalverskynsels in die Afrikaans op mikroblogs identifiseerbaar is. Ten slotte is hierdie studie se bydrae ’n proses wat gebruik kan word vir die identifisering van outeurs op mikroblogs

INHOUDSOPGAWE

VOORWOORD .... **II**

ABSTRACT .... **V**

TSHOBOKANYO .... **VII**

OPSOMMING .... **X**

HOOFSTUK 1: AGTERGROND EN KONTEKSTUALISERING **1**

1.1 Kontekstualisering **1**

1.2 Probleemstelling **5**

1.3 Navorsingsvrae **6**

1.4 Navorsingsdoelwitte **6**

1.5 Sentrale teoretiese stelling **6**

1.6 Navorsingsmetode **6**

1.6.1 Literatuurstudie 7

1.6.2 Empiriese studie 7

1.6.3 Dataversameling 8

1.6.4 Instrumente 9

1.6.5 Data-analise 10

1.6.6 Etiekkwessies 10

1.7 Hoofstukindeling **11**

HOOFSTUK 2: DIE TEORETIESE RAAMWERK **13**

2.1 Inleiding **13**

2.2 Taalbeskouings **15**

2.3 Die ontwikkeling van Afrikaans 18

2.3.1 Historiese oorsig van Afrikaans 19

2.3.2 Variëteite van Afrikaans 23

2.3.2.1 Kontakvariasie 25

2.3.2.2 Interne variasie 27

2.3.2.3 Idiolek 29

2.4 Afrikaans in 'n digitale era 32

2.5 Kommunikasie en taal 40

2.5.1 Kommunikasie 40

2.5.2 Kommunikasie en die internet 42

2.5.2.1 Rekenaarbemiddelde kommunikasie (RBK) 44

2.6 Sosiale media 49

2.6.1 Web 3.0: die mobiele web 49

2.6.2 Sosialenetwerkplatforms 51

2.6.2.1 Mikroblogs 54

2.6.2.2 Twitter 56

2.7 Samevatting 65

HOOFSTUK 3: FORENSIESE LINGUISTIEK IN 'N DIGITALE ERA 68

3.1 Inleiding 68

3.2 Forensiese linguistiek 70

3.3 Forensiese linguistiek: 'n definisie 70

3.4 Ontstaan en ontwikkeling van forensiese linguistiek 73

3.4.1 Forensiese linguistiek: die ontstaan 74

3.4.2 Forensiese linguistiek in Suid-Afrika 77

3.4.3 Forensiese linguistiek en informele tekste 80

3.5 Outeurskapidentifikasie 84

3.5.1 Benaderings tot outeurskapidentifikasie 86

3.5.1.1 Forensiese stilistiek 89

3.5.1.2 Forensiese stilometrie 90

3.6 Outeurskapidentifikasie in ʼn digitale era 92

3.7 Samevatting 96

HOOFSTUK 4: KORPUSLINGUISTIEK 99

4.1 Inleiding 99

4.2 Korpora … 100

4.2.2 Soorte korpora en korpusgroottes 101

4.2.2.1 Verskillende korpora 103

4.2.2.2 Korpusgroottes 106

4.2.3 Die web as ʼn korpus 109

4.2.4 Korpora in Afrikaans 112

4.2.5 Samestelling van ʼn korpus 114

4.3 ʼn Historiese oorsig van korpuslinguistiek 118

4.4 Korpuslinguistiek gedefinieer 126

4.5 Korpuslinguistiek: metode of teorie 129

4.6 ʼn Korpuslinguistiese analise 132

4.6.1 Korpusgebaseerde benadering 133

4.6.2 Korpusgedrewe benadering 134

4.7 ’n Forensiese en korpuslinguistiese benadering 136

4.8 Samevatting 139

HOOFSTUK 5: METODOLOGIE 141

5.1 Inleiding 141

5.2 Navorsingsontwerp 142

5.3 Deelnemers 143

5.4 Data-insameling en -verwerking 148

5.5 Navorsingsmetodes 151

5.6 Kwalitatiewe analise 153

5.6.1 Stilistiese analise 154

5.7 Kwantitatiewe analise 157

5.7.1 Stilometrie 159

5.8 Instrumente 162

5.8.1 Rekenaarprogrammatuur 162

5.8.1.1 WordSmith Tools 163

5.8.1.2 *ATLAS.ti* 163

5.8.2 Terminologie van programmatuur 163

5.8.2.1 Woordelyste 164

5.8.2.2 Konkordansie 165

5.8.2.3 Sleutelwoorde 166

5.8.2.4 Frekwensie 167

5.8.2.5 *ATLAS.ti*-funksies 168

5.8.3 Statistiese konsepte 170

5.8.3.1 Nulhipotese 170

5.8.3.2 Chi-Kwadraat en p-waardes 170

5.8.3.3 N-gram 173

5.9 Kleinkorpora en outeurskapidentifikasie 173

5.10 Oorsig van metodes vir outeurskapidentifikasie 174

5.11 Etiekkwessies 176

5.12 Samestelling van die korpora 179

5.12.1 Deelnemers 179

5.12.2 Dataverwerking 182

5.12.3 Datastelle 186

5.12.3.1 Die verwysingskorpus 186

5.12.3.2 Die gespesialiseerde korpus 187

5.12.3.3 Verdagte tekste 190

5.13 Samevatting 191

HOOFSTUK 6: EMPIRIESE ANALISE: STILOMETRIES 193

6.1 Inleiding 193

6.2 Sleutelwoorde 195

6.2.1 Gespesialiseerde korpus: sleutelwaardes van funksiewoorde 197

6.2.2 Verdagte tekste: sleutelwaardes en inhoudswoorde 200

6.3 Ratiobepalings ... 203

6.3.1.1 Gespesialiseerde korpus: lidwoorde ... 211

6.3.1.2 Verdagte tekste: lidwoorde ... 214

6.3.2 Ratiobepalings: voornaamwoorde ... 215

6.3.2.1 Gespesialiseerde korpus: voornaamwoorde ... 216

6.3.2.2 Verdagte tekste: voornaamwoorde ... 218

6.3.3 Ratiobepalings: voorsetsels ... 220

6.3.3.1 Gespesialiseerde korpus: voorsetsels ... 221

6.3.3.2 Verdagte tekste: voorsetsels ... 223

6.3.4.1 Gespesialiseerde korpus: hulpwerkwoorde ... 225

6.3.4.2 Verdagte tekste: hulpwerkwoorde ... 229

6.3.5 Ratiobepalings: voegwoorde ... 231

6.3.5.1 Gespesialiseerde korpus: voegwoorde ... 232

6.3.5.2 Verdagte tekste: voegwoorde ... 234

6.4 Sinne, woorde en karakters per twiet ... 235

6.4.1 Gespesialiseerde korpus: sinne, woorde en karakters per twiet ... 237

6.4.2 Verdagte tekste: sinne, woorde en karakters per twiet ... 240

6.5 Taalverhouding in die twiets ... 241

6.5.1 Gespesialiseerde korpus: taalverhouding ... 242

6.5.2 Verdagte tekste: taalverhouding in twiets ... 245

6.6 Leesbaarheidstoetse ... 247

6.6.1 Gespesialiseerde korpus: Misindeks ... 248

6.6.2 Verdagte tekste: Misindeks 249

6.7 Sintese van die stilometriese analise 251

6.8 Samevatting 257

HOOFSTUK 7: EMPIRIESE ANALISE: STILISTIESE EN TEKSANALISE 259

7.1 Inleiding 259

7.2 Algemeen 262

7.2.1 Twitter-spesifieke tekselemente 262

7.2.1.1 Hutswoorde 263

7.2.1.1.1 Gespesialiseerde korpus: hutswoorde 266

7.2.1.1.2 Verdagte tekste: hutswoorde 272

7.2.1.2 Verwysing na gebruikers 274

7.2.1.2.1 Gespesialiseerde korpus: verwysings na gebruikers 276

7.2.1.2.2 Verdagte tekste: verwysings na gebruikers 279

7.2.1.3 Hiperskakels 281

7.2.1.3.1 Gespesialiseerde korpus: hiperskakels 283

7.2.1.3.2 Verdagte tekste: hiperskakels 285

7.2.2 Leestekengebruik 287

7.2.2.1 Gespesialiseerde korpus: die gebruik van leestekens 290

7.2.2.2 Verdagte tekste: die gebruik van leestekens 295

7.2.3 Hoofletters 298

7.2.3.1 Gespesialiseerde korpus: hooflettergebruik in twiets 301

7.3 Sintaksis 307

7.3.1 Sinstipes 308

7.3.1.1 Stelsinne 312

7.3.1.1.1 Gespesialiseerde korpus 312

7.3.1.1.2 Verdagte tekste 313

7.3.1.2 Vraagsinne 313

7.3.1.2.1 Gespesialiseerde korpus 313

7.3.1.2.2 Verdagte tekste 316

7.3.1.3 Uitroepsinne 317

7.3.1.3.1 Gespesialiseerde korpus 317

7.3.1.3.2 Verdagte tekste 318

7.3.1.4 Bevelsinne 319

7.3.1.4.1 Gespesialiseerde korpus 319

7.3.1.4.2 Verdagte tekste 320

7.3.2 Woordsoorte 320

7.3.2.1 Adjektiewe 322

7.3.2.2 Konjunkte 324

7.3.2.3 Naamwoorde 325

7.3.2.4 Voornaamwoorde 327

7.3.2.5 Voorsetsels 329

7.3.2.6 Werkwoorde 331

7.3.2.7 Lidwoorde 333

7.3.3 Die invloed van Engels 335

7.3.3.1 Gespesialiseerde korpus: die invloed van Engels 336

7.3.3.2 Verdagte tekste: die invloed van Engels 337

7.4 Morfologie 337

7.4.1 Afkortings 339

7.4.1.1 Gespesialiseerde korpus: foutlose afkortings 340

7.4.1.2 Verdagte tekste: foutlose afkortings 345

7.4.1.3 Gespesialiseerde korpus: idiolektiese afkortings 346

7.4.1.4 Verdagte tekste: idiolektiese gebruik van afkortings 353

7.4.1.5 Gespesialiseerde korpus: akronieme 354

7.4.1.6 Verdagte tekste: akronieme 356

7.4.1.7 Gespesialiseerde korpus: SMS-afkortings 358

7.4.1.8 Verdagte tekste: SMS-afkortings 362

7.4.2 Inkorting 362

7.4.2.1 Gespesialiseerde korpus: inkorting 363

7.4.2.2 Verdagte tekste: inkorting 364

7.4.3 Weglating 365

7.4.3.1 Gespesialiseerde korpus: weglating 366

7.4.3.2 Verdagte tekste: weglating 367

7.5 Semantiek 369

7.5.1 Emotikons 369

7.5.1.1 Gespesialiseerde korpus: emotikons 371

7.5.1.2 Verdagte tekste: emotikons 376

7.6 Kodewisseling 379

7.6.1 Gespesialiseerde korpus: IT en MT 381

7.6.2 Verdagte tekste: IT en MT 384

7.6.3 Gespesialiseerde korpus: dubbelmorfologie 386

7.6.4 Verdagte tekste: dubbelmorfologie 387

7.7 Tussenwerpsels 388

7.7.1 Gespesialiseerde korpus: tussenwerpsels 389

7.7.2 Verdagte tekste: tussenwerpsels 390

7.8 Kraswoorde 390

7.8.1 Gespesialiseerde korpus: kraswoorde 391

7.8.2 Verdagte tekste: kraswoorde 394

7.9 Foute-analise 395

7.10 ’n N-gramanalise 398

7.11 Sintese 401

7.11.1 Kenmerke van Afrikaans op mikroblogs 401

7.11.1.1 Algemene kenmerke 402

7.11.1.2 Sintaktiese kenmerke 404

7.11.1.3 Morfologiese kenmerke 406

7.11.1.4 Semantiese kenmerke 407

7.11.1.5 Kenmerke soos vanuit die ondersoek na kodewisseling 408

7.11.1.6 Gebruik van tussenwerpsels 408

7.11.1.7 Gebruik van kraswoorde 409

7.11.1.8 Foute analise 409

7.11.1.9 N-gramanalise 410

7.11.2 Stilistiese en teksanalise en outeurskapidentifikasie 410

7.11.2.1 Algemene kenmerke 411

7.11.2.2 Sintaktiese ondersoek 412

7.11.2.3 Morfologiese ondersoek 414

7.11.2.4 Semantiese kenmerke 415

7.11.2.5 Kodewisseling 416

7.11.2.6 Tussenwerpsels 417

7.11.2.7 Kraswoorde 417

7.11.2.8 Foute-analise 418

7.11.2.9 N-gramanalise 418

7.12 Samevatting 419

HOOFSTUK 8: SAMEVATTING 423

8.1 Inleiding 423

8.2 Oorsig van die studie 423

8.2.1 Doelwitte van die studie 424

8.2.2 Literatuurstudie 425

8.2.3 Empiriese ondersoek 428

8.3 Beantwoording van die navorsingsvrae 429

8.3.1 Taalstilistiese en stylkenmerke van Afrikaans op mikroblogs 429

8.3.2 Outeurskapidentifikasie op mikroblogs 429

8.3.2.1 Outeurs van die verdagte tekste 430

8.3.2.2 Metode vir die outeurskapidentifikasie van mikroblogouteurs 430

8.4 Beperkings van die studie 435

8.5 Aanbevelings vir verdere navorsing 435

8.6 Samevatting en slot 436

Bronnelys 437

Bylaag A: Etiekklaring 479

Bylaag B: Brief om toestemming vir gebruik van Twitter-data 480

Bylaag C: Sleutelwaardes 482

Bylaag D: Ratio's 484

Bylaag E: Algemene Kenmerke 487

Bylaag F: Sintaksis 493

Bylaag G: Morfologie 496

Bylaag H: Emotikons 501

Bylaag I: JGAAP 506

Bylaag J: Misindeks 508

LYS VAN TABELLE

Tabel 2.1 Verhoef (2001:8-11) se merkers van geskrewe Internetafrikaans 35

Tabel 2.2 Aspekte van Internetafrikaans soos ondersoek deur Jansen van Vuuren (2007:155-287).......38

Tabel 2.3 Verskil tussen die eerste en tweede media-eras (vertaal vanuit Holmes, 2005:10) 43

Tabel 2.4 Voorstelling van die verskille tussen Web 1.0, 2.0 en 3.0 (saamgestel vanuit O'Reilly & Battelle, 2009:2)..50

Tabel 2.5 Eie voorstelling van die verskillende mikroblogplatforms...55

Tabel 3.1 Verskille tussen formele en informele tekste (saamgestel uit Grant (2008:217) 80

Tabel 3.2 Kenmerke van ʼn ondersoek in ʼn stilometriese analise (Barry & Luna, 2012:2) 90

Tabel 3.3 Beskrywing van die uiteensetting van n-gramme in tydens ʼn stilometriesee ondersoek gebruik kan word (geneem vanuit Nikhri *et al.,* 2015:371)...92

Tabel 4.1 Korpusontwikkeling (saamgestel uit Kennedy, 1998; Nesselhauf, 2005; Stubbs, 2007).........102

Tabel 4.2 Verskil tussen die web as ʼn korpus en die web as ʼn bron vir die samestelling van ʼn korpus (geneem vanuit Wunderlich, 2012:44). 109

Tabel 4.3 Voordele van die web as ʼn korpus (Fletcher, 2013:1339-1340) 110

Tabel 4.4 Voorstelling van Svartvik se soektog in 2007 sowel as my eie soektog (2016) om die toenemende gewildheid van korpuslinguistiek as navorsingarea uit te beeld 121

Tabel 5.1 Voorstelling van die twee verskillende groepe deelnemers vir my studie...............................147

Tabel 5.2 Voorstelling van rou data wat gebruik word vir chi-kwadraatberekeninge..............................171

Tabel 5.3 Berekeninge van die verwagte waardes (geneem vanuit Levon, 2010:78) 171

Tabel 5.4 Uiteensetting van korpora 181

Tabel 5.5 Uiteensetting van die gespesialiseerde korpus. 182

Tabel 5.6 Uitbeelding van die Microsoft Excel-lêers waarin die deelnemers se inligting gestoor is........184

Tabel 5.7 Opsomming van die gespesialiseerde korpus 187

Tabel 5.8 Opsomming van die deelnemers se datastelle 188

Tabel 5.9 Opsomming van die individuele deelnemerkorpus 189

Tabel 5.10 Uiteensetting van die Afrikaanse twiets vanuit die verdagte tekste wat in die studie gebruik word. 190

Tabel 6.1 Voorstelling van die Tabelle waar die sleutelwaardes van die grammatikale woorde bepaal is deur middel van WordSmith Tools 197

Tabel 6.2 Voorstelling van die Tabelle waar die frekwensies van die voorkoms van die hulpwerkwoorde gedokumenteer is vanuit O1....................207

Tabel 6.3 Voorstelling van die Tabelle waar die frekwensies van die hulpwerkwoorde gedokumenteer is vanuit die gespesialiseerde korpus 207

Tabel 6.4 Voorstelling van die Tabelle waar die ratio's van die hulpwerkwoorde gedokumenteer is vanuit O1 208

Tabel 6.5 Voorstelling van die wyse waarop die prakties beduidende verskille (al dan nie) aangeteken is 209

Tabel 6.6 Uitbeelding van die ratioberekeninge van die lidwoorde in die korpora wat nie ŉ prakties beduidende verskil getoon het nie. 212

Tabel 6.7 Hulpwerkwoorde soos gebruik deur O5 en O9 228

Tabel 6.8 Voorstelling van ŉ korpus in Microsoft Excel waardeur die woorde en karakters per twiets bepaal is 236

Tabel 6.9 Sinne, woorde, karakters en sinne in die gespesialiseerde korpus 237

Tabel 6.10 Sinne, woorde, karakters en sinne in die verdagte tekste 240

Tabel 6.11 Gemiddelde aantal sinne, woorde en karakters vanuit die gespesialiseerde korpus wat met die verdagte tekste ooreenstem....................240

Tabel 6.12 Voorbeeld van die taalverhouding (gespesialiseerde korpus) 242

Tabel 6.13 Verhouding tussen Afrikaanse en Engelse twiets en hertwiets in die gespesialiseerde korpus 243

Tabel 6.14 Taalverhouding in die verdagte tekste 245

Tabel 6.15 Ooreenkomste tussen die verdagte tekste en tekste vanuit die gespesialiseerde korpus ten opsigte van taalverhouding....246

Tabel 6.16 Moontlike outeurs van die drie verdagte tekste na gelang van die sleutelwaarde-analise....252

Tabel 6.17 Moontlike outeurs van die drie verdagte tekste na gelang van die ratiobepalings 253

Tabel 6.18 Moontlike outeurs van die drie verdagte tekste na gelang van die ondersoek na die sinne, woorde en karakters per twiet 255

Tabel 6.19 Moontlike outeurs van die drie verdagte tekste na gelang van die ondersoek na die taalverhouding in die twiets 256

Tabel 6.20 Opsomming van die moontlike outeurs van die verdagte tekste na gelang van die stilometriese analise 257

Tabel 7.1 Die gebruik van hutstekens in die onderskeie korpora 264

Tabel 7.2 @-gebruikerskonvensies in die onderskeie datastelle....274

Tabel 7.3 Die gebruik van hiperskakels in die onderskeie datastelle 281

Tabel 7.4 Die gebruik van leestekens in twiets 287

Tabel 7.5 Die gebruik van meer as een leesteken in twiets....288

Tabel 7.6 Die gebruik van hoofletters in die onderskeie korpora....299

Tabel 7.7 Gebruik van hoofletters ten einde woorde te beklemtoon....299

Tabel 7.8 Hoofletters wat nie volgens die taalreëls gebruik is nie 300

Tabel 7.9 Frekwensie van die voorkoms van die sinstipes....309

Tabel 7.10 Ooreenstemmende gebruik van adjektiewe....323

Tabel 7.11 Ooreenkomste met betrekking tot die gebruik van konjunkte....325

Tabel 7.12 Ooreenstemmende gebruik van naamwoorde....326

Tabel 7.13 Ooreenstemmende gebruik van voornaamwoorde....329

Tabel 7.14 Ooreenstemmende gebruik van voorsetsels....330

Tabel 7.15 Ooreenstemmende gebruik van werkwoorde 332

Tabel 7.16 Ooreenstemmende gebruik van lidwoorde 334

Tabel 7.17 Gebruik van afkortings in die gespesialiseerde korpora 339

Tabel 7.18 Afkortings (foutloos) wat in die gespesialiseerde korpora aangetref word....341

Tabel 7.19 Idiolektiese afkortings wat in die gespesialiseerde korpora voorkom 347

Tabel 7.20 ʼn Opsomming van die idiolektiese gebruik van afkortings deur die outeurs van die gespesialiseerde korpora 347

Tabel 7.21 ʼn Opsomming van die gebruik van akronieme....354

Tabel 7.22 Die gebruik van akronieme in die verdagte tekste en die ooreenstemmende akronieme in die gespesialiseerde korpora 357

Tabel 7.23 Voorkoms van inkortings in die verdagte tekste....364

Tabel 7.24 Weglatings soos geïdentifiseer in die verdagte tekste....367

Tabel 7.25 Emotikongebruik in die verdagte tekste 377

Tabel 7.26 IT en MT soos gevind in die verdagte tekste 384

Tabel 7.27 Spelfoute 396

Tabel 7.28 ʼn N-gramanalise....400

Tabel 7.29 Twitter-spesifieke kenmerke: samevatting van die moontlike outeurs van die verdagte tekste 411

Tabel 7.30 Samevatting van die sintaktiese analise 412

Tabel 7.31 Samevatting van die morfologiese analise....414

Tabel 7.32 Samevatting van die gebruik van emotikons....415

Tabel 7.33 Samevatting van die semantiese analise....416

Tabel 7.34 Samevatting van die ondersoek na die gebruik van tussenwerpsels 417

Tabel 7.35 Samevatting van die ondersoek na die gebruik van kraswoorde....418

Tabel 7.36 Moontlike outeurs van V1....420

Tabel 7.37 Moontlike outeurs van V2 420

Tabel 7.38 Moontlike outeurs van V3 421

LYS VAN FIGURE

Figuur 2.1 Sosialemediagebruik in Suid-Afrika (geneem en aangepas vanaf Shezi, 2015**)**......................14

Figuur 2.2 Samenstelling van Afrikaners in 1807(geneem vanuit Giliomee & Mbenga, 2007:70).............20

Figuur 2.3 Taalpolitiese veranderinge gedurende die Tweede Taalbeweging (saamgestel vanuit Deumert, 2004:293)..21

Figuur 2.4 Ontwikkeling van Afrikaans (Den Besten, 1989:226)..22

Figuur 2.5 Vertakking van variasietaalkunde (geneem vanuit Du Plessis, 1995:19).................................25

Figuur 2.6 Voorbeelde van mikrobloginskrywings waar ’n Engelse hutswoord (#) gebruik word om die boodskap oor te dra en in konteks met wêreldgebeure te plaas (geneem vanaf www.twitter.com)..........27

Figuur 2.7 Vereenvoudigde voorstelling van Jakobson se eenvoudige kommunikasieproses (Smuts, 1998:11)..41

Figuur 2.8 Eie voorstelling van een-tot-een kommunikasie 46

Figuur 2.9 Eie voorstelling van een-tot-talle-kommunikasie..47

Figuur 2.10 Twitter-logo (geneem vanaf http://www.freelargeimages.com/twitter-logo-hd-png-4557/).....57

Figuur 2.11 Voorstelling van die belangrikste funksies op die Twitter-gebruikersprofiel tuisblad..............58

Figuur 2.12 Voorstelling van hoe die profiele wat die gebruiker volg, aangedui word...............................59

Figuur 2.13 Voorstelling van die manier waarop karakters op Twitter aangedui word 63

Figuur 4.1 Voorstelling van die verskillende soorte korpora (geneem en vertaal vanuit Cabrio, 2010)...103

Figuur 4.2 Toename in die publikasie van artikels oor korpuslinguistiek vanaf 1965-1990 (geneem en vertaal vanuit Johansson, 2008:47) 122

Figuur 4.3 Vertaalde weergawe van Renouf (2007:28) se voorstelling van die verskillende ontwikkelingsvlakke van korpuslinguistiek 125

Figuur 6.1 Interne sleutelwaarde vergelykings van die grammatikale woorde 198

Figuur 6.2 Sleutelwaarde vergelykings van die grammatikale woorde...199

Figuur 6.3 Vergelyking van V1 se korpora met die gespesialiseerde korpora..200

Figuur 6.4 Sleutelwaardevergelyking van V2 se korpora met die gespesialiseerde korpora 201

Figuur 6.5 Sleutelwaardevergelyking van V3 se korpora met die gespesialiseerde korpora 202

Figuur 6.6 Voorstelling van die persentasie van die prakties beduidend verskil van die ratiobepalings van die lidwoorde 211

Figuur 6.7 Voorstelling van die persentasie van prakties beduidende verskil van die ratiobepalings van die lidwoorde 214

Figuur 6.8 Voorstelling van die prakties beduidende verskil van die gebruik van voornaamwoorde (persentasie)..................216

Figuur 6.9 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die voornaamwoorde 218

Figuur 6.10 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die voorsetsels 221

Figuur 6.11 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die voorsetsels..................223

Figuur 6.12 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die hulpwerkwoorde 225

Figuur 6.13 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die hulpwerkwoorde..................230

Figuur 6.14 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die hulpwerkwoorde 232

Figuur 6.15 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die voegwoorde 234

Figuur 6.16 Misindeksreslutate vir die gespesialiseerde korpus..................248

Figuur 6.17 Die misindeks van die verdagte tekste 250

Figuur 6.18 Misindeks-vergelyking tussen die verdagte tekste en die gespesialiseerde korpus..................250

Figuur 7.1 Die verhouding tussen die gebruik van die verskillend sinstipes deur die onderskeie outeurs 310

Figuur 7.2 Woordsoortgebruik..................321

Figuur 7.3 Adjektiewe 322

Figuur 7.4 Konjunkte 324

Figuur 7.5 Naamwoorde 325

Figuur 7.6 Voornaamwoorde 327

Figuur 7.7 Voorsetsels...330

Figuur 7.8 Werkwoorde: gespesialiseerde korpus 331

Figuur 7.9 Lidwoorde: gespesialiseerde korpus 333

Figuur 7.10 Invloed van Engels op sinne 336

Figuur 7.11 Die gebruik van afkortings (gespesialiseerde korpora en verdagte tekste) 339

Figuur 7.12 Die persentasie waarby (foutlose) afkortings in die onderskeie korpora voorkom 340

Figuur 7.13 Die gebruik van die afkorting SA 342

Figuur 7.14 Genormaliseerde waarde van die afkortings wat deur meer as een outeur vanuit die gespesialiseerde korpora gebruik word...342

Figuur 7.15 Die kenmerkende gebruik van (foutlose) afkortings in die gespesialiseerde korpora 343

Figuur 7.16 Persentasie (idiolektiese) afkortings wat in die gespesialiseerde korpora voorkom............346

Figuur 7.17 Idiolektiese afkortings wat as uniek tot die korpus van O2 beskou kan word.......................349

Figuur 7.18 Die unieke gebruik van afkortings deur O6 351

Figuur 7.19 Genormaliseerde gebruik van akronieme deur O1 tot O10 354

Figuur 7.20 Die gebruik van SMS-afkortings in die gespesialiseerde korpora 358

Figuur 7.21 Gewildste SMS-afkortings (genormaliseer tot 1 000 woorde) in die gespesialiseerde korpora 359

Figuur 7.22 Gewildste inkortings wat deur die onderskeie outeurs in die gespesialiseerde korpus gebruik word 363

Figuur 7.23 Inkortings wat slegs aangetref word in die korpus van ’n enkele outeur vanuit die gespesialiseerde korpus..364

Figuur 7.24 Kenmerkende weglatings vanuit die gespesialiseerde korpus 366

Figuur 7.25 Gebruik van emotikons in die gespesialiseerde korpus sowel as die verdagte tekste 370

Figuur 7.26 Genormaliseerde waarde per 1 000 woorde vir die verskillende soorte emotikons wat in die gespesialiseerde korpus gebruik word 373

Figuur 7.27 Emotikons soos gebruik per kategorie in die gespesialiseerde korpus 374

Figuur 7.28 Verspreiding van emotikongroepe 375

Figuur 7.29 Kodewisseling in die gespesialiseerde korpus............380

Figuur 7.30 Die voorkoms van IT en MT in die gespesialiseerde korpus 381

Figuur 7.31 Genormaliseerde waarde per 1 000 woorde vir *boys* soos gebruik deur die gespesialiseerde korpus............382

Figuur 7.32 Tussenwerpsels 388

Figuur 7.33 Kraswoorde in die gespesialiseerde korpus en verdagte tekste............391

Figuur 8.1 Proses van data-insameling en -verwerking 431

Figuur 8.2 Stilometriese analise 432

Figuur 8.3 Stilistiese en teksanalise 433

Figuur 8.4 Metode van outeurskapidentifikasie............434

LYS VAN SKERMSKOTE

Skermskoot 2.1 Voorstelling van die tuisblad wanneer ’n Twitter-profiel geskep word (geneem vanaf https://www.twitter.com/signup?lang=en) 57

Skermskoot 2.2 Voorstelling van die Twitter-tuisblad vir ’n nuwe gebruiker 58

Skermskoot 2.3 Moontlike profiele wat gevolg kan word 59

Skermskoot 2.4 Voorstelling van die chronologiese wyse waarop twiets op ’n gebruiker se stroom verskyn (geneem vanaf www.twitter.com) 60

Skermskoot 2.5 Voorbeeld van ’n twiet gerig aan ’n spesifieke @-gebruiker (geneem vanaf www.twitter.com) 61

Skermskoot 2.6 Voorbeeld van ’n twiet met ’n hutswoord 62

Skermskoot 2.7 Voorbeelde van ’n hertwiet 62

Skermskoot 4.1 Voorstelling van die Wikipedia-korpus (http://corpus.byu.edu/wiki/) 111

Skermskoot 4.2 Voorstelling van die GloWbE-korpus (http://corpus.byu.edu/glowbe/) 111

Skermskoot 4.3 Voorstelling van die Time-korpus (http://corpus.byu.edu/time/) 112

Skermskoot 4.4 Tuisblad van die VivA-korpusportaal (http://korpus.viva-afrikaans.org/page/search) 113

Skermskoot 4.5 Voorbeeld van ’n eenvoudige soektog in die VivA-korpus na die woord “vermiste” (https://youtu.be/SC4xMk98Pdc) 113

Skermskoot 4.6 Afrikaanse korpora wat beskikbaar is op VivA se korpusportaal 114

Skermskoot 5.1 Voorstelling van data wat vanaf ’n Twitter-profiel in Microsoft Word gekopieer is 149

Skermskoot 5.2 Skermskoot van ’n Microsoft Excel lêer met ’n deelnemer se inligting 149

Skermskoot 5.3 Voorstelling van data wat skoongemaak is in ’n Microsoft Word-dokument 150

Skermskoot 5.4 Statistiese inligting vir woordelyste in WST 164

Skermskoot 5.5 Woordelysfunksie van ’n klein korpus 165

Skermskoot 5.6 Voorstelling van die konkordansiefunksie in WST 166

Skermskoot 5.7 Sleutelwoordfunksie van WST 166

Skermskoot 5.8 Voorstelling van die HU in *ATLAS.ti* 168

Skermskoot 5.9 Voorstelling van die aanhalingsfunksie in *ATLAS.ti* 169

Skermskoot 5.10 Voorstelling van die kodes in *ATLAS.ti* 169

Skermskoot 5.11 Voorstelling van die p-waardes (geneem vanuit Levon, 2010:81) 172

Skermskoot 5.12 Voorstelling van ŉ deelnemer se rou data vanaf Twitter 183

Skermskoot 5.13 Voorstelling van onverwerkte twiets 183

Skermskoot 5.14 Voorstelling van die lys van kodes waarmee die emotikons vervang is. 185

Skermskoot 6.1 Uitbeelding van die sleutelwoordlyste in WordSmith Tools 195

Skermskoot 6.2 Sleutelwoordlyste volgens p-waarde gesorteer 196

Skermskoot 6.3 Grammatikale woorde in die sleutelwoordlyste 197

Skermskoot 6.4 Voorstelling van die grammatikale woorde volgens sleutelwaarde gesorteer 197

Skermskoot 6.5 Uitbeelding van die onverwerkte data wat in die ondersoek na die taalverhouding gebruik is 242

Skermskoot 7.1 Voorstelling van ŉ twiet waarin die hoofletter aan die begin van ŉ sin met *ATLAS.ti* gekodeer is 298

Skermskoot 7.2 ŉ Voorbeeld van die manier waarop die emotikons in *ATLAS.ti* gemerk is 370

Skermskoot 7.3 JGAAP-koppelvlak 398

Skermskoot 7.4 Txt-dokumente in JGAAP 399

Skermskoot 7.5 Afvoer vanaf JGAAP 399

Hoofstuk 1
Agtergrond en kontekstualisering

Twitter is changing the way we talk to one another and the way businesses relate with clients, all within 140 characters or less. (130 characters)

Morris (2009:8)

1.1 Kontekstualisering

#taal #niekonstant #verander #behoefte #taalgebruiker #anonimiteit #identifikasie

Statistiek (Internet Live Stats, 2015; Statista, 2016) toon aan dat die internet as ’n gevestigde (en snel groeiende) platform beskou kan word waardeur die wêreld kommunikeer en idees deel. Hierdie massiewe kommunikasieplatform skep ook die geleentheid vir anonimiteit en is daarom ook die ideale platform vir negatiewe kommunikasie soos laster en afknouery (die sogenaamde *trolling*-verskynsel). Veral in die lig van laster en soortgelyke negatiewe kommunikasie, is die identifisering van outeurs (wie se ware identiteit nie bekend is nie) van belang en dit is juis hier waar forensiese linguistiek van toepassing is.

Die vrae ontstaan dus: wie is werklik aan die woord en is dit moontlik om die ware outeurs van kort elektroniese tekste te bepaal? Forensiese linguistiek word gebruik om te bepaal wie die outeurs van bepaalde tekste is en in ’n digitale era waar anonimiteit aan die orde van die dag is, is dit noodsaaklik om te bepaal of hierdie beginsels ook suksesvol op korter Afrikaanse tekste − soos mikrobloginskrywings − toegepas kan word om sodoende outeurskapidentifikasie uit te voer.

Outeurskapidentifikasie (vgl. Kotzé, 2010:185) verwys na die beginsel waarvolgens outeurs van bepaalde tekste geïdentifiseer kan word. Coulthard (2004:431) voer aan dat die linguistiese benadering tot die bevraagtekening van outeurskap van ’n teks as die teoretiese uitgangspunt beskou kan word en hierdie kenmerkende taalstyl van individue sal deurgaans in tekste manifesteer. Coulthard (2004: 431) stel verder dat sprekers elk oor ’n eie kenmerklike en individuele manier beskik waarop taal gebruik word − ’n idiolek. Hierdie manier van ondersoek word onder andere toegepas in die veld van teologie, letterkunde, opvoedingswetenskappe, inligtingsdienste sowel as forensiese linguistiek (vgl. Coulthard, 2004; Coulthard & Johnson, 2010; Lancashire, 1999; Louwerse, 2004; Olsson, 2008). Juola (2008:6) noem ook dat outeurskapidentifikasie na enige poging verwys om die karaktereienskappe van die skeppers van tekste (ongeag die genre) te identifiseer. Olsson (2008:26) stel dat die unieke taalgebruik

van individue nie ’n doelbewuste handeling is nie − dit is ’n wetenskap om die taal van individue te ontleed.

Van Rooy (2014:16) postuleer dat die taalkunde-soeklig al hoe meer op die taalvorme in elektroniese kommunikasie val, byvoorbeeld die beskrywing van taal soos dit gebruik word op verskillende platforms soos teksboodskappe, e-posse en sosiale media (soos byvoorbeeld Facebook, Twitter, Instagram). Crystal (2006:243) argumenteer dat die internet die geleentheid vir nuwe ontwikkelinge in ’n taal skep en dat dit die kreatiwiteit van ’n taal verhoog.

boyd[1] en Ellison (2007:12) is van mening dat sosialenetwerkplatforms soos Twitter baie belangrike data oor die natuurlike optrede van mense kan oplewer. Volgens Crystal (2006:244) bied internetplatforms[2] soos blogs en mikroblogs voorbeelde van ’n skryfstyl “which has never been seen in a public, printed form, outside of literature, and even there it would take an ingenious novelist indeed to capture its innocent spontaneity and unpredictable thematic direction”. Dit blyk dus asof ŉ nuwe taalstyl (wat aan die gebruikers en die platform se spesifieke behoeftes moet voldoen) besig is om te ontstaan.

Vanuit navorsing wat reeds gedoen is (Crystal, 2004; Crystal, 2006:244; Plester *et al.,* 2008; Tannen, 2013:100; Thurlow, 2013:243), blyk dit dat die taalstyl van hierdie sosialemediaplatforms spontaan en onvoorspelbaar is − outeurs word nie deur taalreëls, spelkonvensies of stylvereistes beperk nie. Verder is hierdie tekste ook ongeredigeer − daar is gewoonlik geen teksversorger, redigeerder of subredakteur by die skryfproses betrokke nie. Aangesien die data vanaf hierdie sosialenetwerkplatforms geredelik bekombaar is, gee dit aanleiding tot die moontlikheid van navorsing oor die taalstyle wat daarin weerspieël word. Crystal (2010) stel dit onomwonde dat nuwe tegnologie die bestaande kommunikasiemoontlikhede van ’n taal wysig en skep só ook die geleentheid vir ’n nuwe taal oftewel ’n nuwe styl, register of selfs taalfunksie in die taal om te ontstaan.

Mikroblogs[3] het in ’n kort tydperk[4] ŉ baie gewilde sosialemediaplatform geword. Al hoe meer internetgebruikers begin van tradisionele internetkommunikasieplatforms soos blogs en forums wegbeweeg na mikroblogs (Bruns & Stieglitz, 2012:163; Chen *et al.*, 2015; Lemahieu *et al.*, 2015; Morris, 2009:17; Papworth *et al.*, 2015). Mikroblogs verwys na platforms wat gebruikers in

[1] dana boyd verkies dat haar naam nie met hoofletters gespel word nie (boyd, 2006).
[2] Vir die doel van hierdie studie verwys ’n internetplatform (of -platforms) na die verskillende toepassings of koppelvlakke – toestelle of programme – wat gebruikers daartoe in staat stel om deur middel van die internet te kommunikeer. Hierdie platforms sluit sosiale media, soos Facebook en mikroblogs, forums, blogs, kletskamers en selfs e-posse in.
[3] Daar is verskeie mikroblogplatforms – die gewildste hiervan is Twitter (http://twitter.com).
[4] Twitter het in 2006 ontstaan en word as die vinnig groeiendste sosialemedia-inisiatief sedert Facebook en MySpace beskou (Morris, 2009:8; Naone, 2008; O’Reilly & Milstein, 2009:5).

staat stel om vinnig en doeltreffend ’n boodskap te plaas. Volgens Kelsey (2010:181), Morris (2009:17) en Williams *et al.* (2013:2) kan mikroblogs gedefinieer word as boodskappe wat uit foto’s, video’s, hiperskakels of slegs tekste bestaan, maar soos in die geval van Twitter kan dit slegs 140 karakters bevat − spasies ingesluit.

Mikroblogs kan as rekenaarbemiddelde kommunikasie[5] (RBK) beskou word. Verder skep mikroblogs ’n platform (sosiale netwerke) om jou stem (onmiddellik) die wêreld in te stuur en jou boodskap aan ’n gehoor oor te dra in ’n taalstyl waarmee jy gemaklik is. Styl (vgl. Du Toit & Smith-Müller, 2015:15) verwys na die keuses wat outeurs uit moontlike formulerings maak om aan gedagtes vorm te gee. In hierdie verband definieer Crystal (2008:460) *stilistiek* as ’n tak van linguistiek waar die eienskappe van die taalkeuses (variëteite) ondersoek word, en waar daar verantwoording gedoen word vir hierdie taalkeuses.

Crystal (2008:460) verwys ook na stylverskuiwings waar taalgebruikers van ’n bepaalde taal sal wissel tussen verskillende style om ’n bepaalde doel te bereik. McMenamin (2002:109-110) stel dat styl eie aan individue is en verwys na die onbewustelike keuses (die keuse om een vorm bo ’n ander te gebruik) wat individue maak wanneer hulle taal gebruik. Wanneer styl vanuit die pragmatiek beskou word, is dit duidelik dat deelnemers oor die vermoë moet beskik om die uitings wat hulle gebruik aan te pas binne die konteks waarin hulle dit gebruik (vgl. De Stadler, 1989:336).

Die vraag kan dus gestel word: Beïnvloed die 140-karakterbeperking van mikroblogs die taalstyl, en is dit moontlik dat ’n stylverskuiwing sal plaasvind? Lindqvist (2011:1) is van mening dat internetgebruikers skryf soos hulle praat en dat dít aanleiding gee tot ’n verkorte, informele taalstyl. Aangesien mikroblogs geen stylvereistes of spelkonvensies volg nie, staan dit gebruikers vry om taal te gebruik soos hulle wil. Is daar unieke kenmerke wat in die Afrikaans voorkom soos dit op mikroblogs gebruik word, en is dit moontlik om outeurs van mikrobloginskrywings te identifiseer?

Elke taal beskik oor ’n stel reëls wat as fundamenteel tot die bepaalde taal se konvensies beskou word (Clark, 2009:117). Hierdie reëls hou ook verband met semantiek, sintaksis, morfologie, fonologie, woordeskat en pragmatiek. In hierdie aanlyn kontaksituasie (waar ’n boodskap vanaf senders aan gehore oorgedra word) sal sprekers gebruik maak van woorde en

[5] RBK verwys na internettaal en kan na beide geskrewe teks of gesproke dialoog verwys (Barasa, 2010:322). Daar word tussen sinchroniese (onmiddellike interaksie) en asinchroniese (interaksie is uitgestel) ruimtes onderskei (Brunette, 2005:9; Crystal, 2010:11).

uitdrukkings wat bekend is aan die gehoor (ontvangers van die boodskap). Hoe lyk hierdie gemeenskaplike grammatika in die Afrikaans wat op mikroblogs gebruik word? Kan outeurs van spesifieke mikrobloginskrywings bepaal word?

Soos reeds genoem, is daar ’n karakterbeperking van 140 wanneer ’n inskrywing op hierdie soort platforms geplaas word. Tekste moet dus aangepas word om in ’n (uiters) beperkte spasie ’n boodskap aan bedoelde lesers te kan oordra. Die kommunikasiebeginsels van hierdie soort boodskappe is dus baie belangrik. Wanneer só ’n boodskap ontleed word, moet daar nie net ondersoek na die taalstruktuur ingestel word nie, maar ook na die teks, die konteks en die deelnemers. Aangesien daar ’n maksimum lengte vir dié soort tekste bestaan, is dit moontlik dat die taal aangepas moet word ten einde die boodskap (verstaanbaar) aan die gehoor oor te dra.

Prinsloo (2002:107) sluit hierby aan deur te stel dat ’n teks dus as geslaagde kommunikasie beskou kan word indien dit ’n samehangende eenheid vorm waar die inhoud op só ’n manier oorgedra word dat die verlangde reaksie verkry word. In hierdie studie is diskoersanalise (die manier waarop nuwe inligting oorgedra word, die omgewing − konteks − van die boodskap sowel as die sosiale konteks) as vertrekpunt gebruik om sodoende die patroonmatigheid van korpora te kon ontleed.

Dit is duidelik wanneer al die bogenoemde gegewens in ag geneem word, die moontlikheid dus bestaan dat Afrikaans ’n “ander” vorm kan aanneem om aan die kommunikasiebehoeftes van mikrobloggebruikers te voldoen. Wanneer daar ’n behoefte aan kommunikasie bestaan en die taal nie aan die gebruikers se behoeftes voldoen nie, kan daar ’n “nuwe” vorm van taal (hetsy tussen twee tale of binne ’n enkele taal self) ontstaan om die verlangde kommunikasiebehoeftes te vervul. Sosiale media gee aanleiding tot só ’n taalbehoefte. Tog gee sosiale netwerke ook aanleiding tot ’n platform waar gebruikers hul opinies anoniem[6] met die wêreld kan deel. Hierdie anonimiteit lei tot vrae oor die mate waarin daar bepaal kan word wie die ware outeurs van aanlyn tekste is.

In navorsing oor outeurskapidentifikasie is daar reeds bepaal dat individue oor unieke taalkenmerke[7] beskik. Unieke taalkenmerke maak dit moontlik om outeurs van bepaalde tekste te kan identifiseer. Aangesien die taallandskap daagliks verander en ontwikkel en siende dat die internet onlosmaaklik deel is van ons daaglikse kommunikasienetwerk, is dit nodig om ondersoek in te stel na die identifikasie van die outeurs van Afrikaanse mikrobloginskrywings.

[6] Anoniem in die sin dat jy jou eie aanlyn identiteit kan skep, en dat niemand ooit hoef te weet wie jy werklik is nie.
[7] Soos in studies deur Chaski (2007); Coulthard (2004); Juola (2006); Kotzé (2007); Louwerse (2004); Stamatatos (2009) asook McMenamin, (2002).

Hierdie studie was eksploratief van aard en het ook ten doel gehad om die beste metode te bepaal om outeurskapidentifikasie van kommunikasie op mikroblogs uit te voer.

Verskeie studies (in ander tale) is reeds uitgevoer op korter tekste soos Brocardo *et al.* (2013) wat stilometriese kenmerke gebruik het om outeurskapidentifikasie op tekste van 500 woorde uit te voer. Soortgelyk aan Brocardo *et al.* (2013) se studie het Barry en Luna (2012) aanlyn forums as data gebruik. Nirkhi *et al*. (2015) het van kort tekste bestaande uit 290 tot 800 woorde gebruik gemaak in hulle studie. Green en Sheppard (2013) het bepaal dat dit moontlik is om outeurskapidentifikasie uit te voer op ’n kort teks met ’n gemiddelde lengte van 12,5 woorde per boodskap.

Verder het Dietrick *et al.* (2012) Inches en Crestani (2011), Lake (2010) asook Sousa *et al.* (2011) ook navorsing oor Twitter gedoen. Alhoewel daar al studies onderskeidelik oor outeurskapidentifikasie en Internetafrikaans onderneem is[8], is daar tot op hede nog nie pertinent na die kenmerke van Afrikaans en die moontlikheid van outeurskapidentifikasie op mikroblogs ondersoek ingestel nie. Daarom is dit duidelik dat daar in die digitale era waarin ons ons bevind ’n leemte in Afrikaans bestaan oor die moontlikheid om outeurs van mikrobloginskrywings te identifiseer.

1.2 Probleemstelling

Vanuit die bespreking (afdeling 1.1) blyk dit dat die karakterbeperking (soos gesien op mikroblogs), die interaktiewe aard van sosiale netwerke sowel as die kommunikasiedoel aanleiding kan gee tot die oorsprong van ’n “nuwe” vorm van Afrikaans wat aan die taalbehoeftes van die gebruikers op hierdie platform kan voldoen. Dit blyk ook duidelik dat ’n digitale era waar anonimiteit aan die orde van die dag is, aanleiding gee tot die vraag: Kan outeurs van kort Afrikaanse boodskappe (mikroblogs) suksesvol geïdentifiseer word?

In die navorsing het ek ondersoek ingestel na eienskappe van die taal om sodoende te bepaal wat die kenmerkende taaleienskappe van Afrikaans is soos dit op mikroblogs gebruik word alvorens daar eksploratief ondersoek ingestel kon word na die suksesvolle identifisering van mikroblog-outeurs en of daar wel ’n raamwerk saamgestel kon word waarvolgens outeurskapidentifikasie op mikroblogs kon plaasvind.

Alhoewel die bogenoemde twee doelwitte in hierdie studie dalk uiteenlopend kan voorkom, is dit

[8] Sien hoofstuk 2 en 3. Hier word daar verwys na navorsing soos gedoen deur Carney (2015), Hubbard (1994, 1995), Jansen van Vuuren (2007), Kotzé (2007, 2010), Malan (1998, 2003), Thiart (2014), Venter (2013) en Verhoef (2001).

belangrik om in ag te neem dat daar nie ’n korpus in Afrikaans bestaan wat mikroblogs insluit nie. Juis daarom was dit noodsaaklik om eerstens die taaleienskappe wat kenmerklik van Afrikaans op mikroblogs is, te identifiseer. Eers nadat hierdie eienskappe geïdentifiseer is, kon outeurskapidentifikasie in aanvang neem.

1.3 Navorsingsvrae

Ten einde die bostaande probleemstelling aan te spreek, is die studie deur die volgende navorsingsvrae gelei:

- Wat is die taalstilistiese en stylkenmerke van Afrikaans op mikroblogs?
- Kan die outeurs van mikrobloginskrywings suksesvol geïdentifiseer word?
- Wat is die forensiese beginsels wat gebruik kan word om outeurskapidentifikasie op mikroblogs toe te pas?

1.4 Navorsingsdoelwitte

Vanuit die bogenoemde navorsingsvrae was die volgende doelwitte vir die studie voorgehou:

- Om te bepaal wat die taalstilistiese en stylkenmerke is − soos dit op mikroblogs voorkom.
- Om te bepaal of die outeurs van mikrobloginskrywings suksesvol geïdentifiseer kan word.
- Om te bepaal wat die forensiese beginsels is wat gebruik kon word om die outeurs van mikrobloginskrywings te identifiseer.

1.5 Sentrale teoretiese stelling

Vanuit die bogenoemde oorsig kan die veronderstelling dus gemaak word dat daar wel kenmerkende taaleienskappe geïdentifiseer kan word wanneer Afrikaans op mikroblogs ondersoek word. Verder kan gestel word dat dit moontlik is om outeurskapidentifikasie op mikroblogs uit te voer en om die outeurs van mikrobloginskrywings suksesvol te identifiseer.

1.6 Navorsingsmetode

In hierdie afdeling verwys ek na die manier waarop die navorsing plaasgevind het. Daar word eerstens na die literatuurstudie verwys, die empiriese studie word voorgehou sowel as data-insameling, instrumente wat in die studie gebruik is en die etiekkwessies verbonde aan my studie.

1.6.1 Literatuurstudie

Ten einde te bepaal of daar ’n unieke taalstyl geïdentifiseer kan word vir mikrobloggebruikers, is dit eerstens belangrik om konsepte soos *sosiale media*, *mikroblog* en *kommunikasie* te beskryf. Aangesien hierdie ’n eksploratiewe studie was om aan te toon dat daar wel ’n kenmerkende styl vir mikroblogs bestaan en dat outeurskapidentifikasie wel op hierdie internetplatform uitgevoer kon word, word daar oorsigtelik na verskeie subdissiplines verwys.

Bogenoemde het aanleiding gegee tot die skep van ’n teoretiese raamwerk wat gebruik is om die stilistiese en stilometriese empiriese resultate te interpreteer en verifieer. Daar word ook na literatuur oor outeurskapidentifikasie verwys veral in terme van die ontwikkelinge in digitale forensiese navorsing. Aangesien hierdie ’n eksploratiewe studie was, het die literatuurstudie die verskillende metodes met betrekking tot stilometriese en stilistiese metodes ondersoek waarna die beste metode vir die studie bepaal is.

1.6.2 Empiriese studie

Nadat die literatuurstudie voltooi is, het die empiriese studie die taal- en stylkeuses van die mikrobloggebruikers (soos in Afrikaans) ondersoek. Daarna is ’n eksploratiewe studie oor die outeurskapidentifikasie van mikrobloggebruikers onderneem. Hierdie studie het die vorm van ’n korpuslinguistiese[9] ondersoek aangeneem. Wanneer korpuslinguistiese en forensieslinguistiese beginsels toegepas word om die outeurs van tekste te bepaal, is dit belangrik dat daar ’n verwysingskorpus bestaan.

Volgens Baker (2010a:12) en Nesselhauf (2005) verwys ’n verwysingskorpus na die prototipiese korpus wat as ’n voorstelling van die norm van ’n bepaalde taal beskou kan word. Hierdie soort korpus is gewoonlik baie groot en is afkomstig van verskeie bronne aangesien dit die norm van ’n taal moet voorstel. Baker (2010a:12) en Nesselhauf (2005) verwys ook na ’n gespesialiseerde korpus. ’n Gespesialiseerde korpus verwys na ’n doelbewus geselekteerde groep tekste wat ondersoek word om ’n bepaalde vraag te beantwoord.

Vir die doel van hierdie studie bestaan daar nie ’n korpus wat as die verwysingskorpus beskou kon word nie. Alhoewel daar wel korpora van Afrikaans saamgetel is wat gebruik kan word as verwysingskorpus[10] bestaan hierdie korpora nie alleenlik vanuit Afrikaanse mikrobloginskrywings nie. Aangesien daar ondersoek ingestel word na kenmerke van die Afrikaans van mikroblogs en

[9] Korpuslinguistiek verwys na die studie waar tekste in terme van woordeskat en grammatikale beginsels ontleed word (Nesselhauf, 2005). McEnery en Wilson (2001:11) sluit hierby aan deur korpuslinguistiek te definieer as die studie van taal soos dit in die “regte lewe” gebruik word. ’n Korpuslinguistiese ondersoek word uitgevoer deur korpora te ontleed. ’n Korpus verwys na ’n versameling linguistiese data (Crystal, 2008:117).

[10] Sien afdeling 4.2.4 (Korpora in Afrikaans) vir ‘n verwysing na die bestaande korpora in Afrikaans.

nie na die wyse waarop Afrikaans soos dit op mikroblogs gebruik word van ander genres van Afrikaans verskil nie, is dit juis daarom nodig dat daar eers ʼn prototipekorpus[11] saamgestel om te bepaal wat as die algemene kenmerke van Afrikaans op mikroblogs beskou kan word. Sodra kenmerke van Afrikaans op mikroblogs geïdentifiseer is, sou dit moontlik wees om die kenmerke te vergelyk met die taalgebruik van individuele deelnemers om sodoende die moontlikheid van outeurskapidentifikasie op mikroblogs te kon ondersoek.

Die studie is dus in twee dele uitgevoer. Eerstens, is daar vasgestel watter taaleienskappe kenmerkend van mikroblogs is. In hierdie analise is vier kategorieë van outeurskapidentifikasie gebruik om te bepaal wat die kenmerklike eienskappe van Afrikaans op mikroblogs is. Die vier kategorieë sluit in: leksikale kenmerke, strukturele kenmerke, kontekstuele kenmerke en sintaktiese kenmerke en hulle kan as subkategorieë van ʼn stilometriese analise beskou word (vgl. Dinh, 2014:9). Die frekwensie van die gebruik van karakters, woord- en sinslengte en woordsoortgebruik kan ook in hierdie soort analise gebruik word.

Hierdie kenmerke is na gelang van die bevindinge vanuit die literatuurstudie oor mikroblogtaal ondersoek. Daar is ondersoek ingestel (volgens verkennende forensiese beginsels) na kenmerkende taaleienskappe van Afrikaans op mikroblogs. Om te bepaal of daar enige taaleienskappe voorkom wat kenmerkend[12] van mikroblogs is, was daar ondersoek ingestel na leksikale items, sintaksis en ortografie (onder andere reduksie, emotikons en die gebruik van hutswoorde). Sodra die kenmerklike taaleienskappe van mikroblogs geïdentifiseer is, kon die tweede deel van die studie aangespreek word.

Tweedens, is ʼn korpus saamgestel uit die tekste van tien deelnemers. Hierdie korpus is geanaliseer (forensieslinguistiese metodes is hier gebruik) om te bepaal of dit wel moontlik kon wees om die outeurs van inskrywing op mikroblogs te bepaal. Volgens navorsing deur Schwarts *et al.* (2013:1881) is dit moontlik om die outeurs van mikrobloginskrywings (Engels) te bepaal met so min as 200 inskrywings mits die vergelykingskorpus uit duisend outeurs bestaan.

1.6.3 Dataversameling

Sosiale netwerke maak ʼn argief van alles wat daarop gepubliseer word. Dus is dit moontlik om terug te werk en sodoende data van 1) bepaalde gebruikers en 2) tydens ʼn bepaalde tydstip in te samel. Vir die doel van hierdie studie het ek die data op twee verskillende maniere

[11]Die prototipekorpus verwys na die verwysingskorpus wat saamgestel word uit mikrobloginskrywings van gebruikers wat volgens die gerieflikheidsteekproef geïdentifiseer word, en wat gebruik is om die kenmerkende eienskappe van Afrikaans op mikroblogs te identifiseer.

[12] Aangesien hierdie ‘n eksploratiewe studie is, gaan daar nie vergelykend na Afrikaans soos dit op ander platforms gebruik word gewerk word nie. Die groter taalkundige vraag, hoe verskil die kenmerke van Twitter-Afrikaans van Afrikaans, kan na ‘n aanbeveling in verdere navorsing hanteer word.

ingesamel. Data vanaf mikroblogs is ingesamel deur sleutelwoorde[13] te soek, en om dan die profiele van die onderskeie gebruikers te ondersoek. Hierdeur is mikrobloggebruikers geïdentifiseer wat gereelde Afrikaanse inskrywings maak. Aangesien die studie ondersoek ingestel het na kenmerklike eienskappe van Afrikaans, was dit belangrik om die data van verskillende individue (mikroblog-profiele) te versamel. Die gerieflikheidsteekproefneming (vgl. Maree & Pietersen, 2011:177) is dus gebruik om outeurs van inskrywings te selekteer[14].

Vir die tweede deel van die studie is tien deelnemers geïdentifiseer wat oor genoegsame Afrikaanse mikrobloginskrywings beskik het (elke ondersoeksubjek moes minstens 1000 inskrywings hê − indien meer as tien geskikte deelnemers geïdentifiseer is, kon daar met minder inskrywings per ondersoeksubjek gewerk word). Die korpora is wel geannoteer (deur middel van woordsoortetikettering) om ’n deeglike korpusanalise (leksikaal en grammatikaal) van die patroonmatigheid van die betrokke korpora te verseker.

1.6.4 Instrumente

Die empiriese studie is met behulp van rekenaarprogrammatuur soos *WordSmith Tools* en *ATLAS.ti* uitgevoer. *WordSmith Tools* en *ATLAS.ti* is gekies as rekenaarprogrammatuur vir hierdie spesifieke studie aangesien die kombinasie van hierdie twee programme die moontlikheid geskep het van ’n analise wat op die taalstruktuur sowel as unieke taalkenmerke kon fokus. *WordSmith Tools* is hoofsaaklik in die studie gebruik vir die sleutelwoord- en konkordansiefunksies.

Hierdie funksies vergelyk woordelyste met mekaar en gee aanleiding daartoe dat die verskille in leksikale en grammatikale eienskappe van die korpora beskryf kan word (vgl. Berber-Sardinha, 2001:1). *ATLAS.ti* is gebruik om onder andere kodes aan spesifieke konsepte toe te ken wat die stilistiese analise vergemaklik het. Die kombinasie van hierdie twee programme het dus die moontlikheid geskep om die korpus grammatikaal en leksikaal te ondersoek.

Verder is daar ook gebruik gemaak van ’n n-gram-analise en chi-kwadraattoetse om die data te analiseer. Die n-gram-analise verwys na ’n spesifieke getal (n) woorde wat in ’n spesifieke volgorde geanaliseer is (Chaski, 2005:5). Nirkhi *et al.* (2015:369) verwys na die gebruik van karakter-n-gramme (’n analise kan op leksikale, sintaktiese en strukturele vlak plaasvind), maar

[13] Hierdie sleutelwoorde het byvoorbeeld #afrikaans, #mytaal en #naweek ingesluit.
[14] Daar is gepoog om vir die doel van die verwysingskorpus ‘n korpus van minstens 20 000 woorde in te samel, die finale woordtelling vir die verwysingskorpus is 23 442

ondersoek ook die moontlikheid van karakter-*unigram, bi-gram*, karakter-*tri-gram*, woord-*uni-gram*, woord-*bi-gram* en woord-*tri-gram*[15].

Die n-gram analise verwys in die eenvoudigste vorm na ŉ analise van die kombinasies van aaneengrensende woorde of letters. Neem die sin "dit is lukraak" as voorbeeld. As daar na woord bi-gramme ondersoek ingestel word sal die sin as # *dit*, *dit is*, *is lukraak* en *lukraak* # beskou kan word (# verwys na die woordgrens). As die woord "lukraak" as karakter bi-gram ontleed word, is dit *#l, lu, uk, kr, ra, aa, ak* en *k#*. Die doel van ŉ n-gram analise is om taal statisties te ontleed en om te bepaal watter woorde of letters geneig is om op mekaar te volg.

Die chi-kwadraattoets word beskou as 'n nieparametriese statistiese prosedure wat verhoudings tussen frekwensies toets (Oakes, 1998:24). Hierdie toets bepaal of tekste oor 'n hoër of laer frekwensie van die bepaalde item wat ondersoek word beskik as die frekwensie van daardie selfde item in die prototipe- en verwysingskorpora.

1.6.5 Data-analise

Die ondersoek is stilisties sowel as stilometries uitgevoer. In die tweede afdeling van die empiriese ondersoek het die studie die vorm van 'n verkennende forensieslinguistiese studie aangeneem waar daar bepaal is of outeurskapidentifikasie wel moontlik is op mikroblogs. Hierdie studie het ook gebruik gemaak van metodes vanuit die korpusanalise. Die bevindinge oor kenmerklike eienskappe van Afrikaans − soos dit op mikroblogs gebruik word − is gebruik om eienskappe uit te skakel wat nie uniek aan 'n ondersoeksubjek is nie, maar wat liewer as kenmerkend van ŉ platform beskou kon word.

1.6.6 Etiekkwessies

Aanlyn netwerke gee aanleiding tot 'n platform wat gebruikers in staat stel om hulpbronne (persoonlike profiele, e-posse, foto-albums, video's ensovoorts.) aan te wend om 'n produk te skep waar inligting weergegee word (vgl. D'Arcy & Young, 2012:534). Hierdie data is geredelik beskikbaar deur middel van verskeie soekfunksies. Is dit aanvaarbaar om hierdie data net so te gebruik sonder toestemming van die individue wat dit op die platform geplaas het?

Navorsing wat reeds gedoen is oor die etiekkwessies van publieke aanlyn platforms (Bornoe & Barkhuus, 2011; boyd, 2007; Viégas, 2005), toon duidelik aan dat alhoewel die inligting geredelik beskikbaar is, individue nie baie ingenome daarmee is indien hulle inskrywings sonder

[15] Hierdie verskillende n-gramme sal vir outeurskapidentifikasie gebruik word – aangesien die studie deur Nirkhi et al. (2015) veral op kort tekste toegepas is – en word volledig in hoofstuk 3 (afdeling 3.5.1.1 - Forensiese stilistiek) verduidelik.

hulle medewete vir navorsing gebruik word nie. Volgens D'Arcy en Young (2012:536) moet moontlike deelnemers genader word, die studie moet aan hulle verduidelik word en hulle toestemming moet verkry word om die data te kan gebruik.

Soos met enige ander navorsing, vereis navorsing op aanlyn platforms ook ingeligte toestemming van die deelnemers (Whiteman, 2012:89-90). Vir hierdie rede is die betrokke deelnemers genader om toestemming te gee sodat hulle aanlyn inskrywings gebruik kon word in hierdie studie[16]. Om die oorspronklikheid van die data te verseker, is slegs data gebruik wat versamel kon word voor die datum waarop toestemming verkry is. Toestemming is van al die deelnemers verkry. Data is ten alle tye vertroulik hanteer. Die korpora is ook ná afloop van die studie deur die promotor geberg en is ook aan die RMA (Language Resource Management Agency) verskaf vir argivering.

1.7 Hoofstukindeling

Hierdie studie was tweeledig van aard. Die eerste gedeelte het uit 'n literatuuroorsig van die relevante teoretiese raamwerke en vertrekpunte (hoofstuk 2, 3 en 4) bestaan, en die tweede gedeelte van die studie was empiries van aard (hoofstuk 5 bevat die uiteensetting van die verskillende metodes wat op hierdie studie van toepassing is terwyl hoofstuk 6 en 7 op die empiriese analise fokus).

In hoofstuk 2 word die konsep van taal gedefinieer ten einde die agtergrond te skep waarteen die veranderinge wat in Afrikaans op sosialemediaplatforms waargeneem is, ondersoek kon word. Daar word vlugtig na die variëteite van Afrikaans verwys ten einde Afrikaans in 'n digitale era voor te hou. Kommunikasie en taal word bespreek aangesien mikroblogplatforms 'n manier van kommunikasie is. Ten slotte word sosiale media bespreek en word daar spesifiek ondersoek ingestel na Twitter as 'n mikroblogplatform.

In hoofstuk 3 word forensiese linguistiek bespreek. Hierdie hoofstuk bied 'n oorsig van die ontstaan en ontwikkeling van forensiese linguistiek. Outeurskapidentifikasie word gedefinieer en daar word ondersoek ingestel na die konsepte *forensiese stilistiek* en *forensiese stilometrie*.

Korpuslinguistiek word in hoofstuk 4 bespreek. Dié hoofstuk bied 'n definisie van die term *korpora*, die verskillende soorte korpora, korpusgroottes, en die web as 'n korpus word ook bespreek. Verder verskaf ek 'n historiese oorsig van korpuslinguistiek alvorens ek ondersoek

[16] Sien bylaag A (Etieseklaring) en Bylaag B (toestemmingsbriewe).

ingestel het na onderskeidelik korpusgebaseerde- en korpusgedrewe analises. Ten slotte verwys ek na die forensiese korpuslinguistiese benadering.

Die metode wat ek in my studie gevolg het, word in hoofstuk 5 uiteengesit. Ek verwys na die maniere waarop die deelnemers geselekteer is, die data-insameling en -verwerkingsproses. Hierdie hoofstuk bied ’n oorsig van die kwalitatiewe en kwantitatiewe metodologie en die onderskeie instrumente wat in my studie gebruik is. Kleinkorpora en outeurskapidentifikasie word ook voorgehou sowel as die etiekkwessies verbonde aan my studie.

Hoofstuk 6 en 7 fokus op die empiriese analise. In hierdie hoofstukke word ’n kwalitatiewe en kwantitatiewe analise op die korpora uitgevoer. Ek ondersoek drie navorsingsvrae: 1) Wat is die taalstilistiese en stylkenmerke van Afrikaans op mikroblogs? 2) Kan die outeurs van mikrobloginskrywings suksesvol geïdentifiseer word, en 3) Wat is die forensiese beginsels wat gebruik kan word om outeurskapidentifikasie op mikroblogs toe te pas?

Ten slotte word die bevindinge van die studie saamgevat en opgesom in hoofstuk 8. Daar word opsommend verwys na die metodes wat gevolg is, die beperkinge van die bepaalde studie sowel as moontlike bykomende navorsingsvrae wat vanuit my studie ontstaan het. Vir die doel van my studie is daar vanuit die veronderstelling gewerk dat daar wel kenmerkende taaleienskappe geïdentifiseer kon word wanneer Afrikaans op mikroblogs ondersoek word. Verder het ek die sentrale teoretiese stelling − dat dit moontlik is om outeurskapidentifikasie op mikroblogs uit te voer en om die outeurs van sulke mikrobloginskrywings suksesvol te identifiseer − beantwoord.

Hoofstuk 2
Die teoretiese raamwerk

… The most valuable of all talents is that of never using two words when one will do...

Thomas Jefferson (in Hayes, 2008:89)

2.1 Inleiding

In hierdie studie het ek die kenmerkende eienskappe van Afrikaans ondersoek soos dit op mikroblogs gebruik word alvorens ek die moontlikheid van outeurskapidentifikasie spesifiek met betrekking tot mikroblogs[17] kon ondersoek. Soos met enige wetenskaplike navorsing, moes daar eers ’n teoretiese raamwerk geskep word voordat die analise kon plaasneem − en dit was dan ook die doel van hierdie hoofstuk[18].

Die teoretiese raamwerk wat in hoofstuk twee geskep is, sluit ’n oorsig van die ontwikkeling van Afrikaans in (om aan te dui hoe Afrikaans verander het om by die behoeftes van gebruikers aan te pas). Verder bespreek ek kommunikasie en taal sowel as sosiale media. Aangesien die taal van individu oor unieke eienskappe beskik[19], en siende dat daar vanuit die veronderstelling gewerk is dat die internet taaleienskappe toon wat nog nooit tevore gesien is nie[20], is hierdie hoofstuk verkennend van aard en word daar ondersoek ingestel na die aard van Afrikaans in ’n digitale era. Sodoende is ’n agtergrond geskep oor die eienskappe wat in die empiriese studie ondersoek is.

Die gewildheid van die internet en sosiale netwerke lei onomwonde daartoe dat navorsing in tale al hoe meer op taalvorme in die elektroniese kommunikasie fokus. “Die beskrywing van die taal van die verskeidenheid teksboodskappe oor mobiele telefone, e-posse, sosiale media en ander tekstipes op die internet” is volgens Van Rooy (2014:16) veral van belang. Van Rooy (2014:16) voer verder aan dat daar besorgdheid is oor die uitwerking van tegnologie op ’n taal.

Die belangstelling in die uitwerking wat tegnologie op ’n taal het, is nie ’n nuwe verskynsel nie. Herring (2002:138) en Baron (1984:145) het reeds geteoretiseer oor die invloed van kommunikasie deur middel van rekenaars sowel as die aard van taalgebruik op hierdie platforms. Gedurende die afgelope paar jaar is daar nie net geteoretiseer oor die invloed van

---

[17] Mikroblogs verwys na platforms wat gebruikers in staat stel om vinnig en doeltreffend ’n boodskap te plaas. Die boodskap kan ’n teks, foto’s, hiperskakels of video’s bevat, maar (in die geval van Twitter) is daar ’n maksimum lengte aan die boodskap verbonde – 140 karakters (Kelsey, 2010:181; Morris, 2009:17; Williams et al., 2013:2). Mikroblogs word volledig in afdeling 2.6 bespreek.

[18] Hoofstuk 2, 3 en 4 vorm deel van die literatuurstudie. Hoofstuk 2 fokus op Afrikaans, kommunikasie en sosiale media terwyl ek in hoofstuk 3 forensiese linguistiek ondersoek. Hoofstuk 4 verwys na korpuslinguistiek.

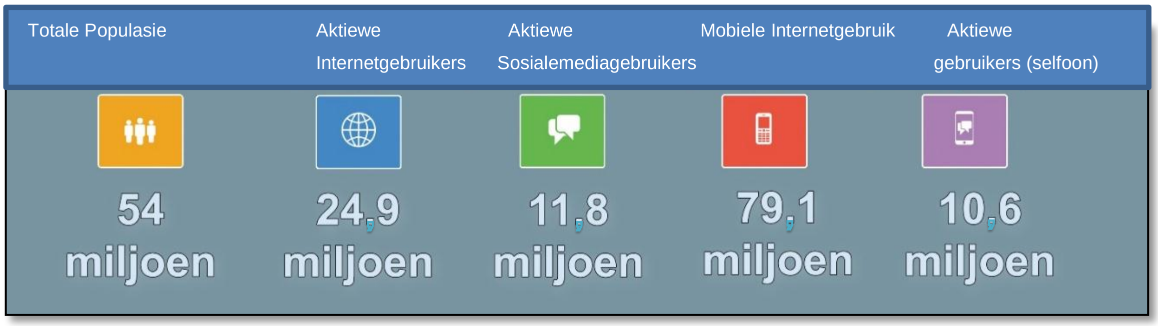
[19] Sien afdeling 2.3.2.2 (Interne variasie en idiolek).

[20] Sien afdeling 2.4 (Afrikaans in ’n digitale era).

RBK[21] nie. Trouens, navorsing oor RBK en veral die outeurskap van anonieme boodskappe word toenemend uitgevoer[22].

Volgens Bhargava *et al.* (2013:37) is daar in die afgelope twee dekades veral aandag gegee aan outeurskapidentifikasie met betrekking tot RBK (e-posse, SMS'e, Twitter-boodskappe, kitsboodskappe word ingesluit). In die verlede (vroeë 19de eeu) moes dokumente wat ondersoek is nie minder as 1 000 woorde bevat nie. Ondersoeke word nou uitgevoer op korter tekste (250 woorde en minder). Hierdie tipe ontledings staan as kleinkorpusontledings bekend[23] aangesien kommunikasievorme al hoe korter word (vgl. Bhargava *et al.*, 2013:37).

Figuur 2.1 verskaf 'n oorsig van Suid-Afrikaanse internetgebruikers en toon aan dat daar sowat[24] 10,6 miljoen aktiewe sosialemediagebruikers is. Dit blyk dus duidelik dat Suid-Afrikaanse sosialemediagebruikers stelselmatig besig is om toe te neem wat dus verder daarop dui dat navorsing oor die identifikasie van outeurs op verskeie sosialemediaplatforms (in Afrikaans) van belang is.



Figuur 2.1 Sosialemediagebruik in Suid-Afrika (geneem en aangepas vanaf Shezi, 2015)[25]

RBK en die spesifieke platform wat vir kommunikasie gebruik word, het 'n uitwerking op die manier waarop taal aangepas moet word om kommunikasie suksesvol te maak. Sekere platforms beskik oor 'n meer formele gebruik van taal (soos LinkedIn) terwyl ander platforms (soos Twitter, Facebook en Instagram) eerder van 'n meer informele "nuwe" vorm van die betrokke taal gebruik maak (vgl. Edwards, 2015). Daar is ook 'n verskil in die taalgebruik wanneer RBK in 'n sinchroniese of asinchroniese ruimte plaasvind. Die doel van kommunikasie en ander sosiolinguistiese faktore het ook 'n invloed op taalgebruik. Die manier waarop mense

[21] RBK verwys na rekenaarbemiddelde kommunikasie en word in afdeling 0 (Kontekstualisering) gedefinieer as 'n konsep wat na beide 'n geskrewe teks of gesproke dialoog verwys (Barasa, 2010:322). Daar word tussen sinchroniese (onmiddellike interaksie) en asinchroniese (interaksie is uitgestel) ruimtes onderskei (Brunette, 2005:9; Crystal, 2010:11). RBK word verder in afdeling 2.5 (Kommunikasie en taal) bespreek.

[22] Sien hoofstuk 3 (afdeling 3.6) waar studies rakende outeurskapidentifikasie in 'n digitale era bespreek word. Ook in hoofstuk 4 word daar na die verskillende studies rakende forensiese linguistiek, korpuslinguistiek, sosiale media en outeurskapidentifikasie verwys.

[23] Sien hoofstuk 3 (afdeling 3.6 -Forensiese linguistiek en 'n digitale era).

[24] Hierdie statistieke verander daagliks soos meer mense sosialemediaprofiele opstel.

[25] In 'n studie deur Ornico is daar bevind dat Twitter-gebruikers nou tot 8 miljoen vermeerder het, terwyl die hoeveelheid aktiwete sosialemediagebruikers vanaf 11.8 miljoen in 2015 tot 15 miljoen in 2017 gegroei het.

met mekaar kommunikeer, word gevolglik al hoe bondiger en dit is daarom belangrik om te bepaal tot watter mate die outeurs van "korter" tekste suksesvol geïdentifiseer kan word.

Vir die doel van my studie was dit nodig om ʼn teoretiese raamwerk oor kommunikasie, taal en sosiale media te skep ten einde die eksploratiewe forensiese analise oor outeurskapidentifikasie en mikroblogs te kon uitvoer.

2.2 Taalbeskouings

Taal is die middel wat kommunikasie moontlik maak hetsy die kommunikasie gesproke of geskrewe van aard is. Hughes (1996:6) beskryf geskrewe taal hoofsaaklik as ʼn visuele/motoriese handeling en alhoewel dit liniêr in tyd plaasvind, beskik dit oor die potensiaal om behoue te bly met die verloop van tyd en kan die taalhandeling telkens weer ondersoek (gelees) word. Van Rooy (2014:5) argumenteer dat taal in die algemeen gedefinieer kan word as "daardie middel wat mense gebruik om met mekaar te kommunikeer en hulle denke uit te druk".

Dit is dus duidelik dat taal sentraal staan tot die kommunikasieproses. Alvorens die verskillende kommunikasiekanale en die manier waarop hierdie kanale taal beïnvloed, ondersoek kon word, was dit noodsaaklik om taal te definieer. Linguiste soos Halliday en Givón fokus juis op die kommunikatiewe aard van taal. Halliday formuleer dat ook die teorie van sistemiese funksionele taalkunde waar ondersoek ingestel word na die wyse waarop taal in ʼn gegewe sosiale konteks gebruik word om bepaalde doelwitte te bereik. Givón (2012) argumenteer dan ook dat twee van die kern funksies van taal is dat taal as die uitbeelding en kommunikasie van informasie beskou kan word. Givón (2012) stel ook dat die verskillende strukturele elemente in kombinasie met mekaar verskillende grammatikale konstruksies kan vorm, wat dan aanleiding tot verskillende kommunikasiefunksies kan gee.

Voorts het ek ondersoek ingestel na die verskillende definisies van taal ten einde taal te kan definieer soos dit vir die doel van my studie gebruik is.

Die *Handwoordeboek van die Afrikaanse Taal* (Odendal & Gouws, 2015:1312) definieer taal soos volg (ek noem die eerste vier definisies wat betrekking het op my studie):

> **Taal** *s.nw*. (**tale; -tjie**) **1 [gew. ekv]** kommunikasiemiddel van die mens, voortgebring deur die spraakorgane; menslike spraak (asook die geskrewe vorm daarvan): *taal gebruik/hoor/aanleer/interpreteer*. **2** stelsel van gesproke en geskrewe woorde wat die mense van ʼn bepaalde land/gebied gebruik om

> met mekaar te kommunikeer: *Die Wes-Kaap het drie omgangstale - Afrikaans, Engels en Xhosa. Afrikaans is my moedertaal* | *'n* **dooie taal** 'n taal wat nie meer 'n spraakgemeenskap het nie, maar slegs in tekste voortbestaan, soos Latyn. **3** spesifieke vorm van 'n taal wat deur 'n bepaalde groep of individue gebruik word: *bendetaal, streektaal, vaktaal*. **4** manier/wyse waarop taalelemente gerangskik/gebruik word; styl: *deftige/informele taal.*

Vanuit die bogenoemde definisies is dit duidelik dat taal na talle aspekte kan verwys, maar dat hierdie aspekte steeds verband hou met mekaar. Vir die doel van my studie verwys taal (vanuit die definisies bespreek) in die breedste sin van die woord na beide die *langue* en *parole* aangesien mikrobloggebruikers die taal moet aanpas om ten spyte van die karakterbeperkings op die platform optimaal te kan kommunikeer. Wat kan vervolgens as die definisie van taal beskou word wanneer taal as die middel wat kommunikasie moontlik maak, gesien word?

Crystal (2008:265) verwys na De Saussure se definisie van taal en postuleer dat taal op die menslike en biologiese proses van spraak dui. Crystal (2008:265) definieer taal op die eenvoudigste vlak as 'n konkrete handeling − om te praat, skryf of om gebaretaal te gebruik. Taal is by uitstek 'n doelbewuste menslike handeling waardeur emosies, idees en gevoelens uitgedruk word. Bloch en Trager (1942:5) dui aan dat taal 'n stelsel van willekeurige mondelinge simbole is waardeur 'n spesifieke groep kommunikeer. Sapir (1921:2) verwys na taal as 'n stelsel van doelbewuste simbole waaraan betekenis gekoppel word. Verder is spraak (vgl. Sapir, 1921:2) 'n menslike aktiwiteit wat van sosiale groep tot sosiale groep verskil.

Davies (1999:2), Du Plessis (1995:1) en Hall (1968:158) sluit by Sapir aan en hou voor dat taal na die menslike handeling van gesproke of geskrewe kommunikasie verwys. Chomsky (1957:13) voer aan dat taal 'n stel onbeperkte sinne is wat onbeperk in lengte is maar wat uit 'n beperkte stel elemente bestaan − die grammatika van 'n taal moet dus beskou word as 'n toestel wat gebruikers in staat stel om sinne in dié spesifieke taal te produseer. Martin en Ringham (2000:79) definieer taal as 'n term wat verwys na 'n betekenisvolle stelsel hetsy verbaal, musikaal, visueel of gebare – 'n taal moet oor vorm en inhoud beskik.

Van Wyk (2003:469) verwys in sy definisie van taal ook na die betekenisdraende aard van taal wanneer hy taal soos volg definieer:

> (1) die aangebore vermoë van die mens om 'n verskeidenheid tekens en reëls tot betekenisdraende elemente te kombineer om sodoende gedagtes te uiter en te kommunikeer, (2) die spesifieke wyse waarop betekenisdraende elemente

gerangskik word (styl), (3) kenmerkende spraak van 'n bepaalde gemeenskap, groep of volk, (4) kommunikasiemiddel van 'n bepaalde dierespesie, (5) middel of simbool om gedagtes en gevoelens mee uit te druk.

Verhoef (1991:10) argumenteer dat taal meer as net "'n blote instrument is wat referensiële betekenis in homself opsluit of ronddra", taal is ook "sosiaal van aard en dit kan 'n spreker in 'n gesprek identifiseer" en in 'n bepaalde verhouding met aangesprokenes (ontvangers van die boodskap) stel. Vir kommunikasie om suksesvol te wees, moet sprekers mekaar verstaan. Wanneer die taal wat gebruik word vir kommunikasie 'n gedeelde taal tussen sprekers is (ongeag daarvan of kommunikasie in een taal geskied en of daar 'n ongesproke reël vir kodewisseling bestaan) ontstaan daar 'n taalgemeenskap tussen sprekers. Hierdie sosiale aard van taal moet in ag geneem word veral wanneer die aard van sosialemediaplatforms ondersoek word.[26]

McKay (2005:284) argumenteer soortgelyk dat 'n taalgemeenskap 'n definitiewe *kode* deel − dit wil sê 'n taal of 'n variëteit van 'n taal (hierdie kan verwys na 'n een- of meertalige taalsituasie) − en dat daar ook norme sal wees waarmee die gebruikers van hierdie *kode* vertroud is. Individue kan op enige gegewe tydstip aan meer as een taalgemeenskap behoort afhangend van waarmee individue op daardie spesifieke stadium wil identifiseer (professie, etnisiteit, geloof, ouderdom, gender en dies meer). Hierdie kode waarna McKay verwys, is belangrik aangesien sekere groepe (ook op sosiale media) oor 'n gedeelde kode kan beskik wat kommunikasie op die platform van hulle keuse kan vergemaklik.

Davies (2005:69) voer aan dat taal na die primêre menslike kommunikasiestelsel verwys wat deur jong kinders aangeleer word as deel van normale ontwikkeling. Verder verwys taal na 'n linguistieke entiteit, maar ook na 'n kulturele, sosiale en historiese entiteit wat volgens sommige navorsers mense van mekaar vervreem terwyl ander − soos De Saussure[27] − van mening is dat taal juis mense verenig (Davies, 2005:69).

Trudgill (1992:43) se mening vind aanklank by Davies se definisie dat taal linguisties, kultureel, sosiaal en histories van aard is wanneer hy sê dat taal buiten die bogenoemde ook polities van aard is. Hy (Trudgill, 1992:43) stel verder dat 'n *Ausbau*-taal 'n samevoeging van linguistieke variëteite is wat bestaan uit 'n outonome variëteit sowel as alle ander variëteite wat van dié variëteit afhanklik is.

[26] Sien afdeling 2.6 (Sosiale media).
[27] Sien afdelinng 2.2 (Taal gedefinieer)

Die bogenoemde definisies toon duidelik aan dat daar verskillende benaderings en sienswyses oor die definisie van taal bestaan. Tog is dit duidelik dat taal hoofsaaklik as ’n kommunikasiemedium beskou word waardeur boodskappe oorgedra word. Om hierdie boodskappe oor te dra, word ’n tekenstelsel (eie aan die betrokke taal) gebruik en betekenis lê in hierdie stelsel opgesluit. Van Rooy (2014:3) voer aan dat taal in sy algemeenste vorm verwys na “die instrument wat eie is aan die menslike spesie, en wat spesifiek die funksie van kommunikasie verrig” en dat die manier waarop individue die instrument gebruik, verskil. Van Rooy (2014:3) stel verder dat “die ordening van elemente binne die kommunikasiesisteem” (die menslike taal) oor verskeie moontlikhede beskik en dat die ordening “as meer of minder goed (of geskik vir ’n spesifieke doel)” beskou kan word.

Vir die doel van my studie oor outeurskapidentifikasie en mikroblogs is daar aangeneem dat *taal* gedefinieer kan word as ’n menslike kommunikasiehandeling (die manier waarop betekenis − gedagtes en emosies − oorgedra word deur middel van simbole) waar daar van ’n tekenstelsel (geskrewe of gesproke) gebruik gemaak word om betekenis aan ’n bepaalde gehoor oor te dra. Hierdie tekenstelsel is kenmerkend van ’n bepaalde gemeenskap (in dié geval mikrobloggebruikers) en die tekens en reëls wat in betekenisdraende elemente gekombineer word, word op ’n spesifieke manier gerangskik wat eie is aan die vereistes (en beperkings) van die medium waardeur die kommunikasie plaasvind. Hierdie taal word ook gekenmerk deur sosiale, kulturele en politiese eienskappe. Ten slotte is dit belangrik om in gedagte te hou dat individue oor kenmerkende taaleienskappe beskik wat eie aan hulle taalgebruik is, en dat hierdie eienskappe van individu tot individu sal verskil. Dus moet ’n definisie van taal ook verwys na die idiolektiese[28] veranderlikes wat daarin kan voorkom.

Alvorens ondersoek ingestel kan word na die kenmerke wat uniek aan Afrikaans is − soos dit op mikroblogs gebruik word − moes daar ’n agtergrond geskep word van sommige van die taalbeginsels wat eie aan Afrikaans is. Hierdie taalbeginsels word verder in hoofstuk 5 gebruik om die metode te vorm waarvolgens outeurskapidentifikasie kon plaasvind. Ten einde die bogenoemde te kon bereik, word daar eers oorsigtelik na die ontstaan en ontwikkeling van Afrikaans verwys.

2.3 Die ontwikkeling van Afrikaans

Om te bepaal hoe die Afrikaans wat op mikroblogs gebruik word daar uitsien en of hierdie Afrikaans oor kenmerkende eienskappe beskik (ten einde te bepaal of outeurs van

[28] Idiolek word in afdeling 2.3.2.2 (Interne variasie en idiolek) bespreek.

mikrobloginskrywings geïdentifiseer kan word), word daar kortliks na die ontstaan en ontwikkeling van Afrikaans verwys om sodoende ’n konteks te skep waarin die Afrikaans op mikroblogs later geanaliseer kon word. Hierdie historiese oorsig het ten doel om ’n agtergrond te skep waaruit dit duidelik sal blyk dat taal (Afrikaans) nie konstant dieselfde is nie en dat dit verander na gelang van die behoeftes van taalgebruikers. Ten einde Afrikaans soos dit in ’n digitale era gebruik word te bespreek, verskaf ek ’n opsommende historiese oorsig van Afrikaans.

2.3.1 Historiese oorsig van Afrikaans

Ten einde die ontstaan en ontwikkeling van Afrikaans (soos ons dit vandag ken) voor te hou, moet verskeie boustene en eksterne invloede op die taal in ag geneem word.

Afrikaans het sonder twyfel ’n sterk Nederlandse agtergrond. Volgens Ponelis (1988:11) en Vikner (1995:5) is Afrikaans deel van die Wes-Germaanse taalfamilie. Dit beteken dus dat Afrikaans − soos Engels, Fries, Vlaams, Nederlands en Duits − genealogies in ’n taalfamilie saamgegroepeer word aangesien hierdie tale vanuit dieselfde grondtaal afstam (almal susterstale van mekaar). Afrikaans het nie net vanuit Germaanse tale ontstaan nie. Die invloed van ander tale soos Maleis, Portugees, Khoi en ander inheemse Afrikatale (soos Xhosa) was ook aanwesig in die vroeë ontstaanjare van Afrikaans. Volgens Raidt (1983:160) kan die oorsprong van die Afrikaanse leksikon grotendeels teruggevoer word na Nederlands (ten spyte daarvan dat daar ook Maleise en Portugese kenmerke aanwesig is).

Afrikaans se fundamentele tipologiese kenmerke plaas die taal dus in die Wes-Germaanse taalfamilie, maar wat is Afrikaans en hoe het die taal tot stand gekom? Volgens Van Wyk (2003:11) verwys die woord “Afrikaans” na ’n “taal wat onder die invloed van Vreemdelinge-Nederlands, en veranderende omstandighede uit die 17de-eeuse volkstaal van Nederlands tot ’n onafhanklike taal in Suid-Afrika ontwikkel het”.

Van Rensburg (2012:14) voer aan dat die eerste periode van die geskiedenis van Afrikaans plaasgevind het nog voordat die VOC (*Vereenigde Oostindische Compagnie)* vir Jan van Riebeeck in 1652 gestuur het om ’n halfwegstasie (verversingspos) by Tafelbaai (Kaap) te vestig. Die inheemse Khoi-stamme het die waarde van handel besef en het Europese tale (Engels, Portugees en Nederlands) aangeleer om as marktaal (handelsjargon) te gebruik.

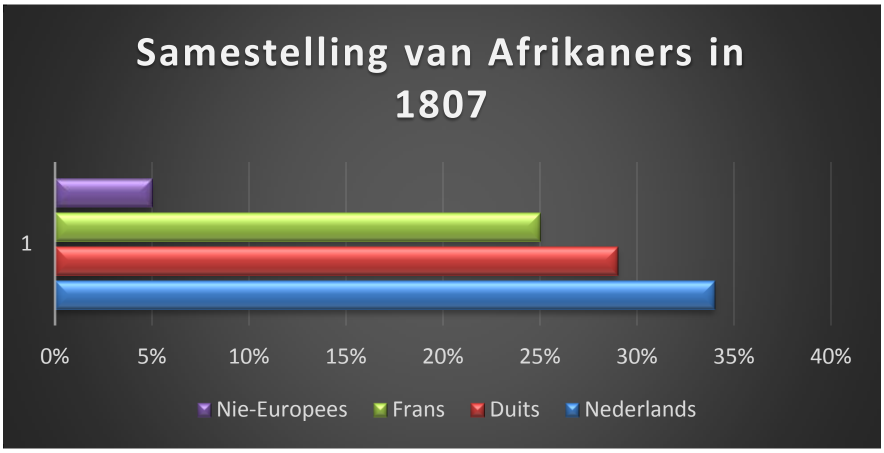
Den Besten (1989:226) sluit by Van Rensburg se argument aan en hou voor dat daar al sowat 60 jaar voor die stigting van die VOC se halfwegstasie (in 1652) ’n Nederlands- en Engelsgebaseerde handelsjargon onder die Khoi in die Kaap ontstaan het. Ná 1652 het daar

baie vinnig ’n Khoi-Nederlandse kontaktaal (ook bekend as ’n pidgin) ontstaan. Den Besten (1989:226) beweer dat daar ten spyte van groot strukturele en woordeskatverskille baie ooreenkomste tussen Khoi en Nederlands was soos die neiging om die werkwoord aan die einde van ’n sin te plaas (SOV: subjek-objek-werkwoord-volgorde).

Vanaf 1658 is slawe (hierdie slawe het veral nie-Germaanse tale gepraat) ingevoer na die destydse Kaapse nedersetting. Die behoefte om te kommunikeer, het daartoe gelei dat hierdie slawe ’n pidgin ontwikkel het en sodoende het die taalkontak vergroot om al die tale wat deur die slawe gepraat is by die Kaaps-Hollandse variëteit van Nederlands in te sluit (vgl. Van Rensburg, 2012:18). Den Besten (1989:224) argumenteer dat die Khoi-pidgin wat reeds in die Kaap aanwesig was as basis vir die slawe-pidgin gebruik is en dat struktuurelemente (soos die SOV-volgorde) behoue gebly het.

Die uitbreiding van die Kaapkolonie se grense het veroorsaak dat die inheemse stamme al hoe verder die binneland ingetrek het en die sogenaamde *Kitchen Dutch* oftewel Kaaps-Hollands saam met hulle geneem het. Den Besten (1989:225) voer aan dat die twee pidginvariëteite (Khoi-Nederlands en die slawe-pidgin) sterk op mekaar ingewerk het en teen 1700 is ’n Nederlandse kreools van Afrikaans gevorm − Proto-Afrikaans I.

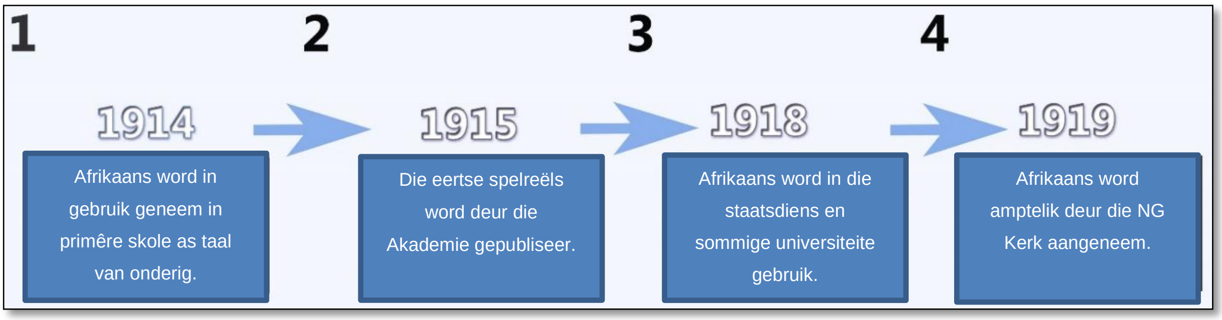
Die Kaapse immigrante was nie net van Nederland afkomstig nie. (Figuur 2.2 beeld die samestelling van Afrikaners [soos in die bron as sodanig beskryf] in 1807 uit Teen 1775 het Kaaps-Nederlands volgens Den Besten (1989:225) as Proto-Afrikaans II tot stand gekom.



Figuur 2.2 Samenstelling van Afrikaners[29] in 1807(geneem vanuit Giliomee & Mbenga, 2007:70)

[29] Volgens Giliomee en Mbenga (2007:70) is daar geen aanduideing dat daar ook na Afrikaanssprekendes van ander afkoms (slawe, bruin mense of mense van gemengde afkoms) as Afrikaners verwys is nie.

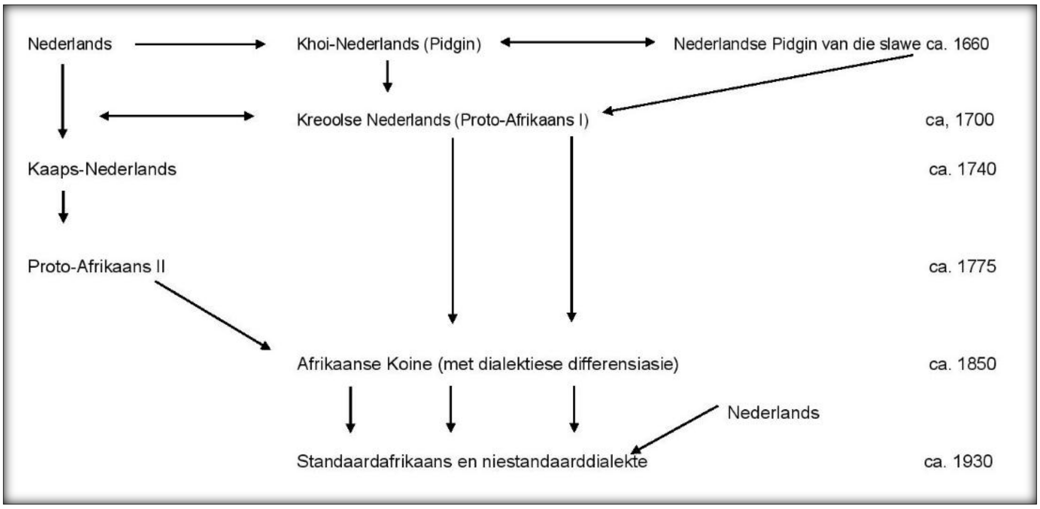
Scholtz (1958:61) en Raidt (1991:116) hou voor dat die skryftaal aansienlik van die spreektaal in dié tyd verskil het aangesien outeurs gedurende die vroeë stadium van Afrikaans probeer het om hulle skryftaal so na moontlik aan die "aanvaarbare" Nederlandse norm te hou. Raidt (1991:116) voeg by dat hierdie vroeë tekste (Kaaps-Nederlands) se taalgebruik misleidend kon wees aangesien taalgebruikers daarna gestreef het om "deftige" Nederlands in hulle skrywes te gebruik en daarom is vorme en konstruksies (wat as eg Afrikaans beskou sou kon word) liefs in geskrewe taal vermy. Den Besten (1989:225) voer aan dat Proto-Afrikaans I en Proto-Afrikaans II intussen geleidelik een geword het en Afrikaans as *lingua franca* in 1850 ontstaan het. In figuur 2.3 word die taalpolitiese veranderinge van Afrikaans gedurende die Tweede Taalbeweging voorgehou.[30]



Figuur 2.3 Taalpolitiese veranderinge gedurende die Tweede Taalbeweging (saamgestel vanuit Deumert, 2004:293)

In 1925 het Afrikaans die status as amptelike landstaal van die Unie van Suid-Afrika (vgl. Roberge, 2003:31) verwerf en is Afrikaans deurlopend uitgebrei in die gebruiksfeer wat ook daartoe gelei het dat standaardisering voortaan deurlopend plaasvind. Figuur 2.4 beeld die ontstaan van Afrikaans (volgens Den Besten, 1989:226) skematies uit:

[30] Die Eerste Taalbeweging word ingelei deur die stigting van die Genootskap van Regte Afrikaners in 1875 onderleiding van S.J du Toit, en die uitbreek van die Anglo Boereoorlog bring die beweging tot 'n einde (Deumert, 2004:55; Giliomee & Mbenga, 2007:191; Nienaber *et al.*, 1963:97-98). Die tweede Taalbeweging het volgens Roberge (2003:30) vanaf 1903 tot 1919 geduur.



Figuur 2.4 Ontwikkeling van Afrikaans (Den Besten, 1989:226)

Gedurende die 1930's het Afrikaans aangebeur en ontwikkel as 'n openbare en literêre taal en die sintaksis en woordeskat het reeds in 'n groot mate die huidige beslag gekry (Giliomee & Mbenga, 2007:289). In die Grondwet van Suid-Afrika word daar bepaal dat Afrikaans vandag tesame met tien ander tale[31] as amptelike landstale van Suid-Afrika erken word (Roberge, 2003:33).

Afrikaans het dus oorspronklik die vorm van 'n jargon aangeneem om die behoefte aan kommunikasie (die behoefte om handel te dryf) te vervul. Hesseling (1923:128) het argumenteer dat Afrikaans in die vroeë stadium tekens van kreolisering[32] getoon het en dat Afrikaans nie uit 'n enkele suiwer dialek ontstaan het nie.

Raidt (1991:145) stel onomwonde dat:

> Ons weet ten minste soveel dat ons met sekerheid kan konstateer: Afrikaans het al aan die einde van die 18de eeu 'n vorm gehad wat baie na gestaan het aan die taal wat ons vandag praat. Ons weet ook dat Afrikaans uit die 17de-eeuse Nederlands ontwikkel het; presieser gesê, dat die Europese Nederlands met baie skakerings wat dit in die 17de eeu gehad het, aan die Kaap al hoe meer "Kaaps" geword het sodat daar naderhand van "Kaaps-Hollands" in teenstelling met "Europese Hollands" (d.w.s. Nederlands) gepraat is.

[31] Die landstale sluit in Sepedi, Suid-Sotho, Tswana, Swati, Venda, Tsonga, Afrikaans, Engels, Ndebele, Xhosa en Zulu.

[32] Aangesien hierdie studie die kenmerkende eienskappe van Afrikaans op mikroblogs ondersoek het sowel as die moontlikheid of outeurskapidentifikasie van kommunikasie op mikroblogs uitvoerbaar is, word daar na die wyse waarop taal verander, verwys. Crystal (1997:362) skryf dat taalsituasies dinamies en veranderlik is en dat taalgebruikers nuwe hibriede tale skep (pidgin en kreools) om aan te pas by die veranderende vereistes wat kommunikasie aan Afrikaans stel.

Vanuit Raidt (1991:145) blyk dit dat Afrikaans deur middel van ŉ proses vanuit (hoofsaaklik) Nederlands ontwikkel het tot ŉ taal wat voldoen aan gebruikers se behoeftes. Conradie en Groenewald (2014:50) postuleer dat die standaardisering van Afrikaans aan die einde van die 19de eeu plaasgevind het. Afrikaans spog dus met 'n ryk geskiedenis en ongeag die feit of Afrikaans as 'n kreools of 'n sosiale dialekkontinuum[33] beskou kan word, het 'n Standaardafrikaans ontstaan. Die verhouding tussen die vroeë Afrikaanse spreektaal en amptelike Nederlands kan ook as diglossie[34] beskryf word. Die aanvanklike standaardiseringproses van Afrikaans het dus Afrikaans vanaf die spreektaal (die laer variëteit) na die amptelike taal (die hoër variëteit) verhef.

Vervolgens verwys ek vlugtig na variëteite in Afrikaans. Hierdie afdeling het ten doel om die basis te vorm wat voorhou hoe Afrikaans aangepas het om aan die behoeftes van gebruikers te voldoen.

2.3.2 Variëteite van Afrikaans

Vanuit afdeling 2.3.1 kan gesien word hoe Afrikaans ontwikkel het tot die taal wat ons vandag gebruik. Tog is Afrikaans nie rigied nie en bestaan daar verskillende vorme van Afrikaans. Du Plessis (1988:1-2) meld dat wanneer mense ŉ variasie in 'n bepaalde taal (of binne tale) aanvaar, aanvaar mense dat die taal nie 'n blote abstraksie is nie, maar dat die taal iets is wat mense gebruik en hierdie sosiale verskynsel − soos dit deur mense in 'n sosiale gemeenskap as ŉ praktiese kommunikasiemiddel gebruik word − verskil soos dit deur elke individu in daardie gemeenskap gebruik word. Individuele variasies verwys na die idiolek[35], terwyl Webb (1989:429) variëteit as *variëteit* as 'n versameling taalvorme wat geassosieer word met 'n bepaalde groep mense, situasies of aktiwiteite beskryf. Crystal (2004:6) verwys na taalvariasie as 'n stelsel van linguistiese uitdrukkings waarvan die gebruik deur situasionele faktore bepaal word.

ŉ Variëteit verwys na die patroonmatige stelsel van linguistiese items met ooreenstemmende sosiale en/of geografiese verspreiding en sluit eenhede groter (verskillende tale) en kleiner (register) as dialekte in (vgl. Du Plessis, 1988:10). Indien daar aanvaar word dat variasie binne elke taal bestaan en dat hierdie variasie deur die situasie, gebruikers en aktiwiteite bepaal word, ontstaan die vraag: Teen wat word hierdie variasie gemeet? Hierdie vraag word deur die

[33] Grebe (2010:21) hou voor dat 'n komplekse sosiale dialekkontinuum gekenmerk word deur stylverskuiwing en -vermenging,

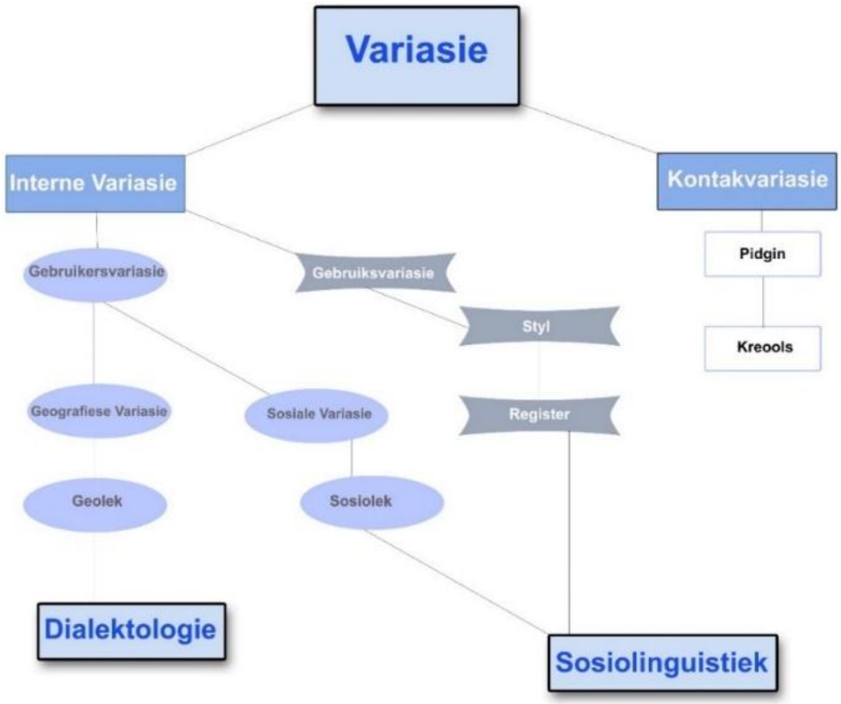
[34] "Diglossie" verwys na 'n term wat in sosiolinguistiek gebruik word om na 'n situasie te verwys waar twee verskillende variëteite van 'n taal naas mekaar bestaan (Conradie & Groenewald, 2014:51). Daar word tussen 'n hoër en 'n laer variëteit onderskei. Die hoër variëteit is die meer formele taal die amptelike taal wat in skole aangeleer word en in amptelike dokumente en kerke gebruik word terwyl die laer variëteit na 'n meer informele omgangstaal verwys (vgl. Crystal, 2008:147).

[35] Sien afdeling 2.3.2 (Variëteite van Afrikaans).

standaardvorm van taal beantwoord. Vir ’n variëteit om te kan bestaan, moet daar ’n standaard gestel word wat as die norm beskou word waarvan die variëteit kan afwyk.

Die term *standaardtaal* is problematies, want hoe kan een vorm van ’n taal bo ’n ander gestel word? Vir die doel van hierdie studie is die term *Standaardafrikaans* gedefinieer as die vorm van Afrikaans wat algemeen in die openbare domein gebruik word (vgl. Kriel, 2014:35) en as die taalvorm wat die algemeenste gebruikswaarde oor die hele Afrikaanse taalgebied het (Carstens, 2011:282). Van den Berg (2000:214) verwys na Standaardafrikaans as “’n sterk heterogene variëteit wat deur standaard-taalgebruikers in die volle registerrepertoire aangewend kan word”, terwyl Carstens (2011:280) na Standaardafrikaans verwys as die aanname dat daar een of ander vorm van taalgebruik is waarna taalgebruikers hulleself kan rig en wat as “goeie” of “korrekte” Afrikaans beskou word. Carstens (2011:281) gaan voort deur te argumenteer dat ’n taal selde homogeen is en dat dit eerder as heterogeen (ongelyksoortig) beskou kan word. ’n Taal sal dus verskeidenheid toon en geen twee mense gebruik die taal op dieselfde manier nie.

Du Plessis (1988:19) stel variasietaalkunde skematies voor (sien Figuur 2.5). Vanuit hierdie Figuur blyk dit dat daar in variasietaalkunde tussen interne en kontakvariasie onderskei kan word. Kontakvariasie verwys na die voorkoms van pidgin en kreools terwyl interne variasie na gebruiks- en gebruikersvariasie verwys. Vir die doel van my studie was interne variasie veral van belang, maar ten einde ’n volledige teoretiese agtergrond te skep waarteen die onderskeie korpusanalises uitgevoer kon word, word kontakvariasie ook vlugtig verduidelik.



Figuur 2.5 Vertakking van variasietaalkunde (geneem vanuit Du Plessis, 1995:19)

Kontakvariasie word volgens Du Plessis (1988:18) beskou as die variasie in ’n taal na gelang van kontak met ander tale. Hierdie soort variasie word vlugtig in afdeling 2.3.2.1 bespreek en die verskillende argumente oor die ontstaan van Afrikaans as ŉ pidgin en later as ŉ kreool word beskryf. Verder is daar in studies oor Afrikaans in ŉ digitale era ondersoek ingestel na Internetafrikaans[36] as ŉ variëteit van Afrikaans en daar is bevind dat dié medium ’n uitwerking op taalgebruik het en dat die Afrikaans wat op die internet gevind word tekens van beide geskrewe en gesproke Afrikaans toon.

2.3.2.1 Kontakvariasie

By kontakvariasie word daar tussen twee belangrike aspekte onderskei, naamlik pidgin en ŉ kreool. Volgens Du Plessis (1995:31) kan ŉ kreool as ŉ primêre taal beskou word aangesien dit die moedertaal van ŉ taalgemeenskap is en in sommige gevalle word ŉ pidgin ŉ kreool. In die sosiolinguistiek kan ŉ pidgin (volgens Akmajian *et al.*, 2010:294, 594; Crystal, 2008:369; Gilmore, 2011:371; Holm, 2000:5; Joseph, 1996:14; Matras, 2009:60) as ’n taal beskou word indien dit oor ’n duidelik gereduseerde grammatikale struktuur, leksikon en stilistiese eienskappe beskik en geen moedertaalsprekers het nie.

[36] Sien afdeling 2.4 (Afrikaans in ’n digitale era).

’n *Pidgin* verwys verder volgens Kriel (2014:40) na ’n tussen- of intertaal, en die taal se status word bepaal deur die geskiedkundige, politieke en sosiale stand van sake in ’n land. ’n Pidgin verwys verder na ’n term wat in sosiolinguistiek gebruik word om na ’n taal te verwys met ’n opvallend ingekorte grammatikale struktuur, leksikale en stilistiese reikwydte wanneer dit met ander tale vergelyk word. Wanneer ’n taalstruktuur só ingekort word, word dit *pidginisering* genoem (Crystal, 2008:369). Holm (2000:5) definieer ’n pidgin as ’n taal wat vereenvoudig word na gelang van kontak tussen groepe wat nie noodwendig ’n taal deel nie en vir wie kommunikasie noodsaaklik is. ’n Pidgin kan dus beskou word as ’n taal wat leksikaal vanuit ander tale afstam en wat struktureel − veral in terme van morfologie − vereenvoudig is (Kouwenberg & Singler, 2008; Siegel, 2008; Todd, 2005).

Elke taal beskik oor ’n stel reëls wat as fundamenteel tot die bepaalde taal se konvensies beskou word (Clark, 2009:117). Hierdie konvensies sluit semantiek, sintaksis, morfologie, fonologie en pragmatiek in. In ’n kontaksituasie sal sprekers gebruik maak van woorde en uitdrukkings wat bekend is aan die gehoor (ontvangers van die boodskap), dus moet daar ’n gemeenskaplike leksikon wees (Clark, 2009:117). ’n Jargon word as spesifieke gedeelde leksikale eienskappe beskou en beskik meestal oor sintaksis wat afwyk van die oorspronklike taal en daarom is dit soms moeilik vir individue wat nie deel van die bepaalde groep uitmaak nie om dié taal te verstaan (Lutz, 1992:543). Volgens Akmajian *et al.* (2010:588) ontstaan ’n jargon gewoonlik om ’n spesifieke rede, dit is linguisties onstabiel en kan as die voorloper of eenvoudiger fase van ’n pidgin beskou word.

Baker (2008:132) hou voor dat ’n jargon na ’n stel individuele maniere verwys waarop kommunikasie plaasvind. Hierdie “individuele taal” word gewoonlik gebruik wanneer daar ’n behoefte aan kommunikasie ontstaan en daar nie ’n taal bestaan wat aan al die vereistes vir hierdie spesifieke kommunikasie voldoen nie. Dit gee dan aanleiding daartoe dat ’n baie vereenvoudigde kommunikasiemedium (taal) ontstaan. Daar is ook geen norme waaraan hierdie taal moet voldoen nie − dit is geskep as ’n oplossing tot ’n kommunikasieprobleem.

’n Jargon het dus ’n vereenvoudigde taalstruktuur, is leksikaal vereenvoudig en kan as “arm” met betrekking tot die leksikale en morfologiese struktuur beskou word. Volgens Holm (2000:1) het dit onlangs duidelik geblyk dat pidgins (en jargons) nie ’n “verkeerde weergawe” van ’n taal is nie, maar eerder ’n nuwe taal wat ontstaan om ’n behoefte aan kommunikasie te vervul. Figuur 2.6 is voorstellinge van mikrobloginskrywings wat op Twitter gevind kan word. Hierdie

afbeeldings toon aan hoe die hutswoord (# − hashtag)[37] gebruik word. Ten spyte van die feit dat dié inskrywings in Afrikaans plaasvind, maak die outeurs gebruik van Engelse hutswoorde om sodoende die boodskap in konteks met wêreldgebeure te plaas en om die soekfunksie te vergemaklik.

Hierdie aspek van mikroblogs word in afdeling 2.6 bespreek, maar vir die doel van kontakvariasie is dit wel belangrik om ag te slaan op die feit dat Afrikaans nie in afsondering op mikroblogs gebruik word nie, maar dat dit wel saam met ander tale gebruik word om kommunikasie te vergemaklik en die toeganklikheid en reikwydte van die boodskap te verbeter en vergroot.

Ek's geensins homofobies maar WEET jy kan geloofsgroepe NIE hul omtrek voorskryf. Wie dink jy is jy? #DemocracyIsABitch

#UWCExams se afstel skep gevaarlike presedent - kniebuiging voor enkele oproerige studente tot nadeel v meerderheid: tinyurl.com/q3943ue

Kyk, daai Skynews aanbieders skraap vanoggend hard aan die bodem van die liberale blik. #France

#ParisAttacks: Aanslag op konsertgangers het sowat 15 minute aangehou.

Figuur 2.6 Voorbeelde van mikrobloginskrywings waar ’n Engelse hutswoord (#) gebruik word om die boodskap oor te dra en in konteks met wêreldgebeure te plaas (geneem vanaf www.twitter.com)

Die voorbeelde in Figuur 2.6 dui aan hoe kontakvariasie op mikroblogs plaasvind. Vervolgens is daar eerstens ondersoek ingestel na die kenmerkende eienskappe van dié medium waarop die taal gebruik word alvorens die idiolek op daardie medium ondersoek kon word. Die terme *interne variasie* en *idiolek* word voorts bespreek alvorens ek die konsep van Afrikaans in ’n digitale era ondersoek.

2.3.2.2 Interne variasie

Soos genoem in afdeling 2.3.2 verwys interne variasie na gebruikers- en gebruiksvariasie. Gebruikersvariasie word breedweg beskou as verskillende dialekte en gebruikers se geografiese ligging of sosiale herkoms is faktore wat ’n invloed op die dialek sal hê. Gebruiksvariasie verwys in teenstelling met gebruikersvariasie na die gebruik van ’n taal oftewel die situasionele gebruik van ’n taal (Du Plessis, 1988:20).

[37] Wanneer daar in terme van mikroblogs na ’n hutsteken (#) verwys word, word daar na die teken self verwys terwyl ’n hutswoord na die teken én die woord wat ná die # gebruik word verwys. **#moeg** is ’n voorbeeld van ’n hutswoord.

In laasgenoemde geval (gebruikersvariasie) kan sprekers dalk verskillende variëteite van ’n bepaalde taal gebruik om kommunikasie so doeltreffend moontlik te laat plaasvind. Sprekers kan dus tussen verskillende dialekte of sosiolekte wissel om die gepaste taal vir ŉ bepaalde situasie te vorm. Hier kan ook verwys word na Carstens (2011:4-7) se argument dat die sosiale konteks van taal die gepaste taalvorm vir daardie bepaalde kontaksituasie sal bepaal en dat hierdie taalvorm nie noodwendig sal ooreenstem met die sogenaamde Standaardafrikaans nie, maar dat dit eerder verwys na die gepaste taalvorm wat binne die gegewe sosiale konteks 100% kommunikasie sal bewerkstellig.

Styl en register is twee belangrike aspekte van gebruiksvariasie. Styl verwys na die keuse wat outeurs uit moontlike formulerings maak en hierdie moontlike formulerings het betrekking op sinsbou, woorde, teksbou, onderwerp, doel, leserspubliek en genre (Du Toit & Smith-Müller, 2015:15). Sinclair (1996) verwys na styl as die patrone van taal wat ooreenstem met eksterne beperkings soos formele of informele gebruike. Volgens Halliday *et al.* (2004:42) word voorgeskrewe stylriglyne oor die algemeen aanvaar om te verseker dat daar ŉ sekere mate van konsekwentheid heers ten opsigte van die formaat vir publikasies (hetsy briewe, akademiese artikels, verhandelings of nuusberigte).

Halliday *et al*. (2004:42) gaan voort deur aan te voer dat dit wat byvoorbeeld as die norm vir die styl van ŉ bepaalde koerant aanvaar word, vir ŉ ander koerant onaanvaarbaar kan wees. Dus is dit duidelik dat styl kan verskil na gelang van die publikasie of die platform. Stilistiek verwys volgens Carstens (2011:335) na gebruikers se vermoë om hulle taal by die taalsituasie aan te pas en dat styl moeilik omskryf word aangesien outeurs elk oor ’n unieke manier beskik waarop hulle hulself uitdruk. Hierdie unieke wyse waarop taalgebruikers hulleself uitdruk, verwys na idiolek en word as *idiolektiese merkers* beskou.

Register verwys daarenteen, volgens Henry en Roseberry (2001:95-96), na taalpatrone wat spesifiek met ŉ bepaalde genre verband hou en sluit woordkeuses, frases en idiome in wat eie aan die genre is. Die register van ŉ taal word volgens Hyon (1996:696) gevorm deur die veld (die aktiwiteit wat plaasvind), die toon (die verhouding tussen die deelnemers) en die manier (die kanaal van kommunikasie).

Wanneer *register* dus vir die doel van my studie definieer word, kan dit beskou word as die sosiale aard van Afrikaans soos dit op mikroblogs waargeneem kan word terwyl *styl* (as die oorkoepelende term wat ook stilistiek insluit) na individue se unieke taalkenmerke en dus ook idiolek verwys. Aangesien die internet nie deur geografiese merkers of grense gestuit word nie, is dit ook duidelik dat daar op enige gegewe tydstip ’n magdom verskillende sosiolekte en

variëteite aanwesig kan wees. Tog is dit ook belangrik om te vra in hoe ʼn mate hierdie variteite/sosiolekte sigbaar is en of daar van ʼn gemeenskaplike mikroblogtaal gebruik gemaak word wat aan die kommunikatiewe vereistes van hierdie platforms voldoen en waar daar moontlik tussen idiolekte onderskei kan word.

2.3.2.3 Idiolek

Daar bestaan verskeie definisies vir die term *idiolek*. Eerstens, moet die term *identiteit* voorgehou word aangesien identiteit nou verwant is aan die idiolektiese keuses wat individue uitoefen. Identiteit is ʼn faktor wat deur verskeie taalkundiges aangespreek word. Volgens Johnstone (2000:406-408) verwys idiolek na vrae oor die konsep van self, selfbewussyn, subjektiwiteit, individualiteit en die manier waarop sprekers hulleself projekteer. Hierdie vrae is van belang om sprekers se identiteit te bepaal en individue maak elk ʼn stel keuses om betekenis te vorm en ʼn boodskap uit te dra. Hierdie keuses lei dan tot die voorkoms van ʼn unieke idiolek. Panicheva *et al*. (2010:1134) definieer idiolek as ʼn voorstelling van individue se eie en unieke taaleienskappe op enige gegewe tydstip waar die taaleienskappe verwys na ʼn kombinasie van sosiolekte, gender-invloede en jargon.

Nguyen *et al.* (2013:439) sluit by Panicheva *et al.* (2010:1134) aan en postuleer dat individue se taalgebruik beskou kan word as ʼn uitbeelding van hulle sosiale identiteit en dat individue se identiteit gevorm word deur die sosiale groepe waaraan hulle behoort. Jessen (2010:391) beklemtoon dat die idiolek van individue as ʼn kombinasie van aspekte van verskeie streeks-, sosiale en meertalige variëteite beskou kan word, en dat sprekers op enige gegewe tydstip hulle idiolek kan aanpas deur eienskappe van ander sosiolekte aan te neem. Dit blyk dus dat ʼn idiolek nie ʼn konstante gegewe is nie, maar dat dit eerder na gelang van individue se omstandighede en behoeftes kan verander.

Die ontstaan van die woord “idiolek” kan beskryf word as ʼn samevoeging van die Griekse woorde *idios* − jou eie − en *lektos* − gekose uitdrukking of woord − en word algemeen gedefinieer as die taalgebruik kenmerkend van bepaalde individue (vgl. Kuhl, 2003:4). Volgens Kuhl (2003) het die konsep idiolek in die afgelope 50 jaar al hoe meer gewildheid verwerf, maar het dit ook die brandpunt van verskeie debatte tussen verskillende dissiplines van taalkunde geword. Bloomfield (1950:45) het in sy navorsing vlugtig na idiolek verwys deur aan te voer dat “if we observe closely enough, we should find that no two persons − or rather, perhaps, no one person at different times − spoke exactely alike” (Bloomfield, 1948:45). Bloomfield gebruik nog nie die term *idiolek* nie, hy verwys net na die voorkoms van ʼn kenmerkende individuele taal.

Die term *idiolek* word deur Bloch in 1948 behandel in die artikel “A Set of Postulates for Phonemic Analysis” en hierdie artikel word dan ook beskou as die oorsprong van navorsing rakende idiolek. Bloch (1948:7) beweer dat die totaliteit van die moontlike uitsprake van sprekers op ’n gegewe tydstip (wanneer taal gebruik word om met ’ ander sprekers te kommunikeer) idiolek genoem word. Bloch (1948:8) wys reeds in hierdie artikel daarna dat idiolek nie ’n eenvoudige konsep is nie wanneer hy aanvoer dat sprekers se idiolek soortgelyk aan ’n variëteit van ’n ander taalgemeenskap kan wees alhoewel sprekers nie deel van daardie taalgemeenskap is nie (idiolek van spreker A = soortgelyk aan dialek van groep A, maar spreker A behoort aan dialekgroep B). Bloch het in die vroeë stadium van navorsing ook nie aandag gegee aan die moontlikheid van ’n interne idiolektiese variasie nie.

Tog is Bloch se vroeë definisie van idiolek van uiterse belang, en kan hierdie definisie in werklikheid as die ontstaan van navorsing oor idiolek beskou word. Paul (1967:100) sluit by Bloch aan en stel dat ongeag of taalkunde diachronies of sinchronies benader word, moet taal beskou word as ’n voorwerp wat oor geordende heterogeniteit beskik. Reeds in 1817 het Coleridge (1817:53) beweer dat “every man’s language has, first, its *individualities*; second, the common properties of the *class* to which he belongs; and third, words and phrases of *universal* use”. Vanuit die bogenoemde stelling blyk dit dat taal en dialek as ’n groepsfenomeen beskryf kan word, en dat taal deur die verskeie groepe wat die taal gebruik, gedefinieer kan word. Só verwys dialek na die taal van ’n spesifieke groep terwyl idiolek na die taalkenmerke van individue verwys. McMenamin (2002:53) definieer idiolek dan ook tereg as ’n persoonlike dialek en argumenteer dat geen twee taalgebruikers taal op dieselfde manier gebruik of ervaar nie.

Izarry (1992:104), Jakobson (1994:104), Lancashire (1999:728), Crystal (2008:235), Kuhl (2003:4), Mollin (2009:369), Wu *et al.* (2009:127) en Panicheva *et al.* (2010:1134) stem ooreen dat *idiolek* in sy eenvoudigste vorm gedefinieer kan word as karaktereienskappe eie aan individue se taalgebruik en dat hierdie unieke eienskappe baie duideliker sigbaar is wanneer geskrewe taal ondersoek word, maar dat idiolektiese taalgebruik in beide tekstuele en ouditiewe tekste sigbaar is. ’n Idiolek verwys dus, wanneer die bogenoemde bespreking in gedagte gehou word, na die eie, unieke taalgebruik waaroor individue beskik (vgl. Coulthard & Johnson, 2010:509; Du Preez, 2008:18; Kraljic *et al.*, 2007; Kuhl, 2003:4; Lancashire, 1999:727; Wu *et al.*, 2009:127-128). Verder word ’n idiolek ook geassosieer met individue se kenmerklike taalgebruik en die manier waarop hierdie persone hulle eie, unieke spreektaal doelbewus gebruik en kan as ’n unieke linguistiese taalafdruk van bepaalde individue beskou word (vgl. Eckert, 2008:456; Izarry, 1992:104; Johnstone, 2000:417; Johnstone, 2011:660; Kuhl, 2003:4; Louwerse, 2004:2; Mollin, 2009:368-372; Newman, 2010:1; Olsson, 2008:29).

In die geval van idiolek moet daar ook na Olsson (2009:119) se argument oor die voorkoms van ʼn linguistiese vingerafdruk ondersoek ingestel word. Volgens Olsson (2008:15) is daar nie so iets soos ʼn linguistiese vingerafdruk nie. Olsson (2008:31) stel dat styl as kenmerkend eerder as uniek beskou moet word. ʼn Kenmerkende styl verwys na die spesifieke styl wat in ʼn groep tekste waargeneem kan word en is volgens Olsson (2008:31) ʼn meer effektiewe doelwit as die konsep *unieke styl* wanneer outeurs van tekste geïdentifiseer moet word.

Olsson voer verder aan dat die konsep *individuele variasie* teenoor die populêre konsep van *linguistiese vingerafdruk* gestel moet word. Individuele variasie kan in dieselfde lig beskou word as die kenmerkende styl van outeurs terwyl linguistiese vingerafdruk en unieke styl taamlik wye konsepte is. Hierdie individuele variasie waarna Olsson (2009:119) verwys, omsluit die verskillende faktore wat ʼn invloed op individue se unieke taalgebruik kan hê. Volgens Olsson (2009:119-120) is daar verskillende aspekte wat individuele variasie kan veroorsaak:

- **Modus**: Dit verwys na die moontlikheid dat die manier en vorm waardeur die taal geproduseer word ʼn invloed op individue se taalgebruik kan hê. Olsson is van mening dat tekste wat op ʼn rekenaar geskep word kan verskil van tekste wat op ʼn slimfoon geskep is en dat tekste wat geskryf word ook ander eienskappe kan toon (op papier, op ʼn swartbord en selfs wanneer die meer moderne slimborde gebruik word).

- **Woordkeuse**: Individue se leksikon (woordeskat) verskil na gelang van die kommunikasiekonteks en die bedoelde ontvangers van boodskappe.

- **Tyd**: Tydsverloop tussen kommunikasie kan dalk daartoe lei dat interne variasie in individue se taalgebruik voorkom. Woordeskat, sinlengtes en sinsamestelling kan met verloop van tyd verander; daarom is dit moontlik dat individue se taal interne variasie kan toon oor ʼn tydperk.

- **Persoonlike omstandighede**: Wanneer individue se omstandighede verander, verander persone se taalgebruik ook. Faktore wat veral ʼn invloed hierop het, is werksomstandighede, verhoudings, kinders en vriende.

- **Kulturele veranderinge**: Alhoewel kulturele veranderinge nie noodwendig daagliks waargeneem kan word nie, verander dit wel. Woorde verval in onbruik en gewoontes verander. Waar hierdie veranderinge volgens Olsson die sigbaarste is, is in SMS-boodskappe. Waar dit voorheen die gebruik was om *dinner* na *dinna* te verkort, word

dit nou nog korter gemaak en is *dina* aanvaarbaar. Hierdie verkorting sluit aan by die tendens dat kommunikasie so vinnig en kort moontlik moet geskied.

Vanuit die bogenoemde is dit moontlik om die aanname te maak dat individue se taalgebruik interne variasie kan toon as gevolg van (onder andere) die bogenoemde faktore. Die bepaalde kommunikasiesituasie sal die outeur/spreker se woordkeuse beïnvloed, dit wil sê die leksikon, leksikon, sinskonstruksie, teksbou, onderwerp, doel, teikengehoor en genre sal verander na gelang van die bepaalde taalkeuses wat uitgeoefen word. Carstens (2011:335) gaan voort deur na hierdie individuele keuses as individue se *stylkeuses* te verwys en beskryf hierdie stylkeuses ook as idiolektiese merkers.

Wanneer kwantitatiewe profiele van leksikale, grammatikale en sintaktiese kombinasies ondersoek word (funksiewoorde, ongewone gebruik van woorde, die gebruik van voorsetsels en affikse) kan individue se idiolek volgens Lancashire (1999:728) bepaal word. Hierdie kwantitatiewe profiele kan deur die gebruik van rekenaarprogramme[38] statisties uitgebeeld word. Lancashire (1999:729) voer verder aan dat die identifisering van idiolek nie beperk word tot geskrewe taal nie, maar dat idiolek op tekstuele sowel as ouditiewe taalgebruik fokus. Individue se geskrewe idiolek kan verbind word met kognitiewe prosesse – dus kan idiolek ook as die uitbeelding van ŉ bepaalde kognitiewe styl beskou word.

Afdeling 2.3 toon aan dat daar tussen verskillende variëteite in ’n taal onderskei kan word, en dat hierdie variëteite as ŉ interne of kontakvariasie beskou kan word. ŉ Kontakvariasie het te doen met die taalkontak tussen verskillende tale terwyl ŉ interne variasie na die stylverskille in ’n bepaalde taal verwys. Hierdie stylverskille is dikwels die gevolg van sosiale faktore. Vanuit hierdie stylverskille is unieke individuele kenmerke van taalgebruikers ook van uiterste belang. Idiolektiese kenmerke is uniek aan elke taalgebruiker en word as ’n unieke individuele taalkenmerk beskou. Dit is dan ook juis hierdie unieke eienskappe wat outeurskapidentifikasie moontlik maak. Wat ook duidelik blyk vanuit afdeling 2.3 is die feit dat ŉ taal konstant verander om aan te pas by die behoeftes van gebruikers. Die vraag wat vervolgens ontstaan, is wat die uitwerking van die medium (tegnologie) op ŉ taal is.

2.4 Afrikaans in ’n digitale era

Dit is onmoontlik om die uitwerking wat die ontwikkeling van tegnologie[39] op taal het te ignoreer. Crystal (2004:viii) het reeds in 2004 aangevoer dat die internet toenemend vanuit ŉ sosiale perspektief beskou moet word, en dat die rol wat taal op die internet speel sentraal in hierdie

[38] Sien hoofstuk 5 (afdeling 5.8 – Instrumente).
[39] Tegnologie verwys vir die doel van hierdie studie na die verskillende aanlyn kommunikasieplatforms.

waarneming moet staan. Hy (Crystal, 2004:viii) is verder van mening dat indien die internet as ʼn revolusie beskou word, ʼn linguistiese revolusie onafwendbaar is. Eisenstein (2013:359) stem met Crystal (2004:viii) se argument saam dat daar wel ʼn linguistiese revolusie kan plaasvind en postuleer dat namate die gewildheid van sosiale media toeneem, voorbeelde van taal waargeneem kan word wat beduidend verskil van dít wat voorheen as die aanvaarbare norm beskou is. Ook Malathi (2015:30) is van mening dat die internet ʼn uitwerking op taal kan hê wanneer taalveranderinge die gevolg van verskeie faktore kan wees, maar dat veral die kommunikasiemedium ʼn invloed op hierdie veranderinge sal hê.

Dit blyk dus sonder twyfel dat navorsers van mening is dat die internet (of tegnologiese ontwikkelinge wat die wyse waarop ons kommunikeer, verander) ʼn verandering in die taal wat ons gebruik, sal meebring. Crystal (2004:4) stel dat individue deur middel van hierdie elektroniese medium (die internet) nou kan kommunikeer op ʼn manier wat fundamenteel verskil van ander semiotiese situasies. Veranderinge vind dus plaas in die manier waarop taal gebruik word.

Peersman *et al.* (2011) skryf dat ʼn algemene karaktereienskap van aanlyn kommunikasie is dat hierdie kommunikasie betreklik kort is, en outeurs maak gebruik van ʼn niestandaardtaalvariëteit oftewel omgangsvariëteit. Ook in Afrikaans kan die invloed van tegnologie waargeneem word, en is daar reeds studies oor die aard van Afrikaans − wanneer dit as ʼn RBK beskou word − uitgevoer. Hierdie studies toon die veranderinge in Afrikaans aan sowel as die uitwerking wat tegnologie op Afrikaans het.

Verhoef (2001:1) postuleer dat die toenemende belangstelling in die internet ʼn nuwe pad vir navorsing oor taal gebaan het. Chenault (1997:1) sê die internet wentel nie net om inligting en tegnologie nie, maar dat dit veral ʼn kommunikasiemedium is. Verhoef (2001:3) is verder ook van mening dat sosiolinguistiese aannames beklemtoon word wanneer die internetomgewing as ʼn virtuele taallaboratorium beskou word en wanneer hedendaagse taalverandering ondersoek word.

Van Heerden (2000) skryf die volgende: "Die internet is vir baie mense nog slegs die smeltkroes waar vinniger-as-die-wind-tegnologie en allerlei *weird* subkulture ontmoet. Maar soos die aantal aanlyn gebruikers toeneem, gaan doodgewone mense al meer as lede van virtuele gemeenskappe aanmeld." Jansen van Vuuren (2007:14) postuleer verder dat die taalgebruik wat eie aan die virtuele karakter van die internet is die krag besit om subkulture te versterk en persone binne die kultuur te definieer. Afrikaans in ʼn digitale era is kortliks ondersoek om

sodoende ’n agtergrond te skep vir die studie oor outeurskapidentifikasie en mikroblogs in Afrikaans.

Verhoef (2001:4) is van mening dat internettaalgebruik om veel meer as net rekenaarjargon gaan, maar dat dit fokus op die invloed van tegnologie op hedendaagse taalgebruik en dat eiesoortige taalgebruik ook in hoofstroom-kommunikasie gevind kan word[40]. Eckert (1995:48) voer aan dat taalgemeenskappe se taalgebruik beskou kan word as ’n weerspieëling van daardie gemeenskap se inherente sosiale gebruike en dat hierdie taalvariasie (sosiolek, geolek, idiolek en genderlek[41]) soms slegs in ’n spesifieke sosiale sfeer gebruik word. Die internet − en die onderskeie sosialemediaplatforms − skep die geleentheid aan gebruikers om te kommunikeer in ’n taal en taalstyl waarmee hulle gemaklik is en wat voldoen aan die vereistes wat daardie spesifieke kommunikasieplatform skep.

Verhoef (2001:7) het data vanaf Afrikaanse kletskamers versamel (die taalgebruik van briewerubrieke van e-tydskrifte is ook ingesluit by die data). Verhoef het ’n versameling van 8 000 woorde saamgestel om sodoende te bepaal tot watter mate opvallende patroonmatige taalveranderinge in die variëteit manifesteer. Vanuit hierdie navorsing het Verhoef bepaal dat daar veral 15 opvallende patroonmatige taalveranderinge voorkom. Hierdie taalveranderinge (sien Tabel 2.1) toon aan dat geskrewe Internetafrikaans al hoe meer tekens van gesproke Afrikaans toon. Hierdie bevinding van Verhoef (2007) sluit egter aan by internasionale navorsing wat onderneem is rakende die aard van RBK, en hoe effektief hierdie kommunikasie is.

Verhoef (2001:11) toon verder aan dat dit van pertinente waarde is om die huidige veranderingsprosesse in Afrikaans aan die hand van die historiese data te begryp en sy verwys byvoorbeeld na die sinkopee van die intervokaliese **d** tussen ’n beklemtoonde en onbeklemtoonde neutrale vokaal in die ontwikkeling van Afrikaans wat daartoe gelei het dat die Nederlandse *zadel* in Afrikaans *saal* geword het[42].

Vanuit Verhoef (2001) se navorsing blyk dit of die sinkopee van die **d** besig is om te verander na ’n algehele assimilasieproses wat wyer strek as intervokaliese kontekste. Hierdie proses blyk duidelik vanuit die data waar daar by die aanduiding van sillabegrense sowel as woordgrense uitstoting is: *anders* − ***anners***, *onder* − ***onner***, *sonder* − ***sonner***, *gaan dit* − ***gaanit***, *al die* − ***allie***, *in die* − ***innie*** en *op die* − ***oppie***.

[40] Hier kan die voorbeeld genoem word van die hutsteken (#) wat in alledaagse gesprekke gebruik word en uitgespreek word as hashtag – amper soos om ’n ekstra gedagte by ’n sin te voeg, byvoorbeeld: “Die eksamen was moeilik #nielekker.”

[41] Sien afdeling 2.3.2 (Variëteite in Afrikaans).

[42] Ander voorbeelde sluit volgens Verhoef (2001:11) in: ader – aar, schouder – skouer en koude – koue. Verder het die intervokaliese **d** in sommige woorde na ’n **j** oorgegaan, as voorbeeld; goede – goeie, dode – dooie en blijde – blye.

Tabel 2.1 Verhoef (2001:8-11) se merkers van geskrewe Internetafrikaans

Merkers	Internetafrikaans in vergelyking met Standaardafrikaans
1. Kodewisseling – **kodewisseling wat algemeen in gesproke taal voorkom en waarvan die gebruik in geskrewe formaat opvallend is.**	1. Ek wil jou komplimenteer op jou site 2. Sal stunning lyk 3. Vat my raad chat eers
2. Fatiese kommunikasie deur gebruikmaking van tussenwerpsels – **gesproke taalsituasies word gestimuleer.**	1. Shame, dit moet hel wees om altyd so dors te wees 2. Toemaar, voor jy nou kan dink ek 'n nors ou koek…
3. Eiesoortige skryftaalkonvensie – **hoofletter en skryftekengebruik is spaarsamig.**	1. Ge-eet (Geëet) 2. Se (sê) 3. Rerig (rêrig)
4. Fonetiese skryfwyse – **word bykans sonder uitsondering gebruik.**	1. Gaanit (gaan dit) 2. Voorie bib (voor die Biblioteek) 3. Biki (bietjie)
5. Gebruik van leksikale items tipies in gesproke taal – **tipies soos gesproke taalkonstruksies.**	1. Soooo 2. Goeters 3. Ekke
4. Die vervanging van woordgrense met morfeemgrense.	1. Gaanit 2. Sallie 3. eks
7. Eiesoortige voorsetselgebruik – **word na analogie van die Engelse patroon gevorm.**	1. Hou *op* die goeie werk 2. Ek wil jou komplimenteer *op* jou site 3. Ons baklei in die huis *vir* die boeke
8. Sintaktiese vereenvoudiging: onderwepsellips.	1. ... maar die naweek moet kom 2. ... moet vir hom 'n memo stuur 3. ... kan net nie aan die gang kom vandag nie
9. Sinskonstruksies lyk soos gesproke taalgebruik – **skakelwerkwoord en voorsetsel dra by.**	1. Loop mailcheck 2. Ons werk op appels
10. Randstukke – **tipies van gesproke taalkonstruksies.**	1. Ek was 'n bietjie suinig gewees *jong* 2. Toemaar, voor jy nou kan dink ek 'n nors ou koek
11. Eenvoudige sintaktiese struktuur met gereelde voorkoms van enkelvoudige sinskonstruksies – **eenvoudiger konstruksies word bewerk.**	1. Ek glo (probeer) net aan matigheid. Maar met 'n man soos myne (dierbare ou siel wat hy is) is matig nie altyd moontlik nie. Nou probeer ek maar net my sanity behou.
12. Sintaktiese konstruksies op die Engelse patroon.	1. Hou *op* die goeie werk 2. Jors *maak* my skater
13. Relatiefkonstruksie – **SVO-volgorde soos vir 'n enkelvoudige sin word in afhanklike sin behou.**	1. ... en ek (die bietjie wat ek weet van haar) ook
14. Betreklike voornaamwoordgebruik – **word ook sterk na analogie van Engels gevorm.**	1. Vir die wie goedkoop reis 2. Hulle wie huis toe verlang
15. Selfrepresentasie van die outeur/spreker – **bevestiging van die gespreksposisie deur sintaktiese manifistasie.**	1. Ek is 'n female fan, nie regtig 'n sport(s) fan nie. Ek word genoeg deur my wederhelf gedwing om sport te kyk. Ek is gek oor jou skryfstyl en sin vir humor.

Hierdie veranderinge word volgens Verhoef (2001:14) as die tussenvlak tussen geskrewe en gesproke taal beskou. Taalveranderingspartone word eerstens in gesproke omgangstaal waargeneem alvorens dit deel van die genormeerde skryftaal word (vgl. Crystal 1997:328; Raidt, 1989:97).

In die skryftaal word daar volgens Verhoef (2001:13) normaalweg nie ruimte gelaat vir die uitdrukking van ekstralinguale aspekte soos klemplasing en intonasie nie, maar in die spreektaal word dit wel gedoen. In hierdie verband kan daar verwys word na navorsing deur Crystal (1997), Danet (1997) en Raidt (1989) waar daar gestel word dat die veronderstelling nie gemaak moet word dat een vorm (geskrewe of gesproke) van ’n taal beter as ’n ander is nie, maar dat vorme eerder komplimenterend tot mekaar beskou moet word en dat digitale kommunikasie tekens van beide geskrewe en gesproke taal toon.

Crystal (2004:26) huldig sewe verskille tussen geskrewe en gesproke taal:

- Spraak is tydsgebonde asook dinamies en kan as deel van ’n interaksie beskou word waar die deelnemers aanwesig is en sprekers ’n definitiewe geselekteerde gehoor het. Geskrewe taal daarenteen is die resultaat van ’n situasie waar outeurs (gewoonlik) nie noodwendig weet presies wie die lesers gaan wees nie en waar die senders en gehoor nie geografies naby aan mekaar is nie.
- Gesproke taal vind onmiddellik en spontaan plaas. Geskrewe taal verg beter beplanning aangesien die gehoor dalk divers kan wees en in detail geanaliseer kan word. Outeurs moet dus deeglik beplan om hulleself korrek uit te druk om te verseker dat die boodskap goed oorgedra word.
- Tydens gesproke taal kan daar van ekstralinguistiese faktore gebruik gemaak word soos gesigsuitdrukkings en gebare om bykomende betekenis oor te dra. Die senders van geskrewe taal moet hulle tot ander hulpmiddels soos grafiese elemente wend om betekenis oor te dra.
- Sekere woorde en konstruksies kan as kenmerkend van gesproke taal beskou word en gesproke taal vereis normaalweg korter en bondiger taalgebruik. Geskrewe taal daarenteen neig om meer omvattend en uitgebreid te wees.
- Gesproke taal is ’n sosiale konstruk waardeur deelnemers sosiale verhoudings uitdruk en geskied meestal in ’n informele konteks. Geskrewe taal is meer formeel van aard.
- Gesproke taal kan nie verander word nadat die sin uitgespreek is nie terwyl geskrewe taal beter beplan kan word.
- Gesproke taal beskik oor kenmerklike aspekte soos intonasie, tempo en ritme wat nie baie suksesvol neergeskryf kan word nie terwyl geskrewe taal oor sekere stilistiese kenmerke beskik wat weer nie tydens spraak weergegee kan word nie.

Die bogenoemde punte dui die verskil tussen geskrewe en gesproke taal aan. Crystal (2004:28) argumenteer dat die internet as ’n vorm van kommunikasie karaktereienskappe toon wat as kenmerkend van beide geskrewe en gesproke taal beskou kan word. Borochovsky-Bar-Aba en

Kedmi (2010:2) is van mening dat RBK as ’n hibride vorm beskou kan word of as ’n variëteit wat op ’n nuwe evolusionêre vlak lê waar gesproke en geskrewe taal geïntegreer word.

Van den Berg (2000:335) beklemtoon in haar studie die verskil tussen die geskrewe en gesproke vorm van Standaardafrikaans. Hierdie twee stelsels word op verskillende maniere gestandaardiseer en gekodeer en volgens Van den Berg boots (ingeligte) taalgebruikers onbewustelik die taalgebruik van openbare rolmodelle na en sodoende word gesproke taal informeel gestandaardiseer. Die verwagting was dus dat daar in die Afrikaanse wat op mikroblogs gebruik word ook só ’n verskynsel sou plaasvind, en dat die taal aangepas sou word om aan die vereistes van die medium te voldoen. Hier word spesifiek na die karakterbeperking verwys en daarom sou daar unieke kenmerke geïdentifiseer kon word vir die Afrikaans wat op mikroblogs gebruik word.

Verhoef (2001:17-18) stel dat wanneer Afrikaans met ander tale vergelyk word, dié taal neig om al hoe meer informeel te wees en die band tussen die variëteite van Afrikaans en gebruikers sterk na vore kom en ook dat die variasiepatrone hoofsaaklik dui op vervaging tussen die gesproke en geskrewe taal. Jansen van Vuuren (2007:16) hou voor dat taalvariëteite ontstaan vanuit ’n sosiale konteks waar ’n nuwe manier van linguistiese uitdrukkings benodig word. Ponelis (2009) se mening sluit aan by Verhoef (2001:17-18) en Jansen van Vuuren (2007:16) se stellings oor die aanvraag na nuwe maniere van linguistiese uitdrukkings in ŉ sosiale konteks, dit wil sê ŉ demotiese Afrikaans. Demotiese Afrikaans is volgens Ponelis (2009) een van twee uiteenlopende variëteite van Afrikaans – ŉ klassieke en ŉ demotiese variëteit. Die demotiese variëteit vertoon die grootste invloed van Engels en dit word volgens Ponelis (2009) as omgangstaal en spreektaal gebruik.

Die sterk aanwesigheid van Engels in demotiese Afrikaans kan as ŉ moontlike kenmerkende eienskap van Afrikaans beskou word soos dit op mikroblogs gebruik word. Soos reeds genoem in afdeling 2.3.2.1 (kontakvariasie) en uitgebeeld in Figuur 2.7, blyk dit dat daar veral voorbeelde op mikroblogs gevind kan word waar gebruikers van Engelse hutswoorde gebruik maak om aan hulle boodskap in die groter kommunikasiekonteks meer trefkrag te verleen. Hierdie gebruik van Engelse woorde word in hoofstuk 7 ondersoek asook die demotiese aard van Afrikaans op mikroblogs[43].

Jansen van Vuuren (2007:66) argumenteer dat elke internetsituasie[44] oor ’n unieke karakter en spesifieke vereistes vir kommunikasie beskik en bepaal dat die anonieme aard van hierdie

[43] Sien afdeling 7.6 (Kodewisseling).
[44] Jansen van Vuuren (2007:65) definieer die verskillende internetsituasies as; e-pos, kletskamers en webtuistes.

platforms bydra tot die voorkoms van inhibisieloosheid in die taalgebruik. In Jansen van Vuuren (2005) se korpuslinguistiese studie: is daar na tien verskillende aspekte ondersoek ingestel (sien Tabel 2.2) en daar is bevind dat Internetafrikaans verskil na gelang van die medium en die aard (sinchronies teenoor asinchronies[45]).

Jansen van Vuuren (2007:303) het bevind dat die Afrikaans wat gebruik word in e-posse, kletskamers en e-tydskrifte nie voldoen aan die konvensies van standaard geskrewe Afrikaans nie. Hierdie bevindinge van Jansen van Vuuren stem ooreen met bevindinge van internasionale navorsers soos Baron (2001), Crystal (2004), Danet *(*1997), Herring (2002) en Werry (1996). Die ontwikkeling van Afrikaans toon sterk ooreenkomste met die veranderinge en ontwikkeling van ander tale ten einde aan die vereistes van rekenaarbemiddelde kommunikasie te voldoen (sien afdeling 2.5.2.1 – Rekenaarbemiddelde kommunikasie (RBK)).

In die onderstaande Tabel (Tabel 2.2) verskaf ek die aspekte wat Jansen van Vuuren in haar studie ondersoek het om Internetafrikaans te analiseer. Hierdie aspekte kan saam met Verhoef (2001) se merkers van geskrewe Internetafrikaans as vertrekpunt dien om Afrikaans te ondersoek soos dit op mikroblogs gebruik word.

Tabel 2.2 Aspekte van Internetafrikaans soos ondersoek deur Jansen van Vuuren (2007:155-287)

Taalaspekte
Leksikale kompleksheid
Lengte van grammatikale eenhede
Grammatikale ondersoek (naamwoorde, lidwoorde en voornaamwoorde)
Verbindingswoorde
Leksikale keuse (kragwoorde, kodewisseling en fatiese kommunikasie)
Eiesoortige skryftaalkonvensies (skryftekens en leestekens)
Spel- en skryftaalkonvensies (eiesoortige spelkonvensies, leksikale verkorting, akronieme, emotikons, fonetiese skryfwyse, geleentheidsvorming en spieëlskrif)
Verlede tyd
Adjunkte (plekadjunkte, tydsadjunkte en voorsetsels)
Sintaktiese strukture (reduksie)

Tog kom Jansen van Vuuren (2007:307-308) tot die uiteindelike gevolgtrekking dat daar nie ’n enkele variëteit vir Internetafrikaans bestaan nie aangesien die variasie verskil na gelang van die internetsituasie − hierdie taal is iets heeltemal nuut. Dis nie ’n gesproke taal nie en ook nie ’n geskrewe taal nie en vervolgens was haar aanbeveling dat Internetafrikaans as ’n vierde

[45] Daar word in afdeling 2.6 (Kommunikasie en taal) na die sinchroniese en asinchroniese aard van kommunikasie verwys.

medium geklassifiseer moet word. Volgens Jansen van Vuuren (2007:308) is die implikasie van ’n vierde medium dat Internetafrikaans as linguistiese medium sal groei ten opsigte van sosiolinguistiese en stilistiese kompleksiteit.

Jansen van Vuuren (2007:303) voer aan dat een van die belangrikste kumulatiewe bevindinge ten opsigte van internetnavorsing is dat rekenaarbemiddelde kommunikasie varieer na aanleiding van die tegnologie waarop dit gebaseer is en ook na aanleiding van die gebruikerskonteks daarvan.

Crystal (2011:21) skryf dat “on the whole Internet language is better seen as writing which has been pulled some way in the direction of speech …”, maar verder maak hy dit ook duidelik dat die internettaal selektiewe en aanpasbare tekens toon in terme van eienskappe van beide die geskrewe en gesproke taal, maar ook dat daar tekens is wat nog nooit in enige van die twee mediums (geskrewe of gesproke) waargeneem is nie.

Namate Afrikaans ontwikkel, is daar deurlopend veranderinge wat in die taal plaasvind. Sprekers gebruik verskillende variëteite van Afrikaans afhangend van die bepaalde taalsituasie waarin hulle hulself bevind en outeurs gebruik die taal om by die karakterbeperking aan te pas. Wat is dus kenmerkende eienskappe van die taalgebruik van Afrikaans soos dit op mikroblogs gebruik word? Sodra hierdie kenmerke geïdentifiseer is, kan outeurskapidentifikasie op hierdie anonieme[46] platform plaasvind.

Trudgill (1992:4) stel onomwonde dat daar ’n onlosmaaklike band tussen ’n spraakgemeenskap en die taal bestaan en dat hierdie band op beide leksikale en semantiese vlak neerslag vind. Verhoef (2001:2) sluit by Trudgill aan deur te stel dat die behoeftes, vooroordele en lewenswerklikheid van ’n taalgemeenskap in die taalgebruik weerspieël word en daarom is dit belangrik om die ontwikkeling van ’n eiesoortige Afrikaans in die moderne taalgebruik te ondersoek.

Wanneer die bogenoemde stelling van Trudgill en Verhoef in ag geneem word (sowel as die navorsing oor Internetafrikaans deur Verhoef en Jansen van Vuuren en die anonieme aard van die internet asook die ontwikkelinge in forensiese linguistiek), is dit duidelik dat daar unieke eienskappe van die Afrikaans op mikroblogs geïdentifiseer moet word. Ondersoek moet verder ook ingestel word na die moontlikheid van outeurskapidentifikasie op hierdie platforms.

[46] Sien afdeling 2.5 (Kommunikasie en taal) en 2.6 (Sosiale media).

Soos gesien in afdeling 2.2 is daar (volgens sosiolinguiste soos Labov, Weinreich en Herzog) ’n interafhanklike band tussen taalkundige aspekte en sosiale faktore. Die internet kan as ’n reuse aanlyn gemeenskap beskou word en in afdeling 2.3.2 (Variëteite van Afrikaans) word daar verwys na die voorkoms van dialekte – gedeelde taaleienskappe van individue in dieselfde gemeenskap. Crystal (2004:6) vra tereg: “Wat is die dialek van die internet, hoe lyk hierdie gedeelde taalkenmerke van hierdie kommunikasiemedium?”

Alvorens daar na kenmerkende eienskappe van die betrokke medium ondersoek ingestel kon word, verwys ek eers na kommunikasie aangesien die aard van kommunikasie sowel as die medium ’n invloed het op die tipe taal wat gebruik word.

2.5 Kommunikasie en taal

In die vorige afdelings is taal bespreek en ook hoe taal gebruik kan word om mee te kommunikeer. Hoe verskil kommunikasie van taal en hoe beïnvloed verskillende kommunikasiemediums die taal wat op die onderskeie mediums gebruik word? In hierdie afdeling ondersoek ek die maniere waarop taal deur kommunikasie beïnvloed kan word.

2.5.1 Kommunikasie

Kommunikasie kan in die eenvoudigste vorm beskou word as die weergee van inligting tussen individue of groepe. ’n Kommunikasiemodel (sien Figuur 2.7) kan beskou word as ’n scenario waarin individue (sprekers) boodskappe aan ander individue/groepe (gehoor) stuur. Die kommunikasieproses begin wanneer individue (sprekers) inligting (die boodskap) deur middel van ’n spesifieke tekenstelsel aan ander (ontvangers) wil oordra met behulp van ’n spesifieke kanaal. Vir ’n boodskap om oorgedra te word, moet dit omgeskakel word in ’n vorm wat deur die gekose kanaal oorgedra kan word. Hierdie vorm verwys na tekenstelsels en behels ’n proses van kodering en dekodering waartydens die boodskap omgeskakel word in ’n gepaste vorm (geskrewe of gesproke) waarna dit gedekodeer word en die betekenis ontleed word.

Daar word onderskei tussen verskillende kommunikasiekanale. Die medium en kanaal verander die manier waarop kommunikasie plaasvind. So kan kommunikasie verbaal (geskrewe of gesproke) of nieverbaal plaasvind. Verbale kommunikasie vind deur middel van linguistiese tekenstelsels plaas.

Figuur 2.7 Vereenvoudigde voorstelling van Jakobson se eenvoudige kommunikasieproses (Smuts, 1998:11)

Die linguistiese teken kan volgens Cruse (2000:10) na so min as ŉ enkele foneem[47] verwys. Kommunikasie is intrinsiek ŉ menslike handeling. Goeie kommunikasie vind slegs plaas wanneer beide senders en die bedoelde gehoor die boodskap verstaan − sprekers se ware bedoeling moet dus aan die gehoor oorgedra word. Mey (2009b:787) argumenteer dat goeie kommunikasie nie noodwendig gepaard gaan met die navolging van taalreëls en grammatika nie. Trouens, Mey (2009b:787) gaan so ver as om te argumenteer dat: "Pronouncing a correct sentence belongs in the realm of the impossible, not to say pure fantasy" en dat die wyse waarop "ware individue" taal gebruik grootliks afwyk van dít wat puriste en taalkenners voorhou – solank die betekenis van die kommunikasie nie verlore gaan nie.

Soms gebeur dit inderdaad dat kommunikasie onsuksesvol is en dat betekenis verlore gaan. Mey (2009b:787) verwys in só ŉ geval na die gedeelde betekenis van tekens − wanneer gesprekvoerders (sprekers) en aangesprokenes (ontvangers) verskillende betekenisse aan ŉ bepaalde woord (of konsep) koppel. Soortgelyk aan Mey (2009b) het Austin (1962:148) aangevoer dat die konteks van uiterse belang is aangesien dit aanleiding daartoe gee dat die totaliteit van spraak (taal) in die groter situasie bestudeer kan word en dit is waaroor die pragmatiek handel − die studie van taal in konteks van die gegewe kommunikasie situasie.

Hymes (1971:5) verwys na die kommunikatiewe bevoegdheid van sprekers en onderskei tussen vier belangrike kenmerke vir optimale kommunikasie:

- Die vermoë om taal te gebruik, berus op die mate waartoe sprekers die gepaste taal in ŉ betrokke konteks kan gebruik.
- Die vermoë om betekenis aan taal te koppel en om die taal te kan gebruik, is nie uitsluitlik afhanklik van grammatikale kennis nie.
- Die gepaste taal vir ŉ gegewe situasie kan meer as een kanaal insluit.
- Gepaste taal word aangeleer deur middel van sosiale interaksie.

[47] ŉ Foneem verwys na die kleinste betekenisonderskeidende (kontrastiewe) eenheid in die klanksisteem van ŉ taal (vgl. Wissing, 2014:96).

Hierdie beskouing van Hymes dui daarop dat kommunikasie nie altyd linguisties korrek hoef te wees om betekenis oor te dra nie, maar dat kennis oor toepaslike kommunikasie in 'n gegewe konteks belangrik is. Sprekers moet die kommuniekasiemedium verstaan om sodoende keuses te kan maak oor die gepaste taal in 'n gegewe situasie.

Kommunikasie kan op verskeie maniere geskied en kan verbaal of nieverbaal van aard wees. Soos reeds in afdeling 2.5 genoem, het navorsing reeds bewys dat Afrikaans verskil na gelang van die kommunikasiemedium wat gebruik word. Tradisioneel word daar onderskei tussen geskrewe en gesproke kommunikasiemediums. Hierdie mediums word ook fisieke en meganiese media genoem. Fisieke media verwys na die kanale (mediums) waartydens senders gesien en/of gehoor kan word deur die gehoor (gesprekke, vergaderings, videokonferensies, radio en televisie). Meganiese media verwys na geskrewe kommunikasie (koerante, tydskrifte, nuusbriewe, e-posse, SMS'e).

In 'n era waar tegnologie hoogty vier, het die internet 'n kommunikasiemedium geword wat veral die aandag van navorsers geniet − spesifiek in terme van die taalveranderinge wat op interaktiewe mediums plaasvind. Tannen en Trester (2013:ix) wys daarop dat ons lewens in en deur elektroniese media geleef word, dat ons afhanklik is van die internet vir nuus, boeke word op elektroniese mediums gelees (Kindle), Facebook is 'n manier waarop vriende opgespoor en kontak behou kan word (daar is verskeie netwerke beskikbaar waardeur moontlike sielsgenote opgespoor kan word), LinkedIn word gebruik om professionele netwerke op te bou, Instagram word gebruik om foto's te deel en Twitter is 'n manier waarop nuus (of inligting wat vir gebruikers nuuswaardig is) daagliks gedeel word. Bogenoemde is kommunikasieplatforms op die internet wat die manier waarop ons daagliks kommunikeer − en dus ook taalgebruik − verander.

Palfrey en Gasser (2008:2) voer aan dat sosiale interaksie, aktiwiteite en selfs vriendskappe deur die sogenaamde digitale tegnologie beïnvloed en gefasiliteer word. Wat is hierdie sogenaamde digitale tegnologie, en hoe verander dit die manier waarop kommunikasie plaasvind? In die hieropvolgende afdeling word kommunikasie en die internet bespreek, ten einde te stel hoe die internet kommunikasie beïnvloed.

2.5.2 Kommunikasie en die internet

Tegnologie − meer spesifiek die internet − het 'n onlosmaaklike deel van ons daaglikse lewenswyse geword. Lilley (2002:vi) stel onomwonde dat ons in 'n digitale era leef (waar *digitaal* na enige aanlyn stelsel verwys), en dat daar talle risiko's aan hierdie digitale, tegnologieryke omgewing verbonde is. Lilley (2002:vii) gaan voort deur te skryf dat die internet ons tegelyk

bevry en gevange neem. Alhoewel daar al hoe meer data en inligting beskikbaar is, is pornografie, haatwebwerwe, terroristebewegings ook vrylik beskikbaar en toeganklik mits jy weet hoe om daarvoor te soek. Deur die kliek van ’n enkele knoppie kan jy met enige persoon op enige plek ter wêreld kommunikeer.

Holmes (2005:2) voer aan dat dit veral die eenvoud van die tegnologiese ontwikkelinge is wat hierdie ontwikkelinge van uiterse belang vir kommunikasie maak. Die gebruikerskoppelvlak van hierdie ontwikkelinge is eenvoudig, toeganklik en verstaanbaar vir gebruikers (veral vir gebruikers wat onbewustelik deel van die inligting-gemeenskap geword het) en van só ’n aard dat gebruikers nie besef dat hierdie koppelvlakke (platforms) deel van hulle daaglikse bestaan geword het nie (vgl. Holmes, 2005:2).

Hierdie tegnologiese ontwikkelinge het tot gevolg dat daar gemeenskappe ontstaan waarin die daaglikse lewe aanlyn bespreek word. Aanlyn besprekings geskied deur middel van ’n proses wat Holmes (2005:3) “moderne kommunikasie” noem. Kommunikasie vind nie in ’n vakuum, of homogene konteks plaas nie. Inteendeel, kommunikasie word deur sosiale interaksie en die vereistes van ’n betrokke platform bepaal.

Gedurende die laat 1970’s het die wêreld van kommunikasie vinnig begin verander. Aanlyn bulletinstelsels is ontwikkel en diegene wat toegang gehad het tot bonkige rekenaars en ’n telefoonverbinding kon dokumente uitruil, nuus lees en boodskappe aan mekaar stuur (Palfrey & Gasser, 2008:2). In die 1980’s het die stuur en ontvang van e-posse ’n gewilde kommunikasiemedium geword, en in 1991 het die wêreldwye web (www) gedebuteer met gebruikersvriendelike soekenjins (Palfrey & Gasser, 2008:3).

Palfrey en Gasser (2008:3) stel verder dat sosiale netwerke teen die 2000’s hulle verskyning gemaak het en dat gebruikers veral begin blog[48] het. Die internet het dus ’n ruimte geskep vir kommunikasie op ’n manier wat voorheen onbekend was. Holmes (2005:10) verdeel die ontwikkeling van die internet as ’n medium van kommunikasie in twee fases (sien Tabel 2.3): die eerste media-era (uitsaai) en die tweede media-era (interaktief).

Tabel 2.3 Verskil tussen die eerste en tweede media-eras (vertaal vanuit Holmes, 2005:10)

Eerste media-era	**Tweede media-era**
• Gesentreerd – enkele individue kommunikeer met ’n groep	• Niegesentreerd – baie individue kommunikeer met baie individue

[48] ’n Blog verwys na ’n inskrywing op die internet waar die gebruiker (blogger) idees en opinies met die lesers deel – hierdie kommunikasiemedium word beskou as ’n persoonlike aanlyn dagboek.

Eerste media-era	Tweede media-era
• Eenrigtingkommunikasie	• Tweerigtingkommunikasie
• Beheerbaar	• Onbeheerbaar
• Instrument van ongelykheid sowel as stratifikasie	• Demokraties van aard – fasiliteer ’n gevoel van universele lidmaatskap
• Deelnemers is gefragmenteerd, en word as ’n massa beskou	• Deelnemers behou individualiteit

Die manier waarop die internet as ’n medium van kommunikasie verander, beïnvloed die manier waarop ons kommunikeer. Vanuit Holmes (2005:10) is dit duidelik dat die media-era aanhoudend verander. Hierdie veranderinge maak dit moontlik vir individue om al hoe makliker te kommunikeer, selfs met ’n gehoor wat in die verlede nie noodwendig bereik kon word nie. Die internet skep ’n globale netwerk, ’n manier om inligting uit te ruil en in te samel, ’n manier om mense te bereik en diskoers te voer wat nie altyd moontlik was nie. Kommunikasie word nie meer deur afstand of tyd geïnhibeer nie. Kommunikasie is vinnig en moontlik deur die druk van ’n enkele knoppie.

2.5.2.1 Rekenaarbemiddelde kommunikasie (RBK)

RBK kan as die integrasie van rekenaartegnologie met ons daaglikse lewens beskou word. Wood en Smith (2005:4) voer aan dat RBK slegs fokus op die rekenaartegnologie wat individue (of groepe) in staat stel om met mekaar in verbinding te tree. Die internet is die kanaal wat hierdie soort kommunikasie moontlik maak en die manier waarop mense met mekaar kommunikeer beïnvloed en die daaglikse bestaanswyse van mense verander.

RBK kan dus gedefinieer word as die studie van die manier waarop menslike gedrag onderhou of verander word as gevolg van die uitruiling van inligting deur middel van masjiene − masjiene verwys hier na toestelle/toerusting wat ’n internetverbinding toelaat sodat kommunikasie kan plaasvind (vgl. Wood & Smith, 2005:5).

Holmes (2009:161) sluit hierby aan deur te stel dat RBK − in die eenvoudigste definisie moontlik − na enige kommunikasie verwys wat deur digitale tegnologie bemiddel word en selfs ’n telefoongesprek of ’n teks wat vanaf ’n rekenaar (of elektroniese toestel) gedruk word op papier kan dus as RBK beskou word.

Bubaš (2002) beskou RBK as ’n interdissiplinêre vakgebied wat ondersoek instel na die verskillende fenomene wat ontstaan deur die gebruik van die internet as ’n menslike kommunikasiemedium. Herring (2001:54) verwys na *computer mediated discourse*

(rekenaarbemiddelde diskoers – RBD) eerder as RBK en beskou die term as ’n spesialisasieveld binne die navorsingsgebied van rekenaarbemiddelde kommunikasie waar kommunikasie plaasvind deur middel van boodskappe wat deur middel van ’n rekenaarnetwerk tussen individue rondgestuur word.

RBK kan eenvoudig as kommunikasie wat met behulp van ’n rekenaar geskied beskou word (Jansen van Vuuren, 2007:29). Tog is bogenoemde definisies nie noodwendig die doeltreffendste nie aangesien slimfone[49] en tablette ook toegang bied tot platforms wat tradisioneel slegs deur middel van ’n rekenaar bereik sou kon word. Crystal (2011:1) argumenteer in dié verband dat die term RBK − wat gedurende die 1990’s inslag gevind het − linguisties as problematies beskou kan word. RBK is as ’n term té wyd en aangesien dit verwys na kommunikasie wat moontlik gemaak word deur middel van ’n rekenaar, is daar nie ruimte gelaat vir ander toestelle wat toegang tot die internet verskaf nie, en wat dus ook kommunikasie (wat as RKB beskou word) kan bewerkstellig.

Crystal (2011:2) voer dus aan dat RBK nie noodwendig die gewenste term is nie, maar dat ander terme soos *digital mediated communication* (DMC) − digitaalbemiddelde kommunikasie − of *electronic mediated communication* (EMC) − elektroniesbemiddelde kommunikasie – eerder oorweeg moet word. Self hierdie twee terme kan volgens Crystal (2011:2) steeds verwarring veroorsaak; daarom stel hy dat ’n eenvoudige term − die taal van die internet (*netspeak*) − vir hom voldoende is.

RBK word vir die doel van my studie gedefinieer as kommunikasie wat moontlik gemaak word deur middel van ’n toestel (hetsy rekenaar, tablet of slimfoon) wat gebruikers in staat stel om toegang tot die internet te verkry om sodoende van mikroblogplatforms gebruik te maak.

Kommunikasie kan óf onmiddellike kommunikasie (sinchroniese kommunikasie) wees óf bemiddelde (asinchroniese kommunikasie) kommunikasie − aktiewe of passiewe kommunikasie. Onmiddellike kommunikasie verwys volgens Wood en Smith (2005:6) asook Holmes (2005:15) na die proses waartydens ’n boodskap min of meer direk oorgedra word sonder die hulp van tegnologie. Bemiddelde kommunikasie verwys na die proses waartydens kommunikasie afhanklik is van tegnologie − dit kan na papier verwys (’n geskrewe boodskap op papier), ’n telefoon of e-pos − en is dus RBK. ’n Voorbeeld van onmiddellike kommunikasie is wanneer persoon A ’n vraag aan persoon B rig terwyl hulle in dieselfde vertrek is. Wanneer

[49] Slimfone word gedefinieer as ’n selfoon wat terselfdertyd dien as ’n klein, ligte rekenaar waarop inligting bewaar kan word, waarmee jy e-posse kan stuur en ontvang, waarmee jy foto’s kan neem, en wat ook toegang bied tot die internet (vgl. HAT, 2015:1182).

persoon A die vraag aan persoon B e-pos (al is die terugvoer per e-pos onmiddellik), is dit ’n voorbeeld van bemiddelde kommunikasie.

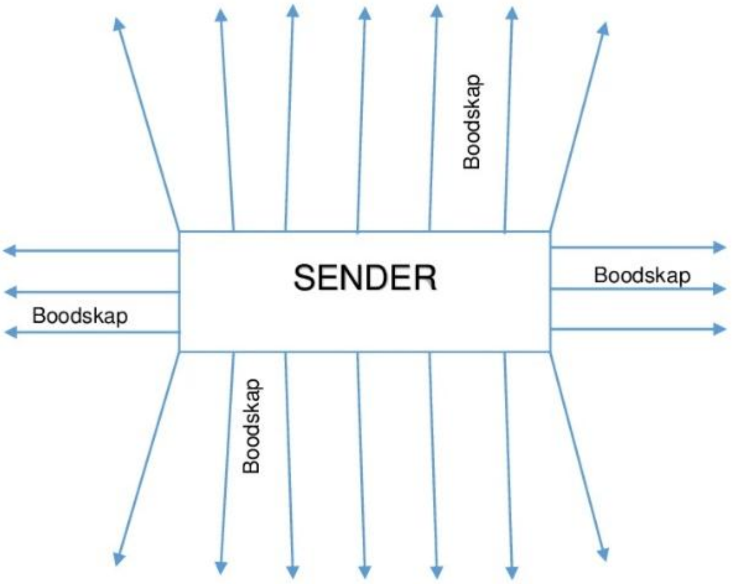
Holmes (2009:162) argumenteer dat RBK dikwels as ’n sinoniem vir internetkommunikasie gebruik word en daarom word e-posse, kletskamers, aanlyn bulletinborde en gesimuleerde wêrelde as RBK beskou, maar die onderskeid moet duidelik getref word tussen vermaak, inligting en kommunikasie. Web 1.0 se primêre gebruik was byvoorbeeld om inligting in te win en word daarom nie as ’n voorbeeld van RBK beskou nie. Web 2.0 daarenteen bemiddel kommunikasie hetsy tussen individue of groepe deur die uitruil van teks, grafika en/of klank (vgl. Holmes, 2009:162). Web 2.0 en die verdere ontwikkelinge van dié web (Web 3.0 en Web 4.0 word nou gebruik, sien afdeling 2.6.1) wel voorbeelde van RBK.

RBK kan sinchronies of asinchronies van aard wees. RBK kan ook as eenrigting-kommunikasie, een-tot-een-kommunikasie of een-tot-talle-kommunikasie geskied. Die onderskeid tussen sinchroniese en asinchroniese kommunikasie word as kontroversieel beskou aangesien sommige RBK-platforms tekens van beide sinchroniese, asinchroniese, een-tot-een en een-tot-talle kommunikasie toon (vgl. Kitade, 2001:4974).

Figuur 2.8 Eie voorstelling van een-tot-een kommunikasie

Figuur 2.8 beeld een-tot-een kommunikasie uit. Wanneer hierdie soort kommunikasie deur middel van ’n rekenaar bemiddel word, verwys dit na privaat boodskappe, e-posse en gesprekke waar daar slegs een bedoelde ontvanger is. RBK (dus ook sosiale netwerke, sien afdeling 2.6) stel ’n enkele gebruiker ook in staat om met etlike individue op ’n gegewe tydstip te kommunikeer.

Figuur 2.9 is ’n voorstelling van kommunikasie waar een gebruiker met talle individue kan kommunikeer. In hierdie figuur is elk van die pyltjies ’n voorstelling van moontlike ontvangers van die boodskap. Met die ontwikkeling van tegnologie het die manier waarop RBK geskied ook verander en afhangende van privaatheidstellings (sien afdeling 2.6) is dit moontlik dat ’n enkele gebruiker se boodskap duisende mense kan bereik. In die geval van RBK kan gebruikers (senders) ook kies om dieselfde boodskap op verskillende platforms te plaas om sodoende die reikwydte van die boodskap te vergroot (ek verwys in afdeling 2.7 verder daarna).



Figuur 2.9 Eie voorstelling van een-tot-talle-kommunikasie

Volgens Holmes (2009:162) is kletskamers soos *Internet Relay Chat* en multigebruikerkerkers (*multi-user dungeons* of *MUDs*) voorbeelde van sinchroniese kommunikasie aangesien dit in die werklike tyd plaasvind. Die meerderheid van RBK is asinchronies. Hierdie asinchroniese kommunikasie (soos e-posse en om aanlyn op sekere sosiale netwerke te verkeer) kan ook as gestruktureerde gesprekke gestoor word (vgl. Holmes, 2009:162).

Asinchroniese kommunikasie verwys na aanlyn kommunikasie waar die tydsverloop tussen kommunikasie lank kan wees. Jansen van Vuuren (2007:33) voer aan dat deelnemers aan asinchroniese kommunikasie nie almal gelyktydig aanlyn hoef te wees vir kommunikasie om plaas te vind nie. E-posse word as een van die eerste vorme van asinchroniese kommunikasie beskou en volgens Jansen van Vuuren (2007:33) toon hierdie soort kommunikasie ooreenkomste met tradisionele boodskapwisseling waar twee persone beurtelings aan mekaar boodskappe stuur − in dié geval geskied dit met behulp van ’n elektroniese medium.

Barasa (2010:31) argumenteer dat gebruikers nie tydens asinchroniese RBK soos e-posse en sosialenetwerkplatforms aanlyn (aanwesig) hoef te wees vir die kommunikasie om plaas te vind nie. Crystal (2004:11) verwys na dié soort kommunikasie as kommunikasie wat met vertragings kan plaasvind. Gebruikers stuur boodskappe, maar ontvangers hoef nie onmiddellik te reageer, te lees of aanlyn te wees vir die kommunikasie om suksesvol te wees nie.

Volgens Crystal (2004:12) verwys asinchroniese gespreksituasies na interaksies wat in ’n bepaalde formaat gestoor word. Dit stel gebruikers in staat om die gesprek (kommunikasie) te volg of daartoe by te dra wanneer hulle tyd het. Hierdie soort kommunikasie vereis dus nie van die deelnemers om aktief (aanlyn) te wees wanneer gesprekke plaasvind nie.

Sinchroniese kommunikasie verwys na kommunikasie wat in die werklike tyd geskied. Alle deelnemers moet aanlyn (in die werklike tyd) wees in die geval van RBK (vgl. Barasa, 2010:31). Voorbeelde van sinchroniese kommunikasie sluit *IM* (*instant messaging*), kletskamers en gespreksforums in. Jansen van Vuuren (2007:33) verwys ook na sinchroniese kommunikasie as aanlyn kommunikasie wat in die werklike tyd plaasvind en waar diskoersbeurte mekaar direk opvolg. Die gespreksdeelnemers moet dus aanlyn wees vir kommunikasie om plaas te vind. Crystal (2004:11) is van mening − wat sinchroniese kommunikasie betref − dat interaksie in die werklike tyd plaasvind en dat alle deelnemers aanlyn en betrokke moet wees vir kommunikasie om suksesvol te wees. RBK kan vanuit die bogenoemde definisies as beide sinchronies en asinchronies beskou word.

’n Sinchroniese kommunikasie situasie kan soos volg plaasvind: gebruikers skakel in (gaan aanlyn) by ’n gespreksforum en word deel (en neem aktief deel) van die gesprek (wat in die werklike tyd plaasvind). Die bydraes van die verskillende deelnemers (elke deelnemer het ’n skermnaam) word uitgebeeld op ’n skerm waar die nuutste bydrae onder aan die skerm verskyn (vgl. Crystal, 2004:11).

’n Belangrike aspek van RBK is die mate waarin gebruikers aangesig-tot-aangesig kommunikasie probeer naboots. Gebruikers probeer om op verskeie kreatiewe maniere nieverbale kommunikasie − wat bykomende betekenis aan ’n boodskap verleen – in RBK te inkorporeer (vgl. Holmes, 2009:163). Ellison en boyd (2013:151) voer aan dat die globale stelsel van rekenaarnetwerke (roeteerders en netwerkbedieners) − beter bekend as die internet − talle aspekte van die moderne samelewing en sosiale interaksie verander het.

Kitade (2001:4974) voer ook aan dat sosialenetwerkplatforms (moontlik gemaak deur RBK) as voorbeelde van multimediakommunikasieplatforms beskou kan word aangesien dit die tradisionele onderskeid tussen sinchroniese en asinchroniese, een-tot-een en een-tot-talle kommunikasie verbreek. Bubaš (2002) beweer dat RBK nie net ’n positiewe ontwikkeling is nie en onderskei tussen vier uitdagings, naamlik:

- ’n beperkte sosiale aanwesigheid;
- anonimiteit − deelnemers se identiteit is nie duidelik sigbaar nie;
- vertraagde terugvoer; en

- onpersoonlike kommunikasie.

Hierdie uitdagings gee aanleiding tot die moontlikheid van misdrywe wanneer RBK gebruik word, en word weer aangespreek in afdeling 2.6. Vanuit hierdie afdeling blyk dit egter duidelik dat daar verskillende vorme van RBK bestaan en dat die ontwikkeling van die web vanaf generasie een tot generasie vier talle veranderinge in die manier waarop kommunikasie plaasvind sowel as die aard van ons daaglike kommunikasie meegebring het. Vervolgens word sosiale media bespreek om ’n geheelbeeld te skep van die platform (mikroblogs) wat in my eksploratiewe forensiese analise van outeurskapidentifikasie in ’n digitale era gebruik is.

2.6 Sosiale media

Vanuit die bespreking oor die internet en kommunikasie en die definisie van RBK sowel as sinchroniese en asinchroniese kommunikasie, is dit duidelik dat die manier waarop kommunikasie plaasvind, verander. Die medium waardeur kommunikasie plaasvind, verander ook die manier waarop taal gebruik word om die kommunikasie suksesvol te laat geskied.

Sosiale media word beskou as internetgebaseerde toepassings wat die ideologiese en tegnologiese beginsels van Web 2.0 volg en wat aanleiding gee tot die skep en uitruiling van gebruikergegenereerde inhoud (vgl. Kaplan & Haenlein, 2010:61). Kietzmann *et al.* (2011:241) stel dat sosiale media van mobiele en webgebaseerde tegnologie gebruik maak om interaktiewe platforms te skep waarop individue kan kommunikeer, idees deel en gebruikergegenereerde inligting wysig. Sosiale media kan dus eenvoudig beskou word as internetgebaseerde platforms wat gebruikers in staat stel om te kommunikeer. Sosialemediaplatforms het ’n invloed op die manier waarop kommunikasie verander. In die volgende afdeling bespreek ek Web 3.0 (die mobiele web) asook sosialenetwerkplatforms ten einde die aard van mikroblogs voor te hou.

2.6.1 Web 3.0: die mobiele web

Web 3.0 word as die semantiese web[50] beskou waar digitale inligting verstaan, geïnterpreteer en gebruik kan word deur “intelligente sagteware” wat gebruikers in staat stel om groot hoeveelhede aanlyn inligting te verwerk. Palfrey en Gasser (2008:202) is van mening dat Web 3.0 op ŉ eenvoudige manier gedefinieer kan word as rekenaars wat aanlyn inligting makliker toeganklik maak. Chun (2013:6375) definieer Web 3.0 − die derde generasie van die web − as

[50] Die semantiese web word as ’n uitbreiding van die wêreldwye web beskou. Die semantiese web stel individue in staat om inhoud te skep en te deel – selfs buite die grense van die aanvanklike toepassings, platforms en webwerwe (vgl. Decker, 2011).

’n mobiele persoonlike web waartydens mobiele toestelle toegang tot die internet en dus internetgebaseerde platforms bied.

O’Reilly en Battelle (2009:2) voer aan dat Web 2.0 eers gedefinieer moet word ten einde Web 3.0 te verstaan en waarheen ontwikkelinge neig. Web 2.0 verwys na die suksesvolle netwerktoepassings wat beskou kan word as stelsels wat kollektiewe intelligensie huisves. Tabel 2.4 is ’n opsomming van die onderskeie kenmerke van Web 1.0, 2.0 en 3.0. Hierdie eienskappe dui die verandering in die aard van die web aan sowel as die betrokkenheid van deelnemers en hoe die rolle verander/ontwikkel het. Waar individue eers nie ’n bydrae kon lewer nie en die web staties was, is dit nou ’n inligtingryke omgewing waar individue bydrae kan lewer en selfs inskrywings en inligting met ander kan deel.

Tabel 2.4 Voorstelling van die verskille tussen Web 1.0, 2.0 en 3.0 (saamgestel vanuit O’Reilly & Battelle, 2009:2)

	Web 1.0	**Web 2.0**	**Web 3.0**
Kommunikasie	Uitsaai	Interaktief	Betrokke
Inligting	Staties (lees)	Dinamies	Mobiel en persoonlik
Fokus	Organisasies	Gemeenskap	Individue
Persoonlik	Tuisblad	Blogs/Wiki’s	Regstreekse uitsendings
Inhoud	Eienaarskap	Gedeelde eienaarskap	Kuratorskap
Interaksie	Webforums	Webtoepassings	Slim toepassings
Soekfunksie	Gids	Sleutelwoorde	Konteks en relevansie
Matriks	Bladsy besoeke	Koste per kliek	Gebruikers is betrokke
Advertensie	Baniere	Interaktief	Kostegebonde
Navorsing	Britannica online	Wikipedia	Die semantiese web
Tegnologie	HTML/FTP	Flash/Java/XML	RDF/RDFS/OWL

Soos gesien kan word in Tabel 2.4 is daar talle verskille tussen Web 1.0, 2.0 en 3.0. Web 3.0 stel gebruikers in staat om inligting te genereer en om op ’n direkte manier hierdie toepassings met ander gebruikers te deel. Individu verkry ’n eie stem en kan as gevolg van die mobiele aard van hierdie generasie van die web enige tyd en plek toegang tot hulle gekose toepassings (platform) kry. Die mobiele aard van die web dra ook by tot die gewildheid van sosialenetwerkplatforms.

Web 4.0 is ook reeds besig om deel van internet-gebruikers se daaglikse lewens te word. Hierdie web (Web 4.0) word volgens Choudhury (2014:8099) as ’n ultra-intelligente elektroniese agent beskou. Choudhury (2014:8099) stel dat die web ook as die simbiotiese en alomteenwoordige web beskou kan word, en dat die interaksie tussen die mens en die masjien van belang is. Hierdie vierde generasie van die web sal die proses van RBK dus nog meer vergemaklik.

Dit is duidelik dat die interaktiewe en mobile aard van die web kommunikasie beïnvloed en vergemaklik.Verder het die toenemende aantal sosialemediaplatforms waarby gebruikers kan aansluit ook ’n effek op wyse waarop kommunikasie plaasvind.

Volgens World Wide Worx (2016) is Facebook die gewildste sosiale media platform in Suid Afrika. Instagram toon weer die meeste groei oor die afgelope jaar. Twitter (volgens World Wide Worx, 2016) het nou sowat 7,4 miljoen gebruikers. Sosialenetwerkplatforms word in die volgende afdeling bespreek alvorens daar spesifiek na mikroblogs ondersoek ingestel word.

2.6.2 Sosialenetwerkplatforms

Die internet het die manier waarop sosiale interaksie en kommunikasie plaasvind onherroeplik verander. Volgens Ellison en boyd (2013:151) het die ontwikkeling van kommersieel georiënteerde tegnologie aanleiding gegee tot wat vandag bekend staan as die “sosiale media” en die belangrikste ontwikkeling van sosiale media is die onstaan van sosialenetwerkplatforms. Voorbeelde van hierdie platforms sluit Facebook, LinkedIn, MySpace, Cyworld, Twitter en Google Plus in. Alhoewel die eerste sosiale netwerk platform reeds in 1997 ontwikkel is, was dit volgens Ellison en boyd (2013:151) eers gedurende 2003 wat sosialenetwerkplatforms as ’n kultureel belangrike fenomeen beskou is.

Riley (2011:1) verduidelik dat die internetgemeenskap dramaties verander het gedurende die afgelope paar jaar. Internetverbindings het verbeter, skootrekenaars en slimfone word met modems (Wi-Fi-kaarte) vervaardig en stadig maar seker is gratis toegang tot Wi-Fi besig om die norm te word. Die internet is ’n netwerk van inligting wat geredelik bekombaar is en dit was nog nooit makliker om in verbinding te bly met vriende, familie, kennisse of selfs individue met gedeelde belangstellings nie. Riley (2011:1) verwys na internetplatforms – sosialemedianetwerke − soos Facebook, LinkedIn en Twitter as ’n manier waardeur kommunikasie al hoe makliker gemaak word. An *et al.* (2011:18) voer aan dat al hoe meer individue weg beweeg van tradisionele media en soek eerder inligting en nuus aanlyn. boyd *et al.* (2010) skryf dat sosialemedianetwerke asinchroniese gesprekke bemiddel wat nie deur ’n geografiese ligging beïnvloed word nie, maar ’n sosiale konteks is nog steeds aanwesig.

Die terminologie waarmee sosialenetwerkplatforms beskryf word, verander ook telkens, maar soos die onderskeie platforms gevestig raak, het daar ook aanvaarbare definisies vir dié term ontstaan. Definisies vir sosialenetwerkplatforms sluit in:

- Volgens boyd en Ellison (2007:211) is sosialenetwerkplatforms ŉ webgebaseerde diens wat individue in staat stel om:

  - ’n publieke of semipublieke profiel saam te stel in ’n gebonde sisteem;
  - ’n lys van ander gebruikers te genereer met wie die individu ’n konneksie deel; en
  - die profiele van die konneksies te kan besoek en met hierdie mense te kan kommunikeer.

- Sosialenetwerkplatforms kan beskou word as “a *networked communication platform* in which participants 1) have *uniquely identifiable profiles* that consist of user-supplied content, content provided by other users, and/or system-provided data; 2) can *publicly articulate connections* that can be viewed and traversed by others; and 3) can consume, produce, and/or interact with *streams of user generated content* provided by their connections on the site” – [beklemtoning in oorspronklike bron] (vgl. Ellison & boyd, 2013:159).

- Sosialenetwerkplatforms verwys na ŉ diens wat gedurende die vroeë 2000’s ontstaan het en sluit platforms in waarop video’s gedeel word, blogs, mikroblogs sowel as enige platform (soortgelyke kanaal) wat gebruikers in staat stel om inligting te genereer en dit te deel (boyd, 2014:6).

- Sosialenetwerkplatforms verwys volgens Blackshaw en Nazzaro (2006) na internetgebaseerde toepassings waar gebruikergegenereerde inhoud geplaas word en waar opinies deur gebruikers gedeel word.

- Kaplan en Haenlein (2010) definieer sosialenetwerkplatforms as toepassings wat gebruikers in staat stel om persoonlike aanlyn profiele te skep, vriende te nooi om deel van hulle sosiale netwerk te word (konneksies) en om boodskappe te stuur en te ontvang.

Met die ontstaan en ontwikkeling van die eerste sosialenetwerkplatforms was die internet ’n generasie twee web (Web 2.0) terwyl die web nou as ŉ generasie drie (Web 3.0) beskou word[51]. Waar die definiërende kenmerke van sosialenetwerkplatforms in 2007 die profiel, die konneksies en die manier waarop interaksie plaasvind was, het dié kenmerke verander.

[51] Web 4.0 is ook reeds stelselmatig besig om in gebruik geneem te word.

Sosialenetwerkplatforms ontwikkel vanaf ’n relatiewe statiese platform (in 2007 was daar slegs twee dominante sosialenetwerkplatforms − profielgesentreerde platforms − soos LinkedIn en MySpace asook mediagesentreerde platforms soos Flickr en YouTube) na nuusgenererende, interaktiewe platforms waar die inhoud wat met die publiek gedeel word in die vorm van ’n status is, maar wat ook die weergee van aktiwiteite, hiperteks foto’s of video’s kan insluit (vgl. Ellison & boyd, 2013:154). Volgens Ellison en boyd (2013:152) het die rol van die profiel veral verander en vervul mediastrome ’n meer prominente rol. Die proses waardeur ’n profiel opgedateer word, is baie makliker as in die verlede. Daar is reeds verwys na die sinchroniese en asinchroniese aard van RBK. In die geval van sosialemedianetwerke kan kommunikasie beide sinchronies en asinchronies wees.

Davenport *et al.* (2014:213) sluit by Ellison en boyd (2013) aan en stel dat die optimalisering van sosialenetwerkplatforms sigbaar is in die aantal uur wat gebruikers op die platforms bestee, die aantal keer wat gebruikers se status verander sowel as die aantal konneksies wat gebruikers het. Daar moet wel onderskeid getref word tussen aktiewe en passiewe gebruikers. Aktiewe gebruikers verwys na gebruikers wat aktief is op ’n gegewe platform − gebruikers wat inhoud skep en deelneem aan gesprekke op die platform (Davenport *et al.* 2014:213). Passiewe gebruikers verwys volgens Davenport *et al.* (2014:213) daarenteen na gebruikers wat op platforms slegs die inhoud lees − hierdie soort gebruikers dra nie self by tot die inhoud van platforms nie.

Die selfverteenwoordiging van individue op sosialenetwerkplatforms kan soms as problematies beskou word alhoewel daar talle kreatiewe maniere is waarop die self uitgebeeld kan word. Ellison en boyd (2013:153) voer aan dat ’n aanlyn identiteit eksplisiet verbind kan word met individue se identiteit wanneer dié individue nie “aanlyn” is nie. Donath (1998) stel dat die gebruikersnaam wat individue vir hulle profiele op ’n sosiale netwerk platform gebruik as ’n skuilnaam kan funksioneer en só kan individue hulle ware identiteit beskerm. Jansen van Vuuren (2007:214) argumenteer dat die anonieme aard van die internet (kletskamers) nie net as ’n masker dien waaragter individue kan wegkruip nie, maar dat dit ook bydra tot ’n hoë gebruiksfrekwensie van sterk taal en dat hierdie anonimiteit deelnemers ’n “vrypas” verskaf om hierdie soort taal te gebruik aangesien die individue se ware identiteit nie bekend is nie.

Die probleem wat hieruit voortspruit, is dat anonimiteit op sosialemedianetwerke tot misdrywe kan lei en die vraag moet gestel word of dit moontlik is om outeurs van boodskappe te identifiseer? Pearson en Watson (2010:64) postuleer dat digitale bewyse − soos enige ander

bewyse − oor unieke kenmerke beskik en dit is juis hierdie bewyse wat verdagtes met die misdaad verbind. Wanneer boodskappe op sosialemedianetwerke geplaas word, is die identiteit van outeurs nie altyd sigbaar nie weens die anonieme aard van die internet en daarom is dit nodig om te bepaal of outeurs van boodskappe op sosiale media wel geïdentifiseer kan word.

Layton *et al.* (2010:1) stel dat dit al hoe belangriker word om die outeurs van kort tekste te kan identifiseer aangesien misdrywe wel op platforms plaasvind. Abbasi en Chen (2005:67) is van mening dat die anonieme aard van sosialenetwerkplatforms, die beperkte lengte van tekste, die mate van geraas sowel as die aantal moontlike outeurs kan veroorsaak dat meer fokus op outeurskapidentifikasie van tekste op sosialemediaplatforms geplaas moet word.

Navorsing is reeds gedoen deur Michell (2013) oor outeurskapidentifikasie op Facebook, maar daar is nog nie ondersoek ingestel na Afrikaans soos dit op mikroblogs voorkom en of outeurs van Afrikaanse mikrobloginskrywings geïdentifiseer kan word nie. Inteendeel, navorsing oor outeurskapidentifikasie en sosialemedianetwerke in Afrikaans is betreklik skaars. Vervolgens word mikroblogs as ŉ sosiale media netwerk bespreek om sodoende die kenmerkende eienskappe van hierdie platform voor te hou.

2.6.2.1 Mikroblogs

Soos genoem in afdeling 2.6.2 bestaan daar verskeie sosialenetwerkplatforms wat gebruikers in staat stel om statusboodskappe op platforms te plaas. Mikroblogs verwys na platforms wat gebruikers in staat stel om vinnig en doeltreffend boodskappe te plaas. Volgens Teevan *et al.* (2011) is mikroblogs een van die gewildste platforms van sosiale media wat gebruik word. Humphreys (2010:3) voer aan dat die beperking op die tekslengte een van die kenmerkendste eienskappe van mikroblogs is terwyl die chronologiese aard en die manier waarop gebruikers alledaagse gebeure kan deel, bydra tot die gewildheid van mikroblogs. Oulasvirta *et al.* (2009) verwys na mikroblogs as ’n term wat in 2006 ontstaan het en gebruik word om ’n vorm van blogs te beskryf wat gebruikers in staat stel om ŉ kort status te skryf en aanlyn te publiseer deur gebruik te maak van slimfone en/of rekenaars.

In teenstelling met blogs, fokus mikrobloggebruikers daarop om inligting weer te gee wat “nuut” is of besig is om te gebeur (vgl. Passant *et al.*, 2008) en boodskappe is gebruikergesentreerd (vgl. Oulasvirta *et al.*, 2009). Ross *et al.* (2011:216) definieer mikroblogs as ’n variëteit van blogs wat gebruikers in staat stel om ŉ vinnige, kort status te plaas wat ’n innoverende manier van kommunikasie is en wat as ’n hibriede manier van blog beskou kan word. Java *et al.* (2007) voer aan dat mikroblogs as ’n relatiewe nuwe fenomeen beskou kan word en dat dit

ooreenstem met tradisionele blogs. Mikroblogs laat gebruikers toe om 'n status te plaas (mikrobloginskrywings bestaan uit 200 of minder karakters) waardeur inligting wat vir die outeur belangrik is, gedeel word.

Volgens Aisopos *et al.* (2012:187) is daar meer geraas aanwesig in kommunikasie wat op mikroblogs geskied. Die taal wat op hierdie platforms gebruik word, word aangepas om kommunikasie (ten spyte van die karakterbeperking) suksesvol te maak. Aisopos *et al.* (2012:187) onderskei tussen vier maniere waarop mikroblogs fundamenteel taalreëls verbreek:

- **Digtheid** − Mikroblogs bestaan uit kort tekste (140 karakters is die norm) en daarom word inligting op 'n bondige en gefragmenteerde manier weergegee.
- **Niestandaardwoordeskat –** Mikrobloginskrywings is informeel, sleng is aan die orde van die dag en die karakterbeperking gee aanleiding daartoe dat verkortings plaasvind.
- **Geraas** − Boodskappe word in die werklike tyd geplaas en die meerderheid gebruikers is nie eng oor die taal wat gebruik word of spelling of sintaksis nie.
- **Multitaligheid** − Gebruikers meng gereeld die taal wat gebruik word met 'n ander taal. 'n Inskrywing kan dus in Afrikaans geplaas word, maar die soekterm (hutsteken[52]) kan byvoorbeeld in Engels geplaas word.

Verskeie mikroblogplatforms (sien Tabel 2.5) bestaan waarvan Twitter die gewildste is (vgl. Dunlap & Lowenthal, 2009; Java *et al.* 2007; McFedries, 2007)[53].

Tabel 2.5 Eie voorstelling van die verskillende mikroblogplatforms

Mikroblog	Webadres	Karakters
Twitter	twitter.com	140
Toeter	toeter.com[54]	160
Tumblr	tumblr.com	4 096
Meetme	meetme.com	1 000
Identi	identi.ca	onbeperk
FindFriend	findfriend.com	350
Friends-hood	Friends-hood.com	200
Plerb	plerb.com	500

[52] Wanneer daar in terme van mikroblogs na 'n hutsteken (#) verwys word, word daar na die teken self verwys, terwyl 'n hutswoord na die teken én die woord wat ná die # gebruik word verwys (**#moeg** is 'n voorbeeld van 'n hutswoord). Sien ook #,(hutsteken) afdeling 2.6.2.2 (Twitter) en afdeling 7.2.1.1 (Hutswoorde).

[53] Alhoewel Instagram ook 'n gewilde sosialenetwerkplatform is, is hierdie platform doelbewus vir die doel van die studie uitgelaat aangesien die doel van Instagram is om deur middle van grafika (foto's) te kommuniker. Alhoewel die gebruiker wel teks kan gebruik om die grafika te verduidelik word daar meestal van sinsnedes en hutswoorde gebruik gemaak om die boodskap oor te dra.

[54] Toeter het intussen gesluit.

Mikroblog	Webadres	Karakters
Plurk	plurk.com	210
Meemi	meemi.com	140
SiteProNews	sitepronews.com	140
LinkedIn	linkedin.com	600

Tabel 2.5 dui etlike van die verskillende mikroblogplatforms aan, sowel as die maksimum aantal karakters waaruit 'n boodskap op die spesifieke platform kan bestaan. Vir die doel van hierdie eksploratiewe forensiese analise is daar besluit om data vanaf Twitter te gebruik, ten einde te bepaal of die outeur van 'n mikrobloginskrywing in Afrikaans wel geïdentifiseer kan word. In die hieropvolgende afdeling word Twitter bespreek, om sodoende die sosialeplatform se basiese kenmerke voor te hou.

2.6.2.2 Twitter

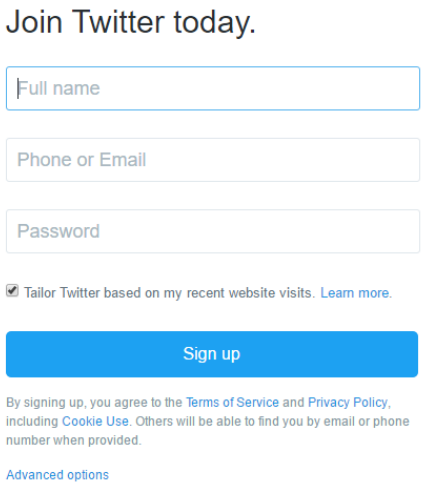
Gedurende 2006 is 'n kommunikasiediens, naamlik Twttr, deur Odeo Inc. geloods (O'Reilly & Milstein, 2009:5). Hierdie aanvanklike kommunikasiediens het gegroei, uitgebrei en is herdoop na die kommunikasieplatform wat vandag as Twitter bekend staan. Aanvanklik is Twitter gebruik as 'n diens vir vriende, familie en kennisse om met mekaar te kan kommunikeer en kontak te behou deur middel van vinnige en kort antwoorde op eenvoudige vrae soos: "What are you doing?" (vgl. Twitter, 2016). Frost (2011) verwys na die verandering in Twitter se definisie gedurende 2011 na 'n eenvoudige: "Twitter is the best way to discover what's new in your world." Die aanvanklike "What are you doing?" is vervang met "What's happening?" Frost (2011) beweer dat hierdie veranderinge kan dui op die ontwikkeling van die platform en die manier waarop gebruikers Twitter gebruik. Waar Twitter se aanvanklike definisie verhoudings en sosiale interaksie beklemtoon het, dui die nuwe definisie op 'n nuusgesentreerde benadering waar gebruikers dit wat besig is om te gebeur met ander deel.

Richardson (2015) verwys na Twitter (sien Figuur 2.10 − Twitter-logo) as 'n webgebaseerde toepassing wat aspekte van sosialenetwerkplatforms, kitsboodskappe en blogs kombineer in 'n enkele, doeltreffende en maklike kommunikasiemedium. Die kommunikatiewe doelwit van Twitter-gebruikers sluit daaglikse gesprekvoering, die deel van inligting, hiperskakels en nuuswaardige gebeure in (vgl. Dunlap & Lowenthal, 2009; Java *et al.*, 2007; Starbird *et al.*, 2010).

Figuur 2.10 Twitter-logo (geneem vanaf http://www.freelargeimages.com/twitter-logo-hd-png-4557/)

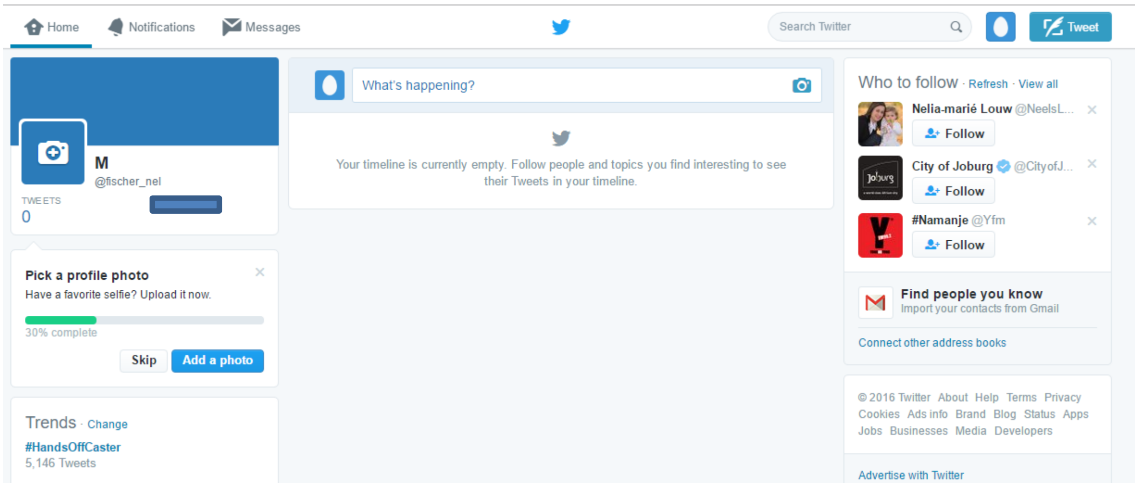
Crystal (2011:51) argumenteer dat dit nie moontlik is om 'n enkele kommunikatiewe funksie aan Twitter te koppel nie aangesien dit vir elke gebruiker anders kan wees. Persoon A kan Twitter dalk as 'n oorredingsplatform beskou, persoon B dalk as 'n blog terwyl persoon C Twitter as net nog 'n sosiale media netwerk beskou. Ongeag wat individue se kommunikatiewe doel op Twitter is, is daar geen ander platform wat taalgebruik en identiteit só kombineer soos Twitter nie (vgl. Crystal, 2011:52).

Twitter is 'n gebruikersvriendelike platform. Volgens Statista (2015) het Twitter ongeveer 305 miljoen aktiewe gebruikers en dit is baie maklik om 'n profiel te skep. Skermskoot 2.1 toon die webblad aan wat voornemende Twitter-gebruikers sal bereik wanneer hulle 'n profiel wil skep. Aanwysings van hoe om 'n profiel te skep, word op die blad aangedui. Ten einde 'n profiel te skep, moet voornemende gebruikers oor 'n geldige e-posadres beskik.



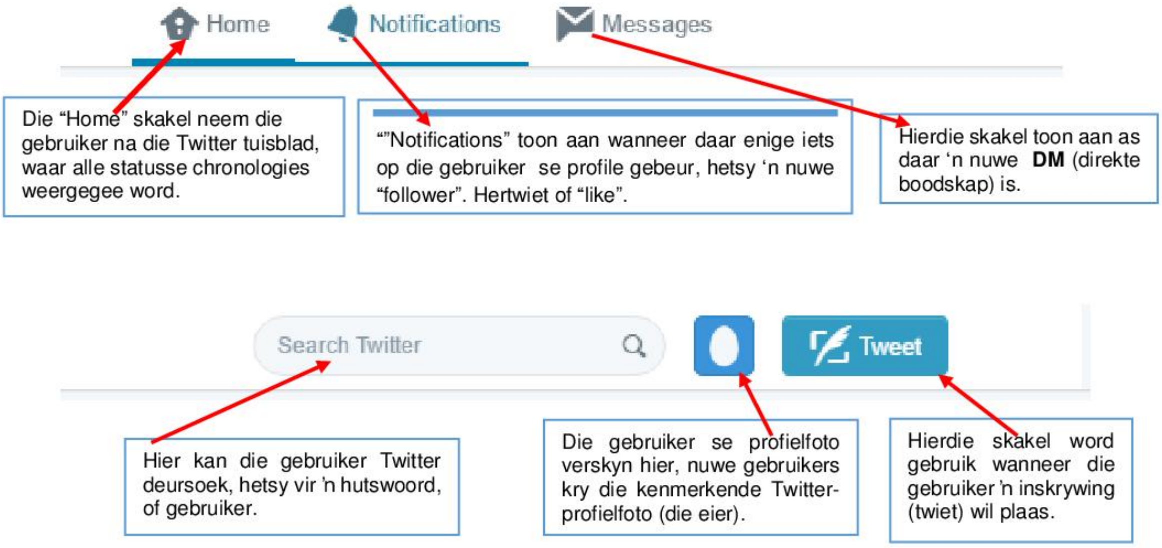
Skermskoot 2.1 Voorstelling van die tuisblad wanneer 'n Twitter-profiel geskep word (geneem vanaf https://www.twitter.com/signup?lang=en)

Sodra die profiel geskep is, sal gebruikers na die Twitter-tuisblad geneem word (uitgebeeld in Skermskoot 2.2).



Skermskoot 2.2 Voorstelling van die Twitter-tuisblad vir ’n nuwe gebruiker

Indien gebruikers oor ’n aktiewe profiel beskik, kan statusse[55] opgedateer word en gebruikers kan begin om ander individue se profiele te volg. Soos reeds genoem, is Twitter gebruikersvriendelik en Figuur 2.11 verskaf ’n oorsig oor die belangrikste funksies op die Twitter-gebruikersprofiel tuisblad.



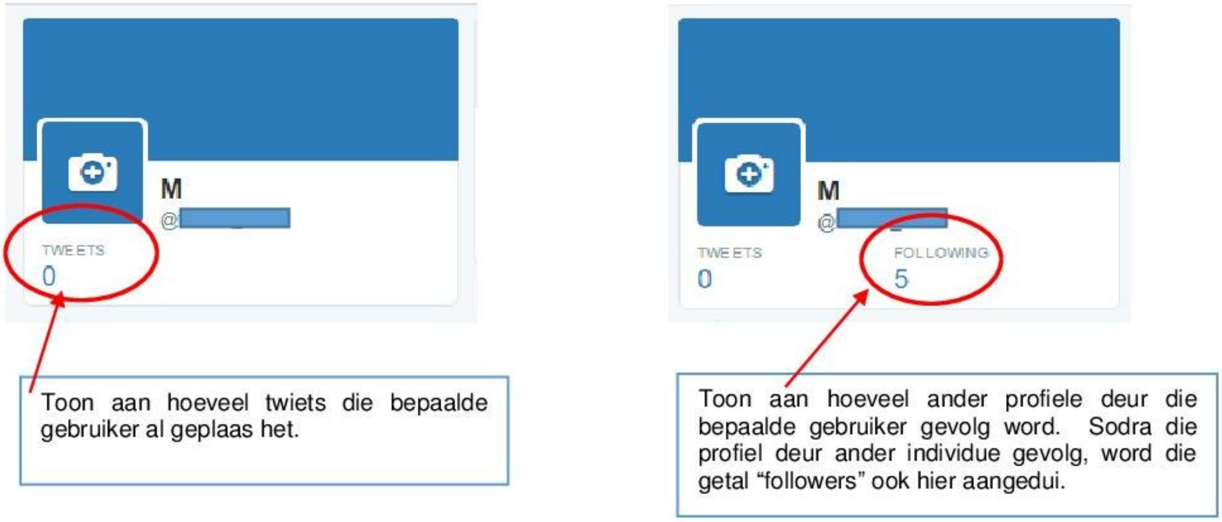
Figuur 2.11 Voorstelling van die belangrikste funksies op die Twitter-gebruikersprofiel tuisblad

Twitter verskaf soos gesien kan word in Skermskoot 2.3 vir gebruikers ’n aanduiding van moontlike profiele om te volg (*follow*):

[55] ’n Status verwys na die boodskap wat gebruikers op Twitter plaas, en kan uit ’n teks, hiperskakels, video’s of foto’s bestaan. Die status beantwoord die vraag (soos vroeër in afdeling 2.6.2.2 genoem) “What’s happening?”

Skermskoot 2.3 Moontlike profiele wat gevolg kan word

Deur profiele te volg, bou gebruikers dus ’n netwerk van gebruikers op wie se statusse op hulle Twitter-stroom sal verskyn. Soos gesien in Figuur 2.12 word die aantal profiele wat hierdie gebruiker volg, aangedui.

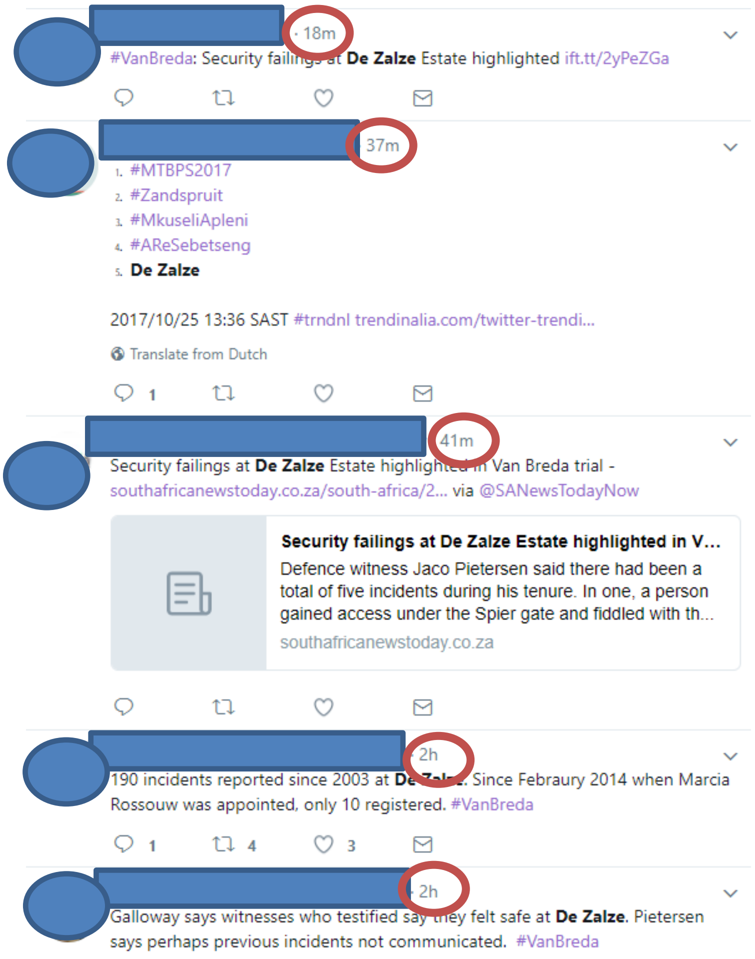


Figuur 2.12 Voorstelling van hoe die profiele wat die gebruiker volg, aangedui word

Twitter-gebruikers kan kort boodskappe (’n status), direkte boodskappe (soortgelyk aan ’n e-pos direk na ’n bepaalde gebruiker se Twitter-inboks), nuus, hiperskakels, foto’s, video’s en nuus op hierdie publieke platform plaas. Gebruikers deel idees, kennis, inligting en beantwoord vrae (vgl. Dunlap & Lowenthal, 2009).

Toegang tot Twitter kan deur enige toestel met internettoegang verkry word (slimfone, die Twitter-webblad, kitsboodskappe of e-pos) en in 'n digitale era begin meer individue van hierdie mikroblogplatform gebruik maak (vgl. Bogers & Björneborn, 2013:197; Dunlap & Lowenthal, 2009; Richardson, 2015). Gebruikers van hierdie platform word dus nie geografies beperk nie.

Bruns en Burgess (2011) postuleer dat alhoewel Twitter-gebruikers baie minder spasie (karakters) as 'n blogger het om hulle gevoelens en denke te deel, toon twiets tekens van spontane, ongesensorde, intydse strome van bewussyn en hierdie boodskappe dui op die werklike taal van gebruikers. Die inhoud wat deur individue gegenereer word wat bepaalde gebruikers volg, word in 'n enkele chronologiese stroomveld (*streaming field*) voorgestel (sien Skermskoot 2.4) (vgl. Conover *et al.*, 2013:2). Die nuutste twiets verskyn bo aan die lys en as gebruikers na hulle eie profiel gaan, sal alle twiets wat gebruikers geplaas het (*@replies*, @gebruikers en ook RT) ook chronologies voorgestel word.



Skermskoot 2.4 Voorstelling van die chronologiese wyse waarop twiets op 'n gebruiker se stroom verskyn (geneem vanaf www.twitter.com)

Twitter se gewildheid het vanaf 2006 beduidend toegeneem (vgl. Bogers & Björneborn, 2013:197; García-Gavilanes, 2015:11; Layton *et al.*, 2010:1). Gebruikers het ’n keuse oor die privaatheidstellings van hulle Twitter-profiel. ’n Profiel kan publiek of privaat wees. Wanneer gebruikers die privaat-opsie kies, sal net ’n geselekteerde groep vriende (*followers*) die twiets kan sien. ’n Profiel wat publiek is, is sigbaar aan almal (vgl. boyd & Ellison, 2008:213; García-Gavilanes, 2015:11). Daar is talle konvensies[56] wat op Twitter gevolg word en wat kenmerkend aan die platform is, neem as voorbeeld:

- @ − Die @-teken word gebruik om spesifieke gebruikers by ’n gesprek in te sluit (vgl. boyd *et al.*, 2010; Ehrlich & Shami, 2010). Volgens Honeycutt en Herring (2009) word die @-gebruikerkonvensie gebruik om ’n twiet aan ’n spesifieke gebruiker te rig, dus kan dit beskou word as ’n aanspreekvorm. Skermskoot 2.5 is ’n voorbeeld van ’n twiet waar die @-gebruikerkonvensie gevolg word. Daar bestaan nie riglyne oor die plasing van die @-gebruikerkonvensie nie, en dus kan gebruikers kies waar in die twiet hulle dit wil plaas. Hierdie konvensie kan ook as ’n kenmerkende eienskap van individuele gebruikers ondersoek word in outeurskapidentifikasie.

Skermskoot 2.5 Voorbeeld van ’n twiet gerig aan ’n spesifieke @-gebruiker (geneem vanaf www.twitter.com)

- Die **hutsteken** (#) − Die hutswoord word volgens Bruns en Burgess (2011) gedefinieer as ’n gebruikersgegenereerde sleutelwoord met ’n hutsteken (#) as voorvoegsel. Die gebruik van die hutsteken op Twitter laat gebruikers toe om inligting te deel en deel te neem aan gesprekke ongeag die feit of al die deelnemers *followers*[57] is (vgl. Ehrlich & Shami, 2010). boyd *et al.* (2010) sluit by die bogenoemde aan en stel dat die hutswoord gebruik word sodat ander gebruikers ’n gesprek met ’n bepaalde onderwerp kan volg.

  Skermskoot 2.6 is ’n voorbeeld van ’n twiet waar ’n hutswoord gebruik word om die sosiale konteks sowel as die betrokkenheid by ’n gesprek wat wêreldwyd plaasvind uit te beeld. Vanuit die voorbeeld is dit duidelik dat gebruikers multitalig twiet. Alhoewel die

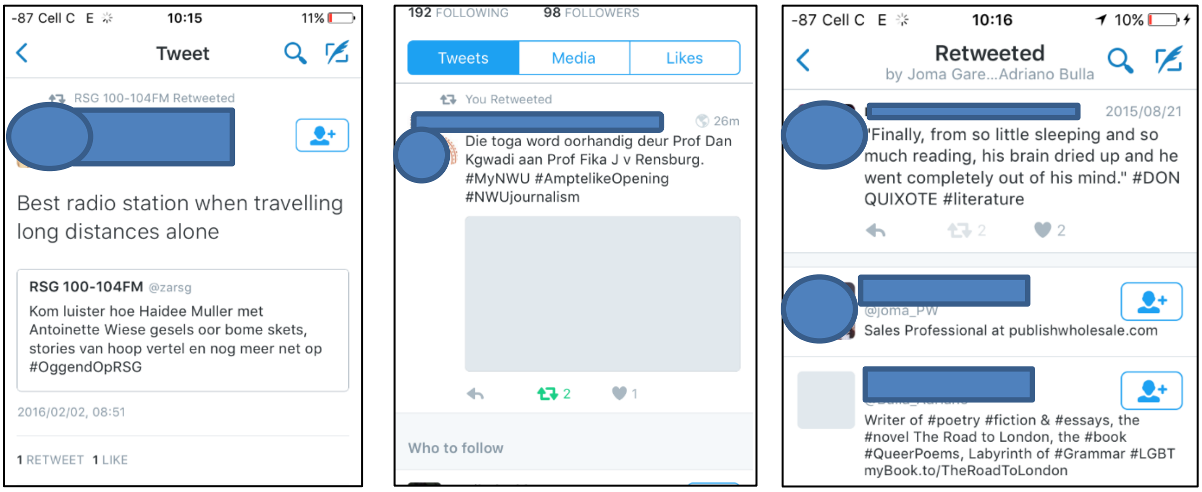
[56] Alhoewel die konvensies wat kenmerkend aan Twitter is (soos die gebruik van hutstekens en verwysings na gebruikers) ook op ander platforms – soos Instagram – gebruik word, word daar slegs op Twitter gefokus vir die doel van hierdie studie.
[57] *Followers* word na volgelinge vertaal.

twiet in Afrikaans is, is die hutswoord (hutsteken en sleutelwoord) in Engels, dus kan enige persoon wat #ParisAttacks soek die twiet lees:

Skermskoot 2.6 Voorbeeld van ’n twiet met ’n hutswoord

- **RT** − *Retweet*[58] word volgens boyd *et al.* (2010) baie onkonsekwent toegepas. Alhoewel dit die gebruik is om die hele boodskap *verbatim* te plaas, word dit nie noodwendig so toegepas nie. boyd *et al.* (2010) gebruik die volgende as voorbeeld:
  A: Hello world!
  B: RT @A: Hello world!

Alle voorbeelde van hertwiets is egter nie so duidelik soos die bogenoemde voorbeeld nie aangesien sommige gebruikers die oorspronklike boodskap aanpas om hulle eie boodskap by die aanvanklike te voeg. Die oorspronklike betekenis en sintaksis kan dus verlore gaan. Skermskoot 2.7 is voorbeelde van die verskillende maniere waarop ’n hertwiet op Twitter sigbaar is:



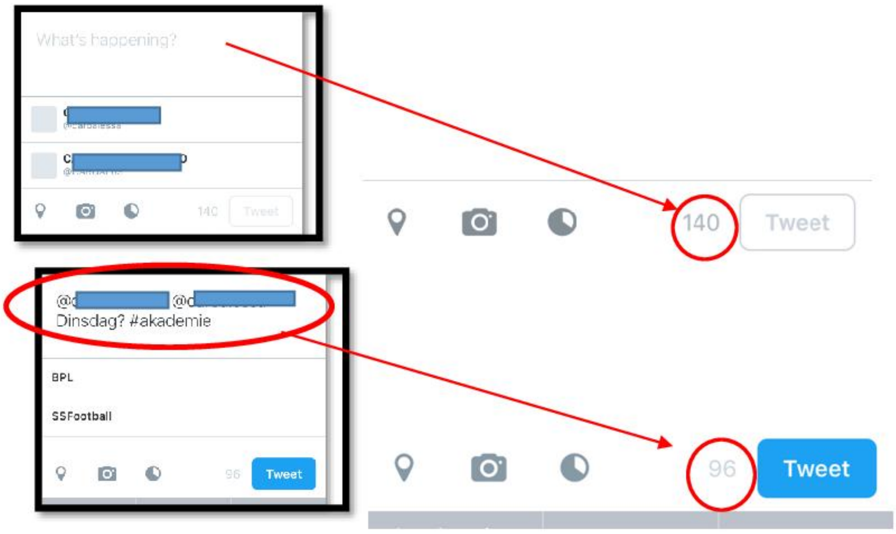
Skermskoot 2.7 Voorbeelde van ’n hertwiet

[58] *RT* (*retweet)* word na hertwiet vertaal.

- Die gebruik van **hiperskakels** kom algemeen op Twitter voor. Omdat die karakterbeperking gebruikers dwing om op kreatiewe maniere staat te maak om bedoelde boodskappe suksesvol oor te dra, is hiperskakels ’n eenvoudige oplossing. Volgens Ehrlich en Shami (2010) word hiperskakels verkort deur gebruik te maak van ’n eenvoudige program (soos bitly.com). Morris (2009:98-100) verwys na Is.gd (http://is.gd) en SnipURL (http://snipurl.com) om hiperskakels te verkort.

Go *et al.* (2009) argumenteer dat daar veral vyf unieke kenmerke is wat navorsing oor Twitter en die aard van taal op Twitter uniek maak:

- **Lengte –** ’n Maksimum van 140 karakters[59] (sien Figuur 2.13) is beskikbaar, maar Go *et al.* (2009) se navorsing bewys dat die gemiddelde lengte van ’n twiet 14 woorde (78 karakters) is. Volgens Silva en Laboreiro (2011:161) is die gemiddelde lengte van ’n twiet slegs tien woorde, maar hulle navorsing is in Portugees uitgevoer. Vanuit Go *et al.* (2009) asook Silva en Laboreiro (2011:161) se bevindinge blyk dit dat die gemiddelde aantal woorde per twiet sal verskil afhangend van die taal van die twiet. Hierdie aspek word in die empiriese analise ondersoek. Figuur 2.13 is ’n voorstelling van die manier waarop karakters op Twitter aangedui word:



Figuur 2.13 Voorstelling van die manier waarop karakters op Twitter aangedui word

[59] Ten tye van die afhandeling van die studie is Twitter se karakterbeperking aangepas van 140 karakters na 280 karakters. Perrez (2017) en Rosen (2017) stel dat daar vanaf September 2017 sprake was om die karakterlimiet te vergroot, en ‘n toetsgroep is saamgestel om hierdie nuwe 280 karakterbeperking te toets. Alhoewel daar talle argumente teen die 280 karakterbeperking was (een van die hoof argumente teen die 280 karakterbeperking is dat die twiets se leesbaarheid beïnvloed sal word, en dat die 140 karakterbeperking een van die redes was waarom Twitter so gewild is – gebruikers moet die boodskappe kort en saaklik hou), is 280 nou (vanaf 09 November 2017) die nuwe karakterbeperking.

- **Beskikbaarheid van data −** Twitter-gebruikers neem toe en so ook die aantal data wat beskikbaar is.

- **Domein −** Gebruikers plaas boodskappe (twiets) oor ’n verskeidenheid onderwerpe. ’n Enkele gebruiker se onderwerpe kan daagliks verskil.

- **Taalmodel −** Gebruikers kan toegang tot Twitter verkry deur ’n verskeidenheid maniere (slimfone, tablette of rekenaars) en daarom is die maniere waarop woorde verkeerd gespel word ook anders (toetsborde verskil, party toestelle beskik oor ’n raakskerm).

- **Emotikons −** Gebruikers maak deurentyd gebruik van emotikons om emosie uit te druk. In Go *et al.* (2009) se navorsing is emotikons vanuit die data verwyder. Vir die doel van outeurskapidentifikasie is dit nie verwyder nie aangesien dit ’n kenmerkende eienskap van individue se taalgebruik kan wees.

Ten spyte van die eienskappe van Twitter wat genoem is, argumenteer MacLeod en Grant (2012) dat die taal- en gebruikersnorme van hierdie platform steeds in die ontwikkelingsfase is en dat wat vandag aanvaarbaar is môre heel moontlik uitgedien en oud kan wees. Dit is immers die inherente natuur van die internet.

Morris (2009:50) stel dat 140 karakters betreklik min is om uiting aan individue se gedagtes te gee. Crystal (2011:39) sluit by Morris aan deur te verwys na die 140 karakterbeperking en die interne variasie wat aanwesig is by gebruikers sowel as die manier waarop die beperkings van hierdie platform gebruikers dwing om ’n spesifieke taalstyl te volg. Crystal (2011:39) argumenteer dat stilistiese kenmerke sigbaar sal wees op hierdie betrokke platform.

Volgens Crystal (2011:42) bevat ’n twiet twee kenmerkende aspekte. Die boodskapelement (linguistiese inhoud is aanwesig) en opsionele identifiseerders met ’n vasgestelde interne struktuur (hutswoorde, hertwiet, @ en hiperskakels). Om ’n boodskap suksesvol oor te dra met slegs 140 karakters tot die beskikking van gebruikers, beïnvloed hulle taalgebruik op verskeie maniere:

- verkorting (wil nie − *willi*)
- logogramme (“&” in plaas van “en”, “@huis” in plaas van “by die huis”)
- afkortings
- enkelspasiëring

- elliptiese sinne (die “ek” word telkens weggelaat)
- weglating van onnodige woorde
- weglating van leestekens
- gebruik van emotikons om gevoel uit te beeld
- niestandaardspelling
- sintaksis

’n Sintaktiese ondersoek is ook van belang wat Twitter betref, maar daar is geen vaste struktuur met betrekking tot sinne nie. Volgens Crystal (2011:45) gebruik sommige gebruikers volsinne terwyl ander twee tot vier gefragmenteerde frases gebruik. Tog is gefragmenteerde sinne koherent en wanneer dit pragmaties ondersoek word, is die betekenis duidelik.

Bykomende kenmerke wat Crystal (2011:46) in die taalgebruik identifiseer, is uitdrukkings soos *yeah!*, *wow* en *haha* wat saam met emotikons as ’n sin beskou kan word aangesien dit na ’n meer komplekse element verwys. Die aantal sinne wat gebruikers in ’n boodskap gebruik kan ook kenmerkend wees sowel as die keuse tussen ’n volsin of die gebruik van twee of meer frases. Volgens Crystal (2011:47) is die gemiddelde sinslengte 7,1 woorde en die gemiddelde aantal woorde 14,7. Die aantal woorde en die sinslengte kan dus gebruik word in die ondersoek na outeurskapidentifikasie op Twitter. Wanneer die taal ondersoek word wat op Twitter gebruik word, is ’n pragmatiese benadering belangrik veral met betrekking tot hertwiet aangesien individue wat ’n oorspronklike boodskap hertwiet die pragmatiese waarde van die oorspronklike boodskap kan verander. Gebruikers kan byvoorbeeld ’n emotikon by die boodskap voeg wat die boodskap positief of negatief kan maak (vgl. Crystal, 2011:48).

2.7 Samevatting

Hoofstuk 2 se unieke bydrae tot die vakliteratuur is ’n oorsig van hoe taal op ’n sosiale netwerk soos ’n mikroblog lyk. Ten einde Afrikaans in hierdie verband te ondersoek, is daar na die wyer konteks van Afrikaans en die betrokke taalvariasie gekyk. Hoofstuk 2 bied ten eerste ’n oorsig oor die ontwikkeling van Afrikaans en hoe Afrikaans figureer in ’n digitale era. Daarna is die ontwikkeling van Afrikaans bespreek. Variasietaalkunde geniet aandag ten einde die verskille wat in ’n taal self kan voorkom uit te beeld. Kommunikasie en taal is bespreek en hoe taal gebruik kan word om pragmaties sinvol te wees en hoe die konteks waarin kommunikasie plaasvind die taalhandeling kan beïnvloed. Die internet en kommunikasie, sosiale media en mikroblogs is bespreek.

Taal is nie rigied nie. Taal verander na gelang van gebruikers se behoeftes. Wanneer daar diachronies na Afrikaans ondersoek ingestel word, is dit duidelik dat Afrikaans ’n taal is wat stelselmatig gevestig en ontwikkel het. Kontak met ander tale, die behoeftes van sprekers en tegnologiese vooruitgang het bykomend ’n invloed op die ontwikkeling van dié taal gehad.

Taal maak ook kommunikasie moontlik hetsy die kommunikasie gesproke of geskrewe is. Taal is ’n menslike kommunikasiehandeling (’n manier waarop betekenis − gedagtes en emosies − oorgedra word deur middel van simbole) waar daar van ’n tekenstelsel (geskrewe of gesproke) gebruik gemaak word om betekenis aan ’n bepaalde gehoor oor te dra.

Die manier waarop individue hierdie tekenstelsel gebruik, verander na gelang van kommunikasievereistes en -beperkings. ’n Spesifieke kommunikasiemedium kan die aard van ʼn taal verander. Individue beskik elk oor ’n idiolek. Idiolek verwys na die unieke kenmerke wat eie aan individue se taal is. ʼn Taal kan verander en aangepas word om aan spesifieke vereistes te voldoen en kan sodoende afwyk van dít wat as die standaardnorm van die taal beskou word.

Norme en reëls is vasgestel wat as die standaard vir Afrikaans beskou word, maar afwykings (variasies) kom voor. Interne variasie verwys na taalgebruikers se vermoë om die gepaste variasie van Afrikaans in ʼn gegewe kommunikasie-situasie toe te pas. Variasietaalkunde verwys veral na styl (keuses wat uit moontlike formulerings gemaak word) en register (taal in sosiale situasies). Wanneer register dus vir die doel van die studie definieer word, kan dit beskou word as die sosiale aard van Afrikaans soos dit op mikroblogs waargeneem word terwyl styl (die oorkoepelende term wat stilistiek insluit) na individue se unieke taalkenmerke en dus ook idiolek verwys. Idiolek verwys na die unieke kenmerke van bepaalde individue se taal, en dit is hierdie kenmerke wat outeurskapidentifikasie moontlik maak.

Die internet het die manier waarop kommunikasie plaasvind onherroeplik verander en dus ook taal. Die invloed wat die internet op taal het, is wêreldwyd sigbaar. Vanuit navorsing wat reeds gedoen is oor Afrikaans en die internet, is dit duidelik dat taalgebruikers Afrikaans aanpas en van verskillende variëteite van Afrikaans gebruik maak om aan die vereistes van die internet te voldoen. Veranderinge wat kan plaasvind sluit in: kodewisseling, fonetiese skryfwyses, leksikale items (tussenwerpsels) wat gebruik word asof in ʼn gespreksituasie (geskrewe taal neig dus na gesproke taal), vereenvoudigde sinskonstruksies, eiesoortige skryfkonvensies, verkorting, leksikale kompleksiteit, logogramme, afkortings en laastens, die gebruik van emotikons.

Met die koms van die generasie drie en vier web word individue beskou as ’n bron van inligting en die aanlyn teenwoordigheid van individue dra by tot die magdom inligting wat geredelik beskikbaar is. Sosialemedianetwerke het die manier waarop kommunikasie (en die gebruik van

taal) plaasvind nog meer verander. Data, diversiteit, verandering en anonimiteit is faktore wat bydra tot die taal wat op die internet gebruik word. Mikroblogs is deel van Web 3.0 en vir die doel van my studie is Twitter as die platform gekies wat ondersoek is. Die verskillende kenmerke van Twitter wat voorgehou word, toon aan dat die platform oor unieke kenmerke beskik en gebruikers pas ’n kombinasie van geskrewe en gesproke taal toe.

Hoofstuk 2 bied dus die teoretiese agtergrond wat gebruik is in die studie. Aangesien my studie ’n eksploratiewe forensiese analise in ’n digitale era was, het ek outeurskapidentifikasie en mikroblogs ondersoek en vervolgens beskryf ek forensiese linguistiek sowel as korpuslinguistiek ten einde die teoretiese agtergrond te verskaf waarteen die metodologie en die empiriese analise in hoofstuk 5 en 6 uitgevoer is.

Hoofstuk 3

FORENSIESE LINGUISTIEK IN ’N DIGITALE ERA

Language is as it is because of what it has to do.

Halliday (1973:34)

3.1 Inleiding

Die fokus van hoofstuk 3 is die forensiese linguistiek en spesifiek ook in die Suid-Afrikaanse konteks sowel as die konsep van outeurskapidentifikasie. Vanuit hoofstuk 1 en 2 blyk dit duidelik dat die internet ’n gevestigde platform is waardeur individue moeiteloos kan kommunikeer. Hierdie kommunikasie is nie beperk tot ’n taal of individue se geografiese ligging nie. Daar is verder ook nie ’n beperking op die reikwydte van individue se kommunikasie deur middel van die verskillende sosiale netwerke[60] nie. RBK word in afdeling 2.5.2.1 gedefinieer as ’n konsep wat volgens Barasa (2010:322) na beide geskrewe tekste of gesproke dialoog verwys en maak kommunikasie moeiteloos moontlik.[61] Daar word tussen sinchroniese (onmiddellike interaksie) en asinchroniese (interaksie is uitgestel) ruimtes onderskei (Brunette, 2005:9; Crystal, 2010:11).

’n Faktor wat nie buite rekening gelaat kan word wanneer die internet en kommunikasie bespreek word, is dat reuse kommunikasieplatform ook die geleentheid skep vir anonimiteit. Die internet kan dus beskou word as die ideale platform vir negatiewe kommunikasie soos laster en afknouery (die sogenaamde *trolling*-verskynsel). Deur die toepassing van forensieslinguistiese beginsels kan individue wat hulleself skuldig maak aan laster en afknouery geïdentifiseer word, en daarom kan forensiese linguistiek as voordelig en noodsaaklik in ’n digitale era beskou word.

Volgens Donath (1999:43) kan *trolling* beskou word as ’n spel waarin individue ’n ander identiteit aanneem. Hardaker (2013:58) verwys na *trolling* as ’n doelbewuste antagonistiese aanlyn handeling en voer aan dat RBK en die mate waartoe ’n gebruiker anoniem kan bly tot *trolling* bydra (Hardaker, 2013:59). Met betrekking tot die internet (oftewel RBK) probeer individue om ’n profiel saam te stel wat deur ander deelnemers op die betrokke platform as oorspronklik en betroubaar beskou sal word. Donath (1999:44-45) skryf verder dat hierdie individue die aanlyn groep waaraan hulle behoort, ontwrig en dat daar altyd ’n moontlikheid van kriminele aktiwiteite (ook die moontlikheid om ander deelnemers te benadeel)

[60] Sien hoofstuk 2 (afdeling 2.6– Sosiale media).
[61] RBK word in afdeling 2.5.2.1 (Rekenaarbemiddelde kommunikasie) bespreek.

aanwesig is. Crystal (2004:53) verwys na *trolling* as die doelbewuste keuse van individue om hulle identiteit weg te steek (dus vermom individue hulle ware identiteit) en vals inligting word weergegee om sodoende ander deelnemers aan die betrokke platform uit te lok om daarop te reageer.

Hierdie soort kommunikasie − *trolling* of kommunikasie waar die identiteit van een of meer individue nie bekend is nie − is veral op sosiale media ’n algemene verskynsel. Laasgenoemde faktore dra by tot vrae oor die ware identiteit van outeurs van hierdie soort kommunikasie (dit kan lasterlik of negatief van aard wees of dit kan ’n profiel wees wat geskep is om vals inligting weer te gee). Crystal (2004:50) voer aan dat die inherente anonimiteit van internetgebruikers hulle gemakliker laat voel – hulle voel minder geïnhibeer − wat bydra tot probleme wat op hierdie medium kan ontstaan. Daar is dus ’n dringende behoefte om anonieme gebruikers wat hulleself aan lasterlike gedrag skuldig maak te identifiseer of om die ware outeurs van tekste te bepaal.

Hoe word die ware identiteit van outeurs dus bepaal? Outeurskapidentifikasie (’n aspek van forensiese linguistiek) verwys na die vasstelling van ware outeurs van gegewe (soms anonieme) tekste (vgl. Castillo *et al*., 2015:135). Castillo *et al*. (2014:1035) voer aan dat die identifikasie van ware outeurs van anonieme tekste veral in die digitale era waarin ons ons tans bevind al hoe belangriker raak vanweë die voortdurende groei van inligting wat op die internet beskikbaar is. Hierdie stelling is ook waar in die Suid-Afrikaanse konteks, en dit is daarom belangrik om metodes te ontwikkel waardeur die outeur van ’n mikrobloginskrywing in Afrikaans geïdentifiseer kan word, sou daar ’n geregtelike situasie ontstaan waar die outeur van so ’n teks geïdentifiseer moet word.

Bhargava *et al*. (2013:37) toon aan dat navorsing in ander tale (soos Engels, Portugees, Nederlands en Spaans) al hoe meer fokus op outeurskapidentifikasie van RBK en dat dokumente korter as 1 000 woorde in die vroeë 19de eeu beskou is as dokumente wat moeilik identifiseerbaar is. Die aantal woorde wat gebruik word om outeurskapidentifikasie uit te voer, daal toenemend en vandag vind identifikasie plaas op tekste vanaf aanlyn platforms soos Facebook en Twitter.[62]

Bhargava *et al.* (2013:37) gaan voort deur te argumenteer dat die ontleding en navorsing van hierdie toenemend korter tekste van kardinale belang is aangesien kommunikasie aanhoudend verkort. Forensiese linguiste moet outeurs van tekste wat bevraagteken word kan aandui deur

[62] Sien afdeling 2.6 (Sosiale media).

gebruik te maak van, onder andere, stylmerkers (Chaski, 2001:3; Grant & Baker, 2001:67; McMenamin, 2001:93).

Deur gebruik te maak van forensieslinguistiese metodes en beginsels, kan antwoorde gevind word op vrae soos wie die ware outeurs van tekste is en of persone met ꞌn aktiewe profiel waarlik bestaan. Vervolgens ondersoek hoofstuk 3 die vakgebied van forensiese linguistiek. Ek verskaf ’n oorsig van die ontstaan van forensiese linguistiek en die geskiedenis daarvan, die verskillende areas van forensiese linguistiek, outeurskapidentifikasie sowel as benaderings tot outeurskapidentifikasie ten einde ꞌn eksploratiewe forensiese analise op outeurskapidentifikasie op mikroblogs in ’n digitale era te kon uitvoer. ’n Kernbydrae van hierdie studie is juis die oorsig van studies oor die forensiese linguistiek in die Suid-Afrikaanse konteks.

Hoofstuk 3 vorm saam met hoofstuk 2 ’n agtergrond vir korpuslinguistiek en die metodologie in hoofstuk 5 sowel as die empiriese analise in hoofstuk 6 en 7.

3.2 Forensiese linguistiek

Forensiese linguistiek verwys na die studie van taal en die reg. Volgens MacLeod (2013:2159) word daar tussen twee navorsingsareas in die vakgebied onderskei: 1) taal en die reg; en 2) die lewering van taalkundige bewyse in straf- en siviele prosesreg − forensiese linguistiek. Volgens Gibbons en Turell (2008:2) is daar talle subdissiplines teenwoordig in die vakgebied van forensiese linguistiek, naamlik die geskrewe taal van die reg; die taal van wetgewing; gesproke regsdiskoers; die taal van hofsake en ondervraging; die lewering van taalkundige getuienis wat betrekking het op outeurskap en identiteit sowel as kommunikasie; onderrig in geskrewe en gesproke regstaal en tolking; en laastens, vertaling van regstaal. Aangesien daar soveel subdissiplines in die vakgebied forensiese linguistiek aanwesig is, gaan daar in afdeling 3.3 na definisies van forensiese linguistiek verwys word, alvorens die ontstaan en ontwikkeling van die vakgebeid bespreek word.

3.3 Forensiese linguistiek: ’n definisie

Die definisies vir forensiese linguistiek is meervoudig en ’n verteenwoordigende seleksie hiervan word hieronder weergegee:

- Olsson (2008:4) voer aan dat forensiese linguiste kennis en tegnieke van die betrokke taal toepas op taal wat 1) in regsake of ondersoeke betrokke is, of 2) as privaat geskille tussen twee of meer partye beskou kan word, en wat dalk regsimplikasies kan inhou.

- McMenamin (2002:67) definieer forensiese linguistiek as die wetenskaplike studie van taal vir toepassing op forensiese ondersoeke.

- Kopka en Vogel (2004:2) verwys na forensiese linguistiek as die ondersoek na patroonmatigheid in taal, en beskou forensiese linguistiek as ’n subdissipline van toegepaste taalkunde. Hulle (Kopka & Vogel, 2004:2) voer aan dat forensiese linguistiek ’n subdissipline van toegepaste taalkunde is, en tref onderskeid tussen forensiese linguistiek[63] en forensiese outeurskapidentifikasie[64].

- Larner (2015:131) verwys na forensiese linguistiek as ’n vertakking van toegepaste taalkunde wat betrekking het op die verhouding tussen taal en die reg. Forensiese linguistiek sluit verskeie onderwerpe in: die taal van polisieondervragings, die taal wat in howe gebruik word, regstaal en outeurskapidentifikasie (hier word data vanaf SMS’e, eisbriewe, selfdoodbriewe, terrorismanifeste en dies meer ondersoek), maar is nie beperk tot slegs dié genoemde onderwerpe nie.

- Forensiese linguiste beantwoord vrae oor taal soos wie outeurs van bepaalde dokumente is en verder word plagiaat, kontrakte, handelsmerke, patente, hofverrigtinge, haatmisdade, dreigbriewe, selfdoodbriewe en bekentenisse ontleed (Leonard, 2005:1447). Leonard (2005:1447) gaan voort deur dit duidelik te maak dat forensiese linguiste bewyse oor geregtelike ondersoeke lewer deur ’n taalkundige analise van tekste betrokke in die ondersoeke te doen.

- Beckman (2007:11) voer aan: “Forensic Linguistics is the study of language within a legal context. Texts, both spoken and written, form the basis of the study, analysis, and measurement of language. Generally speaking, linguistic questions regarding crime, judicial matters, and legal disputes constitute most forensic linguistic cases.”

  Beckman (2007:11) argumenteer ook dat linguiste verskeie aspekte van taal ondersoek om sodoende te verstaan hoe mense kommunikeer − wat individue sê sowel as hóé dit gesê word. Wanneer ’n forensieslinguistiese ondersoek uitgevoer word, kan dit onder andere (vgl. Beckman, 2007:11) fonetiek, sintaksis, sosiolinguistiek, fonologie, morfologie, semantiek, psigolinguistiek, rekenaarlinguistiek en neurolinguistiek insluit.

---

[63] Forensiese linguistiek kan as die oorkoepelende term beskou word wat gebruik word om die verhouding tussen taal en die reg te ondersoek en sluit in, die taal van polisieondervragings, die taal wat in howe gebruik word, regstaal asook outeurskapidentifikasie.

[64] Verwys na die identifikasie van outeurs.

- Ciardullo en Romito (2015:2) definieer forensiese linguistiek as ’n dissipline wat deel vorm van toegepaste taalkunde en forensiese wetenskappe en stel dat die forensiese wetenskaplike aspek verband hou met beide geskrewe en gesproke taal (of tekste) wat betrokke is in ’n geregtelike geskille. Enige teks wat tersaaklik is in ’n geregtelike ondersoek, wat deel vorm van bewyse of wat individue se betrokkenheid in ’n misdryf, kan bewys (of onskuld) in enige wetlike of kriminele konteks, word dus as ’n forensiese teks beskou.

- Coulthard (2005:10) definieer forensiese linguistiek deur die vraag “Wat doen ’n forensiese linguis?” te beantwoord. Hy voer aan dat forensiese linguiste een (of albei) van die volgende vrae moet beantwoord, naamlik 1) Wat is die betekenis van bepaalde tekste en 2) Wie is die outeurs van betrokke tekste. Om die bogenoemde vrae te kan beantwoord, maak taalkundiges gebruik van kennis vanuit een of meer van die subdissiplines van beskrywende taalkunde soos fonetiek, fonologie, sintaksis, semantiek, pragmatiek en diskoers- en teksanalise. Coulthard (2005:10) onderskei ook verder tussen forensiese fonoloë[65] en forensiese linguiste[66].

- Gibbons en Turell (2008:1) voer aan dat forensiese linguistiek vandag in sy aangepaste breër definisie uit ’n menige aspekte bestaan, maar dat die hoofareas van navorsing die geskrewe taal van die reg; die taal van wetgewing; regsgeskille met betrekking tot gesproke diskoers; polisieondervraging en die lewering van linguistiese bewyse (insluitend outeurskapidentifikasie); bewyse oor kommunikasie; die onderrig en leer van gesproke en geskrewe regstaal; en laastens, wetlike vertaling en tolking insluit.

- Cotteril (2013:1249) onderskei tussen ondersoekende en beskrywende forensiese linguiste. Beskrywende forensiese linguiste fokus op die analise van taal wat voorkom (enige tyd) in die regsproses om sodoende ’n onderskeid tussen verskillende genres en tekstipes te kan tref (Cotterill, 2013:1249) − onderhoude, die taal van die reg en kontrakte, handelsmerke en waarskuwingstekens word hierby ingesluit.

---

[65] Wanneer ’n gegewe teks fonologies ondersoek word, sal linguiste volgens Coulthard (2005:10) die woorde en frases dekodeer (in hierdie geval word daar met gesproke taal gewerk). Wanneer die kwaliteit van die opname byvoorbeeld swak is, sal linguiste ondersoek instel na wat werklik gesê is. Coulthard (2005:10) noem as voorbeeld ’n geval waar ’n man met ’n Wes-Indiese aksent klaarblyklik gesê het: “shot a man to kill”. Ná afloop van die forensiese analise is daar bepaal dat die man gesê het: “ showed a man ticket.” Die fonologiese ondersoek van ’n teks kan ook die outeurs van tekste identifiseer. Die uitspraak en sprekers se stem word veral in dié geval ondersoek (Coulthard, 2005: 13).

[66] Forensiese linguiste fokus volgens Coulthard (2005:11) op die betekenis van tekste en die identifikasie van outeurs van tekste. Coulthard (2005:13) stel dat idiolek die fokusarea van hierdie tipe ondersoeke vorm, en dat daar verskeie aspekte is wat ondersoek word.

Ondersoekende forensiese linguiste ondersoek volgens Cotterill (2013:1250) die taal wat betrokke is in kriminele sake en is ’n ontwikkelende veld. Ondersoekende forensiese linguiste maak veral gebruik van korpusanalises om antwoorde op kwelvrae te verkry en tekste wat ondersoek word, sluit dreigbriewe, selfdoodbriewe, vals verklarings, afpersing, kommunikasie van terroriste, e-posse en SMS-boodskappe in (Cotterill, 2013:1251).

- Ariani *et al.* (2014:223) skryf: “[f]orensic linguistics involves the application of scientific knowledge to language in the context of criminal and civil law. Forensic linguists have an interest in understanding the language of the written law, its complexity and its origin, as well as the use of language in forensic procedures They also study the judicial process from point of arrest, and through the interview, charge, trial and sentencing stages.”

Vanuit die bogenoemde lys van definisies is dit duidelik dat daar talle omskrywings van forensiese linguistiek bestaan. Daar kan tussen twee hoofvertakkings in hierdie vakgebied onderskei word, naamlik die taal van die reg en outeurskapidentifikasie. Dit blyk ook duidelik vanuit die genoemde definisies dat daar ondersoek ingestel word na die manier waarop taal gebruik word sowel as identifiseerbare patroonmatigheid in taal wat kan lei tot die identifisering van outeur van bepaalde tekste. Cotterill (2013:1250) voer aan dat forensiese linguistiek nog ’n ontwikkelende veld is wanneer dit in kriminele sake gebruik word en juis daarom is dit belangrik om navorsing oor moontlike metodes (wat vir outeurskapidentifikasie gebruik kan word) te doen.

Vir die doel van my studie (’n eksploratiewe forensiese analise in ’n digitale era oor outeurskapidentifikasie op mikroblogs), het ek op outeurskapidentifikasie gefokus. Die taal van die reg is dus nie ondersoek nie. In die bogenoemde gedeelte is forensiese linguistiek gedefinieer, en dit is duidelik dat forensiese linguistiek talle tekstipes ondersoek en uit verskillende vakgebiede leen. Vervolgens word daar na die oorsprong en ontwikkeling van hierdie vakgebied ondersoek ingestel alvorens forensiese linguistiek in ’n digitale era en spesifiek outeurskapidentifikasie bestudeer kon word.

3.4 Ontstaan en ontwikkeling van forensiese linguistiek

Soos met omtrent alle ander wetenskappe, argumenteer Olsson (2008:4) dat dit nie moontlik is om met sekerheid ’n datum aan die ontstaan van forensiese linguistiek te koppel nie. Vanaf die vroegste tye is vrae oor die ware outeurs van tekste gevra. Kotzé (2007:386) verwys spesifiek na Augustus de Morgan wat in 1851 ondersoek ingestel het na die outeurs van Bybelboeke deur na die lengte van woorde wat gebruik is te kyk. Volgens Olsson (2008:5) word plagiaat al

vir jare ondersoek, en hy verwys na die antieke Griekse dramaturge wat reeds hierdie knelpunt ondersoek het.

Love (2002:15) verwys in hierdie verband na dispute oor die outeurskap van William Shakespeare se werk. Olsson (2008:18) verwys na Wilmot wat in 1785 aangevoer het dat Francis Bacon die ware outeur van Shakespeare se werk was. Tog is dié gerugte nooit bevestig nie aangesien tekstuele eienskappe wat uitstaan dikwels die doelbewuste keuse van outeurs is, en dus relatief maklik nagedoen kan word. Om hierdie rede is die onbewuste taalkeuses wat individue maak van kardinale belang vir outeurskapidentifikasie.

3.4.1 Forensiese linguistiek: die ontstaan

Die eerste ondersoeke in forensiese linguistiek het meestal gehandel oor moontlike plagiaat, maar met verloop van tyd is hierdie metodes ook aangewend om outeurs van bepaalde tekste te identifiseer. Grant (2008:216) argumenteer dat sekere tekste (na gelang van die inhoud en konteks) van belang is vir geregtelike ondersoeke. Hierdie tekste sluit in dreigbriewe, eisbriewe en eksplisiete seksuele gesprekke op die internet waar minderjariges betrokke is (Grant, 2008:216; Kotzé, 2010:185).

In 1964 het Mosteller en Wallace ondersoek ingestel na die outeurskap van die *Federalist Papers* − ’n reeks politieke opstelle wat gedurende 1787-1788 deur drie verskillende outeurs, Alexander Madison, James Hamilton en John Jay geskryf is, maar almal is onderteken as *Publius* (Juola, 2006:10; Kotzé, 2007; Mosteller & Wallace, 1965:275; Olsson, 2008:19). Beide Madison en Jay het aangevoer dat hulle 12 van die opstelle geskryf het. Mosteller en Wallace het ’n ondersoek na funksie[67]- en inhoudswoorde[68] uitgevoer waarin stylmerkers van dié individue geïdentifiseer is. Mosteller en Wallace (1965:302) stel dat dit ná die ondersoek duidelik geblyk het dat die bewyse vanuit die data aantoon dat Madison wel die outeur van die 12 opstelle was.

Die term *forensiese linguistiek* is in 1968 vier jaar ná die publikasie van die *Federalist Papers* in gebruik geneem toe Jan Svartvik die *Evans Statement: A case for Forensic Linguistics* gepubliseer het (Ciardullo & Romito, 2015:2; Coulthard, 2004:431; Coulthard & Johnson, 2007:5; Kotzé, 2007; MacLeod, 2013:2159; Olsson, 2005:4; Olsson, 2008:4). Svartvik het vier verklarings van Timothy Evans geanaliseer (vgl. Svartvik, 1967:17). Svartvik (1967:17)

[67] Michell (2013:50) definieer funksiewoorde as woorde wat selfstandig min/geen betekenis het nie. Koppel en Schler (2003:1) voer aan dat funksiewoorde onafhanklik van die konteks is, en dus sal die woorde nie noodwendig deur die onderwerp van die teks bepaal word nie.
[68] Inhoudswoorde word ook grammatikale woorde genoem en word in hoofstuk 5 (afdeling 5.7.1) bespreek.

verduidelik dat die analise van Evans se taal as problematies beskou kon word aangesien die korpus klein was (4 861 woorde) en omdat Evans ongeletterd was.

Evans is in 1950 tereggestel nadat hy skuldig bevind is van die moord op sy vrou en babadogter. In die analise het Svartvik talle ongerymdhede in die taalgebruik van die verklarings uitgewys. Svartvik het van beide kwalitatiewe en kwantitatiewe metodes gebruik gemaak om die stilistiese ongerymdhede wat die outeurskap van die dokumente bevraagteken het, uit te wys (Kotzé, 2007:387; MacLeod, 2013:2159; Olsson, 2005:4).

Coulthard en Johnson (2007:5) verwys met betrekking tot die ontstaan van die term *forensiese linguistiek* ook na Philbrick wat dié term sowat 20 jaar voor Svartvik gebruik het in sy boek *Language and the Law: the Semantics of Forensic English* (1949), maar die term het ongelukkig toe nie veel neerslag gevind nie, en dit was eers ná Svartvik se gebruik van die term dat dit gewildheid verwerf het.

Die *Unabomber*-saak is ook een van die bekendste sake waar forensiese linguistiek (outeurskapidentifikasie) gebruik is om ’n individu skuldig te vind aan ’n strafregtelike misdryf deur middel van die identifisering van die individu se idiolek. Die *Unabomber* het tussen 1978 en 1995 briefbomme aan talle Amerikaanse universiteite en lugrederye gestuur (vgl. Kotzé, 2007:387; MacLeod, 2013:2159). In 1995 is ’n brief aan die *New York Times* gestuur waarin die outeur beloof dat hy sal ophou om die bomme te stuur mits ’n manifes van 35 000 woorde gepubliseer word (vgl. Grant, 2008:220). Die manifes is toe deur die *New York Times* en die *Washington Post* gepubliseer.

David Kaczynski het die briewe wat deur die koerante gepubliseer is, gelees en besef dat die taalgebruik soortgelyk aan sy broer (Ted Kaczynski) se skryfwyse was. Ted Kaczynski is in hegtenis geneem, en daar is deur middel van forensieslinguistiese metodes bepaal dat hy wel die outeur van die briewe en die manifes was (Ciardullo & Romito, 2015:2; Coulthard en Johnson, 2007:162; Grant, 2008:220; MacLeod, 2013:2159). Coulthard (2004:433) verwys na die *Unabomber*-saak as ’n bepalende voorbeeld van ’n saak wat die gebruik van idiosinkratiese taalkenmerke by die identifisering van ’n outeur bevestig.

Buiten die bogenoemde gevalle waar forensieslinguistiese metodes gebruik is om strafregtelike sake deur middel van outeurskapidentifikasie op te los, is daar talle ander voorbeelde van gevalle in die afgelope tien jaar waar outeurskapidentifikasie gebruik is om ware outeurs van tekste te identifiseer hetsy die teks selfdoodbriewe, dreigbriewe, eisbriewe, testamente of biegbriewe was (Kotzé, 2010:185). Spuy (2007:100) verwys na Hannes Kniffka (Duitsland),

Malcom Coulthard (Engeland) en homself (Amerika) as drie navorsers wat gereeld in geregtiglike ondersoeke as getuies opgetree het ten einde forensies linguistiese getuienis te lewer. Spuy (2007:106) stel verder dat dit die forensiese linguis se werk is om die ware betekenis en die skrywer van tekste te kan identifiseer.

Smith en Spuy (2002:16-17) argumenteer dat alhoewel die proses van kriminele profilering as ŉ hulpmiddel beskou word waardeur daar ŉ verband tussen oortreders en hul misdade getrek kan word, word die oortreder se taalgebruik dikwels nie ondersoek nie. Taal kan inligting rakende ŉ oortreder se geografiese oorsprong, etnisiteit en ras, ouderdom, geslag, vlak van geletterdheid en selfs geloofsoortuigings openbaar, en daarom is dit volgens Smith en Spuy (2002:20) belangrik om na die psigolinguistiese aspekte van forensiese linguistiek in te stel.

Olsson het in 2009 ’n boek gepubliseer: *Wordcrime: Solving Crime Through Forensic Linguistics* waarin hy 23 sake bespreek waar hy as getuie opgetree het en van forensieslinguistiese metodes gebruik gemaak het om die sake te probeer oplos. Voorbeelde sluit in:

- **SMS-boodskappe** − Olsson het SMS-boodskappe geanaliseer om te bepaal of die vermiste/verdagte persoon die ware outeur van die boodskappe was (vgl. Olsson, 2009:6-10, 11-16, 57-61).

- **Plagiaat** − Olsson het in hierdie sake ondersoek ingestel na die moontlikheid van plagiaat. Die bekendste saak hier is Dan Brown se *The Da Vinci Code*. In hierdie saak het Lew Perdue aangevoer dat Dan Brown se *The Da Vinci Code* aangepas is vanuit sy (Perdue se) boeke (*Daughter of God*, *The Da Vinci Legacy* en *The Linz Testament*). Die hof het bevind dat Perdue se aantyging ongegrond was (vgl. Olsson, 2009:22-34).

- **Telefoonoproepe** − Oproepe is geanaliseer om 1) die persoon te identifiseer; 2) die ware boodskap te ontsyfer; en 3) om te bepaal of daar ’n ander betekenis aan die kommunikasie geheg kon word (vgl. Olsson, 2009:17-21, 41-45).

- **Selfdoodbriewe** – Daar moes bepaal word of die oorledene werklik self die selfdoodbrief geskryf het, en of die brief nie deur iemand anders geskryf is nie wat dus kan beteken dat ’n saak van moord en nie selfdood ondersoek moet word (vgl. Olsson, 2009:51-56, 113-122).

- **Dubbelsinnige taal** – Die weiering van verbruikersinligtingsake waar organisasies taal op só ’n wyse aanwend dat wanneer individue navraag doen, hulle nie die nodige inligting kan bekom nie (vgl. Olsson, 2009:62-67).

- **Kodes** − Die ontsyfering van kodes dui veral op die ontleding van tekste waar daar moontlik in ŉ geheime kode geskryf is. Olsson (2009:68) verwys in dié geval na ŉ individu wat in die tronk was, maar nog steeds (in ŉ kode) aan ŉ minderjarige meisie briewe (seksueel van aard) geskryf het.

Ander sake wat Olsson (2009) in sy boek bespreek, sluit dreigbriewe, foutiewe vertalings en outeurskapidentifikasie in. Die verskeidenheid van sake waarmee Olsson werk, dui op hóé wyd die vakgebied van forensiese linguistiek eintlik is. Ook Spuy publiseer belangrike boeke wat betrekking op forensiese linguistiek het. In 2010 verskyn *The Language of Defamation cases*, en in 2014 volg *The Language of murder cases: Intentionality, predisposition, and voluntariness.* In beide hierdie boeke hou Spuy voorbeelde van geregteklike ondersoeke voor waar die gebruik van forensiese linguistiek deur Spuy ŉ onbetwisbare bydrae tot die afhandeling van die sake gelewer het. Hierdie twee boeke volg op Spuy se eerste boek (*Language Crimes: The use and abuse of language evidence in the Courtroom*) wat in 1993 verskyn waarin die outeur verduidelik hoe hy in forensiese linguistiek betrokke raak, en die verskillende sake waarin hy as ŉ kenner getuig.

Alhoewel daar talle individue is wat (internasionaal) in die vakgebied werk, word outeurskapidentifikasie ook toenemend in Suid-Afrika ondersoek. Vervolgens verwys ek oorsigtelik na forensiese linguistiek in Suid-Afrika.

3.4.2 Forensiese linguistiek in Suid-Afrika

In die Suid-Afrikaanse konteks verwys Kotzé (2010:186) na twee sake waar forensieslinguistiese beginsels gebruik is om outeurs van tekste te bepaal. Volgens Kotzé (2010:186) het E.L. Hubbard in 1989 getuienis gelewer in ŉ saak van afpersing. Vir dié saak het Hubbard tien eisbriewe (soos ontvang deur ŉ Johannesburgse tak van ŉ landswye kettingwinkel) geanaliseer. In hierdie briewe het die sender(s) aangevoer dat kos op die betrokke winkel se rakke vergiftig sal word, en dat die media daarvan in kennis gestel sal word tensy ŉ bedrag van R1,5 miljoen betaal word. Kotzé (2010:186) gaan voort en stel dat die verdagte in die saak nie ŉ Engelse moedertaalspreker was nie, en dat ŉ fout-analise uitgevoer is op die twee tekste (Teks X) wat deur die betrokke winkel ontvang is sowel as op sewe tekste (Teks A) wat deur die verdagte geskryf is.

Alhoewel daar beweer is dat ŉ fout-analise alleen nie genoeg bewyse kon lewer om die verdagte skuldig te bevind nie, het Hubbard (1994:1) aangevoer dat ŉ korrelasietoets tussen

die tekste aangedui het dat die ooreenkomste tussen Teks X en A groot genoeg was om die aanname te kon maak dat die partoonmatigheid van spelfoute wat in die tekste gevind is (ten spyte van die beperkte lengte van die gegewe tekste) daarop dui dat Teks X en A deur dieselfde individu geskryf is.

Nog ’n saak in Suid-Afrika waar forensiese linguistiek gebruik is, is die *Father Punch*-saak. In hierdie saak (Staat v. Kerr Hobo) is die beskuldigde aangekla en skuldig bevind aan verskeie klagtes van naamskending met betrekking tot rekenaargegenereerde dokumente wat die beskuldigde onder die skuilnaam *Father Punch* gepubliseer het (Kotzé, 2010:188). E.F. Kotzé is genader om die saak te ondersoek. Die tekste het bestaan uit drie memorandums – 5 482 woorde in totaal - (gerig aan die Provinsiale Wetgewer van die Oos-Kaap) waarvan die outeur nie bekend was nie en ’n stel van 11 kronieke van 25 431 woorde geskryf deur die verdagte. Kotzé (2010:188) verduidelik dat hy beide ’n stilistiese en stilometriese analise[69] gevolg het in die outeurskapidentifikasieproses. Kotzé (2010:191) het bevind dat die teks (geskryf deur ’n onbekende outeur) ooreenstem met die teks wat deur die verdagte geskryf is en die hof kon die verdagte skuldig bevind (SCSA, 2004).

Verdere navorsing in die gebied van forensiese linguistiek in Suid-Afrika sluit in:

- Carney (2015) − Ondersoek is ingestel na semantiese kwessies in Afrikaanse hofsake. Carney (2015) stel dat woordeboekseleksie soms lukraak geskied en dat die aanwending nie met verantwoording gepaard gaan nie. In die studie bevind Carney dat die gebruik van korpora vir die soek van woorde en betekenisse in die bepaalde konteks die antwoord dalk kan wees.

- Cornelius (2013;2015) – Doen navorsing rakende die lekeleser en regstaal, die verstaanbaarheid van die Afrikaanse regstaal, en sinskonstruksies wat kenmerkend aan die Afrikaanse regstaal is,

- Herbig en Joubert (2006) – Engelse semantiek en kriminologie. In hierdie studie van Herbig en Joubert (2006) is ondersoek ingestel na ’n nuwe misdaadklassifikasie en is semantiek vir die ondersoek gebruik.

- Kaschuka en Ralarala (2012) – Ondersoek die taal scenario in die Suid-Afrikaanse Howe, en fokus op die taalregte van die individu.

[69] Sien Hoofstuk 5 (Metodologie).

- Klopper (2009) − Navorsing is gedoen oor etiese skryfwerk en plagiaat sowel as redes om kuberforensiese linguistiek as ’n nuwe Engelse navorsingsterrein te ondersoek.

- Malan (1998; 2003) – Afrikaans as ’n taal in die hof is ondersoek.

- Michell (2013) − Ondersoek is ingestel na outeurskapidentifikasie op Facebook (Engels). In sy studie is outeurskapidentifikasie op Facebook ondersoek deur stilistiese en stilometriese metodes.

- Moeketsi (1997; 1999; 2000) – Engels is ondersoek as ’n taal wat in die hof gebruik word.

- Ralarala (2012; 2014; 2015) – Linguistiese en kulturele uitdagings verbonde aan die Suid-Afrikaanse Krimminele Regsisteem, vertaling en die Regsisteem.

- Sanderson (2007) – Engelse handelsmerkdispuutsake is ondersoek.

- Thiart (2014) – Outeurskapidentifikasie van Afrikaanse SMS-taal deur middel van die gebruik van stilometrie en n-gramontledings is ondersoek.

- Venter (2013) − Die manier waarop outeurskapidentifikasie in Afrikaanse forensiese rekeningkunde aangewend kan word as alternatief vir handskrifontleding is ondersoek.

Die bostaande lys dui navorsing oor forensiese linguistiek aan wat in Suid-Afrika gedoen is. Dit blyk duidelik vanuit hierdie lys dat die meerderheid van die analises op Engelse tekste uitgevoer is wat dus daarop dui dat die Afrikaanse vakgebied van forensiese linguistiek nog baie geleenthede vir navorsing bied − veral as die ontwikkeling van forensiese linguistiek in ander tale in ag geneem word. Vervolgens verwys ek na ondersoeke in forensiese linguistiek wat nie betrekking het tot die taal van die reg nie.

3.4.3 Forensiese linguistiek en informele tekste

Soos dit duidelik blyk vanuit afdeling 3.2 en 3.3, is daar menige tekste wat geanaliseer word wat nie deel vorm van forensiese geregtelike ondersoeke nie, en wat nie beskou word as moontlike kriminele tekste nie (vgl. Grant, 2008:216). Grant (2008:216) argumenteer dat hierdie ondersoeke uitgevoer word om byvoorbeeld ’n alibi te verifieer (in onder andere dagboekinskrywings en briewe). Daar moet ook onderskeid getref word tussen tekste wat geanaliseer word ten einde ’n bydrae te lewer in ’n kriminele ondersoek en die analise van literêre tekste.

Grant (2008:216) voer aan dat daar verskille bestaan tussen formele tekste (literêre werk) en informele tekste (SMS’e, briewe, selfdoodbriewe en dagboekinskrywings). Hierdie verskille word in Tabel 3.1 weergegee:

Tabel 3.1 Verskille tussen formele en informele tekste (saamgestel uit Grant (2008:217)

	Formele tekste	**Informele tekste**
Tyd	Neem lank om te skryf	Onmiddellik, word relatief vinnig geskep
Gehoor	Bedoelde gehoor is groot	Beperkte gehoor
Inhoud	Storiegebonde	Sterk emosioneel
Lengte van teks	Langer tekste	Korter tekste

Vanuit Tabel 3.1 blyk dit duidelik dat daar ’n duidelike verskil is tussen formele en informele tekste. Vir die doel van hierdie studie word daar volgens die onderskeid deur Grant (2008:217) aangetoon, ondersoek ingestel na informele tekste. Vanuit afdeling 3.4.2 dit duidelik dat die ontstaan van forensiese linguistiek oor ’n lang geskiedenis beskik en dat daar talle voorbeelde van sake is waar forensiese linguistiek ’n belangrike rol in die regsproses gespeel het. Ook in Suid-Afrika bestaan daar reeds voorbeelde van hierdie soort sake (Kotzé, 2007; Hubbard, 1994) alhoewel die fokus tot dusver op navorsing in Engels en die taal van die reg geval het.

Volgens Butters (2011:196) het navorsers (wat nie self regstudente is nie) wêreldwyd veral gedurende die afgelope 40 jaar al hoe meer op die reg begin fokus en navorsingsareas sluit in: literêre teorieë en retoriek[70], antropologie[71] en linguistiek. Butters (2011:196) voer aan dat taalkundiges deur regskenners genader word om 1) te help om die taalaspekte van sake (in beide die straf- en siviele prosesreg) na te gaan en voor te berei; en om 2) hulp te verleen ten opsigte van die evaluasie van bewyse in kriminele ondersoeke.

[70] Fish (1989, 1994) het geskryf oor die interpretasie van kwessies in die regstelsel.
[71] Navorsers van antropologie met ’n belangstelling in taal en kultuur het ondersoek begin instel na genres en registers van regstaal.

Forensiese linguistiek het soos reeds genoem (sien afdeling 3.2) nie net betrekking op die regstaal nie, maar die identifikasie van outeurs; die identifisering van plagiaat; forensiese fonetiek; handelsmerkdispute; bedrog en diskoersanalise is ook deel van hierdie studieveld (Butters, 2008:232; Coulthard & Johnson, 2007:187; Eggington, 2008:249; MacLeod, 2013:2160; Turell, 2008:265). Soos gesien vanuit Ariani *et al*. (2014:222-224)[72] word daar ook tussen outeurskapidentifikasie; forensiese stilistiek; forensiese stilometrie; diskoersanalise; taalkundige dialektologie; forensiese fonetiek; forensiese transkripsie; en variasie in dié vakgebied onderskei.

Ciardullo en Romito (2015:3) voer aan dat forensiese linguistiek uit twee teenstrydige, maar aanvullende afdelings bestaan, naamlik ’n geskrewe en ’n gesproke afdeling[73]. Dus kan data (wat ondersoek word) óf geskrewe óf gesproke wees, maar gesproke data sal ook getranskribeer moet word vir analise. Grant (2008:216) voer aan dat fonetiese getuienis (gesproke taal) meer robuus as geskrewe tekste is aangesien daar gewoonlik ook ’n opname van die betrokke teks beskikbaar is wat dus aanleiding daartoe gee dat die analise van gesproke tekste ook foneties ondersoek (hierdie sal die primêre ondersoek vorm) kan word.

Geskrewe tekste daarenteen, kan volgens Grant (2008:216) handgeskrewe of digitale (getikte) dokumente wees of dit kan selfs ’n transkripsie van gesproke taal wees. Geskrewe dokumente is meer duursaam as gesproke tekste en kan konteksafhanklik (byvoorbeeld testamente en besigheidsbriewe waar die inhoud en struktuur van die tekste grotendeels deur die soort teks bepaal word) of konteksonafhanklik wees (SMS’e − gebruikers bepaal self die inhoud en die formaat/skryfstyl). Geskrewe tekste se verskeidenheid dra daartoe by dat daar verskeie metodes bestaan om ’n analise op geskrewe dokumente uit te voer (Grant, 2008:216).

Volgens Gong en Ooi (2008:920) word daar telkens verwys na die manier waarop RBK as ’n kombinasie van gesproke en geskrewe taal beskou kan word. Belmore en Collot (1996:25-27) het bevind dat elektroniese kommunikasie tekens toon van beide gesproke en geskrewe taal en dat die taal wat op die medium (in die geval van Belmore en Collot se studie is daar ondersoek ingestel na die taalgebruik op aanlyn bulletinborde) gebruik word meer gebruikersvriendelik voorkom as tradisionele geskrewe taal, maar steeds meer formeel as gesproke taal.

Baron (1998:139-141) sluit by die ondersoek rakende RBK-taal aan en verwys na die taal van e-posse as ’n hibriede vorm wat as ’n kombinasie van gesproke en geskrewe taal beskou kan word. Dit blyk dus duidelik dat aanlyn kommunikasie as ’n kombinasie van geskrewe en

[72] Sien afdeling 3.5 (Outeurskapidentifikasie).
[73] Sien afdeling 2.4 (Afrikaans in ’n digitale era).

gesproke taal beskou kan word, en juis daarom moet die metodologie waarvolgens analises op hierdie mediums uitgevoer word ook die eienskap van RBK-taal oorweeg[74].

Ariani *et al*. (2014:224) voer aan dat wanneer forensiese benaderings op beskrywende taalkunde toegepas word, kan daar tussen drie belangrike fasette onderskei word:

- **Forensiese linguiste**: Individue wat spesialiseer in forensiese linguistiek word genader om hoofsaaklik twee tipe vrae in die hof twee vrae te beantwoord: 1) wat sê die teks; en 2) wie het die teks geskryf.

- **Teks**: Tekste word ondersoek in ’n forensieslinguistiese analise. Vir fonetikusse behels hierdie soort ondersoek die dekodering van woorde en frases in opnames. Forensiese linguiste moet nie net betekenis kan aflei nie, maar ook die bedoelde betekenis kan aandui.

- **Die linguistiese ondersoek na outeurskap**: Linguiste kan ook genader word om onduidelikhede oor ware outeurs van tekste uit die weg te ruim. Volgens Ariani *et al*. (2014:224) word daar vanuit die teoretiese veronderstelling uitgegaan dat individue elk ’n unieke weergawe (idiolek) van hulle moedertaal praat, en dat hierdie idiolek sigbaar sal wees wanneer sprekers se taal geanaliseer word.

Ariani *et al*. (2014:222-224) gaan verder en voer aan dat taalkundige getuienis op verskeie maniere vir die analise van ’n bepaalde teks aangewend kan word:

- **Outeurskapidentifikasie**: Die identifikasie van individue se eie taalstyl is verwant aan die identifisering van hulle idiolek.

- **Forensiese stilistiek**: Hierdie subdissipline gebruik wetenskaplike analises om gesproke en geskrewe tekste te analiseer om sodoende outeurs van betrokke tekste te identifiseer, om te bepaal of plagiaat gepleeg is, om die inhoud te analiseer en laastens, om betekenis te bepaal.

- **Diskoersanalise**: Die analise van geskrewe, gesproke of gebaretaal of enige belangrike semiotiese gebeurtenis.

[74] Sien afdeling 3.6 (Outeurskapidentifikasie in ’n digitale era).

- **Taalkundige dialektologie**: Die studie van dialekte in 'n metodologiese wyse na gelang van antropologiese data.

- **Forensiese fonetiek**: Die ondersoek na die voortbring van akkurate transkripsies van dít wat gesê is. Transkripsies kan inligting verskaf oor sprekers se agtergrond (sosiaal en geografies) en verder kan forensiese fonetiek onderskeid tref tussen verskillende sprekers of selfs ooreenkomste aandui in verskillende opnames.

- **Forensiese transkripsie**: Daar word tussen twee soorte transkripsies onderskei: 1) geskrewe dokumente; en 2) video- en klankopnames.

- **Variasie**: Interne variasie (individuele variasie) verwys na die manier waarop bepaalde individue se tekste van mekaar verskil[75].

- **Forensiese stilometrie**: Die manier waarop individue oor 'n kenmerkende skryfstyl beskik.

Forensiese linguistiek verwys dus na die analise van tekste (ongeag die lengte van tekste) met die doel om die ware outeurs van hierdie tekste te identifiseer. Hierdie analise ondersoek aspekte van die taal wat as kenmerkend van individue beskou kan word. Ten einde die outeurskapidentifikasie uit te voer, kan daar van 'n korpusanalise[76] gebruik gemaak word. Aangesien outeurskapidentifikasie nie beperk word tot 'n spesifieke tekstipe nie, is tekste vanaf 'n mikroblogplatform (RBK) vir die doel van my studie geanaliseer en soos gesien in hoofstuk 2[77] word RBK-taal as 'n hibriede kombinasie van geskrewe en gesproke taal beskou.

Vanuit hierdie afdeling blyk dit duidelik dat forensiese linguistiek (outeurskapidentifikasie en kwessies rakende plagiaat) al vir jare ondersoek word, maar dat forensiese linguistiek as vakgebied eers met Svartvik (in 1968) werklik tot stand gekom het. Sedertdien het die vakgebied se gewildheid stelselmatig toegeneem, en het die navorsingsareas uitgebrei.

Aangesien my studie oor 'n eksploratiewe forensiese analise rakende outeurskapidentifikasie en mikroblogs gehandel het, ondersoek ek vervolgens outeurskapidentifikasie wat vanuit hierdie afdeling blyk om 'n subdissipline van forensiese linguistiek te wees.

---

[75] Sien 2.3.2.3 (Idiolek).
[76] Sien Hoofstuk 5 (Metodologie).
[77] Sien afdeling 2.5 (Kommunikasie en taal) en afdeling 2.6 (Sosiale media).

3.5 Outeurskapidentifikasie

Coulthard (2004:431) huldig die teoretiese uitgangspunt dat sprekers oor ’n eie kenmerklike en individuele manier beskik waarop taal gebruik word − ’n idiolek[78] − en dat hierdie kenmerkende taalstyl van individu deurgaans in tekste sal manifesteer.

Love (2002:14) postuleer dat argumente oor die outeurskap van tekste al vanaf die vroegste tye voorkom, en dat dit veral gesproke werk was (wat weergegee is onder die naam van moontlike outeurs) wat vrae oor ware outeurskap geskep het. In die Westerse wêreld is daar reeds gedurende die vierde en vyfde eeu ondersoek na outeurskap ingestel. ’n Belangrike projek was dié ondersoek na die werk van Homeros − om te bepaal watter van die tekste wat deur Homeros gepubliseer is, waarlik deur Homeros geskryf is (Love, 2002:15). Juola (2006:5) sluit by die voorafgaande aan deur te skryf dat outeurskapidentifikasie as een van die oudste én nuutste probleme beskou kan word, en dat problematiek oor outeurs van tekste sowel as die vraagstuk “watter afleidings kan vanuit individue se taalgebruik gemaak word” al vanaf die tyd van die Ou Testament aanwesig is.

Kotzé (2007) definieer outeurskapidentifikasie as ’n manier waarop plagiaat blootgelê kan word sowel as ’n manier waarop onbekende betrokkenheid van individue by die skep van tekste aan die lig kan kom of die identifikasie van ware outeurs van tekste kan plaasvind. Castillo *et al*. (2015:135) beskryf outeurskapidentifikasie as die herkenning van patrone in tekste met die doel om sonder twyfel outeurs van dié tekste te identifiseer.

Oakes (2014:1) verwys na outeurskapidentifikasie as ’n konsep wat op die idee gebaseer is dat ’n spesifieke outeur se styl beskryf kan word as ’n unieke stel tekstuele kenmerke (die frekwensie van die gebruik van woorde en die gebruik van hoër-vlak linguistiese kenmerke) en dat sommige van dié kenmerke onbewustelik deur ’n outeur gebruik word. Naga Prasad *et al*. (2015:58) onderskei tussen drie definisies van outeurskapidentifikasie: 1) die identifisering van die outeur van ’n bepaalde teks vanuit ’n gedefinieerde stel outeurs; 2) om te bepaal of ’n spesifieke outeur wel die bepaalde teks geskryf het; en 3) wie die outeur van ’n teks is. Vanuit Naga Prasad *et al*. (2015:58) se definisies is dit duidelik dat outeurskapidentifikasie na die identifisering van outeurs van tekste verwys.

Johnson en Wright (2014:37) sluit by die bogenoemde aan en stel dat forensiese outeurskapidentifikasie verwys na die identifikasie van outeurs hetsy van betwisbare dokumente of anonieme dokumente wat moontlik as bewyse in regsake gebruik kan word deur die analise

[78] Sien afdeling 2.3.2.3 (Idiolek).

van die linguistiese leidrade wat outeurs in ŉ teks nalaat. Johnson en Wright (2014:37) sê verder dat forensiese linguiste die probleem oor outeurskap vanuit ’n teoretiese oogpunt benader en verwys na die konsep van idiolek[79] − moedertaalsprekers gebruik elk ’n eie, unieke weergawe van ŉ betrokke taal.

Nirkhi *et al*. (2015:369) voer aan dat outeurskapidentifikasie van aanlyn boodskappe as uniek beskou kan word aangesien die analise gebaseer is op ’n skryfstyl waar meer as een voorafgedefinieerde kategorie gebruik moet word om outeurs te identifiseer. Chaski (2008:1) definieer outeurskapidentifikasie as ’n probleem wat opgelos moet word deur die herkenning van ’n patroon − analiste moet dus linguistiese gedragspatrone in ŉ teks (soos geskryf deur ’n onbekende outeur of in ’n dokument waarvan die outeur bekend is) identifiseer om sodoende te bepaal hoe die patrone in die tekste van mekaar verskil of met mekaar ooreenstem.

Dinh (2014:9) stel dat outeurskapidentifikasie verwys na die moontlikheid dat ’n teks deur ’n bepaalde individu geskryf is deur ander tekste (wat ook deur die bepaalde individu geskryf is) te analiseer. Love (2002:40-50) verwys na vier verskillende soorte outeurskap:

- **Inleidende outeurskap**: Wanneer ’n aansienlike aantal teks van ’n ander outeur by ŉ nuwe teks ingewerk word, word daar na die eerste outeur as ’n inleidende outeur verwys. Volgens Love (2002:40) is die inleidende outeur enige persoon wat as ’n bron of invloedryke verwysing in die teks dien.

- **Uitvoerende outeurskap**: Dit verwys na die konstruksie van die teks. ŉ Outeur kan volgens Love (2002:43) gedefinieer word as die samesteller van ŉ teks tot op die punt waar die teks gereed is om gepubliseer te word (ongeag die vorm − gesproke of geskrewe).

- **Verklarende outeurskap**: Hierdie tipe outeurskap verwys na ’n outeur wat gesag aan ŉ teks verleen. Die individu is nie noodwendig die outeur nie − in dié geval sal die ware outeur ’n kombinasie van inleidende outeurskap en ’n vorm van borgskap (*sponsorship* of *fostering*) behels (Love, 2002:44). Love (2002:45) verwys verder na die moontlikheid dat sekere individue “eienaarskap van die teks neem”. Individue sal in die publiek as die outeur optree en ook verantwoordelikheid vir die teks neem.

[79] Sien afdeling 2.3.2.3 (Idiolek).

- **Hersieningsouteurskap**: Hierdie tipe outeurskap verwys volgens Love (2002:46) na teksredigering.

Vanuit Love se definisies kan daar dus tussen vier verskillende soorte outeurskap onderskei word. Alhoewel die tipe outeurskap nie altyd bekend is nie, is dit moontlik om outeurs van tekste te identifiseer mits die betrokke tekste deeglik geanaliseer word. Na gelang van die bogenoemde is dit duidelik dat outeurskapidentifikasie na die analise van tekste verwys ten einde ŉ antwoord op die vraag “Wie het die tekste geskryf?” te vind. Daar is ook verskeie faktore wat tekste kan beïnvloed, en die soort outeurs het ook ’n invloed op die bevindinge. Voorts verwys ek na benaderings tot outeurskapidentifikasie.

3.5.1 Benaderings tot outeurskapidentifikasie

Daar is verskeie maniere waarop die outeurs van tekste bepaal kan word en alhoewel verskillende metodologieë in hoofstuk 5 bespreek word, kyk ek in hierdie afdeling oorsigtelik na van die benaderings tot outeurskapidentifikasie. Brocardo *et al*. (2013), Naga Prasad *et al*. (2015:59) en Nirkhi *et al*. (2015:369) stel dat outeurskapanalise vanuit drie verskillende perspektiewe benader kan word:

i. **Outeurskapidentifikasie**: Dui op die vasstelling van die waarskynlikste outeur van ’n teks.
ii. **Outeurskapverifiëring**: Verwys na die bevestiging dat ’n teks deur ’n bepaalde outeur geskryf is.
iii. **Outeurskapkarakterisering**: Dui op die samestelling van ’n profiel (moontlike ouderdom, geslag, etnisiteit en dies meer) van die (onbekende) outeur van ’n anonieme teks.

Koppel en Schler (2004) voer aan dat die verifiëring van outeurs moeiliker is om uit te voer as ŉ basiese outeurskapidentifikasie. Coulthard en Johnson (2007:14) argumenteer dat wanneer linguiste ’n teks analiseer, daar menige diverse interpretatiewe metodes, teorieë en programmatuur is wat in ŉ analise gebruik kan word. Wanneer ’n teks forensieslinguisties ondersoek word, is die metodes, teorieë en programmatuur wat in die analise gebruik kan word nog meer divers.

Coulthard en Johnson (2007:14) stel dat forensiese linguiste eerstens moet bepaal wat die ooreenskomste in die bepaalde teks (wat ondersoek word) is; tweedens, hoe die teks van ander verskil en derdens, wat die gepaste metodes, programmatuur en benaderings vir die spesifieke teks se analise is. Wanneer ’n teks forensieslinguisties ondersoek word, is daar dus verskeie

vrae wat gevra moet word alvorens die korpusanalise van die teks in aanvang kan neem. Die vrae sluit in:

- Hoe is die teks geskep?
- Hoeveel persone was betrokke by die skep van die teks?
- Watter soort persoon het die teks geskryf?
- Wat is die verhouding van die teks in terme van vergelykende tekste?

Grant (2008:221) argumenteer dat elk van die bogenoemde vrae ʼn spesifieke linguistiese benadering vereis ten einde die spesifieke vraag te kan beantwoord. Lawrence en Solan (2004:457) verwys na drie benaderings tot outeurskapidentifikasie wat effektief blyk te wees.

i. ʼn **Eklektiese benadering**: Hierdie benadering verwys na ʼn metode wat die samevloeiing van ooreenkomste ondersoek om sodoende outeurs te identifiseer. Spelling, punktuasiegebruik, taalfoute en dies meer kan ondersoek word. Tekste word met mekaar vergelyk ten einde hierdie kenmerke te kan identifiseer. Dokumente soos deur die betrokke (verdagte) individu geskryf, word geanaliseer met die doel om ongewone linguistiese kenmerke te identifiseer (vgl. Lawrence & Solan, 2004:458).

ii. ʼn **Diagnostiese benadering**: Hierdie benadering probeer volgens Lawrence en Solan (2004:259) om stilistiese kenmerke te vind wat kenmerkend is van individue se taalgebruik. In teenstelling met die eklektiese benadering kan die diagnostiese benadering as ʼn “onder-na-bo”-benadering beskou word. Chaski (2001) het veral in hierdie gebied navorsing gedoen en stel voor dat ʼn korpus saamgestel word. ʼn X-aantal dokumente word dan voorgehou as verdagte tekste en een dokument is die verdagte se ware dokument. Al die dokumente word dan met mekaar vergelyk (verskillende kenmerke word ondersoek) en daar word bepaal watter dokument stem die meeste ooreen met die verdagte se ware dokument.

Chaski (2001) voer verder aan dat spelling, soort-ratio, teken-ratio en woordkeuse nie noodwendig diagnosties van aard is nie, maar dat sintaktiese verskille en die gebruik van punktuasie wel ʼn aanduider van outeurskap is. Kredens (2000) voer ook aan dat sintaktiese verskille (of ooreenkomste) en diskoersmerkers belangrik is in die ondersoek na outeurskapidentifikasie.

iii. ’n **Korpusbenadering**[80]: Volgens Lawrence en Solan (2004:260) kan hierdie benadering vergelyk word met die gebruik van DNS in identifikasieprosesse. Wanneer hierdie benadering gebruik word vir outeurskapidentifikasie, is dit van kardinale belang om oor ’n verwysingskorpus[81] te beskik wat gebruik kan word om die ontledingskategorieë mee te vergelyk. ’n Korpus verwys volgens Crystal (2008:117) na ’n versameling linguistiese data wat as beginpunt vir ’n linguistiese beskrywing kan dien of die basis waarop ’n bepaalde hipotese oor taal geverifieer kan word.

Chaski (2008:1) voer aan dat die metodes wat in outeurskapidentifikasie gebruik word in kontras met mekaar staan, en verwys spesifiek na:

- **Kwalitatief teenoor kwantitatief**: Word die linguistiese kenmerke aangeteken of getel?
- **Beskrywend teenoor voorskrywend**: Word die dokumente deur ’n implisiete of eksplisiete standaard geëvalueer of word dit beskryf deur enige van die inhoud of deur alles?
- **Analise teenoor masjienontleding**: Moet die metode deur ’n kundige uitgevoer word of kan die metode geoutomatiseer word?
- **Teorie teenoor toepassing**: Berus die metode op ’n taalteorie of verwys dit na analiste se eie ondervinding met ’n taal of ’n teks?

Ongeag die keuse wat gemaak word in terme van die metode, kan daar tussen verskillende tipes bewyse in outeurskapidentifikasie onderskei word. Love (2002:51) onderskei tussen eksterne en interne bewyse. Eksterne bewyse verwys volgens Love (2002:51) na:

- kontemporêre eienskappe wat implisiet, eksplisiet of in titels voorkom sowel as in dokumente wat inligting oordra van gebeure of omstandighede (soos dagboeke, rekords, briewewisseling en notas van uitgewers);
- biografiese bewyse wat bewys kan lewer oor die beweerde outeur; en
- die geskiedenis van vorige weergawes van die teks en die omstandighede waartydens die teks geskep is.

In teenstelling met eksterne bewyse verwys interne bewyse volgens Love (2002:51) na:

- stilistiese bewyse;
- verwysing na die “self” in tekste; en

[80] Die korpusbenadering word in hoofstuk 4 bespreek.
[81] ’n Verwysingskorpus verwys na die prototipiese korpus wat as ’n voorstelling van die norm van ’n bepaalde taal beskou kan word. Hierdie korpus is gewoonlik baie groot en is afkomstig van verskeie bronne aangesien dit die norm van ’n taal moet voorstel (vgl. Baker, 2010:12; Nesselhauf, 2005). Sien ook korpuslinguistiek (hoofstuk 4 – Korpuslinguistiek).

- bewyse wat sigbaar is na aanleiding van die tematiek, idees en opinies in die teks sowel as die genre van die betrokke teks.

Eksterne bewyse word gebruik om ’n teks in ’n bepaalde tydskonteks te plaas en daar word ook na die kulturele agtergrond van die teks gekyk − die teks word dus gekontekstualiseer (Love, 2002:52). Interne bewyse verwys na die gebruik van ’n interne metode (kwantitatief) of wanneer daar statisties bewys word dat ’n bepaalde individu ’n gegewe teks geskryf het. In die meeste gevalle word ’n stilometriese benadering tot die analise gevolg. Chaski (2008:3) onderskei tussen drie benaderings tot forensiese linguistiek, naamlik: ’n kwalitatiewe, rekenaarstilistiese en ’n rekenaarlinguistiese benadering en elk van hierdie benaderings beskik oor ’n spesifieke formulering[82].

Coulthard en Johnson (2007:121) sluit by bogenoemde aan en postuleer dat forensiese linguiste ’n analise foneties, fonologies, morfologies, sintakties, pragmaties, tekstueel of as diskoers kan benader terwyl Naga Prasad *et al*. (2015:60) van mening is dat daar tussen vier kategorieë onderskei kan word wanneer kenmerke vir ’n outeurskapidentifikasie-analise geïdentifiseer word, naamlik: leksikale, sintaktiese, konteksspesifieke en strukturele kenmerke.

Vanuit die bogenoemde bespreking blyk dit duidelik dat daar verskeie maniere is waarop outeurskapidentifikasie benader kan word. Hierdie benaderings kan gekombineer word ten einde die bevindinge te verifieer. Daar blyk twee benaderings te wees wat die leksikale, sintaktiese, konteksspesifieke en strukturele kenmerke as geheel ondersoek en dus as oorkoepelend beskou kan word, naamlik forensiese stilistiek en forensiese stilometrie.

3.5.1.1 Forensiese stilistiek

Stilistiek verwys volgens McMenamin (2010) na die bestudering van styl in taal en kan na literêre stilistiek of linguistiese stilistiek verwys. Linguistiese stilistiek kan gedefinieer word as ’n analise van sigbare stylmerkers soos dit deur ’n groep of individue gebruik word en wat as ’n kwalitatiewe analise beskou kan word. McMenamin (1993; 2002) het veral navorsing oor hierdie veld gedoen, en het bevind dat taal beskou kan word as keuses wat deur individue gemaak word in terme van individuele en sosiale voorkeure. Hy (McMenamin, 2002:97) argumenteer dat stylvariasie inherent is aan individuele taalvermoëns.

Linguistieke stilistiek verwys dus na ’n analise van individuele stylmerkers waar die variasie van ’n norm of die afwyking van ’n norm die aanwesigheid en identifisering van ’n norm aandui. ’n

[82] Sien hoofstuk 5.

Stilistiese analise het dus betrekking op die identifisering van stylmerkers in ’n bepaalde teks, en die moontlikheid dat hierdie stylmerkers kan afwyk van dít wat as die norm beskou word. Aspekte wat ondersoek word in 'n stilistiese analise sluit in teksformaat, getalle en simbole, afkortings, leestekengebruik, hoofletters, spelling, woordvorming, sintaksis, diskoers, frekwente woorde en frases asook foute en regstellings[83].

Volgens Thiart (2014:59) is 'n stilistiese analise ’n belangrike eerste stap in outeurskapidentifikasie aangesien kenmerke geïdentifiseer word wat verder in 'n stilometriese analise ondersoek word. Die metode en die kenmerke wat in my stilistiese ondersoek gebruik is, word verder in hoofstuk 5 (Metodologie) bespreek.

3.5.1.2 Forensiese stilometrie

Volgens Burrows (1992) kan daar veronderstel word dat alle outeurs oor 'n kenmerkende skryfstyl beskik. Burrows (1992) gaan voort en verduidelik dat hierdie skryfstyl sigbaar is in individue se woordeskat, sinskonstruksie (kompleksiteit van sinne) en woordorde, en dat hierdie kenmerke onbewustelik in individue se taalstyl figureer. Hierdie kenmerke kan geanaliseer en getel word volgens Burrows (1992).

Wanneer stilometrie gebruik word om outeurskap te bepaal, is ’n verwysingskorpus[84] nodig, en hierdie soort analise word as kwantitatief beskou. Volgens Grant (2008:224) is 'n stilometriese benadering meer statisties van aard. Chaski (2008:5) verwys na stilometrie as ’n kwantitatiewe benadering waar rekenaarprogrammatuur gebruik kan word om berekenbare of telbare taalverskynsels (soos woordlengte, sinslengte, frekwensie van woordgebruik en dies meer) te ondersoek en te identifiseer.

Verskillende kenmerke kan in 'n analise ondersoek word wanneer ’n stilometriese benadering gevolg word. Barry en Luna (2012:2) identifiseer 12 kenmerke wat in aanlyn tekste ondersoek kan word (sien Tabel 3.2):

Tabel 3.2 Kenmerke van 'n ondersoek in ’n stilometriese analise (Barry & Luna, 2012:2)

Kenmerk	**Beskrywing**
Punktuasie	Persentasie karakters wat lees of skryftekens is
Tyd	Tyd van opdatering
Afkappings per woord	(Aantal afkappings) / (Aantal woorde)
Hoofletters	Persentasie woorde wat in hoofletters geskryf is
Karakters per woord	(Aantal karakters) / Aantal woorde)

83 Hierdie kenmerke wat in die stilistiese analise ondersoek word, word in hoofstuk 5 (afdeling 5.6.1 – Stilistiese analise) bespreek word.
84 Sien hoofstuk 5 (Metodologie).

Kenmerk	Beskrywing
Letters	Persentasie karakters wat letters is
Kompleksiteit	(Aantal unieke woorde) / (hoeveelhheid woorde)
Karakters	Aantal karakters in die teks
Woorde	Aantal woorde in die teks
Witspasies	Persentasie karakters wat witspasies is
Syfers	Persentasie karakters wat syfers is
Bigram	Frekwensie van twee karakter kombinasies

Hierdie 12 kenmerke − soos gesien in Tabel 3.2 − kan as ’n riglyn dien vir ’n stilometriese ondersoek. De Vel *et al*. (2001) het na meer as net 12 kenmerke ondersoek ingestel ten einde die bevindinge van hulle stilometriese ondersoek te verifieer. Die kenmerke wat hulle ondersoek het, sluit in:

- aantal reëls in die teks asook aantal oop reëls in die teks
- gemiddelde sinslengte
- gemiddelde woordlengte
- woordeskat (soort-/teken-ratio)
- totale aantal funksiewoorde (teken)
- verspreiding van die funksiewoordfrekwensie
- *hapax legomena* (woorde wat slegs een maal voorkom) en *hapax dislegomena* (woorde wat twee maal voorkom)
- karakters per woord
- karakters per teks
- hooflettergebruik
- totale aantal syfers per woord

Brocardo *et al*. (2013) sluit by Koppel en Schler (2004) en De Vel *et al*. (2001) aan en stel dat die verifiëring van outeurskap veral belangrik is wanneer daar na RBK of aanlyn dokumente ondersoek ingestel word.

Nirkhi *et al*. (2015:370) argumenteer dat outeurskapidentifikasie deur middel van karakter-n-gramontleding ’n waardevolle bydrae kan lewer in ’n stilistiese analise in digitale forensiese ondersoeke. Volgens Nirkhi *et al*. (2015:370) kan hierdie n-gramme belangrike kenmerke aandui op leksikale, sintaktiese en strukturele vlak. Nirkhi *et al*. (2015:370-371) onderskei tussen ses verskillende (sien Tabel 3.3) maniere waarop n-gramme gebruik kan word tydens ’n leksikale ondersoek (hierdie leksikale ondersoek vorm deel van die stilometriese ondersoek).

Tabel 3.3 Beskrywing van die uiteensetting van n-gramme in tydens ’n stilometriesee ondersoek gebruik kan word (geneem vanuit Nikhri *et al.,* 2015:371)

Kenmerk	**N-gramvlak**	**Beskrywing**	**Kenmerktipe**
Karakter-unigram	1-gram	Individuele karakter	Karakters
Karakter-bigram	2-gramme	Twee opeenvolgende karakters	Karakters
Karakter-trigram	3-gramme	Drie opeenvolgende karakters	Karakters
Woord-unigram	1-gram	Enkele woord	Leksikaal
Woord-bigram	2-gramme	Twee opeenvolgende woorde	Leksikaal
Woord-trigram	3-gramme	Drie opeenvolgende woorde	Leksikaal

Die stilometriese ondersoek kan dus ook fokus op n-gramme soos in Tabel 3.2 beskryf word. Thiart (2014:100-101) verwys na leestekens, aanspreekvorme, lagtekens, logogramme, ongewone gebruik van hoof- en kleinletters, verkorting, niestandaardspelling, Engelse woorde/sinne, funksiewoorde en individuele eienskappe (herhalende woorde, frases en/of leestekengebruik) in haar forensieslinguistiese ondersoek na Afrikaanse SMS-taal.

Afdeling 3.5 verwys vlugtig na van die benaderings tot outeurskapidentifikasie. Wat duidelik uit bogenoemde na vore kom, is dat outeurskapidentifikasie in die verlede veral gefokus het op langer tekste. Ontwikkelinge in tegnologie en die maniere waarop individue kommunikeer, het ’n behoefte geskep om metodes te ontwikkel wat op korter tekste toegepas kan word. Vervolgens verwys ek oorsigtelik na outeurskapidentifikasie in ’n digitale era.

3.6 Outeurskapidentifikasie in ’n digitale era

Bhargava *et al.* (2013:37) toon aan dat navorsing in tale al hoe meer fokus op outeurskapidentifikasie van elektroniese tekste en dat dokumente korter as 1 000 woorde in die vroeë 19de eeu as “moeilike dokumente om te identifiseer” beskou was. Die aantal woorde wat gebruik word om outeurskapidentifikasie toe te pas neem toenemend af, en vandag vind identifikasie plaas op tekste vanaf aanlyn platforms soos Facebook en Twitter[85]. Bhargava *et al.* (2013:37) gaan voort deur te argumenteer dat die ontleding van en navorsing oor hierdie

[85] Sien afdeling 2.6 (Sosiale media).

toenemende korter tekste van kardinale belang is aangesien kommunikasie voortdurend korter word.

Crystal (2011:10) wys daarop dat navorsing oor die taal van die internet verskeie uitdagings aan navorsers stel:

- **Data**: Die aantal data beskikbaar op die internet skep die eerste uitdaging. Wanneer die internet as 'n korpus beskou word, is dit inderdaad die grootste korpus ooit. Die internet beskik oor meer voorbeelde van geskrewe taal as die somtotaal van alle biblioteke in die wêreld. Dis belangrik om in gedagte te hou dat die aantal data wat op die internet beskikbaar is daagliks vermeerder.

- **Diversiteit**: Crystal (2011:10) verwys na die diversiteit van tale wat op die internet gevind word. Hy voer aan dat die taal wat gevind word in kletskamers, op Facebook, blogs, IM en dies meer, elk 'n unieke voorbeeld van kommunikasieperspektiewe, -eienskappe, -strategieë en -verwagtinge bied.

- **Verandering**: Die spoed waarteen die kommunikatiewe moontlikhede van die internet verander, dra by tot die komplekse aard van navorsing oor die taalgebruik wat op die internet beskikbaar is.

- **Anonimiteit**: Die anonieme aard van die internet skep verder volgens Crystal (2011:13) ook nuwe vraagstukke wat navorsers van hierdie interaktiewe medium moet beantwoord. Waar daar tradisioneel sterk gefokus is op die ouderdom, geslag, sosiale klas en etnisiteit kan hierdie gegewens nou moeiteloos deur gebruikers verbloem of vervals word. Alhoewel anonimiteit gebruikers aanspoor om meer gemaklik en ongeïnhibeerd te kommunikeer, laat dit vanuit 'n sosiolinguistiese benadering ruimte vir talle onbeantwoorde vrae.

- **Etiese vraagstukke**: Die internet skep ook talle etiese vrae. Wie gee byvoorbeeld toestemming dat data vanaf die internet vir navorsing gebruik kan word[86]?

Vanuit die bogenoemde blyk dit duidelik dat die internet as kommunikasiemedium verskeie uitdagings aan navorsers bied, maar ook dat ons hier met taalvorme te doen het wat nog nooit voorheen gesien is nie. Navorsing het in die verlede die verskil tussen geskrewe en gesproke

[86] Sien afdeling 5.4 (Data-insameling en -verwerking).

taal duidelik aangedui, maar Crystal (2011:21) argumenteer dat die onderskeid tussen geskrewe en gesproke taal – sigbaar in internettaal − besig is om te vervaag.

Barry en Luna (2012:1) voer aan dat een van die internet se mees integrale kenmerke die moontlikheid tot absolute anonimiteit is, en hoewel hierdie anonimiteit dalk voordelig vir individue kan wees indien hulle hul in lande/situasies bevind waar hulle onderdruk word, is anonimiteit nie wenslik vir regstoepassing nie. Gressel *et al*. (2014:1148) sluit by bogenoemde aan wanneer hulle argumenteer dat alhoewel aanlyn sosiale media kommunikasie vergemaklik en individue nader aan mekaar bring, gee dit ook aanleiding daartoe dat gebruikers kwesbaar is vir misdade soos identiteitsdiefstal, die ontvang van onwaar inligting en vervalsde dokumente. Menige individu skep ’n anonieme aanlyn profiel en raak betrokke in ’n verskeidenheid van kubermisdade (vgl. Gressel *et al.* 2014:1148). Die anonieme aard van die internet skep dus die noodsaaklikheid om ondersoek in te stel na metodes waardeur anonieme outeurs op hierdie platforms geïdentifiseer kan word.

Daar is reeds verwys na Grant (2008:216) se argument dat sekere tekste geanaliseer word om lig te werp op ’n bepaalde kwessie om sodoende ’n spesifieke ondersoek te voltooi of om ’n alibi te verifieer. In ’n digitale era is hierdie bydra tot geregtelike ondersoeke van kardinale belang. Kommunikasie is oorwegend rekenaarbemiddeld van aard in die digitale era waarin ons ons bevind, en dus is daar aspekte wat geregtelike ondersoeke van hulp kan wees. Die manier waarop elektroniese kommunikasie se oorsprong (geografiese ligging) vasgestel kan word, dien as voorbeeld.

Verdagtes (of selfs individue wat as vermis aangemeld is) kan ’n inskrywing op ’n sosiale media platform plaas wat individue se geografiese ligging kan bevestig ten tyde van misdrywe. In die geval van vermiste persone of wanneer daar ’n kriminele inskrywing op ’n spesifieke gebruiker se profiel geplaas word, ontstaan die vraag dus of die vermiste persoon of die verdagte die ware outeur van die teks[87] is?

Wanneer daar na navorsing oor metodes van outeurskapidentifikasie van Afrikaanse tekste op mikroblogs ondersoek ingestel word, blyk dit duidelik dat daar nog nie so ’n metode vir Afrikaans bestaan nie. My studie het dus ondersoek ingestel na ’n moontlike metode. Vervolgens verwys ek na navorsing wat reeds gedoen is oor outeurskapidentifikasie in ’n digitale era met ’n spesifieke fokus op die benaderings wat gebruik word om outeurskapidentifikasie uit

[87] In die geval kan daar veral na Olsson (2009) se forensieslinguistiese analises in die boek Wordcrime: Solving crime through forensic linguistics gekyk word (sien ook afdeling 3.4.1 Forensiese linguistiek: die ontstaan).

te voer. Die metodologie word breedvoerig in hoofstuk 5 bespreek. In afdeling 3.5 word daar verwys na outeurskapidentifikasie sowel as benaderings tot outeurskapidentifikasie.

Green en Sheppard (2013:64) en Schwarts *et al*. (2013:1880) stel dat navorsing oor outeurskapidentifikasie meestal gefokus het op lang tekste (soos letterkundige tekste), maar dat hierdie minimum lengte van tekste wat gebruik word in outeurskapidentifikasie nie meer van toepassing is nie, aangesien tekste deesdae al hoe korter word. Verder is dit belangrik dat tegnieke en metodes wat gebruik kan word vir outeurskapidentifikasie van hierdie korter, informele tekste (soos e-posse, forums, bulletinborde, blogs en sosialemediaplatforms) ondersoek word aangesien hierdie tekstipes vrae oor ware outeurs (die anonieme aard van tekste), sekuriteitsvrae en bemarkingsvrae (soos wie die teikengehoor is) kan beantwoord (Green & Sheppard, 2013:64).

Die beperkte lengte van mikrobloginskrywings (twiets) is volgens Green en Sheppard (2013:64) ʼn uitdaging vir navorsers aangesien daar vir lank beweer is dat daar ʼn minimum tekslengte bestaan waarin identifiserende karaktereienskappe nie meer sigbaar sal wees nie. Ledger en Merriam (1994:240) voer aan dat die minimum tekslengte 500 woorde is terwyl Forsyth en Holmes (1996) die getal na 250 verminder het. Green en Sheppard (2013:64) stel dat hierdie minimum tekslengte steeds aansienlik hoër is as die gemiddelde lengte van ʼn twiet − die gemiddelde lengte van ʼn twiet is minder as 25 woorde (gewoonlik selfs minder as 10 woorde per twiet).

Ten spyte van die uitdagings verbonde aan outeurskapidentifikasie waar daar ʼn beperkte aantal teks beskikbaar is vir analise, is daar steeds individuele taaleienskappe wat in die skryfstyl van individue voorkom wat gebruik kan word in die identifikasie van ʼn bepaalde outeur.Coulthard (2005:15) verwys na die manier waarop individue taalreëls verontagsaam en verbreek, en onderskei tussen twee kategorieë, naamlik:

- **Prestasiefoute**: Hierdie taalfoute verwys na reëls wat verbreek word waarvan taalgebruikers onbewus is en wat konsekwent verbreek word. Hier word verwys na algemene taalfoute.

- **Bevoegdheidsfoute**: Hierdie taalfoute verwys na reëls wat doelbewus verbreek word wanneer taalgebruikers te doen het met ʼn stel reëls (of vereistes wat daargestel word deur die kommunikasiemedium) wat niestandaard is, maar wat steeds deurlopend gebruik word.

Wanneer die aard van RBK (en spesifiek die vereistes van mikroblogplatforms) ondersoek word, is dit duidelik dat daar veral na **bevoegdheidsfoute** ondersoek ingestel moet word wanneer outeurskapidentifikasie plaasvind. Die niestandaarduitdrukkings ten opsigte van die taalreëls van mikroblogs sal dus geïdentifiseer moet word om sodoende die algemene foute van dié platform te bepaal alvorens unieke kenmerke van individue ondersoek kan word. Aangesien my studie op mikroblogs gefokus het, het ek nie vir die doel van die studie ondersoek ingestel na ander platforms (soos Facebook of Instagram) om te bepaal of die foute beperk word tot mikroblogs en of dit ook op ander sosialemediaplatforms voorkom nie[88].

’n Bykomende aspek wat Coulthard (2005:16) uitlig in verband met die ondersoek van foute wanneer outeurskapidentifikasie op elektroniese tekste uitgevoer word, is die moontlikheid dat hierdie foute dalk slegs tikfoute kan wees. Wanneer die aard van RBK in ag geneem word, kom hierdie soort fout gereeld voor aangesien tekste wat ondersoek word nie geskep word met die doel om as hardekopieë gelees te word nie. Inteendeel, die kommunikasie op mikroblogs is beknop en vinnig en bedoel om op die platform (aanlyn) gelees te word. Vanuit Coulthard se stelling oor tikfoute sal foute (wat as tikfoute beskou word) herhaaldelik moet voorkom alvorens dit as deel van individue se idiolek of kenmerkend aan die taal van die platform beskou kan word.

Dit is duidelik dat daar talle ontwikkelinge in die veld van outeurskapidentifikasie voorkom − veral as gevolg van die feit dat tekste verkort. Mikroblogs (soos gesien in hoofstuk 2) is boodskappe wat uit ’n maksimum van 140 karakters bestaan. Om outeurs van hierdie korter tekste te identifiseer, sal daar van ’n kombinasie van metodes gebruik gemaak moet word. Hierdie metodes word in hoofstuk 4 bespreek.

3.7 Samevatting

Hierdie hoofstuk het ’n kritiese literatuuroorsig van forensiese linguistiek as dissipline aangebied. Die bydrae van hierdie hoofstuk kan beskou word as die oorsig van die ontwikkeling van forensiese linguistiek, wêreldwyd en dan ook in Suid-Afrika, en hoe hierdie vakgebied verander het ten einde tegnologiese ontwikkelinge – veral ten opsigte van kommunikasie – te kan inkorporeer in die analise. Die term *forensiese linguistiek* is gedefinieer en ’n tematiese oorsig van navorsing in dié veld is aangebied (ook met betrekking tot navorsing in Afrikaans). Outeurskapidentifikasie is gedefinieer en bespreek, só ook verskillende benaderings tot outeurskapidentifikasie. Ten slotte is daar na outeurskapidentifikasie in ’n digitale era gekyk, en belangrike navorsing in dié veld (spesifiek met betrekking tot outeurskapidentifikasie en

[88] Hierdie tipe ondersoek kan as bykomende navorsing beskou word.

mikroblogs) is hanteer. Hierdie hoofstuk bied dan tesame met hoofstuk 2 en 4 die agtergrond wat benodig word om in hoofstuk 5 die metodologie voor te hou en die toepassing van die metode in hoofstuk 6 en 7.

Die vraag “wie is die werklike outeur van die teks” wek al vir jare opspraak. Die term *forensiese linguistiek* is egter eers in 1968 deur Jan Svartvik begin gebruik. Forensiese linguistiek is dus in wese nog ’n jong dissipline, en die ontwikkeling van taal, die wyse waarop die kommunikasiemediums taalgebruikers noodsaak om hulle taal aan te pas sodat dit by gebruikersbehoeftes aanpas sowel as die hibriede aard van RBK skep talle geleenthede vir die ontwikkeling van nuwe metodologieë in forensiese linguistiek.

Dit blyk duidelik vanuit hierdie hoofstuk dat daar al talle navorsing oor forensiese linguistiek gedoen is. Ook in Suid-Afrika is daar etlike navorsers (ook in Afrikaans) wat in dié vakgebied spesialiseer. Tog blyk dit duidelik dat daar nog heelwat leemtes in die vakgebied bestaan − veral met betrekking tot Afrikaans in ’n digitale era, en dit is hier waar hierdie studie ’n bydrae kan lewer.

Daar word tussen twee hoofstrominge in forensiese linguistiek onderskei, naamlik die taal van die reg en outeurskapidentifikasie. Vir die doel van my studie het ek op outeurskapidentifikasie gefokus. Outeurskapidentifikasie word beskou as die analise van ’n teks (hetsy geskrewe of gesproke) ten einde die outeur van die betrokke teks te identifiseer. Outeurskapidentifikasie kan ook na óf outeurskapverifiëring óf outeurskapkarakterisering verwys[89]. Wanneer daar ondersoek ingestel word na die outeurs van mikrobloginskrywings, is dit moontlik dat al drie hierdie vertakkings ondersoek kan word aangesien die presiese aard van die navraag nie noodwendig gedefinieer kan word nie − wie is die outeur van hierdie teks, het persoon A die teks geskryf of is persoon B werklik ’n tiener?

Die anonieme aard van die internet gee aanleiding tot die dringende beantwoording van bogenoemde vrae. Dit word al hoe makliker vir individue om profiele te skep waar hulle nie hul ware identiteit bekend maak nie of profiele word geskep wat vir *trolling* gebruik word. Dit is ook duidelik dat die taal van RBK as ’n hibriede kombinasie van gesproke en geskrewe taal beskou kan word en dat hierdie aspek van RBK ingedagte gehou moet word wanneer metodes vir die identifikasie van outeurs ondersoek word.

Daar is verskeie benaderings wat gevolg kan word om outeurs van tekste te identifiseer. Ten einde die resultate te verifieer, is dit nodig om van benaderings gebruik te maak wat in kontras

[89] Sien afdeling 3.5.1 (Benaderings tot outeurskapidentifikasie).

met mekaar staan hetsy kwalitatief teenoor kwantitatief, beskrywend teenoor voorskrywend, analiserend teenoor masjienontledings of teorie teenoor toepassings. Die identifikasie van outeurs in ’n digitale era bied ook talle uitdagings aan navorsers. ’n Groter aantal data is geredelik beskikbaar, die diversiteit van tale, die spoed waarteen veranderinge op die internet plaasvind sowel as etiese vraagstukke is maar net enkele van die uitdagings wat voorkom.

Waar navorsing in die verlede op langer tekste uitgevoer is, moet die minimum lengte van ’n korpus wat vir outeurskapidentifikasie gebruik word ook telkens verminder word. ’n Mikrobloginskrywing bestaan byvoorbeeld uit ’n maksimum van 140 karakters. ’n Digitale era het dus tot gevolg dat kommunikasie al hoe bondiger word. Individue wil kort boodskappe lees. Daar is nie meer tyd vir langdradige en omvattende tekste nie. Inteendeel, hoe korter en meer kompak die kommunikasie, hoe beter.

Soos genoem in afdeling 3.6, stel Green en Sheppard (2013:64) dat die beperkte lengte van mikrobloginskrywings (twiets) ’n uitdaging stel aan navorsers om te bepaal wat die minimum tekslengte is om outeurs van sulke tekste te identifiseer. Bhargava *et al*. (2013:37) het verwys na navorsing wat reeds in ander tale uitgevoer is, en dat die aanvanklike minimum tekslengte van 1000 woorde nie meer van toepassing is nie. Ek het ook verwys na Forsyth en Holmes (1996) asook Ledger en Merriam (1994) wat aanvoer dat die minimum tekslengte vir outeurskapidentifikasie onderskeidelik 250 en 500 woorde is. Die vraag wat nou ontstaan is wat die minimum tekslengte is waardeur die outeur van ’n Afrikaanse teks geïdentifiseer kan word.

Dit blyk dat die manier van kommunikasie verander het, en dat RBK dit moontlik maak vir gebruikers (in dié geval mikrobloggebruikers) om met korter tekste al hoe vinniger te kommunikeer. ’n Hibriede taal word geskep deur korter tekste − ’n taal wat op die internet sigbaar is en wat as ’n kombinasie van gesproke en geskrewe taal beskou kan word. Wanneer die aard van mikroblogs sowel as die tradisionele minimum lengte vir tekste wat geanaliseer word in ag geneem word, is dit duidelik dat daar ondersoek ingestel moet word na moontlike metodes vir die outeurskapidentifikasie van Afrikaanse mikrobloggebruikers.

Hoofstuk 2 en 3 skep saam die teoretiese agtergrond wat gebruik is in hoofstuk 5 (Metodologie) en in hoofstuk 6 en 7 (Empiriese analise) ten einde die studie te kon uitvoer. Alvorens daar na die metodologie ondersoek ingestel kon word, moes die korpuslinguistiek voorgehou word. Soos gesien word vanuit hoofstuk 4 is daar talle opinies beskikbaar oor die aard van korpuslinguistiek en die manier waarop korpuslinguistiek en forensiese linguistiek gebruik kan word ten einde outeurs van tekste te identifiseer.

Hoofstuk 4

Korpuslinguistiek

What cannot be said above all must not be silenced but written.

Derrida (1974)

4.1 Inleiding

Hierdie hoofstuk het ten doel om korpuslinguistiek te bespreek ten einde 'n volledige agtergrond te skep oor die beginsels wat in die metodologiese en empiriese hoofstukke (5, 6 en 7) gebruik en toegepas is.

In hierdie hoofstuk definieer en verduidelik ek eerstens die fenomeen korpus en verwys tweedens na die verskillende soorte korpora. Nadat die konsep korpus verduidelik is, verskaf ek 'n oorsig van die historiese ontwikkeling van korpuslinguistiek (wat nie sonder 'n korpus uitgevoer kan word nie) en die verskillende definisies van korpuslinguistiek word bespreek.

Daar bestaan talle opinies oor die aard van korpuslinguistiek, veral wat betref die vraag of korpuslinguistiek as 'n teorie of 'n benadering beskou moet word. In hierdie hoofstuk word dié debat ook vlugtig ondersoek ten einde korpuslinguistiek te kon definieer vir my studie.

Ten slotte verwys ek na wat 'n korpuslinguistiese analise behels alvorens ek korpuslinguistiek in perspektief plaas met betrekking tot forensiese linguistiek. Hierdie hoofstuk dien saam met hoofstuk 2 en 3 as 'n teoretiese oorsig wat toegepas is in die samestelling van die metodologie in hoofstuk 5 alvorens die empiriese analise in hoofstukke 6 en 7 kon plaasvind.

Korpuslinguistiek het vanaf die 1990's snel ontwikkel[90] en vandag word korpuslinguistiek in talle navorsingsareas gebruik. Taalstudie gebaseer op die taalgebruik van individue of groepe soos dit in die werklikheid gebruik word, is volgens Baker (2010a:15) die eenvoudigste definisie van korpuslinguistiek. Hierdie korpuslinguistiese[91] benadering het in die afgelope 20 jaar ongelooflik baie gewildheid verwerf. Die impak wat korpusstudies[92] op toegepaste taalkunde het, is enorm − dit verander die manier waarop taal verstaan en bestudeer word (vgl. Hyland *et al.*, 2012:3). Anthony (2013:141) sluit by die bogenoemde aan en postuleer dat korpuslinguistiek 'n benadering is wat een van die dominante metodes geword het om taal mee te analiseer. Hyland

[90] Sien afdeling 4.3 ('n Historiese oorsig van korpuslinguistiek).
[91] Sien afdeling 4.4 (Korpuslinguistiek gedefinieer).
[92] Die term korpus word in afdeling 4.2 (Korpora) gedefinieer, maar kan voorlopig beskou word as 'n versameling van natuurlike tekste vanuit 'n taal (vgl. Tavakoli, 2012:113).

et al. (2012:3) argumenteer verder dat studies met betrekking tot korpora ’n empiriese dimensie by taalkunde voeg (wat nie voorheen aanwesig was nie) en dat dit navorsers in staat stel om ’n intuïtiewe hipotese deur middel van die interpretasie van die geanaliseerde data te versterk en bewys.

4.2 Korpora

Om korpuslinguistiek as ’n geheel te verduidelik, beskryf ek stelselmatig in die volgende afdelings die onderskeie konsepte wat deel van korpuslinguistiek uitmaak, en ek verskaf ’n agtergrond van die historiese ontwikkeling van korpuslinguistiek ten einde korpuslinguistiek soos dit vandag beskou word te kan verduidelik. Die eerste konsep wat dus van belang is, is korpora.

Lüdeling en Kytö (2008:1) voer aan dat alhoewel die definisies oor wat ’n linguistiese korpus werklik behels kan wissel, stem die meerderheid linguiste saam dat ’n korpus beskou kan word as ’n versameling van tekste of dele van ’n teks waarop algemene linguistiese analises uitgevoer kan word. Daar word tussen twee tipes korpora onderskei, die pre-elektroniese korpora en die elektroniese korpora[93]. Die pre-elektroniese korpora verwys na ’n korpus wat sonder die hulp van rekenaars saamgestel en ontleed kan word terwyl die elektroniese korpus na ’n korpus verwys wat masjienleesbaar is en ook met behulp van rekenaarprogrammatuur geanaliseer kan word.

In ’n digitale era word daar nou van masjienleesbare korpora − elektroniese korpora − gebruik gemaak. Vir die doel van my studie was dit belangrik om na die masjienleesbare korpora ondersoek in te stel. Vervolgens verduidelik ek wat ’n korpus is, hoe ’n korpus saamgestel word, en ek verwys ook na die verskillende soorte korpora.

4.2.1 Definisie van korpus

’n Korpus kan beskou word as ’n versameling tekste wat doelbewus geselekteer is volgens ‘n voorafbepaalde ontwerp (wat dit verteenwoordigend van ‘n spesifieke taalvariëteit maak) en wat in ’n elektroniese formaat gestoor word en beskikbaar is vir ’n masjienmatige analise ten einde spesifieke vraagstukke oor die aard van taal soos dit werklikgebruik word te beantwoord (vgl. Baker, 2010a:7; Baker *et al.,* 2006:48; Cheng, 2012:30; Hoover *et al.*, 2014:2; Kennedy, 1998:3; Kübler & Zinsmeister, 2015:4; McEnery & Wilson, 2001:117; Meyer, 2008:1; Sinclair, 1991a:171; Sinclair, 2005:16; Tognini-Bonelli, 2001:2; Tognini-Bonelli, 2010:18).

[93] Sien afdeling 4.3 (’n Historiese oorsig van korpuslinguistiek).

Aangesien ’n korpus uit doelbewus geselekteerde tekste bestaan wat taal soos dit daagliks gebruik word, uitbeeld, kan daar tussen verskillende tipes korpora onderskei word. In die daaropvolgende afdelings word daar eerstens na die verskillende soorte korpora en korpusgroottes ondersoek ingestel en tweedens word daar na die samestelling van ’n korpus verwys.

4.2.2 Soorte korpora en korpusgroottes

Vanuit afdeling 4.2.1 blyk dit dat ’n korpus na ’n versameling (elektroniese) tekste verwys wat gebruik word om ’n taal te analiseer. Korpora is reeds in gebruik voor die sogenaamde rekenaarera ontstaan het. McEnery en Hardie (2012:37-51) argumenteer dat daar tussen drie verskillende generasies korpora onderskei kan word, naamlik:

- **eerstegenerasiekorpora**: Die gebruik van rekenaars is ’n nuwe toevoeging tot korpuslinguistiek. Konkordansies word saamgestel en geanaliseer en die besef het ontstaan dat data gestandaardiseer moet word en ’n behoefte aan rekenaarprogrammatuur het dus ontstaan.

- **tweedegenerasiekorpora**: Persoonlike rekenaars is ’n algemene verskynsel. Die ontwikkeling van rekenaarprogrammatuur stel navorsers in staat om meer data (vinniger) te analiseer en korpuslinguiste het nie meer groot spanne personeel nodig om die data te help ontleed nie.

- **derdegenerasiekorpora**: Rekenaarprogrammatuur soos *WordSmith Tools* en *Antconc*[94] word algemeen gebruik in korpuslinguistiese analises. Groot datastelle word nou ontleed en daar bestaan meer instrumente om die data te analiseer en ’n groter aantal tekenstelsels kan ondersoek word.

Hierdie duidelike onderskeid tussen verskillende generasies in korpuslinguistiek kan ook gesien word in Tabel 4.1. Hierdie Tabel bied ’n oorsig van belangrike korpora (soos gesien kan word in die Tabel is die meerderheid van die korpora Brits of Amerikaanse Engels) en dui ook verder op die ontwikkeling van korpora en die benaderings wat in korpuslinguistiek gebruik word.

[94] Daar bestaan verskeie rekenaarprogrammatuur, maar aangesien *WordSmith Tools* en Antconc in my studie gebruik is (sien afdeling 5.8.1, Rekenaarprogrammatuur) dien hierdie programme as die voorbeeld vir rekenaarprogrammatuur in korpuslinguistiese analises.

Tabel 4.1 Korpusontwikkeling (saamgestel uit Kennedy, 1998; Nesselhauf, 2005; Stubbs, 2007)

Korpora	**Tydperk**	**Grootte**	**Samesteller(s)**	**Taal en tekssoort**	**Ontwikkelingsfase**	**Doel**
Cruden's Concordance	1737	777 746	Cruden	Engels	Pre-elektronies	Konkordansie
Oxford English Dictionary	1857-1928	50 000 000	Murray, Bradley, Johnsone	Engels	Pre-elektronies	Leksikografie
Kaeding's corpus	1890	11 000 000	Keading	Duits	Pre-elektronies	Pedagogies
The English Dialect Dictionary	1898-1905	17 519	Wright	Engels	Pre-elektronies	Dialek
Thorndike's corpus	1921	4 500 000	Thorndike	Engels	Pre-elektronies	Pedagogies
Thorndike en Lorge's corpus	1944	18 000 000	Thorndike en Lorge	Engels	Pre-elektronies	Pedagogies
The Structure of English	1952	250 000	Charles Carpenter Fries	Engels	Pre-elektronies	Grammatikaal
Brown's Corpus	1961-1964	1 000 000	Francis en Kucera	Amerikaanse Engels geskrewe teks	Eerste rekenaarkorpus	Linguisties
London-Lund Corpus of Spoken English (LLC)	1975-1988	500 00	Randolph Quirk	Getranskribeerde taal van Londense tieners	Elektronies	Grammatikaal
LOB	1979	1 00 000	Leech, Johansson, Hofland en Garside	Engelse geskrewe teks	Eerstegenerasie	Grammatikaal
Survey of English Usage	1653-1987	1 000 000	Quirk	Engels	Eerstegenerasie	Grammatikaal
Birmingham Corpus	1980-1986	20 000 000	Universiteit van Birmingham	B/A Engels	Superkorpus	Elektronies
Bank of English	1990	500 000 000	Universiteit van Birmingham	B/A Engels	Superkorpus	Elektronies
ARCHER	1990-2013	3 298 080	Doug Biber en Edward Finegan	Engels	Superkorpus	Elektronies
British National Corpus (BNC)	1991-1995	100 000 000	Paul Baker – Universiteit van Lancaster	Britse Engels	Tweedegenerasie – Mega	Elektronies
British Academic Spoken English Corpus (BASE)	2000-2005	1 644 942	Nesi en Thompson	21-eeuse Engels	Tweedegenerasie – Mega	Elektronies
British Englsh 20016 (BE06)	2003-2008	1 010 996	Lancaster Universiteit	Britse Engels	Tweedegenerasie – Mega	Elektronies
Yahoo-based Contrastive Corpus of Questions and Answers (YCCQA)	2008-2009	29 400 000	De Smet, Katholieke Universiteit Leuven	Engels, Frans, Spaans en Duits	Tweedegenerasie – Mega	Elektronies

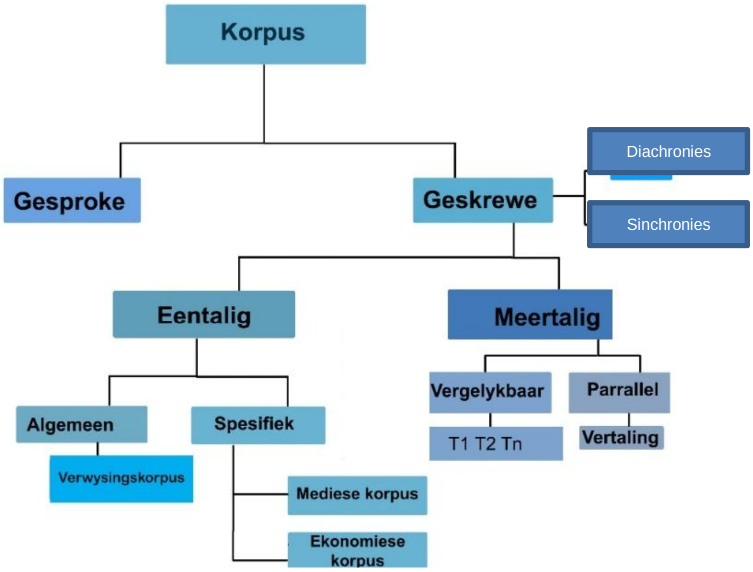
Vanuit Tabel 4.1 is dit duidelik dat korpuslinguistiek dramaties ontwikkel het gedurende die afgelope 50 jaar, en dat die aantal korpora wat beskikbaar is, radikaal toegeneem het. Korpora word vandag uit veelvuldige tale saamgestel, en daar is ook 'n verskeidenheid korpora wat aanlyn beskikbaar is. Vervolgens verwys ek na die verskillende soorte korpora en bespreek ek

die korpusgroottes alvorens daar spesifiek na die web as korpus[95] en die samestelling van ’n korpus ondersoek ingestel word.

4.2.2.1 Verskillende korpora

Kübler en Zinsmeister (2015:3) voer aan dat linguiste vandag van korpora gebruik maak nie net om natuurlike voorbeelde van taal in te samel nie, maar ook om hipoteses te toets. Die aantal data (linguisties geanaliseerde en publiek beskikbare korpora) het beduidend toegeneem. Hundt *et al.* (2007:1) voer aan dat die (amper dramatiese) tegnologiese ontwikkelinge sedert die aanvangsjare van korpuslinguistiek aanleiding daartoe gegee het dat daar nie net ŉ verskeidenheid rekenaarprogrammatuur vir die analise van korpora bestaan nie, maar ook dat korpusgroottes drasties verander het.

Volgens Baker (2010a:12) is ’n korpus nie net ’n versameling tekste wat lukraak gekies is nie. Trouens, wanneer ’n korpus saamgestel word, verg dit ’n deeglik voorafbeplande proses wat gevolg moet word ten einde ’n spesifieke navorsingsvraag te kan beantwoord[96]. Juis omdat korpora gebruik word om spesifieke navorsingsvrae te beantwoord, is dit voor die hand liggend dat daar verskillende soorte korpora sal bestaan. Die eerste onderskeid wat getref moet word, is die keuse tussen geskrewe en gesproke teks as basis vir die korpus. Volgens Cabrio (2010) is hierdie keuse een van die eerste oorwegings wanneer ’n korpus saamgestel word. Figuur 4.1 beeld die verskillende soorte korpora volgens Cabrio (2010) uit:



Figuur 4.1 Voorstelling van die verskillende soorte korpora (geneem en vertaal vanuit Cabrio, 2010)

95 Renouf (2007:42) verwys na tekste vanaf die wêreldwye web (die sogenaamde aanlyn korpora) as ’n kuberkorpus (sien afdeling 4.3.2 - Die web as korpus).
96 Sien afdeling 4.2.5 (Samestelling van ’n korpus).

Vanuit Figuur 4.1 is dit duidelik dat daar verskillende soorte korpora bestaan. Elk van hierdie korpora het ’n unieke doel wat die volgende insluit:

- ŉ **Gespesialiseerde korpus**: Cabrio (2010) tref onderskeid tussen ’n korpus wat vir algemene taaldoeleindes saamgestel is, en ’n korpus wat vir gespesialiseerde taaldoeleindes saamgestel is. Volgens Baker *et al.* (2006:147) word ŉ gespesialiseerde korpus vir ’n definitiewe doel geskep (byvoorbeeld, leksikografie vir woordeboekontwerp; om genres van ’n taal te bestudeer; of om die taal van kinders te ondersoek). ŉ Gespesialiseerde korpus verwys dus na ’n doelbewus geselekteerde en saamgestelde korpus (vgl. Baker, 2010a:12; Gries, 2009:8; Nesselhauf, 2005).

- ŉ **Verwysingskorpus**: Wanneer ’n teks deur middel van frekwensie-gebaseerde tegnieke geanaliseer word, is dit noodsaaklik om iets te hê waarmee die teks vergelyk kan word. Baker *et al.* (2006:137) stel dat indien daar byvoorbeeld vasgestel moet word of ’n woord uniek tot ’n sekere teks is, moet ’n verwysingskorpus gebruik word. Hierdie verwysingskorpus word as ’n prototipe vir die taal waarin die ondersoek plaasvind, beskou en Baker (2010a:12) en Nesselhauf (2005) voer verder aan dat dié soort korpus ook as ŉ algemene gestratifiseerde korpus kan bekend staan.

- ŉ **Meertalige korpus**: Hierdie korpus verwys na ’n korpus wat saamgestel is vanaf tekste vanuit meer as een taal (vgl. Baker *et al.,* 2006:119; Gries, 2009:9; Nesselhauf, 2005). Volgens Cabrio (2010) staan ŉ meertalige korpus in teenstelling met ’n korpus wat slegs uit tekste vanuit een taal saamgestel is. Nesselhauf (2005) stel dat ’n meertalige korpus uit tekste van minstens twee tale moet bestaan.

- ŉ **Parallelle korpus**: Volgens Baker *et al.* (2006:126) en Gries (2009:9) bestaan ’n parallelle korpus uit twee of meer korpora wat op dieselfde manier in verskillende tale versamel is. Hy (Baker *et al.,* 2006:126) gaan voort deur te verduidelik dat ŉ prototipe parallelle korpus gewoonlik uit ’n teks en dié teks se vertalings bestaan (aanwysings, regeringsdokumente, inligtingspamflette en dies meer).

  Daar kan ook na ’n tweede tipe parallelle korpora verwys word, naamlik ŉ **vergelykbare korpus**. ŉ **Vergelykbare korpus** word op dieselfde manier as ŉ parallelle korpus saamgestel, maar met die doel om ondersoek in te stel na die grammatika van verskillende tale, die taalkundige eienskappe van vertaalde tekste en veral ook na masjienvertalings.

Baker *et al.* (2006:127) voer ook aan dat alhoewel die term *parallelle korpus* gewoonlik na ’n korpus verwys wat bestaan uit tekste van verskillende tale, kan dit ook na ’n korpus verwys wat uit tekste van verskillende variëteite van dieselfde taal bestaan. Nesselhauf (2005) verwys ook na ’n **streekskorpus** wanneer die verskillende streekvariëteite van ’n bepaalde taal ondersoek word.

- ŉ **Leerderkorpus**: Hierdie tipe korpus word hoofsaaklik saamgestel vanuit tekste wat deur studente geskryf is. Volgens Baker *et al.* (2006:103) word geskrewe opstelle vir dié korpus gebruik en is korpora van hierdie aard nuttig in navorsing oor taalverwerwing, die aanleer van ’n tweede taal en waar foutanalises uitgevoer word. Volgens Nesselhauf (2005) kan daar verder ook na **tweedetaalsprekerkorpora** (in vergelyking met die alternatief **moedertaalsprekerkorpora**) verwys word. Hierdie twee korpora is verteenwoordigend hetsy van die taalproduk van tweedetaalsprekers of moedertaalsprekers (Nesselhauf, 2005).

- ŉ **Diachroniese korpus**: Volgens Cabrio (2010) is daar ŉ verskil tussen ŉ diachroniese korpus en ’n sinchroniese korpus. Baker *et al.* (2006:56) en Gries (2009:9) verwys na ŉ diachroniese korpus as ’n korpus wat saamgestel word met die doel om die ontwikkeling van ŉ taal oor ’n bepaalde tydperk aan te toon. Die *Helsinki Corpus of English Text: Diachronic Part* bestaan byvoorbeeld uit 400 voorbeelde van tekste wat vanaf 750-1700 strek (vgl. Baker *et al.*, 2006:56).

  ŉ Diachroniese korpus is dus ’n **dinamiese korpus** − ’n korpus wat reëlmatig kan groei. Tekste word bygevoeg sodat dié korpus die ware aard en verandering van taal kan uitbeeld om sodoende taalveranderinge binne ’n taal aan te toon (Baker *et al.*, 2006:64). Nesselhauf (2005) en Lee (2010:113) verwys ook na ’n **historiese korpus** – ŉ korpus wat verteenwoordigend van ’n spesifieke tydperk in die geskiedenis van ŉ taal beskou kan word.

- ŉ **Sinchroniese korpus**: Soos reeds genoem, verskil ŉ sinchroniese korpus van ŉ diachroniese korpus. ’n Sinchroniese korpus verwys na ’n korpus waar al die tekste gedurende dieselfde tydperk geskryf is. Hierdie soort korpus verskaf dus navorsers ’n blik op die taal van ŉ betrokke tydstip (vgl. Baker *et al.*, 2006:153; Gries, 2009:9). ’n Sinchroniese korpus is staties – geen verandering of groei vind plaas nadat dit saamgestel is nie.

- ŉ **Geskrewe korpus**: Hierdie soort korpus verwys na korpora wat saamgestel word uit slegs geskrewe tekste − tekste wat in geskrewe formaat gepubliseer of versprei word. Baker *et al.* (2006:171) stel dat geskrewe tekste ook na elektroniese tekste kan verwys soos e-posse en inskrywings vanaf webblaaie en dat daar soms “grys” areas bestaan wat die klassifikasie van tekste betref wat geskryf is met die doel om voorgedra te word soos toesprake of radiodramas.

- ŉ **Gesproke korpus**: Baker *et al.* (2006:148) en Nesselhauf (2005) verduidelik dat hierdie soort korpora uit getranskribeerde tekste bestaan. Mello (2014:60) sluit by die bogenoemde outeurs aan deur te beklemtoon dat gesproke korpora getranskribeer moet word alvorens dit geanaliseer kan word. ’n Gesproke korpus kan verwys na spontane, informele gesprekke, vergaderings of radiodramas.

Vanuit hierdie verduideliking blyk dit dat daar heelwat verskillende soorte korpora voorkom, en dat korpora elk ’n unieke doel het. Wanneer hierdie verskillende soorte korpora voorgehou en ondersoek word saam met die ontwikkeling van korpora is dit duidelik dat korpusgrootte deur die jare ook verander het. Vervolgens verwys ek na die verskillende korpusgroottes.

4.2.2.2 Korpusgroottes

Baker (2010a:7) voer aan dat ŉ korpus as verteenwoordigend beskou moet word, en dat dit belangrik is dat ŉ korpus groot genoeg is sodat taalfrekwensie en -patrone uitgebeeld en statisties weergegee kan word. Hoe groter ŉ korpus is, hoe makliker is dit om unieke of onnatuurlike taalpatrone te identifiseer. Indien die korpus wat gebruik word te klein is, sal die afwykings nie noodwendig sigbaar wees nie (vgl. Baker, 2010a:7).

Volgens Stamatatos (2009:541) het outeurskapidentifikasie beduidend ontwikkel gedurende die afgelope dekade. Alhoewel die meerderheid van die navorsing ondersoek ingestel het na die identifisering van outeurs van lang tekste, het tegnologiese ontwikkelinge (veral die toenemende gewildheid van sosiale media) daartoe aanleiding gegee dat kleiner korpora ondersoek word.

Hundt *et al.* (2007:1) argumenteer dat die standaardgrootte vir ’n korpus nie meer 1 miljoen woorde is nie, maar eerder 100 miljoen woorde[97]. O’Keeffe *et al.* (2007:4) voer aan dat ’n gesproke korpus as “groot” beskou kan word indien dit meer as ’n miljoen woorde bevat terwyl enigiets kleiner as 5 miljoen woorde vir ’n geskrewe korpus as “klein” beskou word.

[97] Vanuit Tabel 4.1 kan die afname in korpusgroottes waargeneem word.

Flowerdew (2004:19) is van mening dat ʼn klein korpus uit sowat 250 000 woorde kan bestaan. Abbasi en Chen (2008:15) het ondersoek ingestel na outeurskapidentifikasie met betrekking tot e-posse en webforums (die korpusgroottes het hier gewissel vanaf 1 422 tot 23 423 woorde per outeur). Koppel *et al. (*2011:85) het ondersoek na blogs ingestel en het van ʼn korpus van 2 500 woorde per outeur (10 000 outeurs) gebruik gemaak (vgl. Koppel *et al*. 2013:324).

Schwarts *et al.* (2013:1880) het van ʼn korpus gebruik gemaak wat uit 50 outeurs met 50 twiets per outeur (50 X 140 = 7 000 karakters per datastel) bestaan het sowel as ʼn korpus wat uit 50 outeurs met 1 000 twiets per outeur (1 000 X 140 = 140 000) bestaan het. Vanuit Schwarts *et al.* (2013) se navorsing blyk dit dat ʼn korpus betreklik klein kan wees − so klein as 1 400 woorde per outeur. Aangesien navorsing oor outeurskapidentifikasie met betrekking tot mikroblogs nog betreklik nuut is, was dit juis nodig om die minimum woorddrempel (minimum karakterdrempel) te bepaal om sodoende Afrikaanse outeurs op ʼn mikroblog suksesvol te kan identifiseer.

Wat duidelik blyk, is dat ʼn korpus nie uit miljoene woorde hoef te bestaan alvorens dit geanaliseer kan word nie. Inteendeel, analises word toenemend op kleiner korpora uitgevoer. Hierdie kleiner korpora kan (soos reeds genoem) beskou word as die gevolg van die ontwikkeling van tegnologie (veral sosiale media) waar kommunikasie al hoe bondiger word en anonimiteit aan die orde van die dag is. Dit het dus veral betrekking op forensiese linguistiek (outeurskapidentifikasie) waar dit belangrik is om te bepaal wat die minimum drempel vir korpora is sodat analises suksesvol uitgevoer kan word.

McEnery en Wilson (2001:187) voer aan dat ons sonder twyfel die era van super korpora betree het, en dat die grootte van beskikbare korpora toenemend sal groei. Die *BNC* is nou klein (100 000 000 woorde) in vergelyking met korpora soos die *GloWbE* (1 900 000 000 woorde) − veral as die invloed van die wêreldwye web in ag geneem word.

Dit is duidelik dat korpora oor die afgelope 30 jaar toegeneem het in grootte. Vanuit Anthony (2013:145) se voorstelling blyk dit dat korpora in 2000 uit sowat 2 000 000 000 woorde bestaan het. Tog voer McEnery en Wilson (2001:191) aan dat die grootte van gespesialiseerde korpora al hoe kleiner word aangesien navorsers van mikro korpora gebruik maak om spesifieke navorsingsvrae te beantwoord. Koester (2010:68) sluit hierby aan en postuleer dat die samestelling van ʼn korpus sowel as die wyse waarop dié korpus as verteenwoordigend beskou kan word van groter belang is as die grootte van ʼn korpus.

Ghadessy *et al.* (2001:xvii) argumenteer dat ʼn klein korpus vir talle navorsers as beuselagtig beskou word, maar dat die stelselmatige ontwikkeling in korpuslinguistiek aanleiding daartoe

gegee het dat bevindinge wat vanuit die data van ’n klein korpus gemaak kan word net so belangrik en soms onmoontlik deur die gebruik van ’n groot korpus is.

Anthony (2009:91) vergelyk die analise van ’n klein korpus met ’n studie van slegs ’n klein groep sterre. Hy (Anthony, 2009) skryf dat meer van die unieke kenmerke van die sterre sigbaar is wanneer slegs ’n klein groep op ’n slag ondersoek word, en dat ’n klein korpus inherent dieselfde is. Ook Meyer (2004:12) is van mening dat ’n kleiner korpus genoegsame data kan lewer om ’n hipotese te bewys.

Koester (2010:66) vra tereg wat die doel van ’n klein korpus is. Talle linguiste (onder andere Sinclair, 2004a:189) argumenteer dat ’n klein korpus beperkend is, en dus geen ware doel dien nie. Terwyl ander soos Carter en McCarthy (1995:143) aanvoer dat die analise van groot korpora nadelig blyk te wees aangesien die aantal data wat geanaliseer moet word om frekwente eienskappe te identifiseer eenvoudig ongemaklik is, as gevolg van tegnologiese beperkings wat voorgekom het.

Volgens Koester (2010:67) bied die analise van kleiner gespesialiseerde korpora baie voordele soos die moontlikheid om ooreenkomste tussen ’n betrokke korpus en die konteks waarin die tekste wat in die korpus gebruik word te analiseer. Verder bied die analise van ’n klein korpus volgens Koester (2010:67) ’n blik op die taalpatrone wat in ’n bepaalde situasie voorkom. Anthony (2013:146) stel dat die waarde van ’n korpus geensins afhanklik is van die grootte nie, maar veel eerder van die inligting wat dié korpus kan lewer. Indien navorsers die regte benadering volg, kan enige korpus − ongeag die grootte daarvan − belangrike taalkundige resultate lewer.

Koester (2010:68) argumenteer dat daar nie ’n “perfekte grootte” vir ’n korpus bestaan nie, en dat die navorsingsvrae die korpusgrootte sal beïnvloed. Ook O’Keeffe *et al.* (2007:198) stel dat indien gespesialiseerde korpora deeglik saamgestel word, klein korpora as meer betroubaar as algemene korpora beskou kan word. Dit is dus duidelik dat die grootte van ’n korpus deesdae van navorsers en die doel van studies afhang, en dat daar sukses behaal word met die analise van beide groot en klein korpora.

Vir die doel van my studie het ek van ’n sinchroniese statiese geskrewe korpus gebruik gemaak. Ek het verder met klein korpora gewerk aangesien outeurskapidentifikasie op tekste met ’n maksimum lengte van 140 karakters toegepas is. Aangesien ek my korpus saamgestel het

vanaf ʼn mikroblog (Twitter) wat ʼn sosiale media platform is, verwys ek vervolgens na die web as ʼn korpus.

4.2.3 Die web as ʼn korpus

Renouf (2007:42) argumenteer dat die web vanaf die 1990's beskou kan word as ʼn bron van voorheen onbeskikbare linguistiese data en dat die web as ʼn toeganklike en hedendaagse voorstelling van die gebruik van taal beskou kan word. Kübler en Zinsmeister (2015:3) voer aan dat al hoe meer linguiste hulle tot die wêreldwye web wend om korpora saam te stel van ʼn taal soos dit daagliks gebruik word.

Baker *et al.* (2006:170) voer aan dat die wêreldwye web (algemeen bekend as die www) ʼn globale netwerk van hiperteks dokumente is wat deur middel van ʼn internetverbinding toeganklik is. Die eksponensiële groei van die web het aanleiding daartoe gegee dat daar vandag miljarde dokumente daarop beskikbaar is − groter as die grootste korpus − en dus die ideale platform waarop tekste versamel kan word vir ʼn korpus (vgl. Baker *et al.*, 2006:170). Daar moet onderskeid getref word tussen die web as korpus en die web as ʼn platform waarop data ingesamel kan word vir ʼn korpus.

Alhoewel sommige navorsing die web in sy totaliteit as ʼn korpus gebruik, kan fragmente van die inligting wat op die web beskikbaar is ook gebruik word[98]. ʼn Benadering wat volgens Kilgarriff en Grefenstette (2003:336) veral vir leksikograwe voordeel inhou aangesien woorde wat ongewoon is slegs sigbaar is in ʼn buitengewoon groot korpus.

Baker *et al.* (2006:99, 170) voer egter aan dat sommige navorsers nie die web as geheel as ʼn korpus gebruik nie, maar dat hulle die web as ʼn platform beskou wat gebruik word om tekste vir gespesialiseerde korpora in te samel aangesien die web as ʼn korpus gebruikers in staat stel om groot hoeveelhede "rou" data in te samel. Tabel 4.2 toon die verskille (en voordele) in elk van die twee benaderings aan:

Tabel 4.2 Verskil tussen die web as ʼn korpus en die web as ʼn bron vir die samestelling van ʼn korpus (geneem vanuit Wunderlich, 2012:44).

Web as korpus	**Web as bron vir die samestelling van 'n korpus**
Bespaar tyd en geld	Beheer oor die inhoud en die tekstipes wat by de korpus ingesluit word
Op datum - die data wat op die web beskikbaar is, word gereeld opgedateer	Kan geannoteerde korpora vir die vergelyking gebruik (gespesialiseerde korpus word vanaf die web saamgestel, en 'n reeds bestaande verwysingskorpus word gebruik)

[98] Fragmente van die web verwys na die gebruik van data vanaf bepaalde platforms, soos byvoorbeeld Twitter, forums of e-posse.

Web as korpus	**Web as bron vir die samestelling van 'n korpus**
Grootte is onbeperk	Vasgestelde grootte
Kortstondige bestaanstydperk	Vasgestelde inhoud

Renouf (2007:42) verwys na tekste wat vanaf die internet as ’n vorm van aanlyn korpora hanteer word as ’n kuberkorpus. Kilgarriff en Grefenstette (2003: 334) argumenteer dat die web sonder twyfel as ’n korpus beskou kan word. Sinclair (2005:13) stel dat die web nie as ’n korpus beskou kan word nie, maar Kilgarriff en Grefenstette (2003:334) is van mening dat indien die algemene definisie dat ’n korpus as ’n versameling tekste beskou word, aanvaar word, dan moet die web wel as ’n korpus beskou word.

Fletcher (2013:1339) postuleer dat miljoene webouteurs nie noodwendig besef dat hulle daagliks bydra tot die ontwikkeling en groei van die magdom inligting wat aanlyn beskikbaar is nie, en dat hierdie inligting geanaliseer en ondersoek kan word vir taalkundige en kulturele inhoud. Fletcher (2013:1339) stel dat die web as ’n korpus beide as voordelig en nadelig beskou kan word. Terwyl Tabel 4.2 as ’n vergelyking tussen die twee benaderings van die web as ’n korpus en die web as ’n bron vir die samestelling van ’n korpus dien, kan die voordele wat die web, volgens Fletcher, as ’n korpus inhou vanuit Tabel 4.3 gesien word:

Tabel 4.3 Voordele van die web as ’n korpus (Fletcher, 2013:1339-1340)

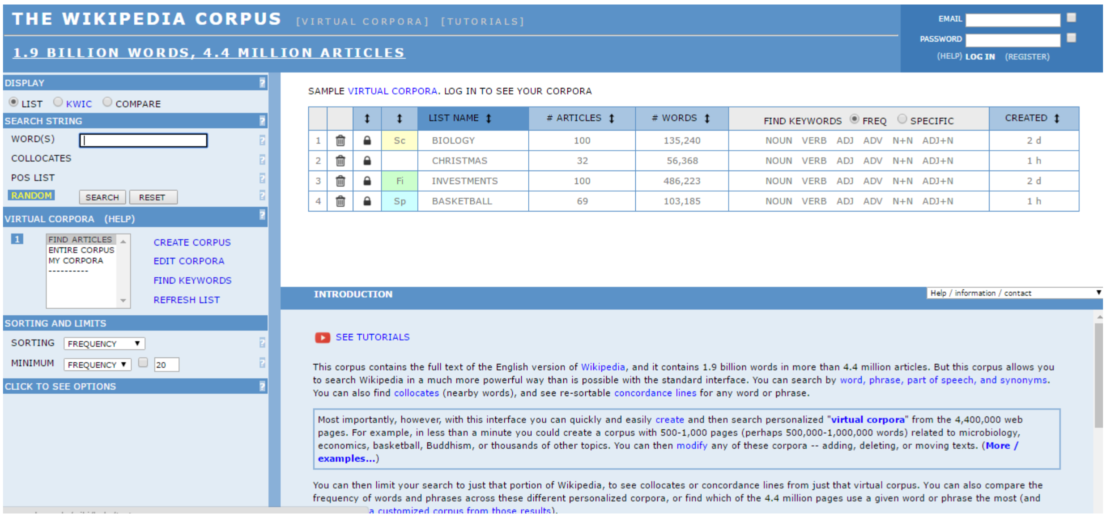
Voordele van die web as korpus	
Grootte	Aangesien die web konstnt uitbrei, is die grootte van die web onbekend, en die aantal data wat vanaf die web gegenereer kan word, is eindeloos.
Reikwydte	Daar is voorbeelde van bykans alle tale beskikbaar op die web (alhoewel Engels as die primêre taal van die web beskou kan word), en sosiale media het aanleiding gegee tot 'n gevoel van 'n aanlyn gemeenskap – wat weer eens tot veelvoudige navorsingsmoontlikhede kan lei.
Op datum	Gebeure geskied in werklike tyd, daar is minimale tydsvertraging en dus is die web 'n spieëlbeeld van die hedendaagse kontemporêre kulture.
Multimodaliteit	Alhoewel die web video en klank insluit, het talle webblaaie ook transkripsies van die klank- en beeldmateriaal beskikbaar gestel, en dus kan hierdie gesproke taal ook by 'n webkorpus geplaas word.
Beskikbaarheid	Die beskikbaarheid van die internet het tot gevolg dar 'n multimiljoenwoordkorpus relatief moeiteloos saamgestel kan word.

Ten spyte van al die voordele wat die web as ’n korpus vir navorsers inhou, verwys Fletcher (2013:1341) ook na ooglopende nadele. Hierdie nadele sluit in dat navorsers webblaaie moet identifiseer vir navorsing aangesien 'n lukrake insameling van data nie spesifieke data sal voortbring nie (veral aangesien elke platform en webblad as 'n eie unieke gemeenskap funksioneer). Vrae wat volgens Fletcher (2013:1341) gevra moet word ten einde tekste vanaf die web vir ’n korpus te kan gebruik, sluit in:

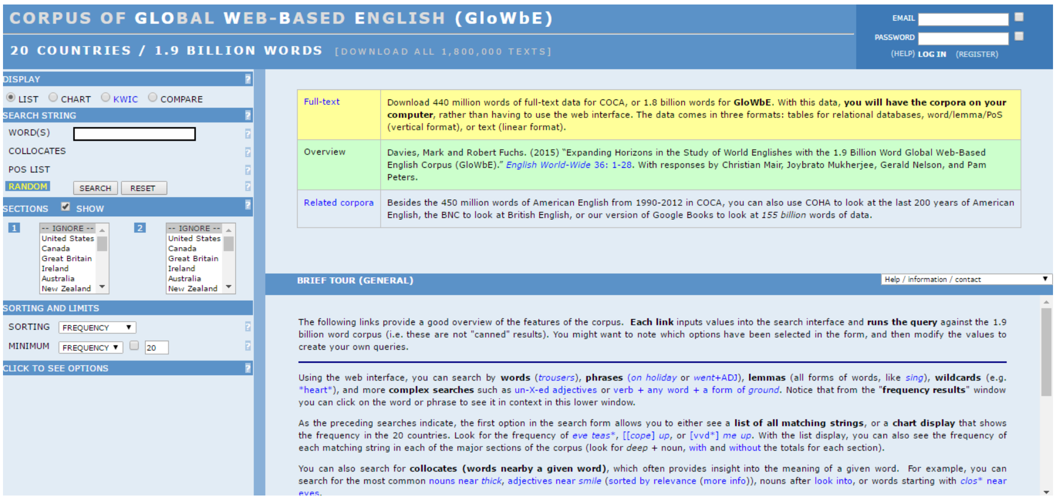
- Wie was die bedoelde gehoor van die teks?

- Wat is die demografiese en geografiese oorsprong sowel as die teikenmark wat die teks verteenwoordig?
- Wie het die teks geskryf − is ’n moedertaalspreker aan die woord?
- Is die teks deur ’n masjien vertaal of gegenereer?
- Wat is die doel van die teks?
- Is die dokument gesaghebbend?
- Is die inhoud akkuraat?
- Is die inhoud verteenwoordigend?

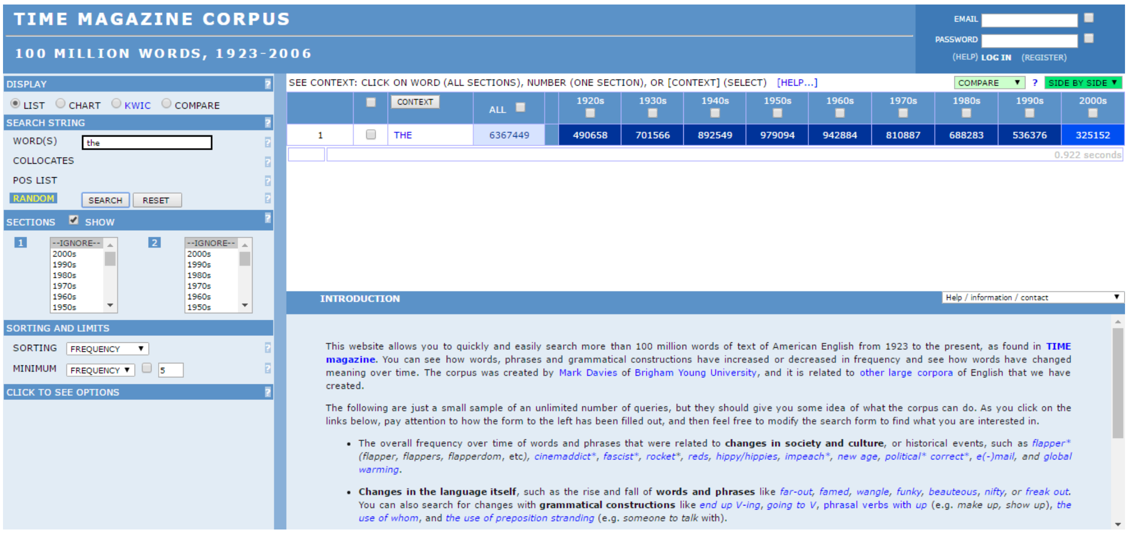
Skermskoot 4.1, 4.2 en 4.3 is voorbeelde van korpora wat aanlyn beskikbaar is:



Skermskoot 4.1 Voorstelling van die Wikipedia-korpus (http://corpus.byu.edu/wiki/)



Skermskoot 4.2 Voorstelling van die GloWbE-korpus (http://corpus.byu.edu/glowbe/)



Skermskoot 4.3 Voorstelling van die Time-korpus (http://corpus.byu.edu/time/)

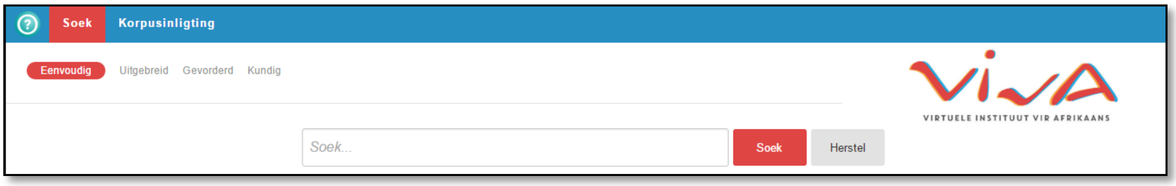
Buiten die voorbeelde wat in skermskoot 4.1 tot 4.3 voorgehou is, bied die web bied ook ŉ geleentheid om taal waar te neem soos dit daagliks gebruik word. Sosialemediaplatforms skep ŉ aanlyn gemeenskap, maar die anonieme aard van die web kan aanleiding tot meer vrae gee soos wie waarlik aan die woord is. Indien navorsers ŉ webkorpus deeglik en sorgvuldig saamstel, kan dit aanleiding gee tot antwoorde oor taal wat nog nooit voorheen beantwoord kon word nie.

Vanuit hierdie afdelings (afdelings 4.2.1 tot 4.2.3) is dit duidelik dat korpora − net soos korpuslinguistiek − ontwikkel het gedurende die afgelope 30 jaar. Nie net het die grootte van die beskikbare korpora verander nie, maar ook die maniere waarop dié korpora saamgestel word. Soos reeds genoem, het ek van ŉ sinchroniese geskrewe korpus gebruik gemaak, en vir die doel van my studie het ek die web as ŉ bron (aangesien mikroblogs sosialemediaplatforms is wat aanlyn beskikbaar is) in die samestelling van my klein korpus gebruik. Vervolgens verwys ek oorsigtelik na korpora in Afrikaans alvorens ek ŉ historiese oorsig van korpuslinguistiek en die talle definisies van hierdie vakgebied voorhou.

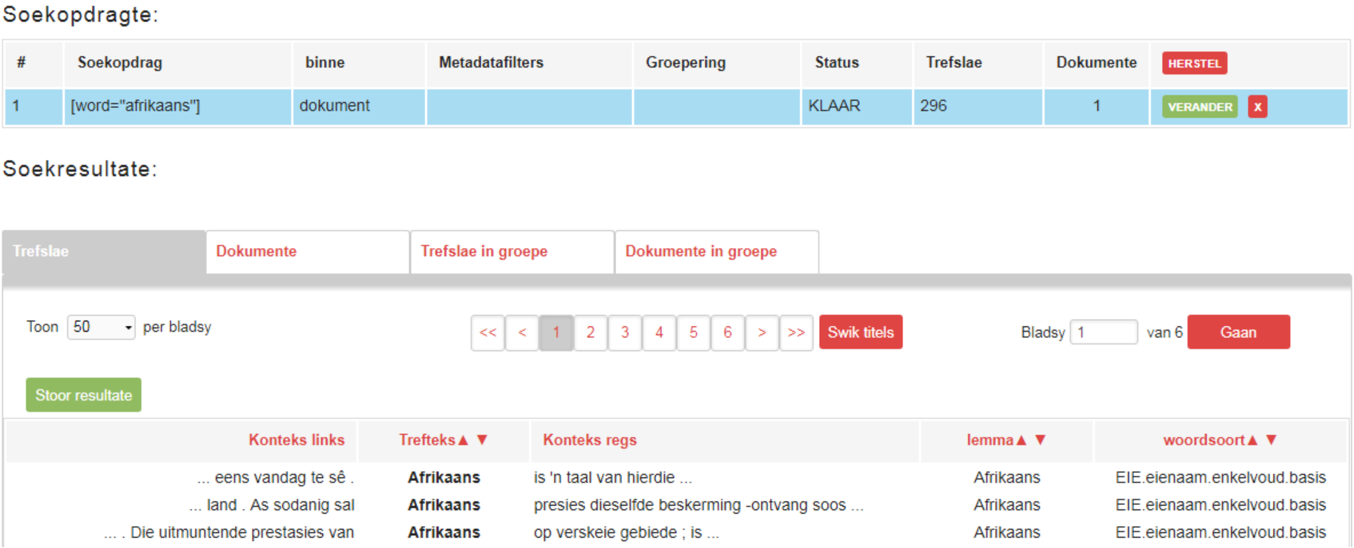
4.2.4 Korpora in Afrikaans

Vanuit die vorige afdelings (afdelings 4.2.1, 4.2.2 en 4.2.3) is definisies van korpora, die rol en grootte van korpora en die verskillende soorte korpora voorgehou. Vanuit hierdie afdelings blyk dit duidelik dat die meeste korpora sonder twyfel in Engels voorkom. Met die toenemende gewildheid van korpuslinguistiek en die gebruik van korpora in talle navorsingsareas, is ander tale besig om stelselmatig ook korpora op te bou.

Ook in Afrikaans is daar talle ontwikkelinge ten opsigte van korpora. VivA (die Virtuele Instituut vir Afrikaans) het aanlyn korpora wat beskikbaar is aan geregistreerde gebruikers (sien Skermskoot 4.4 en 4.5).



Skermskoot 4.4 Tuisblad van die VivA-korpusportaal (http://korpus.viva-afrikaans.org/page/search)



Skermskoot 4.5 Voorbeeld van ’n eenvoudige soektog in die VivA-korpus na die woord “vermiste” (https://youtu.be/SC4xMk98Pdc)

Die VivA-korpus is saamgestel uit verskeie bronne, en bestaan uit verskillende korpora. Die bekendste korpus in Afrikaans is die **Taalkommissiekorpus** (TK-korpus) van 2011. Hierdie korpus is saamgestel deur die Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (Taalkommissie, 2011). Die TK-korpus bestaan uit ongeveer 60 miljoen Afrikaanse woorde vanuit 'n verskeidenheid geskrewe tekste wat prosatekste (6 miljoen woorde), akademiese tekste (24 miljoen woorde) en niefiksie sowel as nie-akademiese tekste (byvoorbeeld koerante, tydskrifte en boeke) (27 miljoen woorde) insluit. Skermskoot 4.6 is ’n opsomming van korpora wat in Afrikaans aangetref word, en wat beskikbaar is op die VivA-korpusportaal.

Vlak	Naam	Aantal woorde	Aantal tekseenhede
OOP	NCHLT- Afrikaanse korpus 1.0	2 229 214	2 489 716
OMVATTEND	Afrikaanse Leipzig-korpus 1.0	28 776 800	32 269 153
OMVATTEND	NWU/Maroela Media-korpus 1.2	9 173 430	10 378 957
OMVATTEND	NWU/Lapa-korpus 1.1	9 804 270	11 639 129
OMVATTEND	PUK/Protea Boekhuis-korpus 2.1	8 022 403	9 255 228
OMVATTEND	RSG-nuuskorpus 2.1	14 829 223	16 160 631
OMVATTEND	Taalkommissie-korpus 1.1	47 321 344	53 622 677
OMVATTEND	Wikipedia- Afrikaanse korpus 1.0	11 523 680	13 119 966
EKSKLUSIEF	Watkykjy.co.za-korpus 1.2	1 236 214	1 406 709

Skermskoot 4.6 Afrikaanse korpora wat beskikbaar[99] is op VivA se korpusportaal

Navorsers in Afrikaans maak wel van korpora gebruik, maar aangesien die aantal korpora beskikbaar nog relatief beperk is, is die aantal verwysingskorpora ook beperk. Navorsers kan gespesialiseerde korpora saamstel om taalverskynsels te ondersoek.

Dit is duidelik vanuit Skermskoot 4.6 sowel as hierdie afdeling (Korpora in Afrikaans) dat Afrikaans nog maar in die ontwikkelingsfase is met betrekking tot die beskikbaarheid en aantal van korpora. Ongeag die beperkte aantal korpora beskikbaar, word daar wel korpora in Afrikaans saamgestel en word korpuslinguistiek gebruik. In afdelings 4.2.1 tot 4.2.4 is daar verwys na die definisies van ’n korpus, die verskillende soorte korpora, die web as ’n korpus, sowel as korpora in Afrikaans. Vervolgens verwys ek na die samestelling van ’n korpus ten einde die teoretiese agtergrond te skep wat ek in hoofstuk 5, 6 en 7 toegepas het.

4.2.5 Samestelling van ’n korpus

Meyer (2004:30) voer aan dat alvorens tekste by ’n korpus ingesluit kan word, moet die korpus deeglik beplan word. Hierdie beplanning verwys (volgens Meyer, 2004:30) na die soorte tekste wat by die korpus ingesluit sal word, die steekproefpopulasie[100] waaruit die korpus saamgestel sal word en ook die doel van die korpus. Meyer (2004:30) gaan voort deur aan te voer dat indien die korpus saamgestel word ten einde diskoerseienskappe te ondersoek, moet die korpus uit volledige tekste (nie net fragmente nie) bestaan. Die lengte van tekste wat by die korpus ingesluit word, is slegs een van die aspekte wat in gedagte gehou moet word wanneer daar op die samestelling van ’n korpus besluit word.

Wanneer ’n korpus saamgestel word, moet daar dus besluite geneem word oor die grootte van die korpus, die interne struktuur van die korpus (dit wil sê die genres wat by die korpus ingesluit word), die lengte van individuele tekste wat ingesluit word, die aantal individue wie se tekste by

---

[99] Daar is ander korpora ook in Afrikaans beskikbaar, maar dit is gewoonlik interne korpora of nie-amptelike korpora. Ek verwys dus vir die doel van my studie na korpora wat vrylik beskikbaar is.
[100] Sien afdeling 5.3 Deelnemers).

die korpus ingesluit moet word asook die demografiese eienskappe van die individue. Die tekstipe moet ook geïdentifiseer word. Gaan jy van geskrewe tekste gebruik maak of word die korpus vanuit gesproke tale saamgestel (wat dus getranskribeer moet word).

Kübler en Zinsmeister (2015:11) stel ook dat die taalkeuse van uiterse belang is. Daar kan tussen eentalige, tweetalige en meertalige korpora onderskei word. Navorsers moet dus ’n keuse maak oor die taal wat in die korpus gebruik gaan word. Meyer (2004:32-53) verwys na ses belangrike oorwegings wat in gedagte gehou moet word wanneer ’n korpus saamgestel word:

- **Grootte van korpus:** Meyer (2004:32) stel dat die samesteller van die korpus die hulpmiddele moet bepaal wat beskikbaar is om die korpus saam te stel − tyd, finansiële aspekte en navorsinghulp. Verder moet die tyd wat dit gaan neem om die tekste te versamel, te verwerk (skoon te maak en te annoteer) sowel as die minimumgrootte van die korpus vasgestel word.

- **Genres:** Daar word tussen twee hoofgenres onderskei: die gesproke en die geskrewe korpus (vgl. Kübler & Zinsmeister, 2015:12; Meyer, 2004:34-38).

  Gaan gesproke taal by die korpus ingesluit word? Indien wel, moet die transkripsie van die data en die tyd wat dit gaan neem in gedagte gehou word. Wanneer daar van geskrewe tekste gebruik gemaak word, is daar talle oorwegings wat in gedagte gehou moet word soos die soort teks (is dit masjienleesbaar) en wat die teks se doel (akademies, informeel of kreatief) is.

- **Lengte van tekste:** Meyer (2004:38-40) postuleer dat die volgende vraag gevra moet word: “Hoe lank moet die tekste van individue wees wat by die korpus ingesluit word?” Daar is talle beperkinge wat die lengte van die tekste wat ingesluit word, kan beïnvloed soos kopiereg, tyd en beskikbare hulpmiddele, maar die ideaal is om volledige tekste in te sluit aangesien dit die konteks en die diskoersanalise kan bevoordeel. Lang tekste maak die korpus lomp, en dus moet die doel van die korpus deurentyd in gedagte gehou word.

- **Aantal en verskeidenheid van deelnemers wat by die korpus ingesluit moet word:** Hierdie aspek van die samestelling van die korpus kan volgens Meyer (2004:40-45) deur ’n eenvoudige metodologiese benadering aangespreek word. Die steekproefneming van die deelnemers kan wetenskaplik uitgevoer word om te verseker dat die korpus as ’n

omvattende uitbeelding van die taalpopulasie (en taalfenomeen) wat ondersoek word, beskou kan word[101].

Die doel van die korpus kan dalk vereis dat ’n spesifieke taalgroep in die samelewing ondersoek word. In dié geval moet die data wat by die korpus ingesluit word, verteenwoordigend van die taalgroep wees wat ondersoek word. Daarom kan ’n korpus uit tekste bestaan van slegs bruin tienerseuns wat in die Noord-Kaap woon. Alhoewel die korpus nie verteenwoordigend van Afrikaans as ’n geheel is nie, is dit verteenwoordigend van die korpus van die studie.

- **Tydsraamwerk van die tekste wat in die korpus gebruik word:** Is die studie diachronies of sinchronies van aard (vgl. Meyer, 2004:45-46)? ’n Sinchroniese korpus sal volgens Kübler en Zinsmeister (2015:12) ondersoek instel na taal wat gedurende ’n beperkte tydperk geskep is (byvoorbeeld in 2010) terwyl ’n diachroniese korpus uit tekste bestaan vanuit verskillende stadiums van die taal se ontwikkeling (byvoorbeeld 1900-2000).

- **Sosiolinguistiese veranderlikes wat die korpus kan beïnvloed:**

    - *Genderbalans* − Alhoewel dit relatief maklik klink om ’n balans te handhaaf ten opsigte van die gender van die deelnemers wat by die korpus ingesluit word (name kan die gender uitwys of ’n eenvoudige vorm kan biografiese inligting aanvra), is daar sommige genres waar die manlike (of vroulike) deelnemers in die meerderheid is. Volgens Meyer (2004:49) is daar geen enkele manier om ’n genderbalans te verseker nie. Navorsers moet bewus wees van beperkings en dit tot die beste van hulle vermoë hanteer.

    - *Ouderdom* − Is daar ’n spesifieke ouderdomsgroep wat ondersoek word? Indien wel, moet ’n afsny-ouderdom vasgestel word. Anders moet die deelnemers verteenwoordigend van die ouderdomspektrum wees van die fenomeen wat ondersoek word (vgl. Meyer, 2004:50).

    - *Vlak van geletterdheid* − Word daar spesifiek ondersoek ingestel na ’n taalaspek wat deur geletterdheid beïnvloed kan word (vgl. Meyer, 2004:50)?

    - *Dialek* − Geskrewe tekste is veronderstel om te voldoen aan die standaardvereistes van ’n bepaalde taal. Daarom stel Meyer (2004:51) dat dialek

[101] Sien afdeling 5.3 (Deelnemers).

nie so maklik in geskrewe tekste sigbaar is nie. Indien die doel van die korpus is om dialek na te vors, kan sprekers vanuit verskillende streke, van verskillende ouderdomme en kulture by die korpus ingesluit word om dialek te probeer bepaal.

- **Sosiale konteks en verhoudinge:** Taal verander na gelang van die sosiale konteks en die verhouding tussen gespreksdeelnemers. Daarom moet daar volgens Meyer (2004:52) na alle gespreksituasies ondersoek ingestel word wanneer die korpus saamgestel word.

Hunston (2008:160) verwys na drie aspekte wat in gedagte gehou moet word wanneer ’n korpus ontwerp word: verteenwoordiging, balans en grootte. ’n Korpus is veronderstel om ’n mikro-uitbeelding van ’n groter fenomeen te wees, en moet dus verteenwoordigend van die fenomeen wees wat ondersoek word (vgl. Hunston, 2008:160). Balans verwys volgens Hunston (2008:163) na die interne samestelling van die korpus − die proporsie van die verskillende subkorpora waaruit die korpus bestaan. Die subkorpora moet proporsioneel wees ten opsigte van die rol wat dit in die groter korpus speel en daar moet veral na die balans tussen die aantal tekste en die aantal tekens gelet word (vgl. Hunston, 2008:160).

Die laaste aspek waarna Hunston (2008:165) verwys, is die grootte van die korpus. Die korpusgrootte kan as ’n paradoks beskou word. Aan die een kant van die argument word daar gestel dat “hoe groter die korpus, hoe beter”, maar aan die ander kant kan ʼn klein korpus taalfenomene uitlig wat nie sondermeer in groot korpora geïdentifiseer kan word nie (vgl. Hunston, 2008:165-166).

Die ontwerp van ’n korpus kan dus relatief maklik voorkom: navorsers besluit wat die komponente van die korpus is, wat die verhouding tussen die verskillende komponente moet wees, verder word die tekstipe geïdentifiseer wat in die korpus gebruik gaan word − en die korpus word saamgestel. Hunston (2008:166) voer aan dat daar talle praktiese aspekte is wat die samestelling van ’n korpus kan beïnvloed: kopiereg, toeganklikheid en beskikbaarheid van tekste; daarom is dit nie so eenvoudig om ’n korpus saam te stel nie.

Volgens Meyer (2004:53) moet navorsers eerstens die doel van die korpus bepaal alvorens die beplanning van die korpus in aanvang kan neem − hierdie proses word inherent gelei deur die doel van die korpus ten einde ’n verteenwoordigende korpus saam te stel. Hunston (2008:156) voer aan dat die samestelling van enige korpus beskou kan word as ’n middeweg tussen dít wat die samesteller beplan het, die korpusverwagting en dít wat moontlik is.

In hierdie afdeling (afdeling 4.2) is daar ondersoek ingestel na wat ’n korpus behels, die verskillende soorte korpora, die web as ’n korpus, korpora in Afrikaans sowel as die basiese oorwegings vir die samestelling van ’n korpus. Dit blyk duidelik dat daar talle aspekte is wat oorweeg moet word wanneer daar met korpora gewerk word, en ook dat die soort korpus die spil vorm waarom ’n korpuslinguistiese analise uitgevoer word. Vervolgens bied ek ’n historiese oorsig van korpuslinguistiek alvorens ek korpuslinguistiek definieer en debatte oor dié vakgebied aanspreek.

4.3 ’n Historiese oorsig van korpuslinguistiek

Hunston (2002:1) stel dat “[i]t is no exaggeration to say that corpora, and the study of corpora, have revolutionised the study of language, and of the applications of language, over the last few decades.” Wilson *et al.* (2006:i) voer aan dat die meerderheid korpuslinguistiese navorsing tot onlangs gehandel het oor die grammatika en woordeskat van standaardtaalvariëteite − meestal Britse en Amerikaanse Engels − maar dat dit drasties verander het oor die afgelope paar jaar, en vandag is korpuslinguistiek ’n snel ontwikkelende vakgebied. McEnery en Wilson (2001:1) stel dat korpuslinguistiek kortliks beskou kan word as die studie van taal wat gebaseer is op voorbeelde van taalgebruik soos dit in die alledaagse lewe gebruik word.

In die volgende gedeelte verwys ek opsommend na die historiese ontwikkeling van korpuslinguistiek met ’n spesiale verwysing na belangrike korpora wat die ontwikkeling en vestiging van dié vakgebied bespoedig het. Sedert die aanvang van die sogenaamde rekenaarera, het korpuslinguistiek vanaf ’n tydrowende proses (waar konkordansies per hand saamgestel word) ontwikkel na ’n benadering wat toegepas kan word in die samestelling van woordeboeke, die interpretasie van literêre tekste, forensieslinguistiese analises, taalvariasie, taalonderrig, taalverwerwing, diskoersanalise en nog vele meer (vgl. Baker *et al.*, 2006:14).

McEnery en Wilson (2001:1) voer aan dat die term *korpuslinguistiek* as relatief modern beskou kan word. Volgens McEnery *et al.* (2006:3) is dié term eers in die vroeë 1980’s in gebruik geneem en het hierdie “moderne term” ’n lang en interessante geskiedenis al was die term nie baie gewild gedurende die 1960’s en 1970’s nie. McCarthy en O’Keeffe (2010:3) beweer dat korpuslinguistiek vandag geassosieer word met linguiste wat voor ’n rekenaar sit en skerm na skerm van konkordansies en woordlyste navors wat deur middel van rekenaarprogrammatuur gegenereer word. Hierdie hedendaagse manier van taalnavorsing kan teruggevoer word na die dertiende eeu toe Bybelwetenskaplikes en hul assistente ’n indeks van die woorde in die Bybel geskep het.

McCarthy en O’Keeffe (2010:3) verwys in dié verband na Anthony van Padua (1195-1231) se konkordansie van die Bybel − *Concordantiae Morales sacrorum Bibliorum* − en Kardinaal Hugo van St. Caro (en sy span van 500 monnike) wat teen 1230 ’n indeks van die Bybel saamgestel het as die oorspronklike korpuslinguiste. Vanaf daardie eerste dae is talle ander konkordansies van die Bybel saamgestel sowel as van literêre tekste (veral van Shakespeare se werk). Leksikograwe en die pre-Chomskyaanse strukturele linguiste word ook as deurslaggewend in die ontwikkeling van korpuslinguistiek beskou (McCarthy & McCarthy, 2010:4) aangesien die insameling van data van kardinale belang vir die voltooiing van hul werk was.

Bonelli (2010:14) voer aan dat historiese linguistiek in der werklikheid altyd korpusgebaseerd was aangesien ’n taalverandering slegs sigbaar is wanneer tekste vanuit verskillende tydsperiodes (diachronies) ondersoek word. Svartvik (2007:12) argumenteer dat alhoewel korpuslinguistiek vandag na elektroniese korpuslinguistiek verwys, was daar ’n tyd toe rekenaars nie ’n alledaagse verskynsel was nie − ’n tyd sonder Google, e-posse, sellulêre telefone en beslis geen rekenaarprogrammatuur wat korpuslinguistiek ondersteun (en vergemaklik) het nie. Svartvik (2007:12) verwys na die *BC*-tydperk (tydperk voor rekenaars) van korpuslinguistiek.

Kennedy (1998:13) sluit by Svartvik aan wanneer hy stel dat alhoewel verskeie individue glo dat korpuslinguistiek in die 1960’s begin het − die sogenaamde aanvang van die elektroniese korpusera. Vergelyk ook Anthony (2009:88), Granath (2007:169), Kennedy (1998:13) asook Svartvik (2007:12). Volgens Kennedy (1998:13) is daar vyf navorsingsareas waarin ’n korpusgebaseerde linguistiese analise reeds gebruik was (in die *BC*-tydperk), naamlik[102]:

1. **Bybel- en literêre studies**: Een van die eerste belangrike korpusgebaseerde navorsingsprojekte het die Bybel as ’n korpus gebruik, en vanaf 1737 is konkordansies saamgestel van woorde wat in die Bybel gebruik word (Kennedy, 1998:13). Stubbs (2007:317) verwys na die werk van Samuel Ayscough in 1790 waar ’n indeks van die woorde wat Shakespeare gebruik het, gepubliseer is.

   Lüdeling en Kytö (2008:2) verwys verder na die Hebreeuse konkordansie wat saamgestel is deur Isaac Nathan ben Kalonymus (ook bekend as Rabbi Mordecai Nathan) in die 15de eeu sowel as twee Engelse konkordansies wat onderskeidelik saamgestel is deur John Marbeck (15de eeu) en Alexander Cruden (18de eeu) en kan as belangrike vroeë werk beskou word wat verband hou met korpuslinguistiek.

[102] Sien Tabel 4.1 vir ’n oorsig van die verskillende korpora wat saamgestel is, beide voor en ná die aanvang van die elektroniese korpus era.

2. **Leksikografie**: Francis (1992:21) stel dat die samestelling van korpora vir leksikografie terugdateer na die 17de eeu. Lüdeling en Kytö (2008:6) voer weer aan dat die oorsprong vir die gebruik van korpora in die samestelling van woordeboeke selfs na die 16de eeu teruggevoer kan word. Svartvik (2007:12) verwys na die *Oxford English Dictionary* − waarvan die eerste uitgawe in 1928 voltooi is. Dié woordeboek is deur middel van leksikografiese korpora saamgestel waar meer as 4 miljoen kaartjies gealfabetiseer moes word.

   McEnery en Wilson (2001:106) sluit by bogenoemde aan en postuleer dat leksikograwe empiriese data gebruik het lank voordat die vakgebied (korpuslinguistiek) ontstaan het. Hulle (McEnery & Wilson, 2001:106) verwys ook na Samuel Johnson en die *Oxford English Dictionary* waartydens sitaatstrokies gebruik is om die inskrywings te formuleer.

   Kennedy (1998:14) voer aan dat die eerste publikasie van die *Oxford English Dictionary* die vrug van sowat 71 jaar se werk was waartydens ’n korpus saamgestel is vanuit die Engelse letterkunde. Kennedy (1998:15) stel verder dat die redakteurs van dié woordeboek (James Murray en Henry Bradley ingesluit) almal gesterf het voordat publikasie kon plaasvind. Sowat 2 000 individue het die ongeveer 5 000 000 sitate versamel ten einde die 50 000 000 woorde saam te stel wat die 414 825 inskrywings gedefinieer het.

   Kennedy (1998:15) verwys ook na Noah Webster se *Webster’s New International Dictionary* (1961) − die derde uitgawe van die *An American Dictionary of the English Language* (1828) − as ’n belangrike leksikografiese werk wat betrekking het op korpuslinguistiek.

3. **Dialekstudies**: Dit verwys na die ondersoek van leksikale variasie met betrekking tot verskillende woordkeuses of spelwyses (Kennedy, 1998:15). In die Verenigde Koninkryk is dit veral werk deur Wright (*The English Dialect Dictionary*, 1898-1905) en Ellis (*The Existing Phonology of English Dialects*, 1889) wat as belangrike bydraes beskou word (vgl. Kennedy, 1998:16).

4. **Opvoedkunde**: Navorsing vir pedagogiese doeleindes het volgens Kennedy (1998:16) van die belangrikste bydraes tot korpusgebaseerde navorsing gelewer. In 1921 het Thorndike ’n korpus saamgestel van 4 500 000 woorde (41 verskillende bronne is

hiervoor gebruik) om ’n frekwensielys saam te stel om Engels aan te leer − die korpus is later uitgebrei met die hulp van Lorge (vgl. Kennedy, 1998:16).

5. **Grammatika**: Navorsers van die grammatika van ’n taal maak al vir etlike jare gebruik van korpora en volgens Kennedy (1998:17) is daar talle navorsers wat in hierdie veld gewerk het (sien Tabel 4.2). Lüdeling en Kytö (2008:4) verwys na die vroeë werk van Jacob Grimm, Robert Lowth, George Curme, Hendrik Poutsma, Henry Sweet, Charles Fries, Otto Jespersen en Herman Paul met betrekking tot grammatika. Volgens McEnery en Wilson (2001:109) word korpora gereeld in grammatikale studies gebruik.

Die bogenoemde vyf navorsingareas verwys na ’n tydperk voor elektroniese korpora, en vandag word korpora in talle ander navorsingsvelde (soos forensiese linguistiek, outeurskapidentifikasie en dies meer) gebruik. Linguiste soos Boas, Firth, Sinclair, Busa (wat as die eerste persoon beskou word wat ’n rekenaarleesbare korpus saamgestel het), Sapir, Newman, Bloomfield en Pike het ’n kardinale rol in die ontwikkeling van korpuslinguistiek gespeel (vgl. Aston, 2011:2-3; McEnery & Wilson, 2001:3, 22).

Lüdeling en Kytö (2008:12) argumenteer tereg dat alhoewel pre-elektroniese korpora vandag as onnodig beskou word, kan ’n mens nie die invloed wat hierdie werk op die ontwikkeling van (moderne) korpuslinguistiek gehad het buite rekening laat nie aangesien dit die fondament geskep het vir verdere navorsing. Svartvik (2007:12) beweer dat linguiste wat op die sogenaamde “korpus-kar” gespring het toe hierdie kar in werklikheid nog maar ’n donkiekar was, sonder twyfel besef het dat rekenaarkorpora vandag as ’n hoofstroombenadering tot taalkunde beskou kan word.

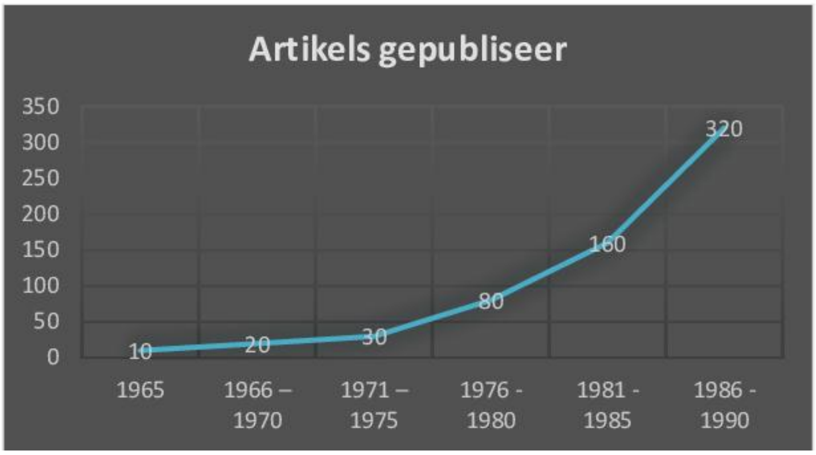
Om die gewildheid van dié term te bewys, het Svartvik (2007:12) ’n eenvoudige Google-soektog geloods waarin hy enkele woorde opgesoek het. Die resultaat van die soektog (soos gesien in Tabel 4.4) toon aan dat *korpuslinguistiek* as ’n gewilde term beskou kan word en dat daar ’n toename in die aantal resultate vir dié term plaasvind.

Tabel 4.4 Voorstelling van Svartvik se soektog in 2007 sowel as my eie soektog (2016) om die toenemende gewildheid van korpuslinguistiek as navorsingarea uit te beeld

Terme soos deur Svartvik nageslaan	**Resultate gevind** (Svartvik, 2007)	**Resultate gevind** (eie soektog in 2016)
Corpus linguistics	34 500	935 000

Ook Johansson (2008:47) het ’n eenvoudige soektog geloods om die toenemende gewildheid

van korpuslinguistiek aan te dui deur na die aantal artikels te verwys wat internasionaal oor die onderwerp gepubliseer word − soos sigbaar in Figuur 4.2:



Figuur 4.2 Toename in die publikasie van artikels oor korpuslinguistiek vanaf 1965-1990 (geneem en vertaal vanuit Johansson, 2008:47)

McEnery *et al.* (2006:3) stel dat korpuslinguistiek gedurende die 1950's verskriklik gekritiseer is en dat dit daarom tot 'n groot mate agterweë gelaat is. Die kritiek was dat korpora na bewering beperk en nie konstant was nie. Vanaf die 1960's[103] het korpuslinguistiek egter elektronies ontwikkel en word daar vandag tussen verskillende soorte elektroniese korpora[104] onderskei.

Svartvik (2007:14) verwys na die *Brown-korpus* in 1963 as die ontstaan van korpuslinguistiek − soos ons dit vandag ken − en stel dat te midde van teenkanting van onder andere linguiste soos Chomsky, korpuslinguistiek ontwikkel en gegroei het. Chomsky (1958:159) was van mening dat korpora oor té veel tekortkominge beskik het, en dat: "any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list."

Ten spyte van teenkanting, het navorsing met betrekking tot korpuslinguistiek steeds voortgegaan. Die *Survey of English Usage Corpus* − 'n korpus van een miljoen woorde wat die geskrewe en gesproke vorm van Britse Engels moes uitbeeld − is in 1959 deur Randolph Quirk geïnisieer (vgl. Meyer, 2008:10; Svartvik, 2007:15; Teubert, 2004:107). Hierdie projek (van Quirk) is in 1985 voltooi. Daar was talle bekende linguiste wat aan die projek gewerk het soos Valerie Adams, John Algeo, Dwight Bolinger, Noël Burton-Roberts, David Crystal, Derek Davy,

[103] Die aanvang van die era van elektroniese korpora volgens Anthony (2009:88), Granath (2007:169), Kennedy (1998:13) en Svartvik (2007:12).
[104] Sien afdeling 4.2 (Korpora).

Jan Fibras, Sidney Greenbaum, Liliane Heageman, Robert Ilson, Ruth Kempson, Geoffrey Leech, Terttu Nevelainen, Jan Rusiecki, Jan Svartvik en Joe Taglicht (vgl. Anon., 2015).

Meyer (2008:10) verwys na die *Survey of English Usage* (SEU) as die invloedrykste pre-elektroniese korpus. Quirk (1974:167) het werk aan die SEU geïnisieer aangesien hy van mening was dat:

> for all their excellence, the big grammars fell short … for two main reasons. They were based on writing, not speech and their generally eclectic use of [written] source materials too often leaves unclear the distinction between normal and relatively abnormal structures and the conditions for selecting the latter.

Vanuit Quirk (1974:167) se opmerking blyk dit dat daar ’n aanvraag was na korpora wat taal uitbeeld soos dit daagliks as omgangstaal gebruik word. Svartvik (2007:16) is van mening dat die projek (*Survey of English Usage Corpus*) − ’n analise van taal soos dit in die daaglikse lewe gebruik word − in skrille kontras staan met die hoofstroom Chomskyaanse beskouing van taal. Hierdie projek word volgens Svartvik (2007:20) as die pioniersprojek van ’n masjienleesbare korpus beskou. McEnery en Wilson (2001:22) verwys ook na die Brown-korpus deur W.N. Francis en H. Kucera in 1963 − ’n korpus wat ongeveer twee dekades geneem het om te voltooi.

Svartvik (2007:20) gaan voort en verduidelik dat die verwerking van data in die aanvangsjare van rekenaarkorpuslinguistiek nie so maklik soos vandag was nie. Brown Universiteit se totale rekenaarhoofraam moes vir die samestelling van die konkordansie gebruik word. Geen ander departement kon dan op daardie dag enige rekenaargebaseerde navorsing uitvoer nie. Hierdie projekte (*Survey of English Usage Corpus* en die *Brown*-korpus) sowel as projekte soortgelyk aan die bogenoemde voorbeelde, het aanleiding daartoe gegee dat korpuslinguistiek toenemend gebruik word vir breër navorsingsvrae en areas soos opvoedkunde, diskoersanalise, literêre stilistiek, forensiese linguistiek, pragmatiek, spraaktegnologie, sosiolinguistiek en kommunikasie (vgl. McCarthy & O’Keeffe, 2010:7).

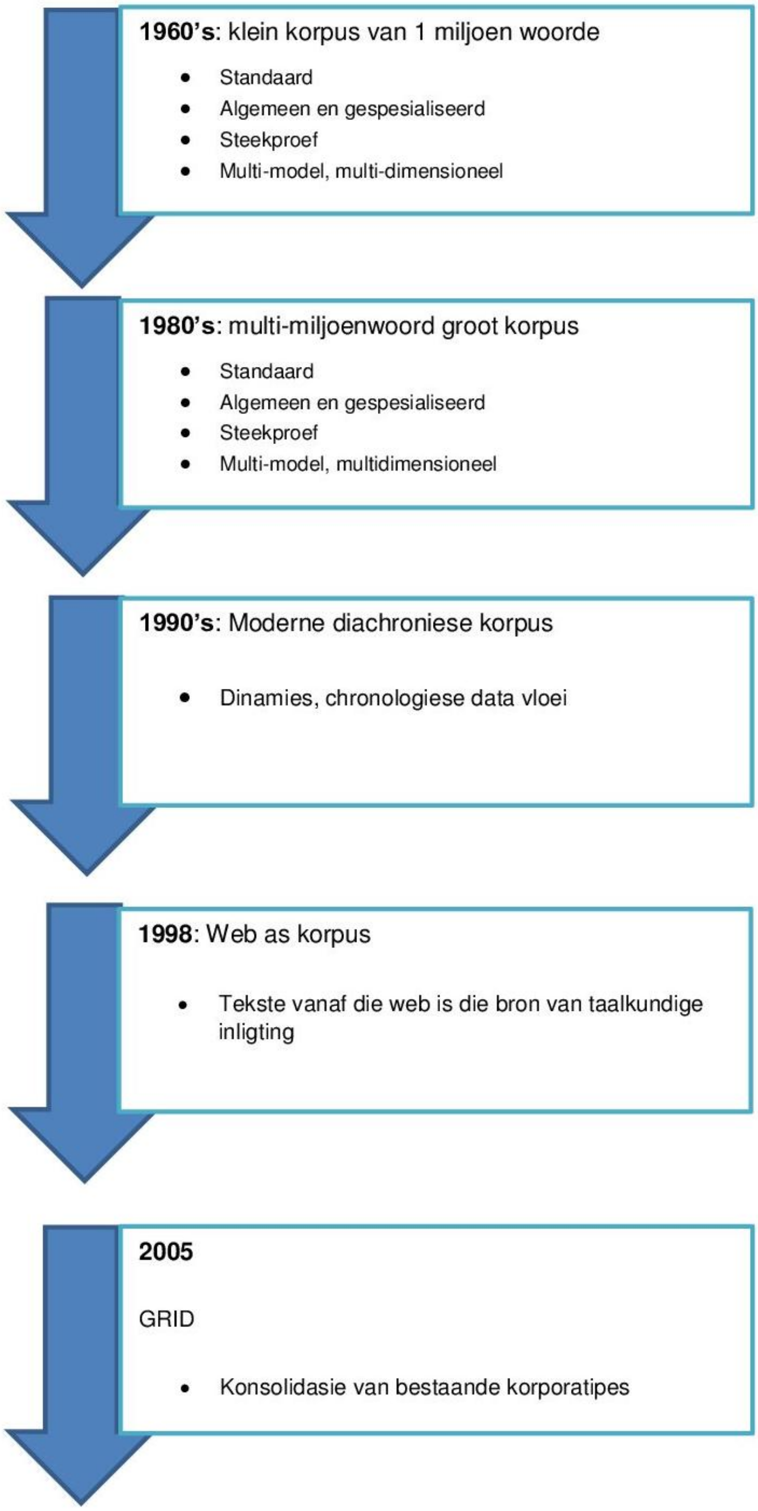
Meyer (2008:12) stel tereg dat die pre-elektroniese korpusanalise ’n tydrowende proses was, en dat dit vandag as onnodig beskou word. Hierdie manier van analises doen, het sonder twyfel ’n belangrike invloed op die ontwikkeling van korpuslinguistiek gehad. Vandag bestaan daar talle

rekenaarprogrammatuur[105] wat vir korpuslinguistiese analises gebruik kan word, maar hierdie programmatuur sou (volgens Meyer, 2008:13) nie bestaan het as dit nie vir individue soos Cruden en Jespersen se visie vir konkordansies en die moontlikhede van verdere navorsing was nie.

Svartvik (2007:22) stel dat rekenaars teen die 1970's 'n meer algemene verskynsel was, en dat die toeganklikheid en beskikbaarheid van rekenaars aanleiding gegee het tot nuwe en opwindende benaderings tot navorsing oor gesproke taal (veral in Engels). Johansson (2008:33) sluit by Svartvik aan en postuleer dat alhoewel die oorsprong van korpuslinguistiek sover as die 12de eeu teruggevoer kan word, kan die deurbraak vir korpuslinguistiek verbind word met die beskikbaarheid van masjienleesbare tekste. Analises wat voorheen as bykans onmoontlik beskou is, kan nou met gemak − met behulp van rekenaars natuurlik − uitgevoer word.

Vanaf die 1970's het die kwantiteit (en kwaliteit) van tekste wat beskikbaar is vir korpuslinguistiese analises merkbaar toegeneem. Volgens Johansson (2008:33) kan die ontwikkeling toegeskryf word aan die feit dat rekenaars nie net 'n alledaagse verskynsel geword het nie, maar ook aangesien rekenaars toenemend meer gebruikersvriendelik is. Renouf (2007:270) stel dat korpuslinguiste teen die 1980's steeds as onortodoks beskou is, en dat navorsers die grense van taaltegnologie moes probeer verskuif. Figuur 4.3 dui die verskillende stadia van die ontwikkeling van korpuslinguistiek volgens Renouf (2007:28) aan:

[105] Sien hoofstuk 5 (Afdeling 5.8 - Instrumente).



Figuur 4.3 Vertaalde weergawe van Renouf (2007:28) se voorstelling van die verskillende ontwikkelingsvlakke van korpuslinguistiek

Vanuit Figuur 4.3 blyk dit dat korpuslinguistiek stelselmatig (maar veral sedert die 1990's) ontwikkel het, en dat daar veral 'n verandering in korpusgrootte[106] plaasgevind het. McEnery en Hardie (2012:37) onderskei (soortgelyk aan Renouf se voorstelling van die stadia van korpuslinguistiek) tussen die pre-rekenaarera (voor 1960's), die oorgangstydperk (1960-1990) en die ontwikkelingsfase (1990's). Renouf (2007:30) argumenteer dat 'n toenemende belangstelling in die "ware" gebruik van taal aanleiding gegee het tot die aanvang van superkorpora in die 1980's (fase 2). Die term *superkorpora* verwys volgens Renouf (2007:30) na groter korpora as wat in die 1960's gebruik is.

Die gebruik van 'n korpus vir navorsing oor taal is dus 'n verskynsel wat al vir jare voorkom (woordelyste word al vanaf 1773 saamgestel) en die ontwikkeling van tegnologie het navorsing aansienlik vergemaklik. Vanuit hierdie afdeling (4.3) is dit duidelik hoe korpuslinguistiek ontwikkel het, hoe korpusgroottes verander het, hoe die tyd verminder het om korpora saam te stel asook hoe die aantal korpora vanaf die 1990's toegeneem het. Wat is korpuslinguistiek? Verwys dit na 'n metode of is dit 'n teoretiese beskouing?

4.4 Korpuslinguistiek gedefinieer

Korpuslinguistiek − en korpora − het ontwikkel vanaf 'n moeisame en tydrowende proses na 'n rekenaargedrewe dissipline waardeur nuwe dimensies van taal ondersoek kan word. Teubert (2008:62) argumenteer dat korpuslinguistiek 'n nuwe perspektief vir die analise van taal skep. Die tradisionele benadering waar taal geanaliseer en gekategoriseer word in 'n omgewing wat ooreenstem met die steriele omgewing waarin 'n nadoodse ondersoek uitgevoer word, word vervang met 'n interpretasie van taal *in vivo* – konteksgebonde met die implisiete en eksplisiete betekenis steeds met diskoersomstandighede verbind.

Korpuslinguiste stel korpora saam en analiseer dan hierdie korpora. Daarom kan korpuslinguistiek in alle eenvoud volgens Cheng (2012:6) gedefinieer word as die samestelling en analise van korpora. Wanneer die historiese oorsprong van korpuslinguistiek in gedagte gehou word, is dit duidelik dat daar 'n behoefte aan 'n manier bestaan het (en steeds bestaan) waarop taal − soos dit daagliks gebruik word − ondersoek en aangeteken kan word.

Chafe (1992:96) vra tereg die vraag "wat is 'n korpuslinguis?" en antwoord deur te postuleer dat hierdie term (*korpuslinguis*) behoort te verwys na 'n linguis wat poog om taal te verstaan deur die waarneming en analise van voorbeelde van natuurlike taal. Váradi (2001:592) sluit by Chafe (1992:96) aan en beweer dat een van die fundamentele doelwitte van korpuslinguiste is om taal

[106] Sien afdeling 4.2. (Korpora).

te ondersoek soos dit in die alledaagse “ware” lewe voorkom. Volgens Teubert (2008:64) beskou korpuslinguiste taal hoofsaaklik as ’n sosiale fenomeen. Taal is ’n openbare verskynsel wat in die publiek geskied en is dus waarneembaar. Sonder taal kan betekenis nie oorgedra word nie.

Hierdie sosiale aard van taal (soos reeds in afdeling 2.2 bespreek) gee volgens Verhoef (1991:10) aanleiding daartoe dat ‘n individu van kleins af leer hoe om hulle bedoelings oor te dra − dit veroorsaak sosiale beheer en beter kommunikasie. Die sosiale aard van taal veroorsaak ook dat individue oor unieke taalkenmerke kan beskik (sien afdeling 0). Hierdie unieke eienskappe – idiolek − kan deur middel van korpuslinguistiek en forensiese linguistiek geïdentifiseer word om sodoende te bepaal wie die outeur van ’n bepaalde teks is. Verder gee die sosiale aard van taal aanleiding tot verskillende variasies van ’n taal (sosiolekte) sowel as ’n keuse vir gepaste taalgebruik in ’n bepaalde situasie ten einde die betekenis van die boodskap suksesvol oor te dra.[107]

Teubert (2004:97, 99, 100) sluit by bogenoemde aan deur te stel dat betekenis ook ’n sosiale fenomeen is, en dat korpuslinguistiek ondersoek instel na betekenis in boodskappe (in die konteks van die diskoers waarin dit voorkom) en dat korpuslinguistiek meer oor betekenis kan aantoon as wat volgelinge van Chomsky se linguistiek ooit kon. Die ondersoek na betekenis selfs in terme van korpuslinguistiese analises word beperk en die volle prentjie kan dus nooit geskep word nie. Dus kan korpuslinguistiek beskou word as die manier waarop taal vanuit ’n nuwe perspektief benader en ondersoek kan word om inligting oor die aard van taal voor te hou.

Fischer-Starcke (2010:62) hou voor dat korpuslinguistiek insig oor taal bied met betrekking tot ’n spesifieke teks wat ondersoek word en dat die taalvariasie en -patroonmatigheid wat in die betrokke korpus sigbaar is, die gevolg is van ’n deeglike analise van die gekose tekste sowel as ’n vergelyking met ander tekste. Hierdie definisie verwys na die belang van die korpus wat ondersoek word, en hoe ’n spesifieke korpus ’n invloed op waarnemings kan hê. Ten einde die korpus te analiseer en gevolgtrekkings te kan maak, moet die korpus nie net verteenwoordigend van aard wees nie, maar daar moet sekere beperkings in plek wees om die geldigheid van die korpus en die ondersoek te verseker[108].

Nesselhauf (2005) definieer korpuslinguistiek eenvoudig as die rekenaaranalise van taal soos dit in die daaglikse lewe deur individue gebruik word, en dat rekenaarprogrammatuur gebruik word om ondersoek in te stel na die frekwensie van voorkoms van sekere aspekte. Deur

[107] Sien afdeling 4.7 (Forensiese en korpuslinguistiese benadering).
[108] Sien 4.2 (Korpora).

ondersoek in te stel na die frekwensie van die voorkoms van sekere (gekose) aspekte in taal, kan afleidings oor die aard van taal gemaak word. Daar kan byvoorbeeld bepaal word dat alhoewel taalreëls bepaal dat eiename met ’n hoofletter geskryf moet word (vgl. AWS, 2009:45 Reël 9.3), dit die norm op mikroblogs is om eiename nie met hoofletters te skryf nie.

Teubert en Čermáková (2004:165) se stelling dat korpuslinguistiek ons in die posisie plaas waar ons ons verder kan inlig oor die manier waarop ander individue taal gebruik − en dat hierdie soort kennis ons in staat stel om meer suksesvol te kommunikeer en funksioneer − kan as ’n betreklik goeie opsomming oor korpuslinguistiek beskou word. Hierdie definisie dui sonder twyfel die nut van korpuslinguistiek aan. Hetsy korpuslinguistiek gebruik word om ’n vreemde taal aan te leer, om die struktuur van ’n taal te ondersoek of gebruik word om outeurskapidentifikasie uit te voer, verander korpuslinguistiek die manier waarop ons taal sien en dit analiseer.

Fischer-Starcke (2010:35) sluit by die bostaande aan en sê:

> *Any statement about langue is based on probabilistic research of language features as represented in a corpus, that is, in a selection of language. As langue is the sum of all potentially possible units of parole, corpus linguistics is a linguistics of parole from which hypotheses about langue are induced.*

Korpuslinguistiek kan dus beskou word as die analise van die konkrete gebruik van ’n taal, en hoe die stelsel in die daaglikse gebruik van dié taal weerspieël word. McEnery en Hardie (2012:1) stel onomwonde dat korpuslinguistiek na ’n verskeidenheid metodes verwys waarvolgens taal bestudeer kan word, en dat korpuslinguistiek nie na vasgestelde metodes of prosedures vir die analise van taal verwys nie en dat alhoewel daar veralgemenings oor die aard van korpuslinguistiek bestaan, dit baie belangrik is om te besef dat korpuslinguistiek as’t ware na ’n heterogene[109] veld verwys.

Kennedy (1998:1) argumenteer, in aansluiting met McEnery en Hardie se stelling oor die heterogene aard van korpuslinguistiek, dat “corpus linguistics is not an end in itself but is one source of evidence for improving descriptions of the structure and use of languages, and for various applications, including the processing of natural language by machine and understanding how to learn or teach a language.” Dit blyk dus dat korpuslinguistiek as ’n benadering beskou kan word om verskeie navorsingsvrae te kan beantwoord hetsy ’n

[109] Hierdie heterogene aard van korpuslinguistiek kan waargeneem word in die verskillende navorsingsareas waar dit gebruik word.

korpuslinguistiese analise gebruik word om bestaande teorieë te ondersoek en te verifieer of om nuwe teorieë te toets.

Baker (2011:22) is van mening dat korpuslinguistiek meer in ander velde van linguistiek geïnkorporeer moet word sowel as in anders sosiale wetenskappe en humaniora soos geskiedenis, sosiologie, religieuse studies of politiek. Baker (2011:22) argumenteer verder dat die onderskeid tussen korpuslinguiste en sogenaamde teenstanders van korpuslinguiste in die toekoms sal vervaag en dat meer navorsers as korpusgebruikers na vore sal tree.

In die bostaande gedeelte is daar na talle definisies van korpuslinguistiek verwys. Definisies waaruit dit duidelik blyk dat korpuslinguistiek ’n heterogene veld is waar ’n korpus saamgestel word ten einde die aard van ’n taal te ondersoek en dat daar verskeie navorsingsvelde is waarin korpuslinguistiek toegepas kan word. Dit blyk ook duidelik dat tegnologiese ontwikkelinge tot die ontwikkeling van rekenaarprogrammatuur gelei het (en steeds ontwikkel word) om die korpuslinguistiese analise van taal − soos dit in die daaglikse lewe gebruik word − te vergemaklik.

Hierdie benadering wat ek vir my studie gebruik het, dui op die samestelling van ’n korpus[110] en ’n rekenaargedrewe analise wat fokus op die ondersoek na die frekwensie van die voorkoms van voorafbepaalde beperkings. Aangesien korpuslinguistiek heterogeen is, is daar nie ’n enkele manier wat alleenlik as die aanslag beskou kan word nie. Korpuslinguistiek is dus ook ’n benadering waar korpora aan die hand van voorafbepaalde grense geanaliseer word ten einde ’n spesifieke navorsingsvraag te beantwoord oor die aard van taal soos dit daagliks gebruik word.

Cheng (2012:7) is egter van mening dat ’n definisie vir korpuslinguistiek nie so eenvoudig is nie aangesien korpuslinguiste dit nie eens is of korpuslinguistiek ’n metode, ’n benadering of ’n teorie is nie. Vervolgens ondersoek ek verskillende standpunte oor die aard van korpuslinguistiek.

4.5 Korpuslinguistiek: metode of teorie

Sinclair (1991a:xvii, 1, 2, 4, 100) voer aan dat die analise van tekste (soos dit natuurlik gebruik word) hetsy geskrewe of gesproke deur middel van rekenaarprogrammatuur patrone in taal aantoon wat voor korpuslinguistiek nie analiseerbaar was nie. Leech (2011:157) stel onomwonde dat korpuslinguistiek as ’n sambreelterm beskou kan word vir linguistiese

[110] Sien 4.2 (Korpora) vir ’n definisie van ’n korpus sowel as ’n bespreking van die verskillende soorte korpora.

navorsing wat afhanklik is van korpora. Hierdie term kan misleidend wees aangesien dit na ’n subdissipline verwys, vergelykbaar met sosiolinguistiek, psigolinguistiek, forensiese linguistiek kliniese linguistiek en dies meer en wat verwys na interdissiplinêre toepassings van taalkunde.

Die beskikbaarheid van data het natuurlik aanleiding daartoe gegee dat groot hoeveelhede tekste geanaliseer kan word, wat taal anders uitbeeld as wanneer dit in ’n vakuum ondersoek word, en korpuslinguistiek asook korpora is dus nie net ’n belangrike hulpmiddel waardeur taal ondersoek kan word nie, maar kan ook as ’n belangrike konsep in linguistiese teorieë beskou word. Wat is korpuslinguistiek dan? ’n Metode of ’n teorie? Vervolgens ondersoek ek oorsigtelik die verskillende opinies oor korpuslinguistiek.

Dit is moontlik om te sê dat daar tot op hede nie uitsluitsel bereik is oor die aard van korpuslinguistiek nie. Volgens Cheng (2012:6) kan daar tussen twee hoofgroepe in die debat oor korpuslinguistiek onderskei word, naamlik diegene wat dit as ’n teorie beskou en dié wat aanvoer dat korpuslinguistiek ’n metodologie is. Volgens Cheng (2012:6) veroorsaak hierdie twee opinies dat korpuslinguiste hulle studies verskillend benader aangesien een groep korpuslinguistiek as ’n hulpmiddel (’n korpusgebaseerde benadering) beskou en die ander groep dit as ’n teorie (korpusgedrewe benadering) sien[111].

Leech (2011:157) skryf verder dat korpuslinguistiek in ’n kombinasie met enige subdissipline van taal gebruik kan word (hetsy sintaksis, pragmatiek of sosiolinguistiek). Leech (2011:158) argumenteer dus dat korpuslinguistiek as ’n metodologie beskou kan word − ’n wetenskaplike sfeer. Die teoretiese implikasies van korpuslinguistiek moet net nie buite rekening gelaat word nie.

McEnery en Wilson (2001:2) sluit by Leech se argument aan dat korpuslinguistiek nie as ’n vertakking van linguistiek beskou kan word nie. Korpuslinguistiek verskil dus van sintaksis, sosiolinguistiek en semantiek aangesien hierdie dissiplines daarop fokus om aspekte van taalgebruik te verduidelik. McEnery en Wilson (2001:2) gaan voort deur te argumenteer dat korpuslinguistiek veel eerder as ’n metodologie beskou kan word aangesien ’n korpusgebaseerde benadering antwoorde op talle taalnavrae oor sintaksis, semantiek en sosiolinguistiek kan bied.

Ook Davies (2011:65), Gries (2011:83) en Meyer (2004:xi) stel dat korpuslinguistiek eerder as ’n metodologie beskou moet word. Nesselhauf (2005) sluit by die bostaande argumente aan en

[111] Die verskil tussen ’n korpusgedrewe en ’n korpusgebaseerde benadering word in afdeling 4.6 (’n Korpuslinguistiese analise) verduidelik.

stel dat korpuslinguistiek na ’n metode verwys waardeur linguistiese analises uitgevoer kan word, en dat dit gebruik kan word om verskillende taalkundige vrae te beantwoord en bied nuwe insigte tot die bestudering van taal.

Vanuit die bostaande stellings is dit duidelik dat talle navorsers ooreenstem dat korpuslinguistiek wel na ’n metodologie verwys, of minstens in teenstelling is met argumente wat stel dat korpuslinguistiek ’n teorie is. Navorsers stem saam dat korpuslinguistiek met betrekking tot teoretiese beginsels ’n bydrae tot ander vakgebiede kan lewer. Tog is nie almal eens dat korpuslinguistiek na ’n metodologie verwys nie.

Tognini-Bonelli (2001:51) argumenteer met betrekking tot die vraagstuk oor die aard van korpuslinguistiek dat dit nie na ’n metodologie met vasgestelde beginsels verwys nie, maar dat dit liewer beskou moet word as:

> … a pre-application methodology. By “pre-application.” we mean that, unlike other applications that start by accepting facts as given, corpus linguistics is in a position to define its own sets of rules and pieces of knowledge before they are applied. [.…] Corpus linguistics has, therefore, a theoretical status and because of this it is in a position to contribute specifically to other applications ...

Johansson (2011:117) voer aan dat korpuslinguistiek nie ingeperk moet word nie, maar dat die klem in korpuslinguistiek wel metodologie is. Ook Scott (2011:214) en Swales (2011:222) argumenteer dat korpuslinguistiek ’n metodologie is, maar dat korpuslinguistiek veel eerder as ’n hulpmiddel beskou moet word aangesien navorsers deur middel van korpuslinguistiek data kan analiseer wat voorheen nie moontlik sou wees nie.

Anthony (2013:142) postuleer dat alhoewel daar min twyfel bestaan oor die aard van ’n korpuslinguistiese benadering, daar verskillende opinies is aangaande die omvang daarvan. Daar kan tussen twee hoofstrome in die debat onderskei word. Die eerste verwys na navorsers (soos Tognini-Bonelli) wat korpuslinguistiek as ’n teorie beskou. Die tweede groep (navorsers soos McEnery en Hardie, Wilson, Davies, Gries, Meyer en Nesselhauf) beskou korpuslinguistiek as ’n metode. Daar is ook navorsers (soos Leech, Johansson, Scott en Swales) wat nie die teoretiese implikasies en die manier waarop korpuslinguistiek as ’n hulpmiddel (instrument) in navorsing gebruik kan word buite rekening laat nie.

Conrad (2011:47) verwys na korpuslinguistiek as ’n benadering waardeur taal bestudeer kan word aangesien alle werk met betrekking tot korpora sekere algemene kenmerke in gemeen het. Sy argumenteer verder dat korpuslinguistiek nie as ’n metodologie beskryf kan word nie aangesien die metodes wat in korpuslinguistiek gebruik word, divers van aard is. Conrad (2011:50) is van mening dat die mees problematiese benaming vir korpuslinguistiek ’n “subdissipline” of self ’n “wetenskap” is aangesien korpuslinguistiese navorsing met ander vakgebiede geïnkorporeer (soos sosiolinguistiek, stilistiek en diskoersanalise) word. Tavakoli (2012:114) verwys na korpuslinguistiek as ’n benadering waardeur taalstrukture en die taalgebruik van ’n taal geanaliseer word soos dit in die alledaagse lewe gebruik word.

Dit blyk dus dat korpuslinguistiek as ’n metode, ’n benadering en ’n teorie beskou kan word. Vanuit navorsing is dit duidelik dat daar talle opinies oor die aard van korpuslinguistiek bestaan. Dit blyk duidelik dat daar grondige redes bestaan vir argumente om korpuslinguistiek as ’n teorie of ’n metode te beskou. Om dit as ’n benadering te erken, omsluit vir my beide die metodologiese en teoretiese aard daarvan.

Vervolgens definieer ek *korpuslinguistiek* as ’n benadering wat ek in my forensieslinguistiese analise gevolg het waardeur ek outeurskapidentifikasie uitgevoer het. Wanneer ’n korpuslinguistiese benadering gevolg word, word daar van ’n korpus gebruik gemaak. Hierdie korpus word geanaliseer ten einde vrae oor die aard van taal (soos dit elke dag gebruik word) te beantwoord. Aangesien ’n korpuslinguistiese benadering in talle navorsingsareas toegepas kan word, is dit vanuit die verskillende definisies van korpuslinguistiek duidelik dat dit ook moontlik om korpuslinguistiek in forensiese linguistiek te gebruik wanneer die outeur van ’n bepaalde teks geïdentifiseer moet word.[112]

Vanuit afdelings 4.1 tot 4.5 is daar ondersoek ingestel na die term *korpora* die verskillende groottes van korpora sowel as die oorsprong en aard van korpuslinguistiek. Vervolgens ondersoek ek ’n korpuslinguistiese analise en verduidelik hoe só ’n analise uitgevoer word.

4.6 ’n Korpuslinguistiese analise

Volgens Conrad (2011:47) is die metodes wat in korpuslinguistiek gebruik word divers, maar daar is ook sekere algemene kenmerke wat korpuslinguistiese analises met mekaar in gemeen het. Daar word tussen twee definitiewe denkrigtings onderskei met betrekking tot korpuslinguistiek, naamlik die korpusgebaseerde en korpusgedrewe benadering. Vervolgens word dié twee benaderings bespreek.

[112] Sien afdeling 4.7 (’n Forensiese en korpuslinguistiese benadering).

4.6.1 Korpusgebaseerde benadering

Wanneer ’n korpusgebaseerde benadering gevolg word, is dit belangrik om te besef dat die analise van ’n korpus nie net na ’n eenvoudige berekening van die voorkoms van taalkundige kenmerke verwys nie, maar dat dit ook ’n kwalitatiewe en funksionele interpretasie van die kwantitatiewe patrone insluit (vgl. Biber *et al.*, 2006:5).Hierdie benadering het ten doel om patroonmatigheid in die tekste (dus die taal) te identifiseer en analiseer. McEnery en Hardie (2012:6) argumenteer dat korpusgebaseerde studies tipies van korpusdata gebruik maak om ’n teorie of hipotese te ondersoek. Hierdie teorie/hipotese bestaan reeds in die literatuur en ’n korpusgebaseerde studie het dus ten doel om die teorie/hipotese te bewys (hetsy reg of verkeerd) of om bykomende navorsing oor die teorie/hipotese uit te voer.

Fischer-Starke (2010:52) definieer ’n korpusgebaseerde analise as ’n analise waar ’n hipotese bewys word deur middel van die gebruik van korpuslinguistiese tegnieke.

Biber *et al.* (2006:4) voer aan dat ’n korpusgebaseerde analise oor die volgende eienskappe beskik:

- Die ondersoek moet die vorm aanneem van ’n empiriese analise wat die patrone in tekste analiseer.
- Die ondersoek moet uitgevoer word op ’n groot aantal tekste wat spesifiek vir die doel van die ondersoek versamel is, en wat as ’n korpus bekend staan.
- In die ondersoek moet daar van rekenaarprogrammatuur gebruik gemaak word ten einde die analise te voltooi.
- Daar moet van beide kwalitatiewe en kwantitatiewe benaderings gebruik gemaak word.

Biber *et al.* (2006:5) gaan voort deur te argumenteer dat wanneer die bogenoemde eienskappe gekombineerd in ’n analise gebruik word, kan die analise as omvangryk en betroubaar beskou word. Volgens Fischer-Starke (2010:52) word die eienskappe wat geanaliseer word vooraf bepaal, en ’n korpus word dus gebruik om te bepaal of die eienskappe wél aanwesig is (of nie) om sodoende ’n hipotese te bewys; daarom kan dié benadering as deduktief beskou word.

’n Korpusgebaseerde benadering verwys dus na ’n metodologie wat sigbaar word in ’n korpus terwyl die navorser ’n teorie wil bewys. Hierdie teorieë verwys na taalteorieë wat geformuleer is lank voor die beskikbaarheid van groot korpora (vgl. Tognini-Bonelli, 2001:65). Verder argumenteer Tognini-Bonelli (2001:65) dat korpuslinguiste wat die korpusgebaseerde benadering volg, in ’n verhouding tussen die teorie en die data glo wat vir ’n betrokke korpus

ingesamel is, en dat dié verhouding fundamenteel die model van taal sal bevestig wat hulle wil beskryf.

In ’n korpusgebaseerde analise word ŉ vraag dus gestel en die korpus word gebruik om dié vraag te beantwoord. Data word dus ingesamel na gelang van die voorafbepaalde vraag wat gestel is, en word dan aan die hand van hierdie navorsingsvrae geanaliseer.

McEnery en Hardie (2012:150) argumenteer dat navorsers wat ŉ korpusgebaseerde benadering volg, gebruik maak van reeds bestaande teorieë en dat ŉ korpus slegs ’n hulpmiddel is om hipoteses te toets. Dit kan gebeur dat wanneer die data geanaliseer word, dit blyk dat die data nie die gewenste resultate lewer nie. In só ’n geval kan navorsers die navorsingsvrae aanpas. McEnery en Hardie (2012:150) verwys na só ’n geval as ’n korpusgebaseerde analise wat dan ŉ korpusgedrewe benadering volg. Alhoewel ŉ korpusgebaseerde en -gedrewe benadering verskillend is, kan dit tog ook oorvleuel.

Beide Sinclair (2004b:53) en Tognini-Bonelli (2001:67-71) voer aan dat navorsers nie data objektief analiseer nie en dit ŉ inherente probleem met ’n korpusgebaseerde benadering is. Inteendeel, korpora kan eerder aangepas word om bestaande teorieë te ondersteun as wat die teorieë verkeerd bewys word. Tognini-Bonelli (2001:68) voer aan dat data nie as sekondêre inligting beskou kan word nie, dit is immers ’n voorbeeld van hoe taal in die alledaagse lewe gebruik word. ŉ Korpusgedrewe benadering kan dus as meer betroubaar en objektief beskou word. Vervolgens hou ek die beginsels van ’n korpusgedrewe benadering voor.

4.6.2 Korpusgedrewe benadering

Anthony (2013:146) voer aan dat wanneer die waarneming van ŉ korpus as die beginpunt van die analise beskou word, die benadering korpusgedrewe is en hierdie benadering het ten doel om ’n linguistiese model te skep. McEnery en Hardie (2012:6) wys daarop dat korpuslinguiste wat van ŉ korpusgedrewe benadering gebruik maak, die karakterisering van korpuslinguistiek as ’n metode verwerp en stel dat ŉ korpus die hoofbron van ’n hipotese oor taal moet wees. Tognini-Bonelli (2004:84) sluit hierby aan en postuleer dat ’n teorie oor taal in ŉ korpus sigbaar sal wees.

Wanneer ŉ korpusgedrewe benadering gevolg word, ondersoek en analiseer navorsers ŉ korpus sonder enige voorafopgestelde aannames of verwagtinge van resultate. Navorsers probeer nie om ’n reeds bestaande teorie te bevestig nie, navorsers stel juis ondersoek in na patrone wat sigbaar word in die analise van ŉ korpus ten einde ontwikkelinge in die taal te verduidelik (vgl. Johansson, 2011:117).

Fischer-Starke (2010:52) dui aan dat ’n korpusgedrewe en ’n korpusgebaseerde analise van mekaar verskil – ’n korpusgebaseerde analise bewys ’n hipotese deur middel van ’n korpuslinguistiese analise en ’n korpusgedrewe analise se doel is om hipoteses te ontwikkel na gelang van die analise van ’n korpus. Taaleienskappe wat in ’n korpusgedrewe analise ondersoek word, is taaleienskappe wat betekenisvol is − eienskappe wat gereeld in ’n betrokke korpus voorkom. Eienskappe word dus eers geïdentifiseer terwyl ’n betrokke korpus geanaliseer word.

Volgens Tognini-Bonelli (2001:84) kan ’n korpusgedrewe benadering gedefinieer word as linguiste se verbintenis tot die integriteit van die data in geheel, en die beskrywende aard van die bewyse wat vanuit ’n korpusgedrewe ondersoek ontbloot kan word . ’n Korpus is dus veel meer as net nog data om ’n teorie mee te bewys – ’n korpus is die hoofbron waaruit nuwe teorieë geformuleer kan word (vgl. Tognini-Bonelli, 2001:85). Sinclair (2004a:191) sluit by die bogenoemde aan deur te verwys na die analise van “rou” data in ’n korpusgedrewe benadering as ’n manier waarop taalpatrone waargeneem word.

’n Korpus is daarom meer as net ’n manier waarop bewyse kan bepaal dat ’n teorie wel gegrond is. ’n Korpus word dus as die hoofbron van data beskou waaruit afleidings gemaak en hipoteses gestel word wat tot ’n uniforme teoretiese stelling veralgemeen kan word (vgl. Tognini-Bonelli, 2001:85). ’n Korpusgedrewe analise kan dus as induktief beskou word aangesien die navorsingsvraag ontwikkel word soos ’n korpus geanaliseer word.

In ooreenstemming met Sinclair (2004b:53) en Tognini-Bonelli (2001:67-71) wat stel dat navorsers wat ’n korpusgedrewe benadering volg nie noodwendig objektief kan wees nie, voer Johansson (2011:117) aan dat ’n korpusgedrewe benadering ook nie sonder aannames en voorafbepaalde idees kan geskied nie. Johansson (2011:117) argumenteer dat aannames onafwendbaar is, en dat die interpretasie van patrone wat deur ’n korpusgedrewe benadering blootgelê word, nie kan geskied sonder ’n intuïtiewe taalkennis wat alleenlik bekom kan word deur die bestudering van bestaande taalteorieë nie.

McEnery en Hardie (2012:150) se argument dat navorsers wat ’n korpusgebaseerde benadering volg tog soms ook korpusgedrewe werk, kan dus ook van toepassing wees op ’n korpusgedrewe benadering aangesien navorsing nooit in ’n vakuum plaasvind nie. Al word ’n korpusgedrewe benadering dus gevolg en word die data (korpus) as die hoofbron van ondersoek en inligting beskou en al word die data nie gebruik om bestaande teorieë te bewys nie, beskik navorsers oor agtergrondkennis. Aannames word dus wel gemaak oor die aard van die data wat vanuit ’n korpus gekry kan word.

Vir die doel van my studie het ek van beide ’n korpusgedrewe en -gebaseerde benadering gebruik gemaak. Aangesien ek Afrikaans soos dit op mikroblogs gebruik word, ondersoek het alvorens ŉ forensiese analise (outeurskapidentifikasie) van die tekste van spesifieke outeurs kon plaasvind, was dit onmoontlik om die data te analiseer sonder om bestaande taalteorieë in ag te neem. Afleidings oor die Afrikaans wat op mikroblogs gebruik word kon wel gemaak word, en die vermoë om die outeurs van hierdie mikroblogtekste te identifiseer, kon slegs geskied as ek ŉ korpus as my hoofbron van inligting beskou het.

Deur bestaande teorieë te erken en steeds ŉ korpus primêr te plaas, kon ek tekste forensieslinguisties volgens ’n korpuslinguistiese benadering analiseer. In die daaropvolgende afdeling (afdeling 4.7) bespreek ek ŉ forensiese en korpuslinguistiese benadering.

4.7 ’n Forensiese en korpuslinguistiese benadering

Hoe skakel korpuslinguistiek egter met forensiese linguistiek? En op watter manier kan korpuslinguistiek voordelig gebruik word in ’n eksploratiewe forensiese analise? Coulthard (1994:27, 30) beantwoord dié vrae moeiteloos wanner hy stel dat korpuslinguistiek as een van die belangrikste hulpmiddels vir forensiese linguistiek beskou kan word, en dat in enige forensieslinguistiese ondersoek wat onderneem word, moet die metodologie die samestelling en analise van ’n korpus insluit.

> Die argument kan dus gemaak word dat ’n forensieslinguistiese analise nie kan plaasvind sonder ’n deeglike oorsig en begrip oor korpuslinguistiek nie. Wanneer die gebruik van korpuslinguistiek in forensieslinguistiese analises ondersoek word, is dit duidelik dat ’n metode waar forensiese en korpuslinguistiek aanvullend tot mekaar gebruik word ’n suksesvolle benadering blyk te wees. Die eerste forensieslinguistiese ondersoek waarin korpora gebruik is, was Mosteller en Wallace (1965) se ondersoek na die *Federalist Papers*[113], waar die tekste geanaliseer is ten opsigte van funksiewoorde om te bepaal wie die ware outeurs van hierdie tekste was. Olsson (2012:84) verwys na Coulthard se *Bentley*-saak as die pioniersaak wat die gebruik van korpuslinguistiek in ’n forensieslinguistiese analise betref.

Soos forensiese linguistiek ontwikkel het, het die manier waarop korpuslinguistiek in die dissipline toegepas word ook ontwikkel. Talle forensieslinguistiese studies maak gebruik van korpora ten einde resultate te verkry. Hierdie studies sluit in:

[113] Sien ook afdeling 3.4 (Ontstaan en ontwikkeling van forensiese linguistiek).

- **Winter (1996)** stel ondersoek in na drie kriminele verklarings (die verklarings wissel tussen 481 en 616 woorde in lengte) om te bepaal of outeurskapidentifikasie wel suksesvol uitgevoer kon word, veral aangesien die lengte van tekste wat forensiese linguiste moet ondersoek en analiseer toenemend korter word. In ’n ander studie ondersoek Winter (1996) twee musikante se idiolek deur “korpora” saam te stel van elk van die musikante se tekste en die korpora forensies linguisties met mekaar te vergelyk.

- **De Klerk (2003)** stel ’n korpus saam van die transkripsies van die hofverrigtinge van amnestieverhore om deur middel van ’n korpusbenadering te bepaal of die taal wat in die hofverrigtinge gebruik was as regverdig beskou kon word.

- **Cotterill (2004)** stel ’n korpus (5 miljoen woorde) saam om die leksikale aspekte van die kruisverhore van getuies te ondersoek.

- **Solan en Tiersma (2004)** bespreek die verskillende benaderings (veral korpuslinguisties) tot outeurskapidentifikasie.

- **Grant (2007)** verwys na moontlike maniere waarop data ingesamel kon word om ’n korpus saam te stel sowel as ’n metode wat in analises met betrekking tot outeurskapidentifikasie gebruik kon word.

- **Jones en Bennell (2007)** onderskei statisties tussen korpora wat uit 33 outentieke en 33 nagebootste selfdoodbriewe bestaan het om ’n metode daar te stel wat gebruik kan word om te bepaal of ’n selfdoodbrief outentiek of nageboots is.

- **Barlow (2010)** stel ondersoek in na die idiolektiese variasie van vyf van die Amerikaanse perssekretaresses in om leksikale en sintaktiese kenmerke te identifiseer.

- **Fitzpatrick en Bachenko (2010)** ondersoek die verskillende maniere waarop ’n korpus saamgestel en geannoteer kon word vanuit verklarings en polisie-ondervragings om sodoende ongerymdhede te identifiseer en inligting oor die betrokke sake te verifieer.

- **Ishihara (2014)** maak gebruik van ’n korpus (saamgestel deur die Nasionale Universiteit van Singapoer) wat uit SMS-tekste bestaan om outeurskapidentifikasie uit te voer met die fokus op n-gramme.

- **Johnson en Wright (2014)** analiseer ’n 2,5 miljoen-woord-korpus (saamgestel vanuit 63 000 e-posse vanaf Enron-werknemers) om ondersoek in te stel na idiolek en

outeurskapidentifikasie. Daar is veral gefokus op n-gramme wat outeurs se skryfstyl kan identifiseer.

- **Larner (2014a)** analiseer ’n korpus van 100 tekste (geskryf deur 20 verskillende outeurs) om ondersoek in te stel na die moontlikheid van die gebruik van ’n formule reeks as ’n metode van onderskeid tussen outeurs.

- **Larner (2014b)** doen navorsing oor die betroubaarheid van die web as ’n korpus, spesifiek met betrekking tot forensiese outeurskap-analises.

- **Hardaker (2015)** ’n korpus op vanaf twee RBK-platforms en kombineer korpuslinguistiek met pragmatiek (korpus-pragmatiek) om sodoende te bepaal hoe gebruikers reaggeer op RBK wat as *trolling* beskou kan word.

- **Hardaker en McGlashan (2015)** ondersoek die toenemende voorkoms van dreigemente (moord en verkragting) op sosiale media. Deur ’n korpus vanaf Twitter saam te stel word daar spesifiek gefokus op dreigemente wat gerig is aan ’n spesifieke Twitter-gebruiker (’n feminis en joernalis) om sodoende die taal in die gegewe sosiale konteks kwantitatief en kwalitatief te analiseer.

Vanuit die voorbeelde van studies waar korpuslinguistiek in forensieslinguistiese studies gebruik is, word dit duidelik dat ’n korpuslinguistiese benadering tot forensiese linguistiek die resultate verbeter. Daar is talle ander studies waar korpuslinguistiek in ’n forensieslinguistiese analise gebruik is, maar bogenoemde voorbeelde bied reeds ’n opsommende oorsig van die soort navorsing waar ’n forensiese en korpuslinguistiese benadering gevolg is.

Olsson (2012:84) voer aan dat korpuslinguistiek veral as ’n suksesvolle benadering beskou kan word wanneer dit in ’n forensieslinguistiese ondersoek toegepas word. Forensiese linguiste bereken die frekwensie van die voorkoms van woorde en frases in ’n verskeidenheid tekstipes – korpora − en dit is ’n tipiese scenario waarin forensiese linguiste die outeurs van bepaalde tekste moet bepaal deur ’n analise. Hierdie analise behels gewoonlik die ontleding van talle tekste van verskeie outeurskandidate (vgl. Olsson, 2012:84).

Vanuit hierdie afdeling blyk dit dat korpuslinguistiek − en veral korpora − ’n kardinale rol in forensieslinguistiese analises speel. Dit sal dus ondenkbaar wees om ’n outeurskapidentifikasie uit te voer sonder die gebruik van ’n korpus.

4.8 Samevatting

In hoofstuk 3 het ek forensiese linguistiek en die kenmerke van hierdie dissipline bespreek. Ek het aangetoon dat ek vir die doel van my studie forensieslinguisties ondersoek ingestel het na outeurskapidentifikasie. Hoofstuk 4 beskryf korpuslinguistiek as ʼn navorsingsveld.

Hoofstuk 4, kan as ʼn oorsig rakende belangrike literatuur van korpuslinguistiek gesien word, en die hoofstuk dui dan ook op die leemte wat daar voorkom nog in hierdie tipe navorsing in Suid-Afrika, en veral Afrikaans, voorkom. Die bydrae van hierdie hoofstuk is om ʼn agtergrond rakende korpuslinguistiek te geskep het, sowel as ʼn forensiese en korpuslinguistiese benadering voor te hou. Daar is spesifiek na die verskillende definisies van korpora ondersoek ingestel, en die verskillende tipes kopora is ook voorgehou. Dit blyk dat die ontwikkeling van tegnologie daartoe gelei het dat die web ook as korpus gebruik kan word, en daar is verwys na die verskillende tipes korpora wat anlyn beskikbaar is. Hier is ook na Afrikaanse korpora verwys.

Ten einde ʼn korpusgebasseerde analise in die empiriese hoofstukke (hoofstukke 6 en 7) te kan uitvoer, is daar verder ook na die ontwikkeling van korpuslinguistiek verwys, sowel as die wyse waarop korpuslinguistiese analises verander het. Navorsing wat reeds gedoen is, waar daar van beide korpus- en forensiese linguistiese benaderings gebruik gemaak is, is ook voorgehou. Dit is veral in die aanbeiding van die korpora in Afrikaans (afdeling 4.2.4) en ʼn forensiese en korpuslinguistiese benadering (afdeling 4.7) wat dit duidelik blyk dat daar in Afrikaans nog talle leemstes op hierdie gebied bestaan, en dit is dan ook hier waar my studie ʼn bydrae tot die ontwikkeling van hierdie vakgebiede in Afrikaans kan lewer.

Vanuit hierdie afdelings (afdelings 4.1 tot 4.7) kon ek aantoon (en verduidelik) dat ek korpuslinguistiek as ʼn benadering beskou wat ek gevolg het ten einde ʼn sinchroniese statiese geskrewe korpus saam te stel vanaf die web (die web word dus as bron van my korpus beskou). Verder is dit ook duidelik dat die onderskeid tussen ʼn korpusgedrewe en ʼn -gebaseerde benadering, alhoewel duidelik, nie altyd gevolg kan word nie. Daarom word daar van beide dié benaderings gebruik gemaak. Afdeling 4.7 bied ʼn oorsig van ʼn forensies- en korpuslinguistiese benadering en hoe korpora veral outeurskapidentifikasie beïnvloed.

Met die voltooiing van hoofstuk 3 en 4 (wat ʼn oorsig oor forensiese linguistiek en korpuslinguistiek bied) is die benaderings wat ek moes volg en in gedagte hou toe ek my metodologie saamgestel het dus duidelik. Daar bestaan verskillende soorte korpora, en wanneer ʼn korpus saamgestel word, is daar verskeie oorwegings wat in gedagte gehou moet

word. Hoofstuk 4 skep saam met hoofstuk 2 en 3 die teoretiese agtergrond wat gebruik is in hoofstuk 5 (die metodologie) en ook in hoofstuk 6 en 7 (die empiriese analise) ten einde die studie oor outeurskapidentifikasie en mikroblogs: ’n eksploratiewe forensiese analise in ’n digitale era uit te voer.

Hoofstuk 5

Metodologie

Language users never choose words randomly, and language is essentially non-random.
Kilgarriff (2005:263)

5.1 Inleiding

Hierdie hoofstuk sit die metodologie van die navorsing (wat in die empiriese hoofstukke gebruik is) uiteen. In die vorige hoofstukke (hoofstuk 1, 2, 3 en 4) is die agtergrond en teoretiese raamwerk vir die studie geskep. Dit is duidelik dat die perfekte forensieslinguistiese scenario met betrekking tot outeurskapidentifikasie met ’n klein aantal (verdagte) outeurs en groot hoeveelhede teks werk. Soos dit in hoofstuk 3 (afdeling 3.5 Outeurskapidentifikasie in ’n digitale era) verduidelik word, blyk dit dat RBK aanleiding gee tot korter tekste wat geanaliseer moet word, en meer spesifiek dat die beperkte lengte van mikrobloginskrywings talle uitdagings aan navorsers bied volgens Green en Sheppard (2013:64).

Daar is reeds studies gedoen waar outeurskapidentifikasie suksesvol uitgevoer is op korter tekste − hierdie tekste was onderskeidelik 500 (Ledger & Merriam, 1994:239) en 250 (Forsyth & Holmes,1996:165) woorde lank. Green en Sheppard (2013:64) argumenteer dat hierdie minimum tekslengtes steeds hoër as die gemiddelde tekslengte van ’n twiet (minder as 25 woorde) is. Schwarts *et al. (*2013:1881) kon egter met 91% akkuraatheid die outeur van ’n twiet identifiseer (in dié studie het hulle 1000 twiets van 50 outeurs geanaliseer)[114]. Hierdie hoofstuk se doel is dus om die metodologiese agtergrond te skep wat die basis van die empiriese studie in hoofstuk 6 vorm waardeur daar bepaal is of outeurs van mikrobloginskrywings wel geïdentifiseer kon word.

Reeds in hoofstuk 3 is daar verwys na Chaski (2008:1) se argument dat die metodes wat in outeurskapidentifikasie gebruik word in kontras met mekaar staan. Daar word deurgaans na hierdie stelling van Chaski verwys aangesien die metodologie wat ek in my studie gebruik het as ’n gemengde metode beskryf kan word. Dit wil sê beide ʼn kwalitatiewe (teksverwante data – ʼn stilistiese analise) en ʼn kwantitatiewe (numeriese data – ʼn stilometriese analise) metode is gebruik (Angouri, 2010:29)[115].

[114] Sien 5.5 (Navorsingsmetodes).
[115] Sien afdeling 5.5 (Navorsingsmetodes).

In hoofstuk 1[116] hou ek die drie navorsingsvrae voor waarin ek dit duidelik stel dat daar ondersoek ingestel word na 1) die taalstilistiese en stylkenmerke van mikroblogs in Afrikaans; 2) die identifisering van die outeurs van mikrobloginskrywings en 3) wat die metode behels wat gevolg is ten einde die suksesvolle identifikasie van mikroblogouteurs[117]. Ten einde hierdie vrae te kon beantwoord, moes ek ŉ korpus saamstel. Wanneer ŉ korpus saamgestel word, is daar etiekkwessies wat in gedagte gehou moet word.

In hierdie hoofstuk bespreek ek eerstens die manier waarop deelnemers vir die studie gekies is. Verder word die verskillende maniere van dataversameling voorgehou, gevolg deur ŉ oorsig oor bestaande navorsingsmetodes. In hierdie geval word daar na outeurskapidentifikasie in die algemeen verwys alvorens daar spesifiek na die identifikasie van outeurs op mikroblogs ondersoek ingestel word.

Die gemengde metodes word ook bespreek met spesifieke verwysing na die verskillende aspekte wat ondersoek is sowel as die verskillende rekenaarprogrammatuur wat gebruik is. Definisies van enkele terme word ook voorgehou. Verder sit ek die raamwerk van die metodologie wat ek in die empiriese hoofstuk gebruik het, uiteen. Ten slotte bespreek ek die etiekkwessies wat verband hou met die studie.

5.2 Navorsingsontwerp

Soos reeds genoem in hoofstuk 1, het my navorsing die vorm van ŉ eksploratiewe studie aangeneem. Ten einde die forensieslinguistiese ondersoek op die korpus uit te voer, het ek dus ŉ gemengde navorsingsontwerp gevolg. Ten einde die gemengde navorsingsontwerp optimaal te kon toepas en verstaan, het ek ŉ pragmatiese wêreldbeskouing gehuldig. Volgens Creswell (2013:11) verbind pragmatiste hulleself nie tot ŉ enkele filosofiese beskouing nie. Trouens, hierdie benadering stel dat navorsers tussen verskillende navorsingsontwerpe kan kies, en sodoende kan die metodes, tegnieke en prosedures gekies en gevolg word wat aan die kriteria vir hulle studie se vereistes voldoen en tot gevolg sal hê dat die doel van die studie bereik word.

Creswell (2013:11) skryf:

> Pragmatists do not see the world as an absolute unity. In a similar way, mixed methods researchers look to many approaches for collecting and analyzing data rather than subscribing to only one way ... Thus, for the mixed methods researcher, pragmatism opens the door to multiple methods,

[116] Sien afdeling 1.3 (Navorsingsvrae).
[117] Om die outeurs van mikroblogs te identifiseer, is daar van beide forensies- en korpuslinguistiese benaderings gebruik gemaak.

> different worldviews, and different assumptions, as well as different forms of data collection and analysis.

Vanuit hierdie aanhaling blyk dit dat die pragmatiese benadering − soortgelyk aan die gemengde metode – probeer om die beste en mees gepaste metodes, teorieë en benaderings te vind om die betrokke probleem op te los. Soos reeds genoem, het ek ’n gemengde metode in my studie gevolg. Die kwantitatiewe en kwalitatiewe aard van die metode word in afdeling 5.6 en 5.7 voorgehou. Ten einde die navorsingsontwerp te begrond kan daar aangevoer word dat die ontwerp van my studie as ’n geïntegreerde gemengde metode beskou kan word.

Hierdie gemengdemetodebenadering betrek volgens Creswell (2013:16) beide die konvergente en sekwensiële gebruik van data. Die kernaspek van hierdie ontwerp is dat die kwantitatiewe of kwalitatiewe data vir ’n groter ontwerp gebruik word (eksperimentele ondersoek) en dat die data ’n ondersteunende rol in die ontwerp van die navorsing speel. Ivankova *et al.* (2011:269) stel dat ŉ geïntegreerde gemengde metode veral nuttig is wanneer daar ’n tweede navorsingsvraag is wat verskil van, maar tog ook verwant is aan die eerste navorsingsvraag. Deur tradisionele kwantitatiewe en kwalitatiewe navorsingsmetodologieë te kombineer, kan vrae beantwoord word met behulp van bykomende data-analises (vgl. Ivankova *et al.*, 2011:270).

Hierdie gemengde navorsingsontwerp van my studie het die dus die volgende vrae aangespreek: Wat is die taalstilistiese en stylkenmerke van Afrikaans soos dit op mikroblogs gebruik word?; en Wat is die forensiese beginsels way gebruik kan word om outeurskapidentifikasie op mikroblogs uit te voer en tot watter mate is elk geslaagd, al dan nie? Omdat ek van ŉ geïntegreerde gemengde metode gebruik gemaak het, is ’n sekondêre datastel gebruik om die bevindinge vanuit die primêre datastel te ondersteun.

In die daaropvolgende afdelings verwys ek na die metodes (data-insameling en -verwerking) van my studie. Ek sal ook die manier waarop deelnemers gekies is, voorhou alvorens ek die metodologieë en instrumente bespreek.

5.3 Deelnemers

Aangesien daar nie ŉ korpus in Afrikaans bestaan waar tekste vanaf Twitter ingesluit is nie, het ek vir die doel van my studie twee afsonderlike korpora saamgestel 1) ’n algemene korpus waardeur ek die taalstilistiese en stylkenmerke van mikroblogs geverifieer het 2) ’n kleiner groep gespesialiseerde korpora bestaande uit tekste van tien deelnemers waarmee ek ondersoek

ingestel het na outeurskapidentifikasie en kenmerke van Afrikaans op mikroblogs. My studie het dus van twee verskillende deelnemergroepe gebruik gemaak.

Die manier waarop deelnemers vir die onderskeie korpora gekies is, verskil aangesien die doel van die onderskeie korpora verskil. Soos reeds genoem, het ek van ’n gemengde metode gebruik gemaak − my navorsingdeelnemers is dus op verskillende maniere gekies. Ivankova en Creswell (2009:147) voer aan dat kwantitatiewe navorsing ’n relatiewe groot datastel vereis wat gewoonlik deur middel van ŉ waarskynlikheidsteekproefneming ingesamel word terwyl kwalitatiewe navorsing meer spesifiek van aard is en daarom word ŉ niewaarskynlikheidsteekproefneming met dié navorsingsmetode geassosieer.

Daar blyk dus verskillende metodes te wees waarvolgens ’n steekproef saamgestel kan word en dat daar ’n verskil is tussen die manier waarop deelnemers vir ’n kwantitatiewe of ŉ kwalitatiewe metode gekies word.

Met betrekking tot die eerste groep (hierna verwys as groep 1) deelnemers, is hulle data gebruik vir die samestelling van die algemene korpus. Ek het dus ’n groot aantal data benodig wat as verteenwoordigend van die Afrikaans op die mikroblog Twitter beskou kon word aangesien groep 1 se data as verwysingskorpus moes dien om die vraag: “Wat is die stilistiese en stylkenmerke van Afrikaans op mikroblogs?” te beantwoord. Ten einde ’n volledige korpus te kon saamstel (met ŉ verteenwoordigende datastel), moes my deelnemers ook verteenwoordigend wees van die verskillende stratawat op Twitter gevind word.

Volgens Williams *et al. (*2013:3) is Twitter ’n medium waardeur nuus en verslae oor gebeurtenisse gedeel word. Hierdie gebeurtenisse kan iets so eenvoudig soos ŉ gebruiker se ontbyt, ’n politiesgelaaide berig of nuus in wording wees (vgl. Muralidharan *et al.*, 2011:175). Volgens Ehrlich en Shami (2010) word mikroblogs gebruik om inligting te deel, om op hoogte te bly met die nuutste nuuswaardige gebeure (plaaslik, nasionaal en internasionaal) sowel as direkte boodskappe aan ander gebruikers. Daar is verskeie maniere waarop die inligting wat op Twitter gedeel word, gekategoriseer kan word. Wu *et al.* (2011) onderskei tussen die media (nuus/joernaliste), vermaaklikheidspersoonlikhede, organisasies en bloggers[118]. Hierdie vier kategorieë kan soos volg gedefinieer word:

- *Media* verwys na die manier waarop nuus gedeel en versprei word. Die mate waarin Twitter die reikwydte van nuuswaardige gebeurtenisse vergroot het, is een van die belangrikste eienskappe van Twitter (vgl. Deller & Hallam, 2011:219; Wu *et al.*, 2011).

[118] Individue wat Twitter as ’n mikroblogplatform gebruik − in dieselfde sin as wat bloggers ’n blog as ’n persoonlike aanlyn dagboek beskou.

- *Vermaaklikheidspersoonlikhede* verwys na individue wat Twitter gebruik om hulle sosiale media teenwoordigheid as vermaaklikheidspersoonlikhede te vestig en uit te brei (Wu *et al.*, 2011).
- *Organisasies* verwys na maatskappye wat Twitter as ’n bemarkingsplatform gebruik (Wu *et al.*, 2011).
- *Bloggers* verwys na individue wat Twitter in hulle persoonlike hoedanigheid gebruik en oor hulle daaglikse doen en late twiet en oor kwessie wat vir hulle van belang is (Wu *et al.*, 2011).

Vir die doel van my studie het ek Twitter-gebruikers in vier groepe geklassifiseer:

i. **Bemarking** − Profiele wat aan groter organisasies behoort wat ten doel het om beriggewing te lewer oor aktuele gebeure en om hulle organisasies te bemark. In hierdie geval kan twiets as amptelike kommunikasie vanaf dié organisasies beskou word[119]. Volgens Biber (2006:130) word institusionele (organisatoriese) kommunikasie dikwels deur komitees saamgestel en finale produkte kan die stilistiese voorkeure van talle outeurs insluit. Na aanleiding van Biber (2006:130) se argument oor institusionele kommunikasie, het ek vir die doel van my studie nie die profiele van organisasies by groep 2 ingesluit nie.

ii. **Nuus** − Alle nuusprofiele is hierby ingesluit. Die bekendste profiel in dié geval is Nuus24, en ook hier kan daar ’n onderskeid getref word tussen sport, finansiële en algemene nuus. Voorbeelde van nuusprofiele sluit Netwerk24, RSG, OFM en SABC Nuus in. In hierdie geval is dit ook belangrik om te beklemtoon dat talle joernaliste vanaf ’n eie profiel twiet, maar in hulle professionele hoedanigheid.

iii. **Vermaak** − In hierdie afdeling verwys ek na vermaaklikheidspersoonlikhede in Suid-Afrika wat hulle Twitter-profiele gebruik om ’n aanlyn teenwoordigheid te skep. Hierdie profiele dien dus as ’n bemarkingsplatform sowel as ’n mondstuk waardeur dié persone met hulle aanhangers kontak het en kommunikeer.

iv. **Persoonlik** − Die laaste kategorie definieer ek as ’n kategorie waarin die meerderheid van die algemene Twitter-gebruikers val. Hierdie profiele verwys dus na gebruikers wat

[119] Vir die doel van my studies het ek nie deelnemers in die gespesialiseerde korpus ingesluit wat as organisasies twiet nie (ek het wel hierdie soort twiets by die algemene korpus ingesluit omdat dit steeds ’n goeie aanduiding van die stilistiese kenmerke van Afrikaans is) aangesien daar nie definitief bepaal kon word dat slegs ’n enkele individu in beheer van die twiets is nie, wat dus daartoe kon lei dat die stilistiese kenmerke van organisasies se twiets verskil. Die ondersoek na die identifisering van outeurs wanneer twee of meer individue dieselfde profiel gebruik, kan ’n verdere navorsingstudie wees.

van Twitter gebruik maak, maar nie in enige van die ander bogenoemde drie kategorieë val nie. Gebruikers in hierdie kategorie kan Twitter as ŉ politiese mondstuk, akademiese platform, vermaak of net ŉ algemene sosialemediaplatform gebruik waardeur daar met ander gebruikers geskakel word.

Hierdie vier kategorieë dien vir die doel van my studie as verwysingsraamwerk om te verseker dat twiets (en deelnemers) verteenwoordigend van die Afrikaanse Twitter-landskap beskou kon word. Vir die data-insameling van groep 1 het ek vervolgens individue geïdentifiseer sodat elk van die bogenoemde vier katogorieë in die korpus verteenwoordig is. Soos reeds in afdeling 5.3 genoem, was daar nie vereistes oor die minimum aantal twiets per deelnemer vir groep 1 nie, aangesien dit in dié geval daaroor gegaan het om ŉ verteenwoordigende korpus van Afrikaans op Twitter saam te stel.

Ten einde die bogenoemde te kon bereik, moes ek ŉ *nieproporsionele gestratifiseerde steekproefneming* en ŉ *sneeubal-steekproefneming* (om moontlike deelnemers te identifiseer) kombineer. ŉ Gestratifiseerde steekproefneming word gebruik wanneer die navorsingpopulasie heterogene eienskappe toon of uit verskillende subgroepe bestaan, en dit dan noodsaaklik is om te besluit watter van hierdie subgroepe (ook bekend as strata) ingesluit word, en of alle subgroepe by die navorsingpopulasie ingesluit moet word (Lombard, 2016:100).

ŉ Sneeubal-steekproefneming verwys volgens Strydom (2012:223) na ŉ metode wat gebruik word wanneer die teikenpopulasie onbekend is aan navorsers. Lombard (2016:103) stel dat ŉ sneeubal-steekproefneming gewoonlik vanuit ŉ enkele bron ontstaan, en dat dié bron navorsers dan na ander bronne lei. Die aanvanklike bron veroorsaak dus ŉ kettingreaksie (sneeubal-effek) na meer bronne.

Vir die samestelling van die tweede korpus (die gespesialiseerde korpus) het ek van tien verskillende deelnemers gebruik gemaak. Hierdie groep (hierna verwys ek daarna as groep 2) verwys na deelnemers wat doelbewus geselekteer is ten einde die navorsingsdoelwitte te bereik. Die navorsingsdoelwitte in groep 2 was om te bepaal wat as kenmerkend van Afrikaans beskou kan word[120], en of outeurskapidentifikasie wel suksesvol op ŉ mikroblog (Twitter) uitgevoer kon word. Ten einde die navorsingsdoelwitte te kon bereik, was daar etlike vereistes waaraan die potensiële deelnemers (van beide groep 1 en 2) moes voldoen:

- Twiets moes hoofsaaklik in Afrikaans wees.
- Twitter-profiele moes aktief wees.

[120] Groep 1 se korpus dien as verwysingskorpus, en word dus gebruik om die resultate vanuit die analise van groep 2 se korpus te verifieer.

- Gebruikers moes die vorm vir ingeligte toestemming (etiekvorm)[121] onderteken en dus ingeligte toestemming verleen dat hulle data in my studie gebruik kon word.

Om die bogenoemde doelwitte te kon bereik en om die deelnemers van groep 2 te kon identifiseer, het ek van ŉ *doelgerigte steekproefneming* sowel as ŉ *sneeubal-steekproefneming* gebruik gemaak. Die sneeubal-steekproefneming is gebruik om moontlike deelnemers te identifiseer terwyl die doelgerigte steekproefneming die mees inligtingryke gevalle of elemente vir die doel van my studie geïdentifiseer het.

ŉ Sneeubal-steekproefneming is baie populêr aangesien Twitter as ’n netwerk beskou kan word waar individue talle ander op die platform “volg” (*following)* en self “gevolg” (*follow)* word. Deur gebruikers se *following* en *follow* profiele te ondersoek, kon soortgelyke profiele (in dié geval ander gebruikers wat in Afrikaans twiet) geïdentifiseer word[122].

In hierdie afdeling is die verskillende maniere van steekproefneming bespreek, en die metodes waarvolgens ek my deelnemers geïdentifiseer het, is ook voorgehou. Die uiteensetting van my deelnemers kan dus soos volg opgesom word:

Tabel 5.1 Voorstelling van die twee verskillende groepe deelnemers vir my studie

Groepe	Inligting
Groep 1	Onbekende aantal deelnemers: Data word ingesamel tot en met die punt van versadiging bereik is. Daar is n-getal deelnemers en die aantal data sal deur middel van die beskikbaarheid bepaal word.
Groep 2 (Minimum van 1 000 twiets per deelnemer)	Outeur 1
	Outeur 2
	Outeur 3
	Outeur 4
	Outeur 5
	Outeur 6
	Outeur 7
	Outeur 8
	Outeur 9
	Outeur 10

Tabel 5.1 toon dus die navorsingspopulasie vir my studie. Nadat die onderskeie groepe se potensiële deelnemers geïdentifiseer is, is die deelnemers van groep 2 gekontak[123] en mits hulle

[121] Sien afdeling 5.11 (Etiekkwessies).
[122] Sien afdeling 2.6.2.2 (Twitter).
[123] Sien afdeling 5.11 (Etiekkwessies) en ook Bylaag B – toestemmingsbriewe.

toestemming verleen het dat hulle twiets in die studie gebruik kon word, is dit in die korpus gevoeg. Vervolgens hou ek die manier waarop die data ingesamel is voor.

5.4 Data-insameling en -verwerking

Soos reeds in hoofstuk 4 verduidelik, kan daar tussen ’n algemene en ’n gespesialiseerde korpus onderskei word. In hierdie afdeling verwys ek na die manier waarop ek die data ingesamel en verwerk het om te verseker dat die onderskeie korpora as verteenwoordigend van die platform beskou kan word.

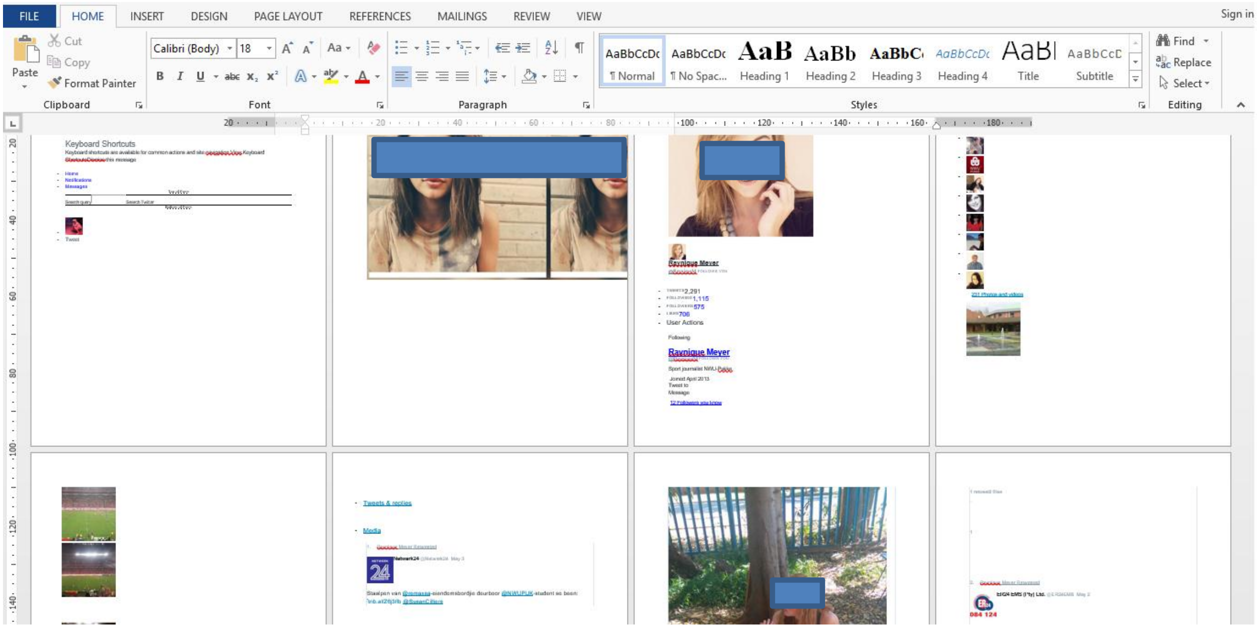
Dit is belangrik om te onthou dat die studie ondersoek ingestel het na die Afrikaans van mikroblogs, en dat daar tot op hede nie ’n korpus in Afrikaans beskikbaar is waar die Afrikaans van mikroblogs voorgehou word nie[124]. Die algemene korpus is dus ’n groot korpus (verteenwoordigend van die Afrikaans wat op Twitter gebruik word). Die doel van hierdie algemene korpus was om as ’n verwysingskorpus te dien wanneer die outeurskapidentifikasie plaasgevind en die analise van die taalstilistiese en stylkenmerke van mikroblogs in Afrikaans te plaasgevind het. Dit wil sê hierdie korpus is gebruik om bevindinge vanuit groep 2 se analises te verifieer en sodoende te bepaal of die kenmerke wel as kenmerkend van Afrikaans soos dit om Twitter gebruik word, beskou kan word.

Afdeling 5.3 verwys na die maniere waarop ek die onderskeie deelnemers gekies het. Ten einde my datastel so verteenwoordigend as moontlik te kon maak, het ek verskillende strata op Twitter geïdentifiseer. Die identifisering van die verskillende strata op Twitter het volgens Wu *et al*, (2011) en Biber *et al.* (2006:130) se navorsing plaasgevind[125].

Nadat die deelnemers geïdentifiseer is, en toestemming verkry is om hulle data te gebruik, het ek die onderskeie deelnemers se twiets vanaf Twitter in ’n Microsoft Word-dokument gekopieer. Ek het aanvanklik gebruik gemaak van ’n “docx.-dokument” aangesien emotikons nie in ongeformateerde teksdokumente (*.txt) wys nie. Emotikons en ander grafiese voorstellinge kan as kenmerkend van dié platform beskou word, en daarom was dit belangrik om hierdie voorstellinge in die data te behou.

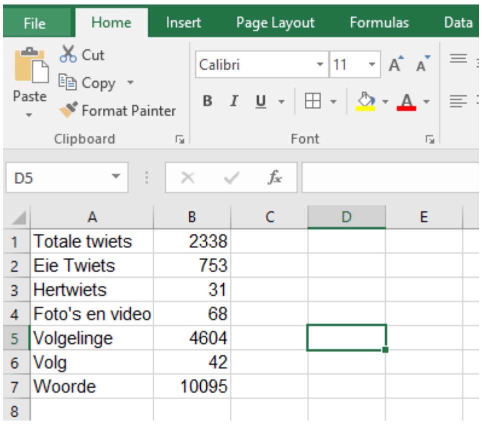
[124] Jansen van Vuuren het wel in 2007 ’n studie, “Die Internet as platform vir 'n nuwe Afrikaanse spraakgemeenskap en 'n nuwe variëteit − 'n korpuslinguistiese ondersoek”, uitgevoer. Alhoewel sy ’n korpus van Internetafrikaans saamgestel het, sluit hierdie korpus nie data vanaf Twitter in nie.

[125] Sien afdeling 5.3 (Deelnemers).



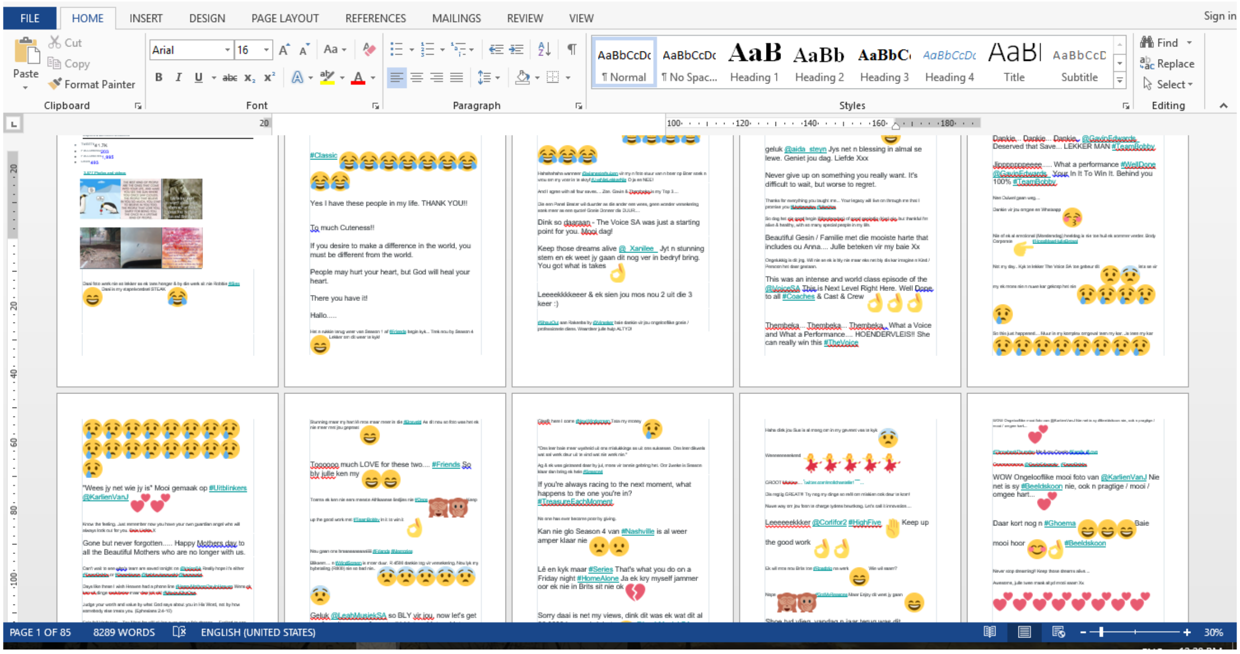
Skermskoot 5.1 Voorstelling van data wat vanaf ’n Twitter-profiel in Microsoft Word gekopieer is

Soos sigbaar vanuit Skermskoot 5.1 is die data wat in die Microsoft Word-dokument gekopieer is nie skoon nie, en kon dit nie net só gebruik word nie. Vervolgens het ek vir elke datastel ’n Microsoft Excel-lêer oopgemaak waarin ek spesifieke profiele se inligting gestoor het. Die inligting wat ek gestoor het sluit in: die aantal twiets, aantal *retweets* (her-twiets), aantal woorde in die individuele korpus, aantal foto’s en video’s wat die individue plaas, aantal keer wat die individue op ’n twiet reageer (kommentaar lewer), sowel as die aantal eie twiets.



Skermskoot 5.2 Skermskoot van ’n Microsoft Excel lêer met ’n deelnemer se inligting

Nadat hierdie data (Skermskoot 5.2) gestoor is, het ek begin om die data in die Microsoft Word-dokument skoon te maak. Skermskoot 5.3 is ’n voorbeeld van hoe die data lyk nadat dit skoongemaak is:



Skermskoot 5.3 Voorstelling van data wat skoongemaak is in ŉ Microsoft Word-dokument

Vanuit Skermskoot 5.3 is dit duidelik dat daar gereeld van emotikons gebruik gemaak word. Volgens Thiart (2014:79) verwys emotikons[126] na prentjies wat gebruik word om emosie in boodskappe uit te druk. Die probleem wat met emotikons ondervind is, was dat dit nie in ’n txt-dokument sigbaar is nie. Ek het dus alle emotikons met ’n kode vervang sodat emotikon-gebruik ook geanaliseer kon word. Nog ’n struikelblok was die feit dat gebruikers foto’s, video’s en hiperskakels na ander webblaaie plaas. Ek het die voorkoms van bogenoemde aspekte aangeteken aangesien dit ook kenmerkend van die styl van dié platform kon wees.

Nadat alle deelnemers in groep 1 se data gekodeer is, is dit saamgevoeg om sodoende die algemene korpus (ook genoem die verwysingskorpus) te vorm. Met betrekking tot die insameling van die data vir die gespesialiseerde korpus was die proses soortgelyk aan die proses wat gevolg is vir die insameling van die data vir die algemene korpus. Groep 2 het uit tien outeurs bestaan wat gekies is in die mate waaraan hulle aan die vereistes (soos deur die navorsingsvrae bepaal) van die studie voldoen het. Soos reeds genoem, moes elke outeur oor minstens 1 000 twiets beskik.

In die geval van groep 2 is die data weer eens vanaf Twitter gekopieer en in ’n Microsoft Word-dokument gestoor. Die deelnemers se inligting is ook in ’n Microsoft Excel-lêer gestoor en die data is verwerk (skoongemaak) en as ongeformateerde teksdokumente (*.txt) gestoor. In dié geval het ek die korpora apart van mekaar gehou aangesien ek outeurskapidentifikasie op die korpora uitgevoer het.

[126] Emotikons word verder in afdeling 7.5 (Semantiek) bespreek.

Om die geloofwaardigheid van die data te verseker, het ek alle twiets van die deelnemers wat ingestem het om deel te neem aan die studie ingesamel tot en met die datum waarop toestemming gegee is. Sodoende het ek gehoop om die moontlikheid van veranderde taalgebruik uit te skakel. Nadat die data skoon[127] was en in txt-formaat gestoor is, is dit deur middel van die rekenaarprogram NCHLT-Tagger geannoteer (POS-tag) om ondersoek in te stel na die woordsoortgebruik en sinskonstruksies[128].

Hierdie afdeling (afdeling 5.4) sowel as afdelings 5.2 en 5.3 het ten doel om my navorsingsontwerp − die maniere waarop ek my deelnemers geïdentifiseer het − sowel as die manier waarop ek my data ingesamel en verwerk het voor te hou. Dit blyk duidelik dat daar verskeie faktore was wat ek in gedagte moes hou sowel as aspekte wat as problematies ervaar is. In die daaropvolgende afdeling verduidelik ek die metodes wat in die analises gebruik is.

5.5 Navorsingsmetodes

Holmes en Meyerhoff (2003:15) voer aan dat daar heelwat debat gevoer word oor die poging om kwalitatiewe en kwantitatiewe metodes te integreer waar patroonmatigheid wat deur middel van ŉ kwantitatiewe metode geïdentifiseer is as noodsaaklike agtergrond dien vir ŉ kwalitatiewe interpretasie van die data. Angouri (2010:33) argumenteer dat ŉ gemengde metode beskou moet word as ŉ metode waar kwalitatiewe en kwantitatiewe metodes gekombineerd gebruik word – eerder as ŉ integrasie van dié twee metodes.

Tashakkori en Creswell (2007:4) voer aan dat kwalitatiewe en kwantitatiewe metodes gekombineer kan word, wanneer:

- daar twee soorte navorsingsvrae is wat deur kwalitatiewe en kwantitatiewe benaderings beantwoord kan word.
- die manier waarop die navorsingsvrae ontwerp is óf deelnemend óf voorafbeplan is.
- daar twee verskillende maniere van steekproefneming in die data aanwesig is.
- daar twee verskillende maniere van data-insameling aanwesig is.
- daar twee verskillende benaderings tot data analise is.
- daar twee verskillende soorte data aanwesig is (numeries en teksverwant).
- daar twee verskillende gevolgtrekkings gemaak kan word (objektiewe en subjektiewe gevolgtrekkings).

[127] "Skoon" verwys na data wat in ŉ gepaste formaat gestoor is, en waar alle emotikons reeds met kodes vervang is, en waar foto's en ander grafika verwyder is.
[128] Sien afdeling 5.5 (Navorsingsmetodes).

Die kombinasie van ŉ kwantitatiewe en kwalitatiewe metode kan bydra tot ŉ beter begrip van die onderskeie fenomene wat ondersoek word, en volgens Angouri (2010:33) sal ŉ kwantitatiewe benadering die navorsingsbevindinge aantoon terwyl ŉ kwalitatiewe benadering diepgaande en inligtingryke data verskaf.

ŉ Gemengde metode kan dus eenvoudig gedefinieer word as ŉ navorsingsmetode waar die kwantitatiewe (numeriese data) en kwalitatiewe (teksverwante data) navorsingsontwerpe geïntegreer of in kombinasie met mekaar gebruik word, en waar die twee navorsingsontwerpe mekaar aanvul binne ŉ enkele studie om die geldigheid van die resultate te verseker en die navorsingsvrae so deeglik moontlik te beantwoord (vgl. Nel & Jordaan, 2016:277). Angouri (2010:33-34) sluit by Nel en Jordaan aan en stel dat ŉ kwantitatiewe en kwalitatiewe kombinasie aanleiding gee tot triangulasie in ŉ studie.

Triangulasie verwys na die kombinasie van metodes binne dieselfde navorsing. Volgens Denzin (1978:291) kan daar tussen vier verskillende soorte triangulasie onderskei word, naamlik: **datatriangulasie** (ŉ verskeidenheid bronne word binne dieselfde studie gebruik); **ondersoektriangulasie** (die idees van verskillende navorsers word gebruik); **teorietriangulasie** (verskillende perspektiewe en teorieë word gebruik om die resultate van die studie te interpreteer); en **metodologiese triangulasie** (verskillende metodes word gebruik om die navorsingsprobleem op te los). Nel en Jordaan (2016:379) argumenteer dat triangulasie die geldigheid van ŉ studie verhoog.

Ivankova en Creswell (2009:141) voer aan dat triangulasie die algemeenste tegniek is om ŉ gemengde metode-ontwerp se bevindinge geldig te verklaar, en dat kwalitatiewe en kwantitatiewe data oor dieselfde fenomeen gelyktydig ingesamel kan word om sodoende die data vanuit ŉ kwalitatiewe én kwantitatiewe benaderings te kan vergelyk en geldige bevindinge te kan maak. Ivankova en Creswell (2009:138) onderskei tussen drie hoofkenmerke van ŉ gemengde metode, naamlik:

- **Tydsberekening**: Hier word na die volgorde van die implementering van die prosedures oor kwalitatiewe en kwantitatiewe data-insameling en -analise verwys waar die stappe op mekaar volg en mekaar beïnvloed. Data-insameling en -analise kan gelyktydig geskied of dit kan opeenvolgend plaasvind.

- **Gewig**: Gewig verwys na die prioriteit wat aan elke datatipe toegeken word. Die gewig kan óf gelyk wees óf een soort data kan as belangriker beskou word. Om te bepaal watter van die twee bogenoemde opsies gebruik moet word, kan vrae gevra word soos:

is die studie verkennend of beskrywend, watter data-insamelingsproses is sentraal tot die studie (kwalitatief of kwantitatief) en watter metode word meer indringend bespreek?

- **Samesmelting**: In dié geval word daar verwys na die manier waarop die twee metodes in die studie geïntegreer word. Die samesmelting van ʼn kwalitatiewe en kwantitatiewe metode kan op enige gegewe tydstip in ʼn studie plaasvind hetsy tydens data-insameling, die analise of die interpretasie van die data. Aangesien my studie eksploratief van aard is, was dit noodsaaklik om bevindinge (vanuit ʼn kwantitatiewe en kwalitatiewe metode) te vergelyk en te verifieer ten einde al twee metodes te toets.

ʼn Gekombineerde kwalitatiewe en ʼn kwantitatiewe metode blyk sonder twyfel die hoekstene van ʼn gemengde metode te wees. Vervolgens bespreek ek elk van hierdie metodes met verwysing na outeurskapidentifikasie en korpuslinguistiek. Alhoewel daar reeds verwys is na benaderings tot outeurskapidentifikasie (afdeling 3.5.1), outeurskapidentifikasie in ʼn digitale era (afdeling 3.6), ʼn korpuslinguistiese analise (afdeling 4.6.1) en ʼn forensiese en korpuslinguistiese benadering (afdeling 4.6.2), fokus ek nou spesifiek op metodes wat in my studie van nut was.

5.6 Kwalitatiewe analise

Kwalitatiewe navorsing verwys na die insameling en interpretatiewe analise van (hoofsaaklik) tekstuele data (vgl. Croker, 2009:5, 9). ʼn Kwalitatiewe metode is inherent eksploratief van aard, en het ten doel om nuwe insigte oor bestaande teorieë te vorm of om selfs nuwe teorieë te vorm (vgl. Croker, 2009:9). Hierdie insigte kan nie noodwendig in ʼn ander konteks toegepas word nie aangesien kwalitatiewe navorsing van navorsers vereis om patroonmatigheid en skakels in die data te identifiseer en daarop voort te bou. ʼn Kwalitatiewe benadering ontwikkel dus ʼn teorie na gelang van die bevindinge wat vanuit die data verkry word (vgl. Perry, 2005:75). Data word dus in ʼn konteks van die "werklike lewe" versamel en ondersoek en dit behels holistiese data-insameling deur waarneming, maar in die afwesigheid van ʼn voorafbepaalde teorie en beperkinge (vgl. Perry, 2005:76).

Volgens Croker (2009:11) is objektiwiteit van navorsers een van die grootste probleme in kwalitatiewe navorsing aangesien navorsers as die hoofnavorsingsinstrument beskou kan word. Navorsers is verantwoordelik vir die insameling van data sowel as die analise daarvan; daarom kan navorsers se eie wêreldbeskouing ʼn invloed op die interpretasie van die data hê. Daarom

is dit ’n goeie beginsel om van triangulasie[129] gebruik te maak om sodoende bevindinge te verifieer.

McMenamin (2002:75) postuleer dat ’n kwalitatiewe analise (wanneer hierdie soort analise gebruik word vir outeurskapidentifikasie) as die identifisering van kenmerke in ’n teks, en die beskrywing van hierdie kenmerke gedefinieer kan word. Rasinger (2010:52) argumenteer dat die induktiewe aard van kwalitatiewe navorsing nie buite rekening gelaat kan word nie aangesien teorieë gevorm word na gelang van die resultate van die navorsing.

Wanneer Afrikaans − soos dit op ’n mikroblogplatform gebruik word − dus ondersoek word, is ’n kwalitatiewe analise van uiterse belang aangesien hierdie aanlyn platform ’n blik op Afrikaans kan bied soos dit daagliks deur Twitter-gebruikers gebruik word en omdat daar afleidings oor die aard van Afrikaans op Twitter gemaak kan word na gelang van hierdie analise. Die sosiale aard van taal word dus in ’n natuurlike (aanlyn) omgewing ondersoek. Thiart (2014:99) verwys na die kwalitatiewe analise as ’n beskrywende stilistiese analise van taal. Daar kan dus vanuit die veronderstelling gewerk word dat ’n kwalitatiewe analise na die stilistiese ontleding van korpora verwys wanneer outeurskapidentifikasie ondersoek word. Vervolgens bespreek ek ’n stilistiese analise.

5.6.1 Stilistiese analise

Linguistieke stilistiek verwys na die wetenskaplike analise van individuele stylmerkers soos waargeneem in die idiolek van ’n enkele outeur of in die taal of dialek van ’n groep outeurs (vgl. McMenamin, 2002:115). McMenamin (2002:115) argumenteer verder dat linguistiese variasie, hetsy die variasie van ’n norm of die afwyking vanaf die norm, die aanwesigheid en identifisering van ’n norm aandui. Volgens Carstens (2011:335) verwys die taalgebruikers se vermoë om hul taal by die bepaalde situasie aan te pas na hul stilistiese vermoë en dat elke individu se styl uniek is - dus kan dit as ’n idiolektiese merker beskou word.

McMenamin (2002:163) argumenteer dat geen twee outeurs se taalstyl dieselfde is nie, en dat daar selfs variasie binne ’n enkele outeur se skryfstyl sigbaar is[130]. Om hierdie stylkenmerke van groepe of individue te ondersoek, kan ’n stilistiese analise uitgevoer word. McMenamin (2002:163) verwys na die ondersoek na patrone van variasie in geskrewe taal as ’n stilistiese analise.

[129] Aangesien ek van ’n gemengde metode (beide kwalitatief en kwantitatief) gebruik gemaak het, vind triangulasie dus outomaties deur middel van die gemengde metode plaas.

[130] Sien afdeling 2.3.2.3 (Idiolek).

ŉ Stilistiese analise het dus betrekking op die identifisering van stylmerkers in ŉ bepaalde teks, en die moontlikheid dat hierdie stylmerkers kan afwyk van dít wat as die norm beskou word. Daar kan onderskei word tussen stylmerkers wat eie aan ŉ dialek of groep is en die stylmerkers wat eie aan ŉ individu is en dus as die idiolek van ŉ individu geïdentifiseer kan word.

Olsson (2008:45) argumenteer dat ŉ stilistiese analise telkens ondersoek instel na eienskappe in ŉ teks wat as "ongewoon" deur navorsers beskou word. Daarom is dit noodsaaklik dat daar teorieë gebruik word wat afwykings van die norm kan beskryf. Olsson (2008:45) gaan voort deur na *gemerktheid*[131] te verwys wat as ŉ weergawe van ŉ ongewone linguistiese teken of struktuur beskou kan word. Kotzé (2012) verwys ook na gemerktheid en argumenteer dat in die geval van ŉ stilistiese analise, sleutelwaardebepalings van belang is. Deur na die sleutelwaardebepalings ondersoek in te stel kan eienskappe geïdentifiseer word wat as gemerk beskou kan word (grammatiese, leksikale, ortografiese en selfs tipografiese kenmerke).

Grant en Baker (2001:69) voer aan dat daar steeds twyfel bestaan oor wat as betroubare outeurskapmerkers beskou kan word sowel as die manier waarop hierdie merkers geïdentifiseer kan word. Vervolgens verwys ek na die verskillende stylmerkers wat in ŉ stilistiese analise ondersoek kan word. Michell (2013:46) onderskei tussen karakter-, woord-, sin-, dokument- en strukturele of sintaktiesgebaseerde eienskappe wat in ŉ stilistiese analise ondersoek kan word. McMenamin (2002:54) verwys na ŉ stilistiese analise as ŉ onder-na-bo benadering waar daar na herhalende patrone, verspreiding en vorme van organisasie ondersoek ingestel word om sodoende patrone wat eie aan die outeur se taalstyl is te identifiseer.

McMenamin (2002, 216-231) het vervolgens 11 stylmerkers geïdentifiseer wat in 80 verskillende gevalle gebruik is om outeurskapidentifikasie uit te voer.

- **Teksformaat** verwys na verskille in die uitleg van die dokument en kan die manier waarop die datum geskryf is, leestekengebruik by aanspreekvorme en selfs lynspasiëring insluit.
- **Getalle en simbole** verwys onder andere na die manier waarop getalle geskryf is, hoe daar na pryseenhede verwys is, en die manier waarop simbole voorgestel is (byvoorbeeld *plus* of +).
- **Afkortings** verwys na die manier waarop afkortings gebruik is (byvoorbeeld, *ENS*. of *ens.* of *ensovoorts*).

[131] Sien afdeling 7.4 (Morfologie).

- **Leestekengebruik** verwys na die manier waarop leestekens in die teks gebruik is, en of daar enige afwykings van die norm voorkom. Voorbeelde daarvan sluit die gebruik van meer as een leesteken (!!!!!!) in.
- **Hoofletters** verwys na die gebruik daarvan wat afwyk van die normale konvensies van hooflettergebruik, byvoorbeeld die skryf van ’n eienaam met ’n kleinletter.
- **Spelling** verwys na afwykings van die aanvaarde spelkonvensies en kan los en vas skryf van woorde, verkeerdelike homoniemgebruik, fonetiese spelling en verkortings insluit.
- **Woordvorming** verwys byvoorbeeld na die vorming van nuwe samestellings.
- **Sintaksis** verwys na die sinskonstruksie en die manier waarop ingewikkelde of gefragmenteerde sinne gebruik word.
- **Diskoers** verwys onder andere na temas wat algemeen in ’n outeur se tekste voorkom, die gebruik van retoriese vrae en dies meer.
- **Frekwente woorde en frases** verwys na woorde en/of frases wat eie aan ’n bepaalde outeur se tekste is. Hierdie woorde of frases kan afwyk van die norm in terme van byvoorbeeld spelling of punktuasie of dit kan na ’n vreemde konstruksie verwys.
- **Foute en regstellings** verwys na die manier waarop foute in ’n teks voorkom, en die manier waarop die outeur hierdie foute regstel.

Hierdie 11 stylmerkers dien as ’n basis vir ’n stilistiese ondersoek en verwys na die verskillende stylaspekte wat ondersoek kan word in outeurskapidentifikasie. Coulthard en Johnson (2007:165) verwys ook na die sinslengte en leksikale rykheid van ’n teks as ’n goeie vertrekpunt vir ’n stilistiese analise. Michell (2013:46) stel dat hierdie stylmerkers nie noodwendig altyd die enigste eienskappe is wat ondersoek kan word nie − veral wanneer daar ondersoek ingestel word na outeurskapidentifikasie op korter tekste soos in die geval van e-posse, SMS’e en kommunikasie op sosiale netwerke. Ook Baayen *et al.* (2000:157) stel dat stylmerkers kan wissel na gelang van die genre en tema van die teks wat geanaliseer word en dat klein korpora veral probleme kan skep.

Kotzé (2015) argumenteer dat die aard van die teks belangriker is as die lengte, en dat korter tekste (wat uit informele taalgebruik bestaan) dikwels meer waardevol as langer, meer formele tekste is aangesien die formele tekste se woordeskat en struktuur meestal deur konvensies bepaal word en “bied min moontlikheid om ’n eie stempel daarop af te druk”. In afdeling 2.4 word daar in Tabel 2.1 (Tabel 2.1 Verhoef (2001:8-11) se merkers van geskrewe Internetafrikaans) en Tabel 2.2 (Aspekte van Internetafrikaans soos ondersoek deur Jansen van Vuuren (2007:155-287)) verwys na verskeie kenmerke van Afrikaans wat met betrekking tot die stilistiese kenmerke van Internetafrikaans ondersoek is.

Ook Thiart (2014) het ondersoek ingestel na outeurskapidentifikasie. In haar studie is SMS’e as ondersoekplatform gebruik. Thiart (2014:100-101) het in haar stilistiese analise ondersoek ingestel na leestekens, aanspreekvorme, lagtekens (*smileys*), logogramme, ongewone gebruik van hoof- en kleinletters, verkortings, niestandaardspelling, Engelse woorde/sinne, funksiewoorde en individuele eienskappe (herhalende woorde, frases en/of leestekengebruik).

Volgens Thiart (2014:59) is ’n stilistiese analise ’n belangrike eerste stap in outeurskapidentifikasie aangesien die kenmerke geïdentifiseer word wat verder in ’n stilometriese analise ondersoek word. Dit blyk dus duidelik dat ’n stilistiese analise gebruik word om stylkenmerke van individue te identifiseer. Aangesien die kenmerke van Afrikaans − soos dit op ’n mikroblog gebruik word − nog nie geïdentifiseer is nie, moes ek in my studie eerstens die kenmerkende taalstilistiese en stylkenmerke van Afrikaans soos dit op Twitter gebruik word, bepaal. Eers ná afloop van hierdie analise kon ek ’n stilistiese analise in terme van outeurskapidentifikasie toepas.

Dit is belangrik om die kenmerkende eienskappe van Afrikaans soos dit op ’n mikroblog gebruik word te identifiseer ten einde die algemene kenmerke te kan uitskakel wanneer outeurskapidentifikasie uitgevoer word. ’n Kwalitatiewe analise het dus ten doel om algemene kenmerke van Afrikaans soos dit op Twitter gebruik word te identifiseer. Soos reeds genoem, het ek van ’n gemengde metode in my studie gebruik gemaak. ’n Gemengde metode gee nie net aanleiding tot triangulasie nie, maar soos Kotzé (2012) dit stel kan problematiek oor die betroubaarheid van ’n stilistiese analise uitgeskakel word wanneer dié analise in kombinasie met of aanvullend deur ’n stilometriese analise gebruik word. Vervolgens bespreek ek ’n kwantitatiewe metode.

5.7 Kwantitatiewe analise

Ivankova en Creswell (2009:137) definieer ’n kwantitatiewe analise as die insameling van numeriese data om ’n spesifieke vraag te beantwoord. Kwantitatiewe navorsing word volgens Tavakoli (2012:508) as die belangrikste voorbeeld van navorsing voorgehou waarin ’n hipotese getoets word, en voer aan dat hierdie soort navorsing met ’n hipotese begin wat dan ondersoek word deur middel van datakwantifisering en numeriese analises. Thiart (2014:7) sluit by die bogenoemde definisies aan en stel dat ’n kwantitatiewe analise na die objektiewe analise van data deur middel van ’n verskeidenheid statistiese tegnieke verwys.

Rasinger (2010:51-51) verduidelik 'n kwalitatiewe analise aan die hand van 'n voorbeeld (hieronder in Afrikaans vertaal) van 'n statistiese ontleding van 'n gesprek wat tussen 'n ma en kind plaasvind.

Kind: Pappa kom ook saam ons. (5 woorde)

Ma: Wie kom saam met ons?

Kind: Pappa. (1 woord)

Ma: Pappa? Nee, waar is pappa?

Kind: Pappa hê, Pappa kom ook. (5 woorde)

Ma: Pappa werk nog.

Kind: Nee, Pappa soek. Hier my bottel. (6 woorde)

Volgens Rasinger (2010:51) kan die bostaande gesprek tussen 'n ma en kind geanaliseer word in terme van patrone wat vanuit dié gesprek duidelik word. 'n Ondersoek na die eienskappe en aard van 'n gesprek verwys na 'n kwalitatiewe[132] analise – 'n gesprek kan ook geanaliseer word ten opsigte van (onder andere) die gemiddelde uitingslengte (MLU − mean length of utterance). Die gemiddelde uitingslengte verwys na 'n metode waarvolgens 'n kind se taalvaardigheid bepaal word deur die gemiddelde lengte van morfeme wat gebruik word te bereken. Die gemiddelde uitingslengte kan soos volg bereken word (vgl. Rasinger, 2010:51):

$$\frac{Totale\ hoeveelheid\ woorde}{Totale\ spreekbeurte} = \text{Die gemiddelde uitingslengte} \quad \therefore \quad \frac{17}{4} = 4.25$$

Hierdie berekening dui aan dat die kind in die voorbeeld gemiddeld 4.25 woorde per spreekbeurt gebruik. Deur die gemiddelde uitingslengte berekening te gebruik, word daar 'n numeriese waarde toegeken aan 'n entiteit wat voorheen slegs uit teks bestaan het − die teks word dus gekwantifiseer. Kwalitatiewe navorsing stel dus ondersoek in na die patroonmatigheid wat in 'n teks gevind kan word terwyl kwantitatiewe navorsing fokus op die aantal keer (kwantifiseerbare numeriese waardes) wat iets voorkom (vgl. Levon, 2010:68; Rasinger, 2010:52; Rose, 2002).

Daar is verskeie maniere waarop die numeriese waarde van linguistiese data wat as 'n korpus voorgehou word, ondersoek kan word. Hierdie kwantitatiewe analises wat betrekking op outeurskapidentifikasie en die identifisering van stylkenmerke het, word vervolgens in die afdeling oor 'n stilometriese analise (afdeling 5.7.1) voorgehou. Aston (2011:7) verwys na die

[132] Sien afdeling 5.6 (Kwalitatiewe analise).

toename in kwantitatiewe analises waar die korpora rekenaargereed is, dit wil sê die kenmerke in die onderskeie korpora kan deur rekenaarprogrammatuur[133] geïdentifiseer en getel word. Wanneer rekenaarprogrammatuur gebruik word om kenmerke in korpora te ondersoek, word deur na onder andere woordvorme, die groepering van woorde (kollokasies) en n-gramme ondersoek ingestel. Die frekwensie van hierdie kenmerke word vergelyk met die frekwensies wat vanuit ’n vergelykingskorpus verkry word om sodoende die statistiese beduidendheid te bepaal (vgl. Aston, 2011:7).

Simpson (2004:39) argumenteer dat daar ’n basis vir vergelyking aanwesig moet wees wanneer kwantitatiewe korpusstudies uitgevoer word, dus word daar van ’n verwysingskorpus gebruik gemaak om die statistiese moontlikhede te ondersoek dat ’n kenmerk as uniek aan ’n bepaalde teks of genre beskou kan word.

Wanneer navorsers kwantitatiewe metodes toepas, moet ’n hipotese bewys kan word. Rasinger (2010:52) voer aan dat ’n hipotese verwys na ’n verklaring oor ’n potensiële of voorgestelde verhouding tussen veranderlikes. In die geval van my studie was die hipotese wat bewys moes word die veronderstelling dat outeurs van mikrobloginskrywings geïdentifiseer kan word. Wanneer kwantitatiewe metodologie gebruik word met betrekking tot forensiese linguistiek (outeurskapidentifikasie), word daar van ’n stilometriese benadering gebruik gemaak. In die volgende afdeling (5.7.1) bespreek ek ’n stilometriese analise.

5.7.1 Stilometrie

Stilometrie verwys na ’n vorm van outeurskapidentifikasie wat staat maak op linguistiese inligting wat vanuit ’n dokument sigbaar word. Alhoewel stilometrie toegepas is voor die toename in die gebruik van rekenaars, verwys dit tans na analises wat patroonmatigheid in korpora ondersoek en hoofsaaklik deur rekenaarprogrammatuur gedomineer word (vgl. Brennan *et al.,* 2011:1). Lakshmi en Pateriya (2012:653) is van mening dat ’n stilometriese analise kan bydra tot die suksesvolle identifisering van ’n outeur. Stilometrie word deur Kotzé (2007:388) gedefinieer as:

> ’n Deeglike kwantitatiewe ontleding, deur middel waarvan die relatiewe frekwensie van identiese woordeskatitems of woordgroepe vergelyk word. Dit word ’n kwantitatiewe ontleding genoem omdat dit gebaseer is op die kwantifisering van tekstuele kenmerke as ’n basis vir verdere berekeninge, wat beteken dat ieder en elke woord opgeteken en getel moet word. ’n Aantal

[133] Sien afdeling 5.8.1 (Rekenaarprogrammatuur).

berekeninge word dan op die data uitgevoer, gevolg deur statistiese beduidenheidstoetse.

Kotzé (2010:187) argumenteer dat ’n stilometriese analise ten doel het om die teenoorgestelde van ’n stilistiese analise te bewys. In stede daarvan om algemene tekstuele eienskappe (linguisties) te identifiseer, probeer die stilometriese analise om te bewys dat die tekste wat ondersoek word in so ’n mate van mekaar verskil dat dit waarskynlik deur verskillende outeurs geskryf is. Indien dit nie sonder redelike twyfel bewys kan word − deur toepaslike statistiese beduidendheidtoetse − dat die tekste wat ondersoek word oor duidelike unieke kenmerke beskik nie, moet analiste dus probeer om deur ’n stilometriese analise te bewys dat die tekste oor genoegsame ooreenstemmende kenmerke beskik dat dit wel deur dieselfde outeur geskryf kon gewees het (vgl. Kotzé, 2010:187).

In aansluiting met Kotzé stel Barry en Luna (2012:2) dat stilometrie uit twee prosesse bestaan. Eerstens, die ondersoek na moontlike kenmerke wat ondersoek kan word en tweedens, die statistiese verwerking van die voorkoms van hierdie kenmerke. Kotzé (2010:388) voer verder aan dat hierdie eerste proses in ’n stilometriese analise na ’n stilistiese analise verwys.

Brennan *et al. (*2011:3) argumenteer dat stilometrie gebaseer word op die veronderstelling dat elke individu oor ’n unieke taalstyl[134] beskik. Die bekendste geval waar stilometrie gebruik is om die outeur van ’n teks te identifiseer, is die *Federalist Papers*-saak (Brennan *et al.,* 2011:13; Juola, 2006:10; Kotzé, 2007; Mosteller & Wallace, 1965:275; Olsson, 2008:19). Stilometrie kan as ’n bo-na-onder model beskryf word waar daar ondersoek ingestel word na die voorafbepaalde taksonomie van stilistiese items om sodoende die stilometriese eienskappe te bepaal (vgl. McMenamin, 2002:54). Krause (2013) verwys na sinslengte, woordeskatrykheid en frekwensie van woorde of woordlengtes as kenmerke wat stilometries ondersoek word en verduidelik dit soos volg:

- **Karakterfrekwensie**: Bereken die relatiewe frekwensie van individuele karakters.

- **Woordlengtefrekwensie**: Bereken die relatiewe frekwensie van woordlengtes. In dié geval is dit noodsaaklik om te onderskei tussen die gebruik van @ (gebruikersnaam) en # (veral in die geval van mikroblogs)[135] en besonderse lang woorde.

- **Sinslengtefrekwensie**: Bereken die gemiddelde frekwensie van die lengte van sinne.

[134] Sien afdeling 2.3.2.3 (Idiolek).
[135] Omdat Twitter-gebruikers gereeld van die hutsteken gebruik maak, is dit moontlik dat sekere hutswoorde, hutsteken-frekwensie of selfs die lengte van hutswoorde kenmerkend van ’n bepaalde individu se idiolek kan wees.

- **Woordsoortfrekwensie**: Annotering − woordsoortetikettering (POS-tag) – word ook uitgevoer ten einde die frekwensie van die afsonderlike woordsoortgroepe in die korpora te kan bepaal.

- **Woordspesifisiteit-frekwensie**: Hierdie soort kenmerk is volgens Krause (2013) nog nie voor sy studie gebruik nie, en verwys na hoe die spesifisiteit van woordgebruik ’n inkriminerende faktor in outeurskapidentifikasie kan wees.

Stamatatos (2009:3-12) definieer die kenmerke wat ondersoek kan word effens anders as Krause, maar tog blyk die kenmerke ooreenstemmend te wees. Stamatatos (2009:3-12) verwys na:

- **Leksikale kenmerke**: Die teks word as ’n opeenvolging van tekens beskou, en outeurskapidentifikasie word ondersoek na gelang van woord- en sinslengtes.

- **Karakterkenmerke**: Die teks word as ’n opeenvolging van karaters beskou, en daar kan na alfabetiese karaktertelling, digitale karaktertelling, hoof- en kleinlettergebruik, frekwensie van letters en leestekengebruik ondersoek ingestel word.

- **Sintaktiese kenmerke**: Outeurs maak onbewustelik van sintaktiese patrone in hulle tekste gebruik en hierdie patroonmatigheid kan geanaliseer word.

- **Semantiese kenmerke**: Indien die korpus geannoteer is (woordsoortetikettering), kan die semantiese kenmerke van die onderskeie outeurs statisties ondersoek word.

’n Ondersoek na funksiewoorde word volgens Holmes (1998:110) as een van die akkuraatste stilometriese analises beskou[136]. Funksiewoorde (ook grammatikale woorde) verwys volgens Crystal (2008:203), Juola (2006:265) en Kotzé (2007:388) na woorde wat semanties oor min of geen onafhanklike betekenis beskik nie, maar wat grammatikale betekenis van die taal oordra en sluit lidwoorde, hulpwerkwoorde, voorvoegsels en voorsetsels in (*’n*, *die*, *en*, asook *is*).

Hubbard (1995:61) voer aan dat alhoewel frekwente funksiewoorde relatief onafhanklik staan van die genre waarin geskryf word, kan die uitwerking van die genre nie buite rekening gelaat word nie. Kotzé (2015) voer aan dat funksiewoorde onder die bewussynvlak van gebruikers lê en dat dit gebruik word in die keuse van hoe om die inhoud te formuleer. Koppel en Schler

[136] Sien afdeling 3.5.1.1 (Forensiese stilistiek)

(2003:1) sluit hierby aan en argumenteer dat funksiewoorde onafhanklik van die konteks is, en dus nie noodwendig deur die onderwerp van die teks bepaal word nie.

Buiten funksiewoorde kan daar ook ondersoek ingestel word na inhoudswoorde. Inhoudswoorde (ook leksikale woorde) verwys na woordsoorte soos werkwoorde, naamwoorde, byvoeglike naamwoorde en sekere bywoorde (vgl. Michell, 2013:50). Kotzé (2010:118) voer aan dat die gebruik van leksikale woorde deur outeurs aansienlik kan verskil. Leksikale woorde staan in kontras met grammatikale woorde en verwys na woorde wat leksikale betekenis het − leksikale woorde beskik oor ’n semantiese inhoud (vgl. Crystal, 2008:279).

ŉ Leksikale ondersoek kan ook kollokasies en leksikale digtheid insluit (vgl. Crystal, 2008:279). Kotzé (2015) stel dat inhoudswoorde veel meer prominent is en dat dié soort woorde deur outeurs gemanipuleer kan word. Inhoudswoorde is ook konteksafhanklik.

Dit blyk vanuit die bogenoemde twee definisies dat funksiewoorde onbewustelik gebruik word, en dat dié soort woorde nie konteksgebonde is nie. Holmes (1998:110) stel dat tydens ’n stilometriese analise kan daar bepaal word hoeveel maal ’n funksiewoord in ’n teks gebruik word. Soos reeds genoem, is ŉ stilometriese analise hoofsaaklik rekenaargebaseerd. Vervolgens verwys ek na die onderskeie programmatuur en relevante terminologie van programmatuur wat gebruik word asook na die belangrikste statistiese konsepte.

5.8 Instrumente

In hierdie hoofstuk is daar tot dusver ondersoek ingestel na maniere waarop die deelnemers vir my studie geselekteer is, hoe data-insameling plaasgevind het sowel as ’n verduideliking van die metodologie wat gevolg is. Dit het duidelik geblyk vanuit hierdie ondersoek dat ’n forensieslinguistiese ondersoek na die identifisering van outeurs sowel as die bepaling van kenmerkende styleienskappe tot ’n groot mate digitaal is. Daarom word verskeie rekenaarprogrammatuur gebruik om die stilistiese en stilometriese eienskappe in korpora te ontleed. Vervolgens verwys ek na die verskillende rekenaarprogrammatuur wat in my studie gebruik is. Ek bespreek ook die belangrikste funksies en terminologie en ten slotte verwys ek na die belangrikste statistiese konsepte.

5.8.1 Rekenaarprogrammatuur

In die digitale era waarin ons ons bevind, is daar menige rekenaarprogrammatuur wat gebruik kan word wanneer ’n korpuslinguistiese benadering in ’n studie gebruik word. In my studie het ek van verskillende programmatuur gebruik gemaak vir die onderskeie analises. Ek het hoofsaaklik van *WordSmith Tools* gebruik gemaak vir die analise van my korpora. *ATLAS.ti*

(2014) is verder as 'n aanvulling tot *WordSmith Tools* gebruik om die stilistiese kenmerke in die korpora te kon identifiseer en analiseer.

5.8.1.1 WordSmith Tools

Scott (2012:2) definieer *WordSmith Tools* (WST) as 'n geïntegreerde reeks programme wat gebruik kan word om ondersoek in te stel na hoe woorde in 'n teks gebruik word. Daar word tussen drie hooffunksies onderskei, naamlik woordelyste (*WordList*), konkordansies (*Concord*) en sleutelterme (*KeyWords*)[137]. Hierdie drie hooffunksies maak dit moontlik om tekste te analiseer en bied verder ook statistiese ondersteuning. Prinsloo en Prinsloo (2011) argumenteer dat *WordSmith Tools* gebruik kan word om patrone in tekste te analiseer.

5.8.1.2 *ATLAS.ti*

Fries (2014:10) beskryf *ATLAS.ti* as 'n program wat gebruik kan word vir die kwalitatiewe analise van groot hoeveelhede teks asook grafiese, klank- en video-data. Hierdie program bied 'n verskeidenheid hulpmiddele wat aangewend kan word in die sistematiese analise van ongestruktureerde data. Fries (2014:10) definieer "ongestruktureerde data" as data wat nie deur middel van die formele statistiese benaderings geanaliseer kan word nie.

In *ATLAS.ti* skep gebruikers 'n hermeneutiese eenheid (HU) wat as die oorkoepelende projek beskou kan word. Binne 'n HU word die onderskeie dokumente (hetsy tekste, video-, klank- of grafiese data) wat in die analise gebruik word as hoofdokumente gestoor (Fries, 2014:15). Wanneer analises dan uitgevoer word, kan navorsers die hoofdokument wat ondersoek moet word, selekteer en daarmee werk. Daar is verskeie funksies in *ATLAS.ti*, maar die koderingsfunksie is veral van belang (hierdie funksies word in afdeling 5.8.3 bespreek).

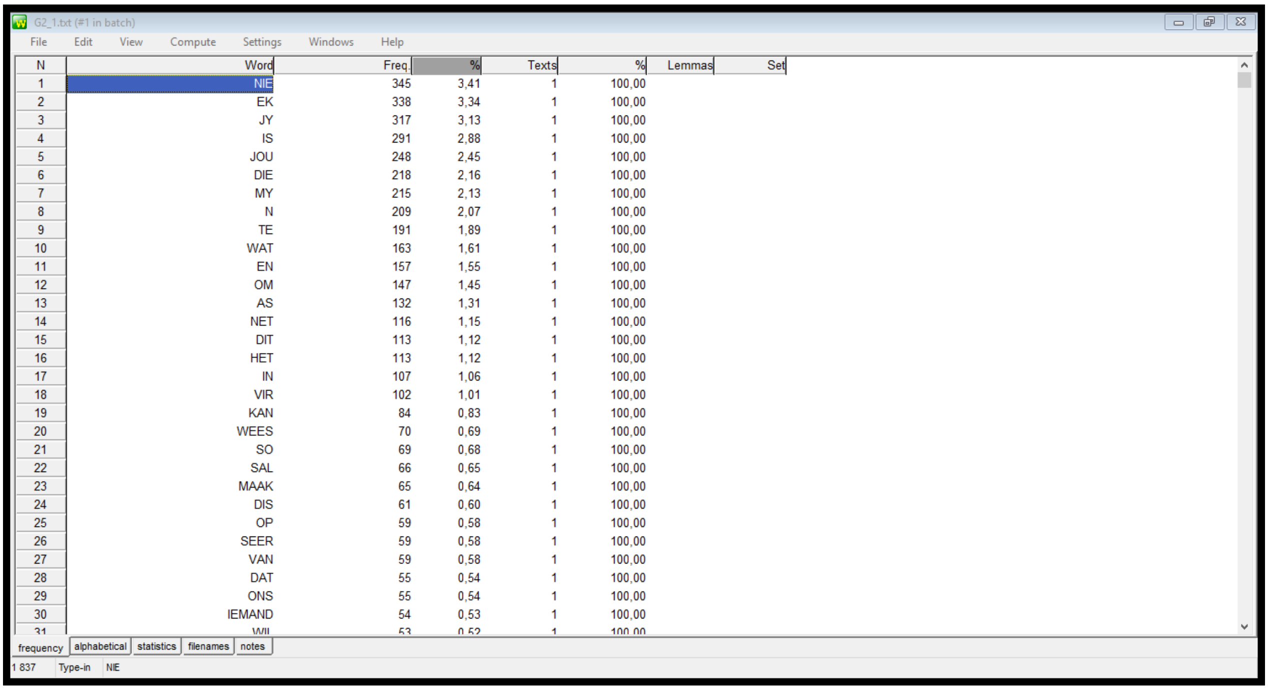
5.8.2 Terminologie van programmatuur

Vanuit die oorsig oor die programmatuur wat ek in my studie gebruik het, is dit duidelik dat daar verskeie funksies is wat deur al die genoemde programme uitgevoer kan word. In hierdie afdeling bespreek ek elk van hierdie funksies om 'n oorsig daaroor te verskaf.

[137] Hierdie funksies word in afdeling 5.8.3 (Terminologie van programmatuur) verduidelik aangesien dit in die meerderheid van die programme voorkom.

5.8.2.1 Woordelyste

Scott (2012:24) verwys na woordelyste as ’n funksie in *WordSmith Tools* wat aandui hoeveel maal elke woord in ’n korpus verskyn asook die totale aantal woorde in ’n teks. Skermskoot 5.5 en 5.6 is onderskeidelik ’n woordelys van ’n klein korpus en die statistiese inligting soos deur *WordSmith Tools* bereken:

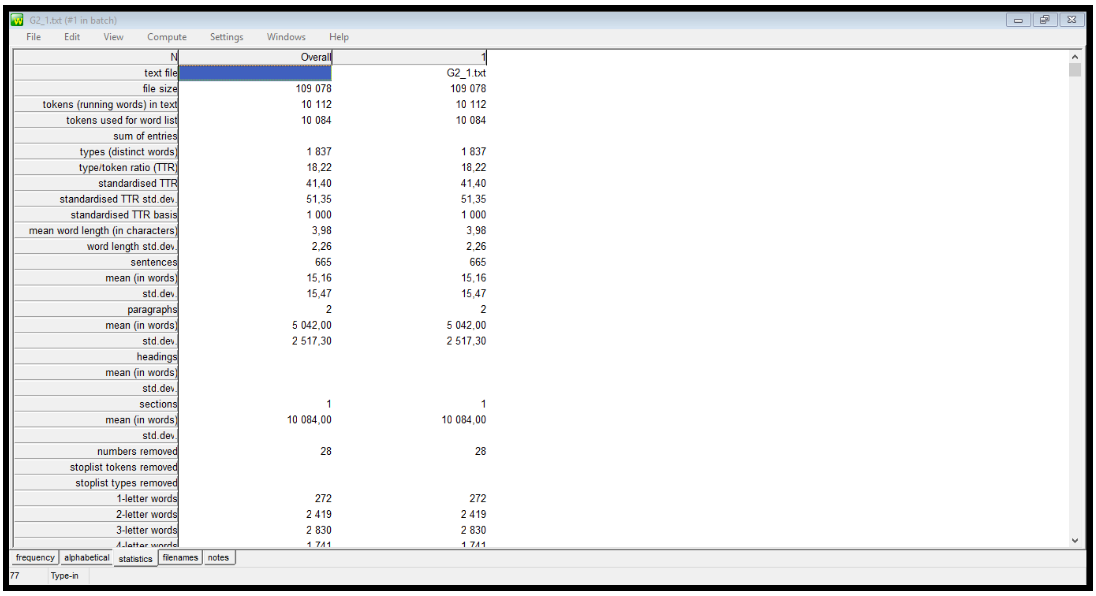


N	Word	Freq.	%	Texts	%	Lemmas	Set
1	NIE	345	3,41	1	100,00		
2	EK	338	3,34	1	100,00		
3	JY	317	3,13	1	100,00		
4	IS	291	2,88	1	100,00		
5	JOU	248	2,45	1	100,00		
6	DIE	218	2,16	1	100,00		
7	MY	215	2,13	1	100,00		
8	N	209	2,07	1	100,00		
9	TE	191	1,89	1	100,00		
10	WAT	163	1,61	1	100,00		
11	EN	157	1,55	1	100,00		
12	OM	147	1,45	1	100,00		
13	AS	132	1,31	1	100,00		
14	NET	116	1,15	1	100,00		
15	DIT	113	1,12	1	100,00		
16	HET	113	1,12	1	100,00		
17	IN	107	1,06	1	100,00		
18	VIR	102	1,01	1	100,00		
19	KAN	84	0,83	1	100,00		
20	WEES	70	0,69	1	100,00		
21	SO	69	0,68	1	100,00		
22	SAL	66	0,65	1	100,00		
23	MAAK	65	0,64	1	100,00		
24	DIS	61	0,60	1	100,00		
25	OP	59	0,58	1	100,00		
26	SEER	59	0,58	1	100,00		
27	VAN	59	0,58	1	100,00		
28	DAT	55	0,54	1	100,00		
29	ONS	55	0,54	1	100,00		
30	IEMAND	54	0,53	1	100,00		

Skermskoot 5.4 Statistiese inligting vir woordelyste in WST

Vanuit Skermskoot 5.4 blyk kan die aantal woorde in die woordelys, gemiddelde lengte van woorde, sinne en ook wat die Tipe/Teken-ratio[138] is, gesien word. Hierdie inligting is dus die statistiese inligting van ’n betrokke korpus.

[138] Sien afdeling 5.8.4 (Statistiese konsepte).



N	Overall	1
text file		G2_1.txt
file size	109 078	109 078
tokens (running words) in text	10 112	10 112
tokens used for word list	10 084	10 084
sum of entries		
types (distinct words)	1 837	1 837
type/token ratio (TTR)	18,22	18,22
standardised TTR	41,40	41,40
standardised TTR std.dev.	51,35	51,35
standardised TTR basis	1 000	1 000
mean word length (in characters)	3,98	3,98
word length std.dev.	2,26	2,26
sentences	665	665
mean (in words)	15,16	15,16
std.dev.	15,47	15,47
paragraphs	2	2
mean (in words)	5 042,00	5 042,00
std.dev.	2 517,30	2 517,30
headings		
mean (in words)		
std.dev.		
sections	1	1
mean (in words)	10 084,00	10 084,00
std.dev.		
numbers removed	28	28
stoplist tokens removed		
stoplist types removed		
1-letter words	272	272
2-letter words	2 419	2 419
3-letter words	2 830	2 830
4-letter words	1 741	1 741

Skermskoot 5.5 Woordelysfunksie van ’n klein korpus

Skermskoot 5.5 toon die woorde waaruit n betrokke korpus bestaan. Woordelyste kan volgens frekwensie of alfabeties gerangskik word − ’n funksie wat veral nuttig is wanneer die outeur van ’n teks geïdentifiseer moet word. Scott (2012:202) beskryf die doel van woordelyste soos volg:

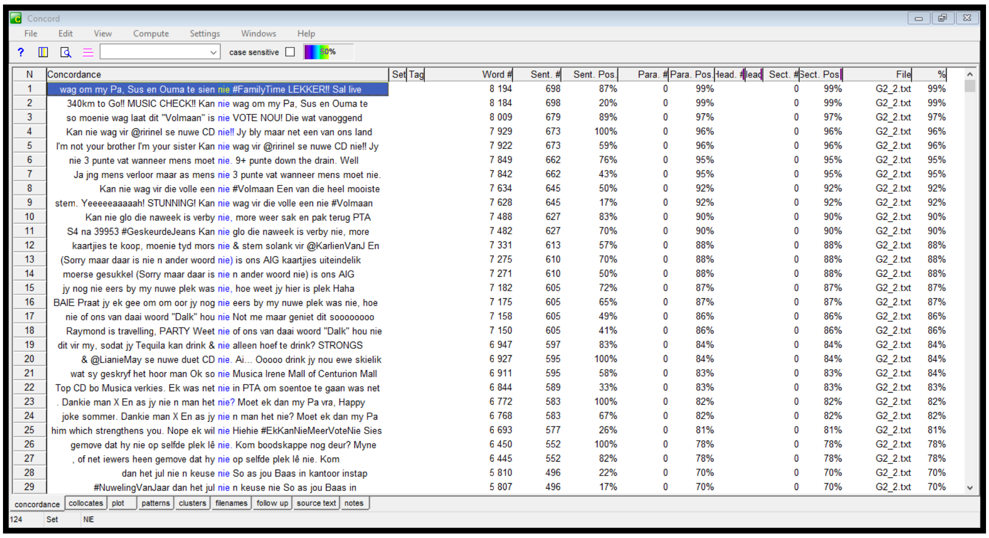
- Analiseer die woordeskat wat gebruik word.
- Identifiseer woorde wat frekwent saam gegroepeer word.
- Vergelyk die frekwensie van ’n woord in verskillende korpora.
- Vergelyk die frekwensie van verwante woorde of vertalings wanneer daar met vertaalde tekste gewerk word.
- Stel konkordansies vanuit die woordelyste saam.

Nog ’n belangrike funksie van woordelyste is die multiwoordeenhede − beter bekend as die n-gram[139].

5.8.2.2 Konkordansie

Konkordansie verwys volgens Baker (2010a:21) na ’n Tabel wat die linguistiese items uitbeeld wat in ŉ bepaalde korpus voorkom in die konteks waarin dit oorspronklik voorkom (sien Skermskoot 5.6). Scott (2010:15) stel dat die konkordansiefunksie elke keer wat ’n woord in ŉ korpus voorkom, lys mits dit deur ’n leesteken of spasie voor en ná gepaardgaan. Baker (2010a:22) dui daarop dat die konkordansiefunksie in ŉ kwalitatiewe analise gebruik kan word sowel as om die semantiese waarde van woorde te bepaal.

[139] Sien afdeling 5.8.4.3 (n-gram).

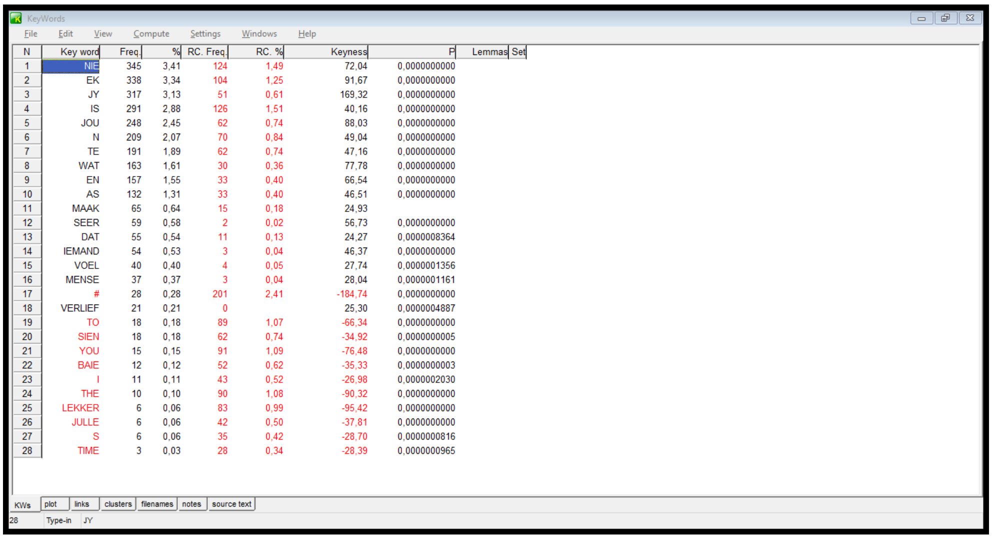


N	Concordance	Set	Tag	Word #	Sent. #	Sent. Pos.	Para. #	Para. Pos.	Head. #	Head	Sect. #	Sect. Pos	File	%
1	wag om my Pa, Sus en Ouma te sien nie #FamilyTime LEKKER!! Sal live			8 194	698	87%	0	99%			0	99%	G2_2.txt	99%
2	340km to Go!! MUSIC CHECK!! Kan nie wag om my Pa, Sus en Ouma te			8 184	698	20%	0	99%			0	99%	G2_2.txt	99%
3	so moenie wag laat dit "Volmaan" is nie VOTE NOU! Die wat vanoggend			8 009	679	89%	0	97%			0	97%	G2_2.txt	97%
4	Kan nie wag vir @ririnel se nuwe CD nie!! Jy bly maar net een van ons land			7 929	673	100%	0	96%			0	96%	G2_2.txt	96%
5	I'm not your brother I'm your sister Kan nie wag vir @ririnel se nuwe CD nie!! Jy			7 922	673	59%	0	96%			0	96%	G2_2.txt	96%
6	nie 3 punte vat wanneer mens moet nie. 9+ punte down the drain. Well			7 849	662	76%	0	95%			0	95%	G2_2.txt	95%
7	Ja jng mens verloor maar as mens nie 3 punte vat wanneer mens moet nie.			7 842	662	43%	0	95%			0	95%	G2_2.txt	95%
8	Kan nie wag vir die volle een nie #Volmaan Een van die heel mooiste			7 634	645	50%	0	92%			0	92%	G2_2.txt	92%
9	stem. Yeeeeeaaaaah! STUNNING! Kan nie wag vir die volle een nie #Volmaan			7 628	645	17%	0	92%			0	92%	G2_2.txt	92%
10	Kan nie glo die naweek is verby nie, more weer sak en pak terug PTA			7 488	627	83%	0	90%			0	90%	G2_2.txt	90%
11	S4 na 39953 #GeskeurdeJeans Kan nie glo die naweek is verby nie, more			7 482	627	70%	0	90%			0	90%	G2_2.txt	90%
12	kaartjies te koop, moenie tyd mors nie & stem solank vir @KarlienVanJ En			7 331	613	57%	0	88%			0	88%	G2_2.txt	88%
13	(Sorry maar daar is nie n ander woord nie) is ons AIG kaartjies uiteindelik			7 275	610	70%	0	88%			0	88%	G2_2.txt	88%
14	moerse gesukkel (Sorry maar daar is nie n ander woord nie) is ons AIG			7 271	610	50%	0	88%			0	88%	G2_2.txt	88%
15	jy nog nie eers by my nuwe plek was nie, hoe weet jy hier is plek Haha			7 182	605	72%	0	87%			0	87%	G2_2.txt	87%
16	BAIE Praat jy ek gee om om oor jy nog nie eers by my nuwe plek was nie, hoe			7 175	605	65%	0	87%			0	87%	G2_2.txt	87%
17	nie of ons van daai woord "Dalk" hou nie Not me maar geniet dit sooooooooo			7 158	605	49%	0	86%			0	86%	G2_2.txt	86%
18	Raymond is travelling, PARTY Weet nie of ons van daai woord "Dalk" hou nie			7 150	605	41%	0	86%			0	86%	G2_2.txt	86%
19	dit vir my, sodat jy Tequila kan drink & nie alleen hoef te drink? STRONGS			6 947	597	83%	0	84%			0	84%	G2_2.txt	84%
20	& @LianieMay se nuwe duet CD nie. Ai... Ooooo drink jy nou ewe skielik			6 927	595	100%	0	84%			0	84%	G2_2.txt	84%
21	wat sy geskryf het hoor man Ok so nie Musica Irene Mall of Centurion Mall			6 911	595	58%	0	83%			0	83%	G2_2.txt	84%
22	Top CD bo Musica verkies. Ek was net nie in PTA om soentoe te gaan was net			6 844	589	33%	0	83%			0	83%	G2_2.txt	83%
23	. Dankie man X En as jy nie n man het nie? Moet ek dan my Pa vra, Happy			6 772	583	100%	0	82%			0	82%	G2_2.txt	82%
24	joke sommer. Dankie man X En as jy nie n man het nie? Moet ek dan my Pa			6 768	583	67%	0	82%			0	82%	G2_2.txt	82%
25	him which strengthens you. Nope ek wil nie Hiehie #EkKanNieMeerVoteNie Sies			6 693	577	26%	0	81%			0	81%	G2_2.txt	81%
26	gemove dat hy nie op selfde plek lê nie. Kom boodskappe nog deur? Myne			6 450	552	100%	0	78%			0	78%	G2_2.txt	78%
27	, of net iewers heen gemove dat hy nie op selfde plek lê nie. Kom			6 445	552	82%	0	78%			0	78%	G2_2.txt	78%
28	dan het jul nie n keuse nie So as jou Baas in kantoor instap			5 810	496	22%	0	70%			0	70%	G2_2.txt	70%
29	#NuwelingVanJaar dan het jul nie n keuse nie So as jou Baas in			5 807	496	17%	0	70%			0	70%	G2_2.txt	70%

Skermskoot 5.6 Voorstelling van die konkordansiefunksie in WST

5.8.2.3 Sleutelwoorde

Berber-Sardinha (2001:7) definieer “sleutelwoorde” as ’n vergelyking tussen die woordelys wat vanuit ’n gespesialiseerde korpus saamgestel word, en die woordelys saamgestel vanuit ’n verwysingskorpus. Ten einde sleutelwoorde te ondersoek, moet die verwysingskorpus groter as die gespesialiseerde korpus wees. In Skermskoot 5.7 kan die resultate van ’n sleutelwoordlys wat in *WordSmith Tools* opgesoek is, gesien word:



N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P	Lemmas	Set
1	NIE	345	3,41	124	1,49	72,04	0,0000000000		
2	EK	338	3,34	104	1,25	91,67	0,0000000000		
3	JY	317	3,13	51	0,61	169,32	0,0000000000		
4	IS	291	2,88	126	1,51	40,16	0,0000000000		
5	JOU	248	2,45	62	0,74	88,03	0,0000000000		
6	N	209	2,07	70	0,84	49,04	0,0000000000		
7	TE	191	1,89	62	0,74	47,16	0,0000000000		
8	WAT	163	1,61	30	0,36	77,78	0,0000000000		
9	EN	157	1,55	33	0,40	66,54	0,0000000000		
10	AS	132	1,31	33	0,40	46,51	0,0000000000		
11	MAAK	65	0,64	15	0,18	24,93			
12	SEER	59	0,58	2	0,02	56,73	0,0000000000		
13	DAT	55	0,54	11	0,13	24,27	0,0000008364		
14	IEMAND	54	0,53	3	0,04	46,37	0,0000000000		
15	VOEL	40	0,40	4	0,05	27,74	0,0000001356		
16	MENSE	37	0,37	3	0,04	28,04	0,0000001161		
17	#	28	0,28	201	2,41	-184,74	0,0000000000		
18	VERLIEF	21	0,21	0		25,30	0,0000004887		
19	TO	18	0,18	89	1,07	-66,34	0,0000000000		
20	SIEN	18	0,18	62	0,74	-34,92	0,0000000005		
21	YOU	15	0,15	91	1,09	-76,48	0,0000000000		
22	BAIE	12	0,12	52	0,62	-35,33	0,0000000003		
23	I	11	0,11	43	0,52	-26,98	0,0000002030		
24	THE	10	0,10	90	1,08	-90,32	0,0000000000		
25	LEKKER	6	0,06	83	0,99	-95,42	0,0000000000		
26	JULLE	6	0,06	42	0,50	-37,81	0,0000000000		
27	S	6	0,06	35	0,42	-28,70	0,0000000816		
28	TIME	3	0,03	28	0,34	-28,39	0,0000000965		

Skermskoot 5.7 Sleutelwoordfunksie van WST

Die resultate wat in Skermskoot 5.7 aangetoon word kan soos volg beskryf word. Sleutelwoorde word in die regterkantste kolom gelys. Hierdie woorde se frekwensie is statisties hoër in die gespesialiseerde korpus as in die verwysingskorpus. Wanneer die sleutelwoorde meer frekwent

in die verwysingskorpus is, word die getal as ’n negatief aangetoon (sien die woord **lekker** in Skermskoot 5.7). Bednarek (2009:21) verwys na sleutelwoorde as leksikale items wat in so ’n mate deur outeurs gebruik word dat dit afwyk van die frekwensie van die verwysingskorpus.

5.8.2.4 Frekwensie

Frekwensie is een van die belangrikste aspekte wanneer *WordSmith Tools* (en outeurskapidentifikasie) gebruik word, en vorm volgens Baker (2010a:19-21) die hoeksteen van korpuslinguistiek. Frekwensie verwys na die aantal keer wat iets (’n woord, sinsnede of sin) in ’n korpus voorkom.

Vanuit Skermskoot 5.4 en 5.5 blyk dit dat *WordSmith Tools* reeds al die berekeninge doen. Soos gesien kan word, is die TTR (Tipe/Teken-ratio) reeds bepaal sowel as die praktiese beduidendheid. Die frekwensie word gebruik om die TTR en die praktiese beduidendheid te bereken[140]. Frekwensie kan deur middel van ’n eenvoudige formule bereken word:

$$\frac{Totale\ keer\ wat\ woord\ in\ korpus\ voorkom}{Totale\ woorde\ in\ korpus} \times 1000 = Frekwensie$$

As die data vanuit Skermskoot 5.5 en 5.6 gebruik word om die frekwensie van die woord **nie** te bereken, lyk die formule soos volg:

$$\frac{345}{10084} \times 1000 = 34{,}21$$

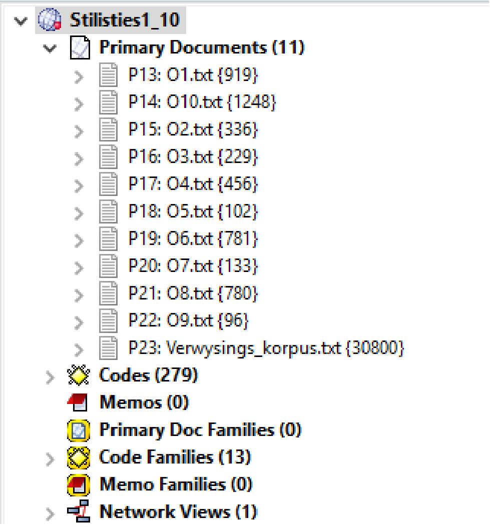
Die woord **nie** se frekwensie in die korpus is dus 34,2. Die frekwensie van woorde vanuit een korpus (gespesialiseerde korpus) kan vergelyk word met die frekwensie van woorde vanuit ’n ander korpus (die verwysingskorpus wat as die taalnorm van die spesifieke konsep wat ondersoek word, dien) om sodoende te bepaal of die woord afwyk van dít wat as die norm beskou word. As voorbeeld: in korpus 1 kom die woord **nie** 345 keer voor (frekwensie = 34,3), en in korpus 2 kom die woord **nie** 124 keer voor (frekwensie = 14,9). Daarom is die woord **nie** meer frekwent in korpus 1. Vervolgens verwys ek vlugtig na belangrike statistiese konsepte wat betrekking het op die metodologieë wat tydens die identifisering van stylkenmerke in ’n korpus sowel as outeurskapidentifikasie toegepas kan word.

[140] Sien 5.8.4 (Statistiese konsepte).

5.8.2.5 *ATLAS.ti*-funksies

Soos genoem in afdeling 5.8.3.5 (*ATLAS.ti*) beskik die program oor 'n verskeidenheid funksies wat in 'n kwalitatiewe analise van groot "ongeorganiseerde" aantal data. Vervolgens verwys ek oorsigtelik na die verskillende funksies ten einde die analise wat in hoofstuk 6 plaasvind te begrond.

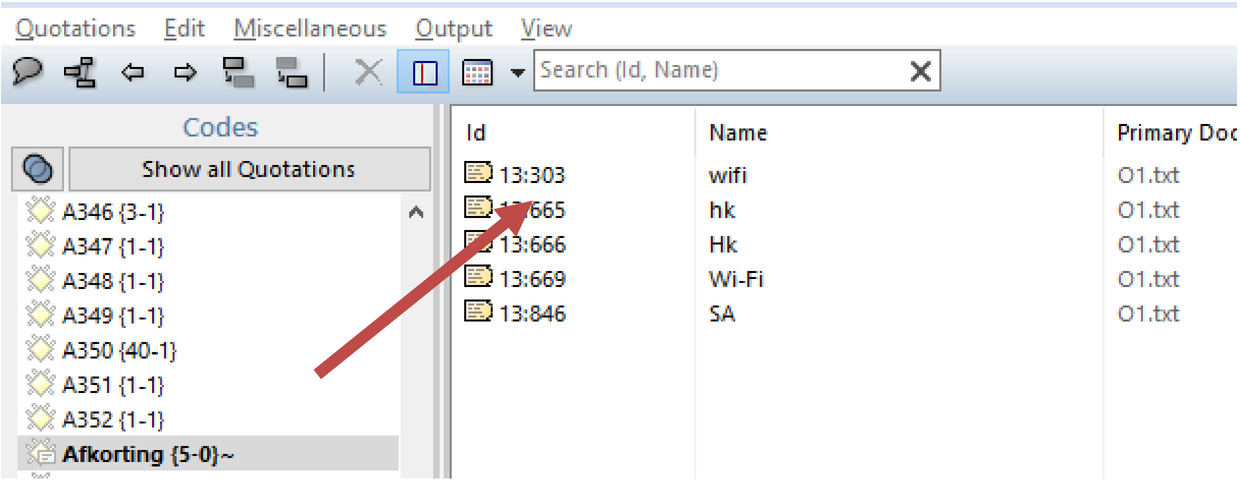
- **Hermeneutiese eenheid**: HU kan as die datastruktuur van die projek beskou word (vgl. Fries, 2014:14) en alles wat relevant tot die betrokke projek is, word in die HU gestoor soos uitgebeeld in Skermskoot 5.8:



Skermskoot 5.8 Voorstelling van die HU in *ATLAS.ti*

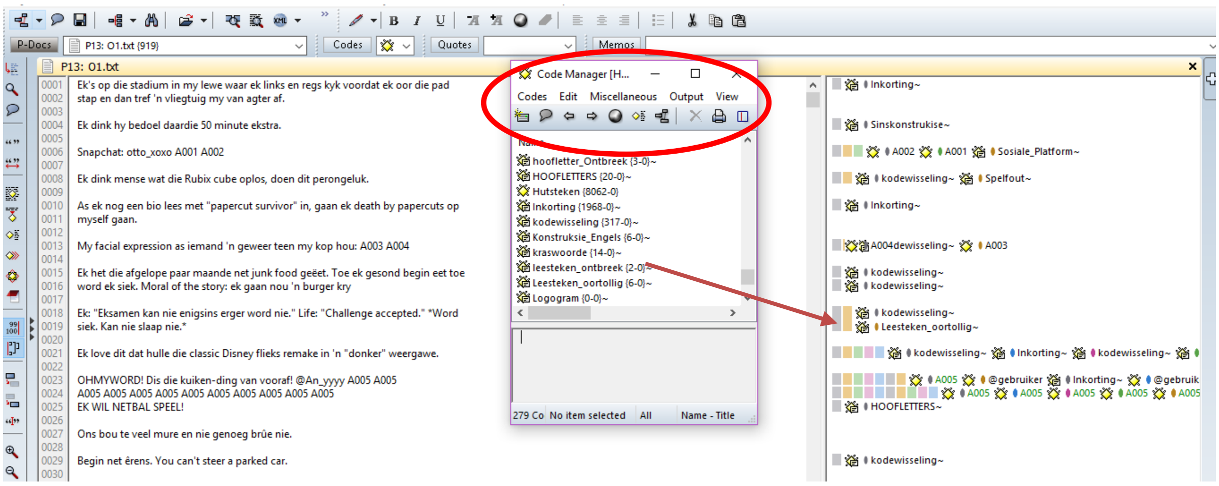
- **Primêre dokumente**: Die primêre dokumente (PD) (of hoofdokumente) verwys na die data wat geselekteer is om deel te vorm van die projek (Fries, 2014:15). Hierdie PD's kan as families saamgegroepeer word ten einde kenmerke te vergelyk en identifiseer.

- **Aanhalings**: Aanhalings is 'n funksie wat die voorkoms van kenmerke wat ondersoek word, uitbeeld en saamgroepeer (Fries, 2014:17). Die ondersoek na kenmerke word dus vergemaklik (sien Skermskoot 5.8). Wanneer bepaalde kodes geselekteer word, dui hierdie funksie alle aanhalings aan wat gekodeer is in die betrokke teks.



Skermskoot 5.9 Voorstelling van die aanhalingsfunksie in *ATLAS.ti*

- **Kodes**: Kodes kan as een van die belangrikste funksies in *ATLAS.ti* beskou word. Dit word gebruik om die kenmerke wat in Skermskoot 5.9 uitgebeeld word te merk. Volgens Fries (2014:18, 19) kan die koderingstelsel op tekste toegepas word en herhalings van dieselfde kenmerke word gemerk en saamgegroepeer.



Skermskoot 5.10 Voorstelling van die kodes in *ATLAS.ti*

Die bogenoemde drie funksies van *ATLAS.ti* kan as die belangrikste konsepte van die program beskou word. Ander funksies van *ATLAS.ti* sluit in ’n netwerkfunksie, families en woordsoektogte. Hierdie funksies is wel in die kwalitatiewe analise van my data gebruik en word in hoofstuk 6 voorgehou wanneer die empiriese analise beskryf word. ’n Belangrike aspek om in gedagte te hou, is dat *ATLAS.ti* my in staat gestel het om die stilistiese kenmerke van die korpus te merk (kodeer) en statistiese voor te stel.

5.8.3 Statistiese konsepte

In afdeling 5.8.2 is die belangrikste funksies van *WordSmith Tools* en *ATLAS.ti* bespreek. Met betrekking tot frekwensie berekeninge is dit duidelik dat daar talle statistiese berekeninge is wat gebruik word om unieke woorde en kollokasies te identifiseer. In afdeling 5.7 (Kwantitatiewe analise) word daar verwys na die manier waarop numeriese waardes aan tekste toegeken word om sodoende ’n konsep statisties te bewys. Vir die doel van my studie was dit veral van belang om die terme p-waardes, chi-kwadraat toetse en n-gram te definieer. Alhoewel die programmatuur wat ek in my studie gebruik het hierdie waardes outomaties bereken het, hou ek in hierdie afdeling die statistiese konsepte opsommend voor om ’n oorsig te verskaf.

5.8.3.1 Nulhipotese

Wanneer daar van ’n kwantitatiewe benadering gebruik word, moet daar ’n hipotese bestaan wat deur navorsers as waar (of onwaar) bewys moet word. Levon (2010:70) postuleer dat om ’n statistiese analise uit te voer, moet daar eerstens ’n hipotese gevorm word. Die hipotese stel navorsers in staat om ’n meer ingewikkelde ondersoek uit te voer as om slegs kenmerke te tel in ’n korpus (Meyer, 2004:102). ’n Hipotese verwys na ’n doelbewus geselekteerde navorsingsvraag wat aanleiding daartoe gee dat die korpus beide kwantitatief en kwalitatief geanaliseer moet word, en waar die kwantitatiewe navorsing die hipotese statisties bewys (Meyer, 2002:102).

Levon (2010:70) verwys na die hipotese wat in navorsing ondersoek word as die eksperimentele hipotese (H) − ’n hipotese wat deur voorbeelde vanuit kwalitatiewe navorsing ondersteun word. Ten einde ’n hipotese te kan bewys, word daar dus van ’n nulhipotese (H_0) gebruik gemaak. Die H_0 verwys na ’n hipotese wat in direkte teenstelling met H staan (vgl. Levon, 2010:72). Daar bestaan verskeie statistiese berekeninge om H en H_0 te bereken.

5.8.3.2 Chi-Kwadraat en p-waardes

Thiart (2014:94) voer aan dat ’n chi-kwadraattoets ’n nulhipotese aanvaar of verwerp. Dit verwys na die vergelyking van twee of meer frekwensies met die doel om te bepaal wat die moontlikheid is dat die verskille in die frekwensies bloot toevallig is. Chi-kwadraattoetse word soos volg bereken:

Tabel 5.2 Voorstelling van rou data wat gebruik word vir chi-kwadraatberekeninge

Deelnemers	X	Y	Totaal
Groep 1	(42 X 177) / 380 = 19,6	(42 X 203) / 380 = 22.4	42
Groep 2	(206 X 177) / 380 = 96	(206 X 203) / 380 = 110	206
Groep 3	(132 X 177) / 380 = 61,5	(132 X 203) / 380 = 70,5	132
Totaal	177	203	380

Tabel 5.2 is ŉ voorstelling van die data[141] wat gebruik word wanneer chi-kwadraatberekeninge uitgevoer word. Hierdie data moet volgens Levon (2010:77) rou data wees − die data moet die getalle wees wat aangeteken is, en nie persentasies of proporsionele berekeninge nie. Data word verkry vanaf waarnemings vanuit ŉ steekproefneming. Tabel 5.4 se data kan dus beskou word as: X kom 34 keer in groep 1 voor terwyl Y 8 keer voorkom. Nadat die roudata opgeskryf is in Tabelvorm word die verwagte waardes bereken (sien Tabel 5.3):

Tabel 5.3 Berekeninge van die verwagte waardes (geneem vanuit Levon, 2010:78)

Deelnemers	X	Y	Totaal
Groep 1	(42 X 177) / 380 = 19,6	(42 X 203) / 380 = 22.4	42
Groep 2	(206 X 177) / 380 = 96	(206 X 203) / 380 = 110	206
Groep 3	(132 X 177) / 380 = 61,5	(132 X 203) / 380 = 70,5	132
Totaal	177	203	380

Die laaste stap is om die verskil tussen die waargeneemde en die verwagte waardes te bereken. Levon (2010:78) verwys na ŉ wiskundige berekening om dié waardes te bereken:

$$\sum \frac{(Waargeneem - Verwagte)2}{Verwagte}$$

Die berekeninge word soos volg uitgevoer:

$$\chi 2 = \frac{(23-19.6)2}{19.6} + \frac{(8-22.4)2}{22.4} + \frac{(177-96)2}{96} + \frac{(89-110)2}{110} + \frac{(26-61.5)2}{61.5} + \frac{(106-70.5)2}{70.5}$$

$$\therefore \chi 2 = \frac{(14.4)2}{19.6} + \frac{(-14.4)2}{22.4} + \frac{(21)2}{96} + \frac{(-21)2}{110} + \frac{(-35.5)2}{61.5} + \frac{(35.5)2}{70.5}$$

$$\therefore \chi 2 = \frac{207.4}{19.6} + \frac{207.4}{22.4} + \frac{441}{96} + \frac{441}{110} + \frac{1260.3}{61.5} + \frac{1260.3}{70.5}$$

$$\therefore \chi 2 = 66.9$$

141 Die data word geneem vanuit Levon (2010:77).

Die berekeninge toon dus aan dat die chi-kwadraatstatistiek vir die data (soos gesien in Tabel 5.2) 66.9 is. Vervolgens moet die p-waarde bepaal word.

Die p-waarde verwys na die graad van vryheid (*df*) en word gedefinieer as wiskundige beperkings wat gebruik kan word om te bepaal of die statistiese berekening geldig en waar is (vgl. Levon, 2010:80). Vir die doel van my studie was dit nie nodig om die manier waarop die p-waardes bereken is aan te dui nie.

Die p-waardes verwys (volgens Michell, 2013:104) na die berekening van die statisties beduidende verskil. Om 'n resultaat te kry wat statisties beduidend is, moet die $p \leq 0.05$ (die nulhipotese word verwerp) of $p > 0.05$ (die nulhipotese kan nie verwerp word nie) wees. Nini (2014:114) stel dat die p-waarde van 0.05 soms té min resultate lewer, en stel dus voor dat 'n p-waarde van 0.001 eerder gebruik moet word as die drempel vir statistiese beduidendheid.

Indien daar weer van die voorbeeld in Tabel 5.2 en 5.3 gebruik gemaak word, kan die p-waarde soos volg bereken word:

df = (aantal rye -1) X (aantal kolomme - 1)

∴*df* = (3-1) χ (2-1)

∴ *df* = 2

Die chi-kwadraatwaarde is dus 66.9 en die *df* is 2. Die volgende stap is om die p-waarde in die Tabel (chi-kwadraatinterpretasie) op te soek.

df	0.10	0.05	0.025	0.01	0.001
1	2,706	3,841	5,024	6,635	10,828
2	4,605	5,991	7,378	9,210	13,816
3	6,251	7,815	9,348	11,345	16,266
4	7,779	9,488	11,143	13,277	18,467
5	9,236	11,070	12,833	15,086	20,515

Skermskoot 5.11 Voorstelling van die p-waardes (geneem vanuit Levon, 2010:81)

Soos reeds genoem, moet $p \leq 0.05$ (die nulhipotese word verwerp) of $p > 0.05$ (die nulhipotese kan nie verwerp word nie) wees. Aangesien die chi-kwadraatwaarde 66.9 is en die *df* 2, kan die p-waarde vanaf die Tabel (Skermskoot 5.11) gelees word. In dié geval is die p-waarde 0.001. Dus is $p \leq 0.05$ en daarom kan daar gestel word dat die nul hipotese verwerp kan word.

5.8.3.3 N-gram

Die n-gram word deur Johnson en Wright (2014:38) voorgehou as klein (twee tot ses woorde) dele van tekste wat kenmerkend van outeurs se skryfstyl is. Hulle (Johnson & Wright, 2014:38) vergelyk die n-gram met DNS en stel dat indien die stukkies teks oor genoegsame data beskik, kan die teks wat gebruik word om ʼn outeur te identifiseer, verminder word tot betekenisvolle sleutelsegmente en sodoende beweeg navorsing nader aan ʼn bewys van die individu se unieke taaleienskappe. Die n-gram is dus soortgelyk aan kollokasies. Hoey (2005:181) skryf: "Everybody's language is unique, because all our lexical items are inevitably primed differently as a result of different encounters, spoken and written. We have different parents and different friends, live in different places, read different books, get into different arguments and have different colleagues."

Volgens Johnson en Wright (2014:38) kan daar deur middel van stilistiese, korpus- en rekenaarbenaderings tot tekste herhalende kollokasies en patrone − n-gramme − in die tekste geïdentifiseer word wat kan lei tot die suksesvolle identifisering van outeurs van tekste.

In hierdie afdeling (5.8.3) is daar oorsigtelik na verskillende statistiese konsepte verwys. Hierdie konsepte word verder in afdeling 5.9 (Kleinkorpora en outeurskapidentifikasie) bespreek om ŉ agtergrond te skep vir die metode wat ek in my ondersoek na die stylkenmerke van Afrikaans op Twitter en die outeurskapidentifikasie van outeurs op hierdie platform gebruik het.

5.9 Kleinkorpora en outeurskapidentifikasie

In afdelings 5.5, 5.6 en 5.7 is daar verwys na die gemengde metode wat ek in my studie gevolg het. In die bespreking daarvan word daar na ŉ kwantitatiewe en kwalitatiewe analise verwys. Hierdie analises is gedefinieer en daar is bepaal dat daar (veral met betrekking tot outeurskapidentifikasie en die identifisering van stylkenmerke) van ʼn stilistiese en stilometriese analise gebruik gemaak kan word. Vanuit hoofstuk 4 het dit ook duidelik geblyk dat my studie oor outeurskapidentifikasie ook ʼn korpuslinguistiese benadering moes volg ten einde die korpora saam te stel waaruit die stylkenmerke geïdentifiseer kon word en waaruit die outeurskapidentifikasie kon plaasvind.

Volgens Sardinha (2011:31) word die aanname gemaak dat korpuslinguistiek in die kwalitatiewe navorsingsparadigma val aangesien daar statisties te werk gegaan word, maar dit is nie die geval nie. Korpuslinguistiek kan in die middel van twee uiterstes geplaas word (kwalitatief en kwantitatief) aangesien navorsing waar korpora saamgestel en geanaliseer word kwalitatief van

aard is aangesien dit die statistiese analise van data vereis, maar ook kwalitatief omdat die data ook geïnterpreteer moet word in die natuurlike omgewing waar dit ingesamel is (vgl. Sardinha, 2011:31).

In die volgende afdeling gaan ek na die verskillende metodes verwys wat reeds gebruik is in outeurskapidentifikasie en die identifisering van stylkenmerke. Hierdie metodes kan óf gemeng óf net kwalitatief óf kwantitatief wees.

5.10 Oorsig van metodes vir outeurskapidentifikasie

Kopka en Vogel (2004:3) voer aan dat die verskeie ontwikkelinge in tegnologie en rekenaars forensiese navorsing dwing om aan te pas by dié situasie. Tekste wat op rekenaars geskep word, beskik nie oor dieselfde "materiële implikasies" as handgeskrewe dokumente (of selfs dokumente wat op tikmasjiene geskep is) nie; daarom moet die inhoud, woordkeuses en gebruik van punktuasie en letters (in die geval van sosiale media en SMS'e, moet daar ook na die emotikon-gebruik ondersoek ingestel word) geanaliseer word (vgl. Kopka & Vogel, 2004:3).

Soos reeds genoem, is daar van beide ŉ kwantitatiewe en kwalitatiewe benadering gebruik gemaak. Wanneer outeurskapidentifikasie uitgevoer word, word dié twee metodes gekombineerd gebruik. McMenamin (2010:488) verwys na drie modelle wat gebruik kan word vir outeurskapidentifikasie, naamlik:

- **Ooreenkomsmodel** (*resemblance model*): Hierdie model word gebruik wanneer eksterne faktore ŉ invloed het op die aantal moontlike outeurs.
- **Konsekwentheidmodel** (*consistency model*): Hierdie model word toegepas wanneer daar bepaal moet word of verskeie tekste deur een outeur geskryf is.
- **Bevolkingsmodel** (*population model*): Hierdie model word gebruik wanneer daar ŉ groot getal moontlike outeurs is.

Al drie modelle kon in hierdie studie toegepas word. Chaski (2008:3) verwys na die gebruik van forensiese stilistiek waar foute en idiosinkrasieë in tekste geanaliseer word. Daar is reeds verwys na die verskillende kenmerke wat ondersoek kan word wanneer ŉ forensies stilistiese analise uitgevoer word. ŉ Stilometriese metode verwys na ŉ kwantitatiewe metode waar telbare taalkenmerke ondersoek word. Volgens Chaski (2008:5) sluit dié kenmerke woordlengte, fraselengte, sinslengte, woordeskat rykheid asook die verspreiding van woorde van verskillende lengtes in, en hierdie metode is beskrywend van aard.

Wanneer daar ondersoek ingestel word na outeurskapidentifikasie met betrekking tot ŉ mikroblog, is daar talle beperkings wat in gedagte gehou moet word. Bhargava *et al. (*2013:37)

voer aan dat outeurskapidentifikasie nie op tekste korter as 1 000 woorde in die vroeë 19de eeu uitgevoer kon word nie. Hierdie minimum teksdrempel het oor die jare afgeneem, en in die vroeë 21ste eeu is 250 woorde as die minimum drempel beskou, maar met die ontwikkeling van RBK is dit van die uiterse belang om hierdie minimum drempel steeds te verlaag (vgl. Bhargava *et al.,* 2013:37).

Etlike studies is al uitgevoer om die outeurs van kort tekste te bepaal, insluitende:

- Abbasi en Chen *(*2005:67-75) het verwys na die manier waarop die anonimiteit van sosialemediaplatforms outeurskapidentifikasie beïnvloed en het probeer om outeurs van hierdie korter tekste te identifiseer.

- Koppel *et al. (*2002) het probeer om geskrewe tekste outomaties te kategoriseer om sodoende die geslag van outeurs te bepaal. Daar is na woordfrekwensies, woordlengtes, aantal sinne en stilometriese kenmerke[142] ondersoek ingestel.

- Silva en Laboreiro (2011:161) het Portugese twiets geanaliseer om outeurskapmerkers te identifiseer deur gebruik te maak van stilistiese merkers. Hulle het bevind dat emotikons en spesifiek kort boodskapmerkers dominant is teenoor tradisionele stilometriese merkers wanneer outeurskapidentifikasie uitgevoer word. In die finale resultate van die analise is daar bevind dat ’n F-telling[143] van 0.63 vir die 100 voorbeelde van elk van die outeurs (wie onder verdenking was) bepaal kon word. Daar is met slegs drie verdagte outeurs per analise gewerk.

- Rasinger (2010) het ondersoek ingestel na die leksikale en sintaktiese karaktereienskappe van tekste. Dié tegniek staan as die *Probalistic Contex-Free Grammars* (PCFG) bekend en daar is na afsonderlike datastelle (poësie, besigheid, reis en krieket) ondersoek ingestel. Vir die datastel oor krieket was die akkuraatheid 95%.

- Bhargava *et al*. (2013) het ondersoek ingestel na outeurskapidentifikasie op Twitter. Hulle het ’n stilometriese analise gebruik waar leksikale kenmerke (woorde per twiet, totale aantal sinne per twiet, woorde per sin, frekwensie van woorde en leksikale diversiteit), sintaktiese kenmerke (hoofletters aan die begin van ’n sin, frekwensie van

[142] Sien afdeling 5.7 (Kwantitatiewe analise).
[143] F-telling verwys na ‘n statiese konsep wanneer ‘n ANOVA toets gebruik word ten einde te bepaal of ‘n enkele veranderlike statisties beduidend verskil (Statistics, 2017).

hooflettergebruik in terme van gemiddelde woordgebruik, die gebruik van syfers in plaas van woorde, gebruik van punktuasie, gebrek aan hoofletters sowel as die gebruik van aanhalings), kenmerke eie aan Twitter (binêre kenmerke − kenmerke van ’n hertwiet – soos die gebruik van hutswoorde in verhouding met die aantal woorde wat gebruik word, aantal verwysings na ander gebruikers, gebruik van webadresse) sowel as ander kenmerkende eienskappe (frekwensie van emotikons, aantal emotikons per sin en die aantal emotikons per karakter) ’n rol gespeel het.

- Schwarts *et al. (*2013) het ondersoek ingestel na outeurskapidentifikasie van twiets deur gebruik te maak van k-handtekeninge (*k-signatures*). Hierdie k-handtekeninge verwys na die kenmerke van outeurs wat in minder as k% van ander outeurs se tekste voorkom. Wanneer die waarde van k hoog is, dui dit daarop dat die k-handtekening van ’n betrokke outeur uniek is. Volgens Schwarts *et al.* (2013:1880) is ’n k-handtekening ’n tipiese kenmerk van die meeste outeurs. Ten einde k-handtekeninge te kon ondersoek, is daar na karakter n-gramme en woord n-gramme gekyk.

Vanuit die bogenoemde blyk dit dat navorsers elk oor ’n unieke aanslag tot die identifisering van outeurs van kort tekste beskik. Die gebruik van n-gramme (*k-signatures*), stilistiese en stilometriese kenmerke is ook ’n algemene verskynsel.

5.11 Etiekkwessies

Hunston (2008:156) skryf dat alle korpora beskou kan word as ’n kompromie tussen die verwagting van navorsers, die beplande metode en dít wat wel moontlik is aangesien daar verskeie praktiese beperkinge is wat die samestelling van ’n korpus kan beïnvloed. Beperkinge sluit rekenaarprogrammatuur, kopiereg, etiekkwessies en die beskikbaarheid van tekste in. Na gelang van Hunston (2008:156) se stelling, blyk dit duidelik dat alle etiekkwessies oor data-insameling vir korpora aangespreek moet word alvorens die samestelling daarvan in aanvang kan neem. In hierdie afdeling (afdeling 5.10) spreek ek die onderskeie etiekkwessies aan. Ek verwys ook na die kwessie oor kopiereg aangesien ek die data vir my studie data vanaf ’n publieke platform ingesamel het.

Volgens Hunston (2008:157) is dit onwettig (in die meeste lande) om elektroniese kopieë van gepubliseerde tekste te stoor mits toestemming daarvoor verkry is. Ook in Suid-Afrika geld hierdie beginsel ingevolge die Suid Afrikaanse Wet op Kopiereg van 1978. Literêre, musikale en kunswerk-kategorieë word outomaties deur hierdie wetgewing beskerm (Griffith & Smalbergen, 2015).

Wanneer korpora vanuit ongepubliseerde dokumente saamgestel word, is die verkryging van kopiereg egter nie ʼn probleem nie, maar daar is steeds etiekkwessies van toepassing. Hunston (2008:157) stel dat daar in die geval van ongepubliseerde dokumente (soos opstelle of die transkripsies van gesprekke) wel toestemming verkry moet word om die dokumente in ʼn korpus te gebruik. Meyer (2002:75) sluit by bostaande aan en stel dat die vertroulikheid van deelnemers van die uiterste belang is en te alle tye verseker moet word wanneer toestemming verkry word. Ook Du Plessis (2016:73) sluit by die bogenoemde aan en stel dat die "vertroulikheid van die resultate en bevindinge van die studie, sowel as die beskerming van die identiteit van die deelnemers" van belang is.

Wanneer data vanaf RBK-platforms verkry word, is daar verskeie etiekkwessies om in gedagte te hou. Volgens Beißwenger en Storrer (2008:300) moet individue se regte in ag geneem word wanneer ʼn korpus vanaf RBK-platforms saamgestel word. Die ideaal sou wees om toestemming te verkry vanaf individue, maar die anonieme aard van hierdie platforms kan die verkryging van toestemming bemoeilik (skuilname word gebruik, gebruikers wissel en soms kan meer as een individu van ʼn bepaalde sosialemediaprofiel gebruik maak). Beißwenger en Storrer (2008:300) gaan voort deur te verwys na die waarnemersparadoks − gebruikers pas hulle taalgebruik aan indien hulle bewus is van die feit dat hulle data vir navorsing gebruik word. Tog blyk dit dat toestemming vanaf die deelnemers die aanvaarde etiese roete is om te volg (vgl. Beißwenger & Storrer, 2008:305).

Die gebruik van die web as ʼn korpus het talle nuwe vraagstukke oor kopiereg voortgebring. Wanneer daar na die kopiereg van mikroblogs (in die geval van my studie spesifiek Twitter) ondersoek ingestel word, blyk dit dat Twitter ʼn publieke platform is waar gebruikers self die beskikbaarheid van hulle profiele bepaal. Sommige gebruikers verander die privaatheidstelling van hulle profiel sodat alle twiets slegs sigbaar is aan ʼn gekose gehoor. Die internet is ʼn publieke platform, en dus bly die kwessie oor die gebruik van data asook die verkryging van toestemming ʼn etiese vraagstuk.

Navorsing wat reeds gedoen is oor etiekkwessies wat verband hou met die gebruik van data vanaf publieke aanlyn platforms toon dat alhoewel hierdie inligting geredelik beskikbaar is, individue nie noodwendig ingenome daarmee is wanneer hulle inskrywings sonder toestemming gebruik word nie (Bornoe & Barkhuus, 2011; boyd, 2007; Viégas, 2005). D'Arcy en Young (2012:536) argumenteer dat moontlike deelnemers genader moet word, die studie moet aan hulle verduidelik word en hulle toestemming moet verkry word om die toepaslike data te gebruik. Whiteman (2012:89-90) sluit hierby aan deur te postuleer dat navorsing op aanlyn platforms soos enige ander navorsing ingeligte toestemming van die navorsingsubjekte vereis.

Griffith en Smalbergen (2015) voer aan dat die outeurs van Twitter-plasings ingevolge die Suid-Afrikaanse wetgewing die eksklusiewe reg het om hulle werk (waarop kopiereg van toepassing is) te beskerm in terme van uitbuiting deur ’n derde party waar ’n derde party hulle werk kan misbruik vir persoonlike gewin. Wanneer daar dus teoreties gesproke na die reg tot die gebruik van twiets ondersoek ingestel word, kan twiets nie (indien dit wel deur kopiereg beskerm word[144]) sonder toestemming gebruik word nie. Die antwoord blyk nie so eenvoudig te wees nie.

Die beperkte lengte van twiets, en die feit dat Twitter ’n platform is waar *retweets, shares* en *comments* aan die orde van die dag is, skep bykomende probleme. In hoe ’n mate kan individue ’n “nuwe, kreatiewe” teks in slegs 140 karakters skep? Griffith en Smalberger (2015) voer aan dat dit nie onmoontlik is om kreatiewe werk in slegs 140 karakters te skep nie, maar dat daar nie ’n besliste antwoord in dié geval is nie aangesien dit verwys na ’n “question that will have to be considered on a case-by-case basis. The short answer is #itdepends #notalltweetsarecreatedequal.” Hulle (Griffith & Smalbergen, 2015) argumenteer dat aangesien twiets wel deur kopiereg beskerm kan word, dit die logiese antwoord is dat derde partye nie twiets kan kopieer sonder outeurs se toestemming nie.

Vir die doel van my studie is etiekklaring[145] vanaf die NWU verkry. Die Noordwes-Universiteit, Potchefstroomkampus, se Fakulteit Lettere en Wysbegeerte se Regulatoriese Etiekkomitee het die navorsing goedgekeur en die etiekklaringsnommer: NWU-00142-16-A8 is aan die studie toegeken. Nadat etiekklaring aan die studie verleen is, is potensiële deelnemers gekontak om toestemming tot die gebruik van hulle twiets te verkry. Ek het reeds onderskeid getref tussen twee verskillende groepe potensiële deelnemers[146] na gelang van die doel van die onderskeie korpora wat in my studie gebruik is. Vir die doel van my studie het ek twee afsonderlike korpora saamgestel. Die eerste korpus – ’n algemene korpus – se doel was as ’n verwysingskorpus te dien. Aangesien ek nie na enige individuele taalgebruik ondersoek ingestel het nie is daar besluit ná konsultasie met die relevante etiekkomitee dat ek nie die deelnemers se toestemming om die data vir hierdie gedeelte te gebruik hoef te kry nie.

Die tweede korpus – ’n gespesialiseerde korpus – is saamgestel om te bepaal of outeurskapidentifikasie wel in Afrikaans op mikroblogs uitgevoer kon word, en wat kenmerkend van die Afrikaans wat op Twitter gebruik word beskou kan word. Vir die samestelling van hierdie korpus was dit belangrik om elke deelnemer se toestemming te verkry alvorens ek die data vir analise kon gebruik. Die verskil tussen die twee korpora (en dus ook die verkryging van

[144] Dit is belangrik om in ag te neem watter soort werk deur kopiereg beskerm word – Griffith en Smalberger (2015) stel dat kopiereg in Suid Afrika ’n outomatiese reg is, en dat dit van toepassing is sodra ’n nuwe werk geskep word en in ’n tasbare vorm aangebied word (daar is nie kopiereg op idees nie). Die term “werk” verwys volgens Griffith en Smalberger (2015) na ’n versameling van woorde, letters of getalle soos in die geval van ’n artikel, brief, roman of gedig en dat dit dus na ’n literêre werk kan verwys.

[145] Sien Bylaag A (Etiekklaring)

[146] Sien afdeling 5.3 (Deelnemers) Groep 1 en Groep 2.

toestemming) is dat die algemene korpus slegs gebruik is om stilistiese kenmerke van Afrikaans op ’n mikroblog te identifiseer en daar dus glad nie na die individue se taalkenmerke verwys is nie. Die gespesialiseerde korpus se doel was om idiolektiese kenmerke van die individue te identifiseer en daarom was die potensiële deelnemers se toestemming vir die gebruik van hulle data noodsaaklik.

Die anonieme aard van Twitter[147] het die verkryging van kontakbesonderhede van die potensiële deelnemers gestrem. Ek moes elke potensiële deelnemer deur Twitter kontak. ’n **d**irekte **b**oodskap (***D**irect **M**essage*) moes aan elke potensiële deelnemers gestuur word waarin ek die beginsels van die studie verduidelik en gevra het of hulle sou belangstel om deel te neem aan die studie.

Indien potensiële deelnemers wel belanggestel en ingestem het om hulle e-posbesonderhede aan my te stuur, is die toestemmingsbrief[148] aan hulle gestuur. Eers nadat toestemmingsbriewe onderteken is, kon ek hulle data in die korpus byvoeg. Om die deelnemers se anonimiteit te verseker, is geen persoonlike inligting in die studie gebruik nie. Die deelnemers se identiteit is weerhou. Twitter-gebruikersname sowel as foto’s is dus verwyder.

5.12 Samestelling van die korpora

Hierdie studie is ŉ eksploratiewe forensiese ondersoek na outeurskapidentifikasie in ŉ digitale era. Vanuit die bogenoemde stelling is dit duidelik dat daar na tekste wat eie aan die digitale era is, ondersoek ingestel sal word. Ten einde hierdie ondersoek uit te voer sal ek dus korpora saamstel vanaf ŉ mikroblogplatform (Twitter). In hierdie afdeling gaan ek dus die wyse waarop die deelnemers vir my studie gekies is, dataversameling, sowel as die verwerking van die data om dit gereed te maak om in die onderskeie rekenaarprogrammatuur te gebruik, voorhou.

5.12.1 Deelnemers

Soos reeds genoem[149], het hierdie studie ten doel het om eksploratief ondersoek in te stel na die identifisering van die outeur van ŉ mikroblogplatform. Daar is dus eerstens na bestaande korpora in Afrikaans ondersoek ingestel. Daar is bevind dat daar tot op hede nie ŉ korpus vir Afrikaans bestaan wat mikrobloginskrywings insluit nie. Die gebrek aan ŉ korpus wat Afrikaanse mikrobloginskrywings insluit, kan dus as ŉ wesenlike probleem in hierdie studie beskou word.

[147] Die anonieme aard van Twitter verwys na die manier waarop individue van gebruikersname gebruik maak wat as ŉ skuilnaam beskou kan word, en dat daar nie noodwendig enige kontakbesonderhede is waardeur die betrokke individue gekontak/geïdentifiseer kon word nie.
[148] Sien Bylaag B (Toestemmingsbrief).
[149] Sien hoofstuk 1 (afdeling 1.1 - Kontekstualisering, afdeling 1.3 – Navorsingsvrae - en afdeling 1.4 – Navorsingsdoelwitte).

Ten einde die bogenoemde probleem rakende die verwysingskorpus op te los, het ek ŉ korpus saamgestel wat vir die doel van my studie as verwysingskorpus kan dien. Dus het ek beide ŉ verwysingskorpus (groep 1), en ŉ gespesialiseerde korpus (groep 2) saamgestel ten einde die navorsingsvrae te kan beantwoord. In hoofstuk 5 (5.3 Deelnemers) het ek die verskillende wyses waarop die deelnemers vir die onderskeie korpora gekies word, voorgehou. Aangesien die doel van die korpora verskil, verskil die wyse waarop die data vir die onderskeie korpora ingesamel word.

Die verwysingskorpus (groep 1) se deelnemers moet verteenwoordigend van alle strata[150] op Twitter wees, en daarom is daar van ŉ kombinasie van die *nieproporsionele gestratifiseerde steekproefneming* (ten einde die gedifferensieerde kategorieë[151] op Twitter uit te beeld) en die *sneeubal-steekproefneming* (om die moontlike respondente te identifiseer) gebruik maak. Aangesien hierdie korpus nie gebruik word om enige individuele taalgebruik te ondersoek nie, is daar besluit, in konsultasie met die relevante etiekkomitee, dat ek nie die deelnemers se toestemming om die data vir hierdie gedeelte te gebruik hoef te kry nie[152].

Ten einde die gespesialiseerde korpus (groep 2) saam te stel, is daar van die *doelgerigte steekproefneming* sowel as die *sneeubal-steekproefneming* gebruik gemaak. Soos reeds in hoofstuk 5 genoem is daar vereistes daargestel ten einde te verseker dat die data van die individue waaruit groep 2 bestaan voldoende is. Hierdie vereistes het ingesluit;

- Twiets moet hoofsaaklik in Afrikaans wees
- Die Twitter-profiel moet aktief wees
- Die gebruiker moet die vorm vir ingeligte toestemming (etiekvorm)[153] onderteken en dus ingeligte toestemming verleen dat sy/haar data in my studie gebruik mag word.

Ek het dus potensiële deelnemers geïdentifiseer en hul data vanaf Twitter gekopieer. Daar is besluit om die data eers vanaf Twitter te kopieer, alvorens die potensiële deelnemers die etiekvorm onderteken het, om sodoende die moontlikheid van ŉ waarnemersparadoks uit te skakel. Indien die potensiële deelnemer nie die etiekvorm onderteken het nie, is hul data eenvoudig by die verwysingskorpus gevoeg. Die eerste tien deelnemers wat die etiekvorm onderteken het, is gebruik om die gespesialiseerde korpus saam te stel.

[150] Sien ook afdeling 5.4 (Data-insameling en -verwerking) waar die vier kategorieë (media, vermaaklikheidspersoonlikhede, organisasies en persoonlik) van Twitter voorgehou en verduidelik word.
[151] Die verskillende strata wat ondersoek word, word in afdeling 5.4 (Data-insameling en -verwerking) voorgehou.
[152] Aangesien daar nie enige voorbeelde vanuit die verwysingskorpus aangehaal word nie, end at die korpus slegs ten doel het om die voorkoms van sekere kenmerke te verifieer, was dit nie nodig om ingeligte toestemming vir die gebruik van die twiets te kry nie.
[153] Sien afdeling 5.11 (Etiekkwessies), sowel as bylaag A en B.

Ten einde die potensiële deelnemers se toestemming te kry om hulle in die studie te gebruik, het ek, waar moontlik, e-posse aan die individue gestuur. Daar was verskeie deelnemers wat ek nie deur middel van e-pos kon kontak nie. Wanneer ek nie kontakbesonderhede van die deelnemers tot my beskikking gehad het nie, het ek hulle deur middel van (ander) sosiale media (Facebook, Twitter en selfs Instagram) gekontak en gevra of hulle bereid sou wees om hul e-posadres aan my te verskaf. Reeds hier was daar enkele potensiële deelnemers wat nie op my boodskappe geantwoord het nie. Indien die potensiële deelnemer wel ʼn e-posadres verskaf het, het ek ʼn e-pos gestuur waarin ek die doel van my studie verduidelik en vra of hulle wel bereid sou wees om deel te neem daaraan. Vanuit die 25 potensiële deelnemers wat ek genader het, het tien ingestem om deel te neem aan my studie. Die korpora wat dus vir die studie gebruik word, kan soos volg opgesom word:

Tabel 5.4 Uiteensetting van korpora

Groep 1 (verwysingskorpus)	Die korpus bestaan uit 246 908 woorde en is saamgestel vanaf Twitter-profiele wat al die genoemde strata (media, vermaaklikheidspersoonlikhede, organisasies en persoonlik) insluit.
Groep 2 (gespesialiseerde korpus)	10 deelnemers, die grootte van elke individuele korpus wissel na gelang van die aantal twiets waaroor die betrokke individue beskik. Deelnemers is gekies na gelang van die vereistes soos genoem, en hul terugvoer op die etiekvorms.

Tabel 5.4 gee ʼn opsomming van die verskillende korpora wat in die studie gebruik word. Dit is belangrik om in gedagte te hou dat die anonieme aard van sosiale media, die gebruikers die geleentheid gee om hul identiteit te verskans. Daarom is dit nie moontlik om met sekerheid te sê in watter ouderdomsgroep my deelnemers val nie, verder kan daar ook nie met sekerheid gesê word wat die ras of geslag van die deelnemers is nie, alhoewel daar sekere afleidings met betrekking tot ouderdom, geslag en ras gemaak kan word na gelang van die individue se twiets.

Ten opsigte van die gespesialiseerde korpora het ek gepoog om so ver as moontlik ʼn omvattende korpus, wat as verteenwoordigend van die wyer Afrikaanse Twitter-populasie beskou kan word, saam te stel. Ek het talle potensiële deelnemers gekontak om toestemming tot die gebruik van hul Twitter-profiele te verkry, ongelukkig het slegs wit deelnemers die etiekvorms onderteken en teruggestuur[154]. Die samestelling van die gespesialiseerde korpus word in Tabel 5.5 voorgehou.

[154] Sien hoofstuk 8 (Afdeling 8.5 - Aanbevelings vir verdere navorsing).

Tabel 5.5 Uiteensetting van die gespesialiseerde korpus[155].

Deelnemer	Aantal twiets	"Followers[156]"	"Following[157]"	"Likes[158]"	Foto's en video's
O1	7 479	5 729	70	3 881	547
O2	739	428	73	42	4
O3	3 431	1 034	1 994	21	439
O4	1 488	15 756	324	2 278	96
O5	602	283	62	1 137	14
O6	3 283	1 978	420	5	26
O7	1 530 000	47 456	595	59	88
O8	2 140 000	16 949	1 772	4 737	2 326
O9	895	270	136	1 721	161
O10	2 397	605	1 141	795	240

Vanuit Tabel 5.5 blyk dit dat die deelnemers verskil ten opsigte van aantal twiets, hou van, volgelinge en volg. In die hier opeenvolgende afdeling gaan ek na die dataverwerking van die korpora verwys.

5.12.2 Dataverwerking

Soos genoem het ek twee verskillende tipes korpora saamgestel: groep 1, die verwysingskorpus, en groep 2 die gespesialiseerde korpus. Die gespesialiseerde korpus bestaan uit tien verskillende deelnemers. In hierdie afdeling gaan ek eerstens na die verwerking van die gespesialiseerde korpus verwys, alvorens die verwysingskorpus se verwerking voorgehou sal word. In afdeling 5.12.1 het ek na die wyse waarop die deelnemers gekies is, en die proses waardeur toestemming verkry is om die data te gebruik verwys. Nadat die data vanaf Twitter gekopieer is, is die rou data aanvanklik in word docx. gestoor, soos uitgebeeld in Skermskoot 5.12 en 5.13.

[155] Die volgelinge vir die onderskeie korpora is op 24 Oktober 2016 nagegaan, maar aangesien Twitter verander mag die getalle nou anders daaruitsien.
[156] Vervolgens sal daar na "followers" as volgers verwys word.
[157] Vervolgens sal daar na "following" as volg verwys word.
[158] "Likes" word vervolgens "hou van".

17h17 hours ago

I've forgotten how hard it is to wake up at 5 in the morning on purpose

1 retweet2 likes

May 18

I am Afrikaans and creative, but I'm not "will one day own my own craft winkel" Afrikaans and creative

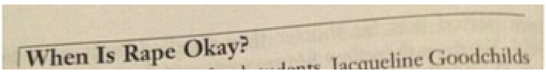
7 retweets16 likes

1d1 day ago

Ek het road rage, maar ek ry ʼn poegie. Dit is waarom ek sterflik is.

2. 1d1 day ago

Wát is fout met die mensdom?



Skermskoot 5.12 Voorstelling van ʼn deelnemer se rou data vanaf Twitter

👏👏Agree well done @c[redacted]

added,

[redacted]

Impressive run by Caster in Monaco. New PB. Rio is hers to lose.

0 retweets1 like

George Germishuys @GeorgeLezer Jul 15

[redacted]

#Saru baklei kaalvuis terug teen "lasterlike aantygings" deur oud-speler //

1 retweet0 likes

Verskoning aanvaar! 😉😉

Skermskoot 5.13 Voorstelling van onverwerkte twiets

Soos gesien in Skermskoot 5.12 en 5.13 moet die rou data wat vanaf Twitter ingesamel is, verwerk word. Die eerste stap in die verwerkingsproses was om alle grafika (foto's en dies

meer) te verwyder. Volgende is alle Engelse twiets verwyder. Hierdie Engelse twiets is in ’n afsonderlike dokument gestoor ten einde te kan bepaal hoeveel Engelse twiets elke individu gebruik in verhouding met die Afrikaanse twiets. Daar is verder ook onderskeid getref tussen die twiets wat hertwiets[159] is, en twiets wat deur die outeur self geskryf is (beide in Afrikaans en Engels). Hierdie stap is uitgevoer ten einde die ratio van die eie twiets en hertwiets te bepaal. Vir die doel van my studie is slegs die eie Afrikaanse twiets geanaliseer. Daar is vir elke deelnemer ’n Microsoft Excel-lêer (soos gesien in Tabel 5.6) geskep waarin hierdie data gestoor is, en wat later in die stilistiese en stilometriese analise gebruik is.

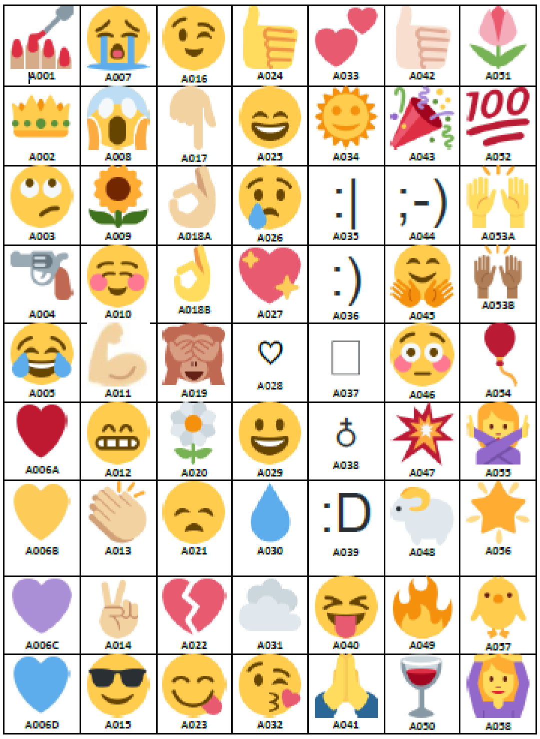
Tabel 5.6 Uitbeelding van die Microsoft Excel-lêers waarin die deelnemers se inligting gestoor is

Twiets	7479
Eie Twiets	721
Hertwiets	175
Foto/video	547
Volgelinge	5729
Volgelinge	70
Woorde	10017
Antwoord op hertwiet	76
Hou van	3881
Aanvangsdatum	2013 Desember
Twiets	721
Afrikaans	669
Engels	53
Hertwiet	175
Afrikaans	120
Engels	55
Emotikons	103
Karakters	68,23
Woorde	13,17
Woordlengte	3,98
Emotikons per twiet	0,15
Gebruiker per twiet	0,035
Hutstekens per twiet	0,035
Woorde	8896
Twiets	670

Die volgende stap was om alle emotikons te kodeer sodat die teks na ’n txt-formaat omgeskakel kon word. Ten einde die emotikons ook te kan analiseer, het ek ’n kode aan elke emotikon toegeken, en die emotikon met hierdie kode vervang (sien Skermskoot 5.14), aangesien

[159] ’n Hertwiet verwys na ’n twiet, (wat nie deur die gebruiker op wie se profiel dit sigbaar is nie, geskryf is. Die gebruiker deel dus iemand anders se twiet met sy volgelinge, en lewer dikwels kommentaar daarop.

WordSmith Tools slegs txt-dokumente lees[160]. Aangesien ek ook die aantal karakters moet tel wat die deelnemers per twiet gebruik, het ek ŉ aparte dokument vir elke deelnemer geskep, waar ek die emotikons met ŉ getaletiket vervang het, aangesien emotikons slegs een karakter opneem en die kodes wat ek aan die emotikons toegeken het langer as ŉ enkele karakter was.



Skermskoot 5.14 Voorstelling van die lys van kodes waarmee die emotikons vervang is.

Ek het na afloop van die verwerkingsproses vir elke deelnemer vyf Microsoft Word docx-dokumente gehad (die rou data, die verwerkte Afrikaanse twiets waar die emotikons met kodes vervang is, sowel as ŉ tweede dokument waar die emotikons met ŉ syfer vervang is, die verwerkte Afrikaanse hertwiets – ook hier is die emotikons met kodes vervang, die deelnemer se eie Engelse twiets, sowel as die Engelse hertwiets). Die verwerkte dokument met die deelnemer se eie Afrikaanse twiets (emotikons met kode vervang) is ook as ŉ txt-dokument gestoor ten einde die data in *WordSmith Tools* te kan analiseer.

Die tien deelnemers se hertwiets is by die verwysingskorpus gevoeg, aangesien ek slegs die deelnemers se eie twiets wil analiseer, ten einde te bepaal of die outeur van ŉ bepaalde mikrobloginskrywing wel geïdentifiseer kan word. Die verwysingskorpus se data is ook verwerk, maar anders as in die geval van die gespesialiseerde korpus het ek slegs die Engelse twiets en grafika verwyder, aangesien die verwysingskorpus net ten doel het om by te dra tot die

[160] Sien bylaag H vir ŉ lys van die emotikon kodes.

ondersoek na die taalstilistiese en stylkenmerke van Afrikaans soos dit op mikroblogs gebruik word. In hierdie afdeling is die wyse waarop die data verwerk is verduidelik, vervolgens gaan ek die onderskeie datastelle verduidelik.

5.12.3 Datastelle

Soos in hoofstuk 1 en 2 bespreek, is Twitter ŉ sosialemediaplatform waar gebruikers boodskappe kan plaas wat maksimum 140 karakters kan bevat. Dus is die boodskappe kort, en die aantal teks wat tot die navorsers se beskikking is uiters beperk. Aangesien ek in hierdie studie onderneem om te bewys dat die outeur van mikrobloginskrywing wel geïdentifiseer kan word, moet die kenmerke en beperkinge van die betrokke platform in ag geneem word, ten einde ŉ denkbeeldige scenario te skep wat so na as moontlik aan die werklikheid is. In die vorige afdelings is die wyse waarop die deelnemers vir die studie gekies is, sowel as die wyse waarop die data verwerk is voorgehou. In hierdie afdeling gaan ek vervolgens die onderskeie datastelle uiteensit.

5.12.3.1 Die verwysingskorpus

Die verwysingskorpus[161] is saamgestel met die doel dat dit verteenwoordigend van Afrikaans soos dit op Twitter gebruik word, beskou kan word. Dus is twiets vanaf elk van die kategorieë[162] (media, vermaaklikheidspersoonlikhede, organisasies en persoonlik) ingesamel om deel te vorm van hierdie korpus. Tabel 5.7 is ’n uiteensetting van die verwysingskorpus.

Tabel 5.7 Uiteensetting van die verwysingskorpus, soos weergegee in *WordSmith Tools*

Twiets	19 499
Woorde in teks	246 908
Tipe/teken-ratio	9.84
Gestandaardiseerde tipe/teken-ratio	40.74
Gestandaardiseerde tipe/teken-ratio standaardafwyking	59.00
Gemiddelde woordlengte (karakters)	4.99
Woord lengte standaardafwyking	3.78

Die verwysingskorpus sal in die studie gebruik word om, tesame met die gespesialiseerde korpus, na die taalstilistiese en stylkenmerke van Afrikaans op Twitter ondersoek in te stel. Verder sal die verwysingskorpus gebruik word om vreemde woorde en bevindinge vanuit die gespesialiseerde korpora te verifieer.

[161] Soos in afdeling 4.2.2 genoem word hierdie verwysingskorpus as ’n prototipe korpus vir die taal waarin die ondersoek plaasvind beskou.
[162] Sien ook afdeling 5.4 (Data-insameling en -verwerking).

5.12.3.2 Die gespesialiseerde korpus

Die gespesialiseerde korpus bestaan uit tien individuele datastelle. In afdeling 5.12.3.1 is daar verwys na die wyse waarop die deelnemers gekies is, en hoe die datastelle verwerk is. Vervolgens gaan ek die datastelle van elk van die tien individuele deelnemers uiteensit. Hierdie uiteensetting het ten doel om die samestelling van elk van die datastelle voor te hou, en die nodige agtergrond te skep vir die stilistiese en stilometriese analise wat in hoofstukke ses en sewe sal plaasvind.

Ten einde die deelnemers se identiteit te beskerm is daar aan elke deelnemer ’n skuilnaam toegeken. Die gespesialiseerde korpus bestaan dus uit tien deelnemers, en word vervolgens soos volg geïdentifiseer; O1, O2, O3, O4, O5, O6, O7, O8, O9 en O10. Tabel 5.7 is ’n opsomming van die data soos ingesamel vir die gespesialiseerde korpus terwyl Tabel 5.8 tot 5.9 die uiteensetting van elk van die deelnemers in die gespesialiseerde korpus aandui.

Tabel 5.7 Opsomming van die gespesialiseerde korpus

Deelnemer	Aantal twiets (Afrikaans en Engels)	Totale twiets Afrikaans	Totale aantal woorde in teks	Tipe/teken-ratio	Gestandardi-seerde tipe/teken-ratio	Gestandardi-seerde tipe/teken-ratio standaardafwyking	Gemiddelde woord lengte (karakters)	Woordlengte standaard-afwyking
O1	896	669	8 942	21,58	73,80	26,07	3,98	2,33
O2	667	589	12 259	26,99	80,43	19,87	4,76	3,02
O3	834	206	3 379	34,49	67,91	29,50	4,90	2,84
O4	900	484	7 110	26,84	77,23	22,74	4,41	2,60
O5	573	333	4 579	23,00	68,58	31,04	3,76	2,03
O6	872	768	16 340	17,98	67,44	31,85	4,15	2,50
O7	1 058	457	5 655	27,78	75,36	24,25	4,02	2,25
O8	995	308	3 690	39,19	74,72	24,16	5,10	3,06
O9	559	275	4 918	30,60	74,90	24,03	4,04	2,39
O10	856	499	7 744	19,62	58,44	40,99	4,99	3,15

Tabel 5.7 gee ’n oorsig van die samestelling van die gespesialiseerde korpus. In hierdie tabel word die totale hoeveelheid twiets (Afrikaans en Engels) aangedui, sowel as die aantal Afrikaanse twiets wat na die verwerking van die data in die samestelling van die korpus gebruik is. Die aantal twiets, die woorde in die teks sowel as onder andere die gemiddelde karakter lengte word hierin aangedui. Tabel 5.8 is uitbeeldings van die individue se datastelle. Die inligting wat hierin weergegee word, word ook in die stilistiese en stilometriese analise gebruik.

Tabel 5.8 Opsomming van die deelnemers se datastelle

Deelnemer	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Totale twiets	896	667	834	900	573	872	1 052	995	559	856
Afrikaanse twiets	669	589	206	484	333	768	457	308	275	449
Engelse twiets	53	38	58	22	13	54	4	54	132	136
Afrikaanse hertwiet	120	4	101	233	185	44	432	483	26	92
Engelse hertwiet	55	36	469	161	42	7	169	150	130	129
Ratio Afrikaanse twiets	92,6	94	78	95,7	96,2	93,4	99,1	85,1	68,7	78,58
Ratio Engelse twiets	7,4	6	22	4,3	3,8	6,6	0,9	14,9	31,3	21,41
Ratio eie twiets	80,46	94	31,65	79,66	60,38	94,26	84,02	79,49	71,7	74,18
Ratio hertwiet	19,53	6	68,34	20,33	39,61	5,73	15,9	20,5	28,26	25,81
Gemiddelde woorde per twiet	13	20,22	15,3	14,14	13,89	18,5	12,1	11,02	17,43	14,53
Gemiddelde karakters per twiet	68	123,16	99,02	82,94	67,78	116,35	62,29	77,33	92,2	96,5

Tabel 5.8 dui die wyse aan waarop die gespesialiseerde korpus saamgestel is, en hoe elke afsonderlike datastel daaruit sien. Die datastelle, soos uiteengesit in tabel 5.7 en 5.8, wissel in grootte aangesien ek slegs die beskikbare twiets op elke deelnemer se Twitterprofiel kon kopieer, en aangesien hierdie twiets (soos uitgebeeld in tabel 5.7) nog verwerk moes word. Nadat die hertwiets en die Engelse twiets verwyder is, was daar ŉ totaal van 4 588 twiets in die gespesialiseerde korpus. Die kleinste datastel (O3) bestaan uit 209 twiets, terwyl die grootste datastel (O6) 768 twiets bevat. Hierdie verskil in die grootte van die korpora is nie ŉ probleem nie, aangesien die doel is om te bepaal of die outeur van ŉ mikrobloginskrywing suksesvol geïdentifiseer kan word. Soos reeds genoem, is mikroblogs kort inskrywings, en sal die data vir so ŉ ondersoek (indien dit in werklikheid uitgevoer moet word) redelik beperk wees. Ten einde te verseker dat die verskil in die grootte van die korpus nie die resultate benadeel nie, word die frekwensie van die voorkoms van die kenmerke wat ek ondersoek genormaliseer. Die onderstaande formule sal telkens gebruik word om die data te normaliseer:

$$\frac{Totale\ kere\ wat\ woord\ (kenmerk)\ in\ korpus\ voorkom}{Totale\ woorde\ in\ korpus} \times 1\ 000 = Frekwensie$$

Aangesien ek in hierdie studie ondersoek instel na die moontlikheid om outeurs op mikroblogs te identifiseer, het ek verder ook elke individuele korpus in twee gedeel. Hierdie stap is uitgevoer ten einde ondersoek in te stel na die mate waartoe die individuele outeur se eie tekste met mekaar ooreenstem, of verskil. Tabel 5.9 is ŉ opsomming van hierdie individuele verdeling van die korpora – of te wel die verdeling van die deelnemerkorpus.

Tabel 5.9 Opsomming van die individuele deelnemerkorpus

Deelnemers	Twiets	Gemiddelde woorde per twiet	Gemiddelde karakters per twiet
O1	699	13	68
O1_1	350	13,44	68,99
O1_2	349	12,83	66,64
O2	589	20,21	123,16
O2_1	298	20,25	124,58
O2_2	291	20,31	121,58
O3	206	15,3	99,02
O3_1	107	14,65	95,32
O3_2	99	16,06	102,77
O4	484	14,14	82,94
O4_1	243	13,6	14,73
O4_2	241	14,73	89,73
O5	333	13,89	67,78
O5_1	164	13,55	66,13
O5_2	169	14,02	69
O6_1	387	18,56	116,41
O6_2	386	18,41	115,84
O7	457	12,1	62,29
O7_1	233	12,27	62,77
O7_2	224	12,02	62,07
O8	308	11,02	77,33
O8_1	151	10,78	75,09
O8_2	157	11,31	79,91
O9	275	17,43	92,21
O9_1	147	16,65	89,19
O9_2	128	18,62	95,79
O10	499	14,53	96,51
O10_1	235	15,49	103,23
O10_2	264	13,81	90,77

Tabel 5.9 dui die aantal twiets per deelnemerkorpus, sowel as die sub-korpus wat vir elke deelnemer saamgestel is aan, die gemiddelde woorde per twiet, sowel as die gemiddelde karakters per twiet word voorgehou. Hierdie uiteensetting is ŉ aanduiding van hoe die individuele datastelle, wat verdeel is ten einde die ooreenkomste en verskille in ŉ enkele deelnemer se tekste te ondersoek, daar uitsien. Die deelnemerkorpora is verdeel ten einde te bepaal of daar ooreenkomste waargeneem kan word wanneer die individue se eie tekste met mekaar vergelyk word. Vanuit Tabel 5.9 is dit duidelik dat die onderskeie deelnemerkorpora

(wanneer dit intern met mekaar vergelyk word) sterk ooreenkomste toon met betrekking tot die gemiddelde woorde per twiet, sowel as die gemiddelde karakters per twiets. Aangesien daar nie een korpus[163] is wat ooreenstem in terme van die gemiddelde karakters per twiet of die gemiddelde woorde per twiet nie, kan die gevolgtrekking gemaak word dat hierdie eienskappe van elke outeur as 'n aanduiding van outeurskap beskou kan word.

Die data, soos uiteengesit in Tabel 5.9, sal voorts in beide die stilistiese en stilometriese analise gebruik word. Elke deelnemer se tekste word dus eerstens intern met mekaar vergelyk, alvorens die onderskeie deelnemer se tekste met mekaar vergelyk word. Nadat die interne vergelykings uitgevoer is, sal die eksterne vergelykings uitgevoer word.

In hierdie afdeling is daar verwys na die wyse waarop die deelnemers gekies is, hoe die data ingesamel is, die dataverwerking en hoe die datastelle daaruitsien. Hierdie inligting het ten doel om die basis te skep waarvan die stilistiese en stilometriese analise vervolgens uitgevoer gaan word. Aangesien ek in my studie ondersoek instel na die identifisering van die outeur van 'n mikrobloginskrywing gaan ek van verdagte tekste gebruik moet maak, ten einde te bepaal of die metode van outeurskapidentifikasie wel geslaagd is.

5.12.3.3 Verdagte tekste

Ten einde 'n scenario te skep wat soortgelyk aan 'n situasie is waar die outeur van 'n mikrobloginskrywing geïdentifiseer moet word, is daar van drie van die deelnemers 'n ekstra datastel elk geskep. Hierdie data is op dieselfde wyse ingesamel en verwerk as verduidelik in afdeling 5.12.3.1 en 5.12.3.2. Die twiets is ingesamel vanaf die datum van die laaste twiet wat gekopieer is vir die aanvanklike datastelle, te einde te verseker dat die twiets nie oorvleuel nie. Die drie deelnemers wie se twiets gekies is om die verdagte tekste saam te stel, is lukraak gekies. Tabel 5.10 toon die uiteensetting van die verdagte tekste aan.

Tabel 5.10 Uiteensetting van die Afrikaanse twiets vanuit die verdagte tekste wat in die studie gebruik word.

Deelnemer	Verdagte teks 1	Verdagte teks 2	Verdagte teks 3
Aantal twiets	62	120	60
Totale eie Afrikaanse twiets	42	91	32
Totale eie Engelse twiets	2	5	3
Afrikaanse hertwiets	5	11	22
Engelse hertwiets	13	13	3
Ratio Afrikaanse twiets	95,45	94,79	91,42
Ratio Engelse twiets	4,45	5,20	8,57

[163] Hier word spesifiek verwys na die 10 individuele datastelle.

Deelnemer	Verdagte teks 1	Verdagte teks 2	Verdagte teks 3
Ratio van eie twiets	70,96	80,00	58,33
Ratio van hertwiets	29,03	20,00	41,66
Gemiddelde woorde per twiet	20,07	14,91	13,25
Gemiddelde karakters per twiet	121,21	76,55	63,12
Woorde in teks	581	1440	408
Tipe/teken-ratio	53,17	38,80	45,32
Gemiddelde woord lengte (karakters)	4,88	4,01	3,70
Woord lengte standaardafwyking	3,59	2,30	1,96

Soos gesien in Tabel 5.10 is die tekste wat die korpora van die verdagte tekste opmaak, baie kleiner as die tekste wat die gespesialiseerde korpus uitmaak. Die grootste datastel is slegs 120 twiets lank, maar aangesien mikroblogs van nature na korter kommunikasie verwys, is dit nodig om te bepaal of die outeur van so ʼn kort teks wel met ʼn beperkte aantal data geïdentifiseer kan word.

In hierdie afdeling is daar verwys na die wyse waarop die deelnemers vir die studie gekies is, die verwerking van die datastelle, die uiteensetting van die onderskeie korpora, sowel as die wyse waarop verdagte tekste saamgestel is. Soos reeds in hoofstuk 5 bespreek, maak ek gebruik van ʼn gemengde metode in my studie. In hoofstukke ses en sewe word die kwantitatiewe en kwalitatiewe analises uiteengesit, die metodes wat gevolg is, word verduidelik en die bevindinge van die ondersoek word weergegee.

5.13 Samevatting

In hierdie hoofstuk is daar verwys na die manier waarop die data ingesamel is, die verskillende navorsingsmetodes en belangrike statistiese konsepte is bespreek, en ten slotte is die data insameling en verwerking voorgehou. Daar is na die rekenaarprogrammatuur verwys wat gebruik is ten einde die studie uit te voer, en daar is ook na navorsing oor outeurskapidentifikasie en Twitter verwys. ʼn Oorsig is verskaf oor die etiekkwessies verbonde aan ʼn studie van taalgebruik soos dit gebruik word op sosialemediaplatforms. Die bydrae van hierdie hoofstuk is dus nie net om die metode wat in my studie gevolg is voor te hou nie, maar ook om te onderskei tussen verskillende metodes wat gevolg kan word.

Aangesien die navorsing rakende tekste op sosiale media (in Afrikaans) nog redelik beperk is, ka die hantering van die etiese kwessies, die manier waarop die korpus saamgestel kan word, en die prosseseering van die data veral as belangrike bydraes beskou word.

Ek het aangetoon dat die beste metode vir my studie ’n gemengde metode was. Ek het eerstens na die stilistiese analise verwys alvorens ek ’n stilometriese analise voorgehou het. Hierdie twee analises blyk vanuit afdeling 5.6 en 5.7 beste analises te wees om te te as ten einde die navorsingsvrae te kon beantwoord. Vanuit die literatuurstudie (hoofstuk 2, 3, 4 en ook 5) is daar verskeie kenmerke geïdentifiseer wat gebruik kon word ten einde die taalstilistiese kenmerke sowel as die outeurskapidentifikasie in hierdie studie te kon uitvoer.

Hierdie kenmerke sluit in: karakterfrekwensie; frekwensie van woordlengte; sinslengte; woordsoortfrekwensie; leksikale kenmerke; karakter, sintaktiese en semantiese kenmerke; en die gebruik van emotikons. Deur die bogenoemde kenmerke te analiseer, sal daar eerstens ’n raamwerk geskep word waardeur die taalstilistiese kenmerke van Afrikaans uitgebeeld word soos dit op Twitter gebruik is alvorens die outeurskapidentifikasie kon plaasvind.

Hoofstuk 1, 2, 3, 4 en 5 se doel is om die agtergrond en konteks te skep waarteen die empiriese analise vervolgens in hoofstuk 6 en 7 uitgevoer is.

Hoofstuk 6

EMPIRIESE ANALISE: STILOMETRIES

While it is possibly true that mistakes made by authorship analysts in the field of literature could lead to red faces and bad press at worst, the same cannot be said of the forensic context, where mistakes could lead to imprisonment or even execution in some countries. The importance of extreme caution before arriving at conclusions can therefore not be overemphasised.

(Kotzé, 2010:186)

6.1 Inleiding

In hierdie hoofstuk word die kwantitatiewe korpusontleding bespreek ten einde die navorsingsvrae (soos gestel in hoofstuk 1) te kon beantwoord. Die navorsingsvrae sluit in:

- Wat is die taalstilistiese en stylkenmerke van mikroblogs in Afrikaans?
- Kan die outeur van ’n mikrobloginskrywing suksesvol geïdentifiseer word?
- Wat is die forensiese beginsels wat gebruik kan word om outeurskapidentifikasie op mikroblogs toe te pas?

Die voorafgaande hoofstukke het ten doel om die teoretiese agtergrond en konteks te skep waarteen die empiriese analise in hierdie hoofstuk plaasgevind het. Die metodes van ’n kwantitatiewe analise word in hoofstuk 5[164] bespreek en verwys na die insameling van numeriese data om ’n spesifieke vraag te beantwoord of om ’n numeriese waarde[165] aan ’n entiteit wat voorheen slegs teks was toe te ken. Wanneer ’n kwantitatiewe analise gebruik word, word daar gepoog om ’n hipotese te bewys. Rasinger (2010:52) voer aan dat ’n hipotese verwys na ’n verklaring oor ’n potensiële of voorgestelde verhouding tussen veranderlikes. In die geval van my studie was die hipotese wat bewys moes word die veronderstelling dat bestaande forensiese linguistiek beginsels en tegnieke van outeurskapidentifikasie aangewend kan word om die outeurs van Afrikaanse mikrobloginskrywings te identifiseer. Wanneer kwantitatiewe metodologie gebruik word met betrekking tot forensiese linguistiek (outeurskapidentifikasie), word daar van ’n stilometriese benadering gebruik gemaak.

---

[164] Sien afdeling 5.7 (Kwantitatiewe analise).
[165] Kotzé (2007:388) stel dat die tekstuele data gekwantifiseer word.

Kotzé (2007:388) sluit die volgende by ŉ kwantitatiewe (stilometriese) analise in: ŉ ondersoek na die frekwensie en verspreiding van woordtipes; die gesamentlike voorkoms van woorde; variasie in die gebruik van funksiewoorde; en die identifisering van sleutelwoorde. Rocha *et al.* (2017:6) verwys na ŉ stilometriese analise as die identifisering van moontlike outeurs van tekste deur gebruik te maak van styleienskappe wat in dié tekste ondersoek word. Hierdie styleienskappe[166] verwys volgens Krause (2013), onder andere, na sinslengte, woordeskatrykheid en die frekwensie van woorde of woordlengtes as kenmerke wat eie aan ŉ stilometriese analise beskou word.

Kotzé (2007:309) voer aan dat alhoewel die gebruik van grammatiese woorde (funksiewoorde) by sinsvorming hoofsaaklik deur die grammatika van die betrokke taal bepaal word, het sprekers (outeurs) steeds ŉ verskeidenheid van keuses ("alternatiewe strukture om dieselfde of soortgelyke inhoude") om iets oor te dra. Kotzé (2007:309) beklemtoon dat dit juis hierdie "verskeidenheid" van maniere is waarop ŉ idee oorgedra kan word waardeur outeurs se onderbewuste keuses uitgebeeld word en sodoende ŉ belangrike rol speel in outeurskapidentifikasie.

ŉ Stilometriese analise verwys dus na die numeriese analise van linguistiese kenmerke. ŉ Stilometriese analise waar die tipiese strukture van die taalgebruik ondersoek word, sal volgens Kotzé (2007:309) eerstens ŉ ondersoek behels na die funksiewoorde wat in die onderskeie tekste gebruik word. Rocha *et al.* (2017:6-7) argumenteer dat wanneer die internet[167] as platform vir die identifisering van outeurs gebruik word, daar verskeie faktore is wat ŉ stilometriese analise kan kompliseer, insluitend (maar nie beperk daartoe nie) onreëlmatige taalgebruik, oorbeklemtoonde leestekengebruik, emotikons en die algemene voorkoms van spelfoute.

Alhoewel bogenoemde faktore op die oog af as bykomende stylmerkers beskou kan word, stel Rocha *et al.* (2017:7) tereg dat dié faktore ook ŉ uitdaging aan navorsers stel aangesien die genoemde faktore aanhoudend verander (nuwe emotikons, afkortings en dies meer word gereeld in gebruik geneem); daarom moet daar probeer word om stilometriese eienskappe te vind wat nie net die diversiteit van die taal op ŉ spesifieke platform identifiseer nie, maar ook outeurs se stylmerkers kan identifiseer. Ten einde die stilometriese analise vir my navorsing te kon uitvoer, is sleutelwoorde, frekwente woorde en funksiewoorde ondersoek. Bykomend is sins- en woordlengtes ook bestudeer.

---

[166] Wybenga (1987:2) argumenteer dat styl as ŉ keuse beskou kan word of as ŉ eksploitering van die norm (sien ook hoofstuk 7).
[167] Hier word veral na sosiale mediaplatforms soos Twitter, Tumblr en Facebook verwys.

6.2 Sleutelwoorde

ŉ Ondersoek na sleutelwoorde verwys na die stilometriese analise wat deur middel van WordSmith Tools uitgevoer is. ŉ Sleutelwoordfunksie in WordSmith Tools vergelyk die frekwensies van twee tekste se leksikale items. In hierdie analise is ŉ chi-kwadraattoets uitgevoer. WordSmith Tools toon ŉ chi-kwadraattoets se resultate as die sleutelwaarde (*keyness*) aan.

Ten einde die analise te kon uitvoer, is sleutelwoordlyste in WordSmith Tools opgestel. Aangesien hierdie analise vergelykend is, is die tekste (soos verduidelik in afdeling 5.12.3 – die gespesialiseerde korpus, die subkorpus van die onderskeie deelnemers, sowel as die verdagte tekste) met mekaar vergelyk. Ten einde die ondersoek na die sleutelwoorde te kon uitvoer, moes daar onderskeid getref word tussen funksie- en inhoudswoorde[168]. Funksiewoorde word as woorde beskou wat konteks onafhanklik is en daarom hou hierdie woorde (vgl. Koppel & Schler, 2003) nie met ŉ spesifieke onderwerp verband nie. Vir die doel van hierdie analise is die inhoudswoorde dus vanuit die teks verwyder en is die funksiewoorde met mekaar vergelyk.

Funksiewoorde kan (volgens Biber *et al.*, 2003:16) gekategoriseer word as woordgroepe wat gewoonlik betekenis en verhoudings tussen leksikale woorde aantoon. Funksiewoorde sluit lidwoorde, hulpwerkwoorde, voornaamwoorde, voegwoorde en voorsetsels in (vgl. Biber *et al.*, 2003:26-33, Carney, 2015:91; Rocha *et al.*, 2017:8). Ten einde die funksiewoorde te kon ondersoek, is die sleutelwoordlyste (soos genoem in afdeling 5.8.3.3) as ŉ Microsoft Excel-lêer gestoor en leksikale woorde is vanuit die lys verwyder. Die lyste het dus slegs uit funksiewoorde (ook genoem grammatikale woorde) bestaan.

N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P
1	EK	383	4,22	95	0,77	283,56	0,0000000000
2	MY	243	2,68	54	0,44	190,13	0,0000000000
3	JY	194	2,14	43	0,35	151,26	0,0000000000
4	JOU	108	1,19	7	0,06	123,59	0,0000000000
5	DIE	272	3,00	165	1,34	71,20	0,0000000000
6	NIE	355	3,91	245	1,99	70,47	0,0000000000
7	NET	96	1,06	37	0,30	47,43	0,0000000000
8	N	218	2,40	159	1,29	36,81	0,0000000000
9	MENSE	59	0,65	26	0,21	24,42	0,0000007699
10	KAN	80	0,88	44	0,36	24,11	0,0000009076
11	IEMAND	33	0,36	8	0,06	22,88	0,0000017197
12	SLAAP	22	0,24	3	0,02	19,50	0,0000100792
13	IS	244	2,69	222	1,80	18,99	0,0000131311
14	AS	134	1,48	104	0,84	18,51	0,0000168611
15	WIL	36	0,40	13	0,11	18,18	0,0000201392
16	HTTPS	17	0,19	1		17,92	0,0000230881
17	LEWE	24	0,26	5	0,04	17,76	0,0000250301
18	WEET	31	0,34	10	0,08	17,23	0,0000331237
19	NOOIT	23	0,25	5	0,04	16,55	0,0000473083
20	INSTAGRAM	13	0,14	0		15,40	0,0000870264

Skermskoot 6.1 Uitbeelding van die sleutelwoordlyste in WordSmith Tools

[168] Sien ook hoofstuk 5 (Afdeling 5.7.1 Stilometrie) en hoofstuk 3 (Afdeling 3.4.1 Forensiese linguistiek: die ontstaan asook Afdeling 3.5.1.1 – Forensiese stilistiek).

Skermskoot 6.1 is 'n uitbeelding van die onverwerkte sleutelwoordlyste in WordSmith Tools en hierdie lyste is as 'n Microsoft Excel-lêer gestoor. Die woorde is volgens die p-waardes gesorteer (sien Skermskoot 6.2) en daarna is die top tien grammatikale woorde vanuit elke datastel gekies (sien Skermskoot 6.3) en volgens sleutelwaarde gesorteer (sien Skermskoot 6.4) soos gebruik deur Kotzé (2007:392-393) en Michell (2013:115). In Skermskoot 6.2 en 6.3 verskyn sommige van die waardes (die *keyness* oftewel die sleutelwaarde) in rooi. Die rooi syfers verwys na sleutelwaardes wat beduidend meer of minder gereeld frekwent in die verwysingskorpus[169] voorkom as die korpus wat vir die vergelyking gebruik is.

	A	B	C	D	E	F	G
1	Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P
2	VOOR	14.00000	0.19768	8.00000	0.08811	2.75461	0.09697
3	HOM	12.00000	0.16944	6.00000	0.06608	2.94851	0.08596
4	POSTS	4.00000	0.05648	0.00000		3.10102	0.07824
5	SMART	4.00000	0.05648	0.00000		3.10102	0.07824
6	WEBSITE	4.00000	0.05648	0.00000		3.10102	0.07824
7	SNORS	4.00000	0.05648	0.00000		3.10102	0.07824
8	UITVERKOOP	4.00000	0.05648	0.00000		3.10102	0.07824
9	RAPS	4.00000	0.05648	0.00000		3.10102	0.07824
10	SUPERMAN	4.00000	0.05648	0.00000		3.10102	0.07824
11	TEAM	4.00000	0.05648	0.00000		3.10102	0.07824
12	SHARKS	4.00000	0.05648	0.00000		3.10102	0.07824
13	RADII	4.00000	0.05648	0.00000		3.10102	0.07824
14	PULL	4.00000	0.05648	0.00000		3.10102	0.07824
15	RAND	4.00000	0.05648	0.00000		3.10102	0.07824
16	SPELERS	4.00000	0.05648	0.00000		3.10102	0.07824
17	CHEETAHS	4.00000	0.05648	0.00000		3.10102	0.07824
18	DAM	4.00000	0.05648	0.00000		3.10102	0.07824
19	DOOGZ	4.00000	0.05648	0.00000		3.10102	0.07824
20	BLACKS	4.00000	0.05648	0.00000		3.10102	0.07824
21	BLIKSEM	4.00000	0.05648	0.00000		3.10102	0.07824
22	LEEEEEEUS	4.00000	0.05648	0.00000		3.10102	0.07824
23	KLOMP	4.00000	0.05648	0.00000		3.10102	0.07824
24	COACH	4.00000	0.05648	0.00000		3.10102	0.07824
25	LAAS	4.00000	0.05648	0.00000		3.10102	0.07824
26	JP	4.00000	0.05648	0.00000		3.10102	0.07824
27	LEEEEEEEEUS	4.00000	0.05648	0.00000		3.10102	0.07824

WordSmith Keyword list

Skermskoot 6.2 Sleutelwoordlyste volgens p-waarde gesorteer

A	B	C	D	E	F	G
Key word	Freq.	%	RC. Freq.	RC. %	Keyness	P
EK	383	4.218061447	95	0.769729376	283.5640869	5.14688E-17
MY	243	2.676211357	54	0.437530398	190.1281433	2.10207E-16
JY	194	2.136563778	43	0.348403811	151.2558746	4.96234E-16
JOU	108	1.189427257	7	0.0567169	123.5915527	1.11247E-15
KAN	80	0.881057262	44	0.356506228	24.10870552	9.07553E-07
IEMAND	33	0.363436133	8	0.064819314	22.88192368	1.71972E-06
OOR	30	0.330396473	106	0.858855963	22.32896233	2.29423E-06
IS	244	2.687224627	222	1.798735976	18.990942	1.31311E-05
OP	71	0.781938314	177	1.434127331	18.8799324	1.3918E-05
AS	134	1.47577095	104	0.842651129	18.51423454	1.68611E-05
WIL	36	0.396475762	13	0.105331391	18.17580795	2.01392E-05
JULLE	25	0.275330395	7	0.0567169	15.32832146	9.03488E-05
WANNEER	23	0.253303975	6	0.048614487	14.7338829	0.000123798
MENS	27	0.297356814	9	0.072921731	14.39727688	0.000148013
HIERDIE	25	0.275330395	8	0.064819314	13.73463535	0.000210533
BY	22	0.24229075	69	0.559066594	11.67272377	0.000634228
DEUR	5	0.055066079	32	0.259277254	11.49593449	0.000697483
HULLE	56	0.616740108	41	0.332199007	8.775728226	0.00305265

169 Vir die doel van die sleutelwaardebepalings is die onderskeie datastelle vergelyk soos in afdeling 5.12.3.2 Die gespesialiseerde korpus). Ten einde die sleutelwaardes te kon bepaal, is die grootste van die twee korpora wat gebruik is as die verwysingskorpus gebruik.

Skermskoot 6.3 Grammatikale woorde in die sleutelwoordlyste

O1_O5	Sleutelwoorde	Frekwensie	%	Verwysingskorpus frekwensie	Verwysingskorpus %	Sleutelwaarde	P
	JYS	12	0.261894	0		20.910	4.81E-06
	EKS	9	0.196421	0		14.988	0.000108
	SY	36	0.785683	32	0.352422893	10.684	0.001081
	WANT	16	0.349193	9	0.099118941	9.102	0.002553
	HOM	11	0.24007	6	0.066079296	6.084	0.013641
	MYNE	4	0.087298	0		5.227	0.02224
	ELKE	10	0.218245	6	0.066079296	4.797	0.028505
	HAAR	14	0.305543	12	0.132158592	3.950	0.046869
	JULLE	4	0.087298	25	0.275330395	4.234	0.039618
	BY	3	0.065474	22	0.24229075	4.289	0.038351
o1_o6	Sleutelwoorde	Frekwensie	%	Verwysingskorpus frekwensie	Verwysingskorpus %	Sleutelwaarde	P
	ENIGE	11	0.06796	4	0.024712715	7.564	0.005955
	SOU	12	0.074138	6	0.037069071	6.113	0.013421
	WAT	117	0.722847	154	0.9514395	5.916	0.015003
	AAN	36	0.222414	36	0.222414434	5.605	0.017913
	ALMAL	24	0.148276	21	0.129741758	5.193	0.022683

Skermskoot 6.4 Voorstelling van die grammatikale woorde volgens sleutelwaarde gesorteer

Skermskoot 6.1 tot 6.4 is voorstellinge van die manier waarop die vergelykings tussen die onderskeie datastelle uitgevoer is. Nadat die bogenoemde stappe uitgevoer is, is die gemiddelde sleutelwaardes vir elk van die datastelle wat met mekaar vergelyk is ook uitgewerk (sien Tabel 6.1).

Tabel 6.1 Voorstelling van die Tabelle waar die sleutelwaardes van die grammatikale woorde bepaal is deur middel van WordSmith Tools

Sleutelwoorde	Frekwensie	%	Verwysingskorpus frekwensie	Verwysingskorpus %	Sleutelwaarde	P
IN	51	0.672	65	1.328	13.176	0.0003
ALMAL	3	0.040	13	0.266	10.170	0.0014
DIT	39	0.514	50	1.021	10.105	0.0015
HUL	17	0.224	0		9.401	0.0022
AAN	41	0.540	11	0.225	6.412	0.0113
HIERDIE	13	0.171	21	0.429	6.349	0.0117
MOET	6	0.079	13	0.266	5.633	0.0176
WIE	3	0.040	9	0.184	5.034	0.0249
VAN	118	1.555	52	1.062	5.031	0.0249
HIER	26	0.343	6	0.123	4.813	0.0282
SO	8	0.105	14	0.286	4.531	0.0333
				Sleutelwaarde gemiddeld	7.332	

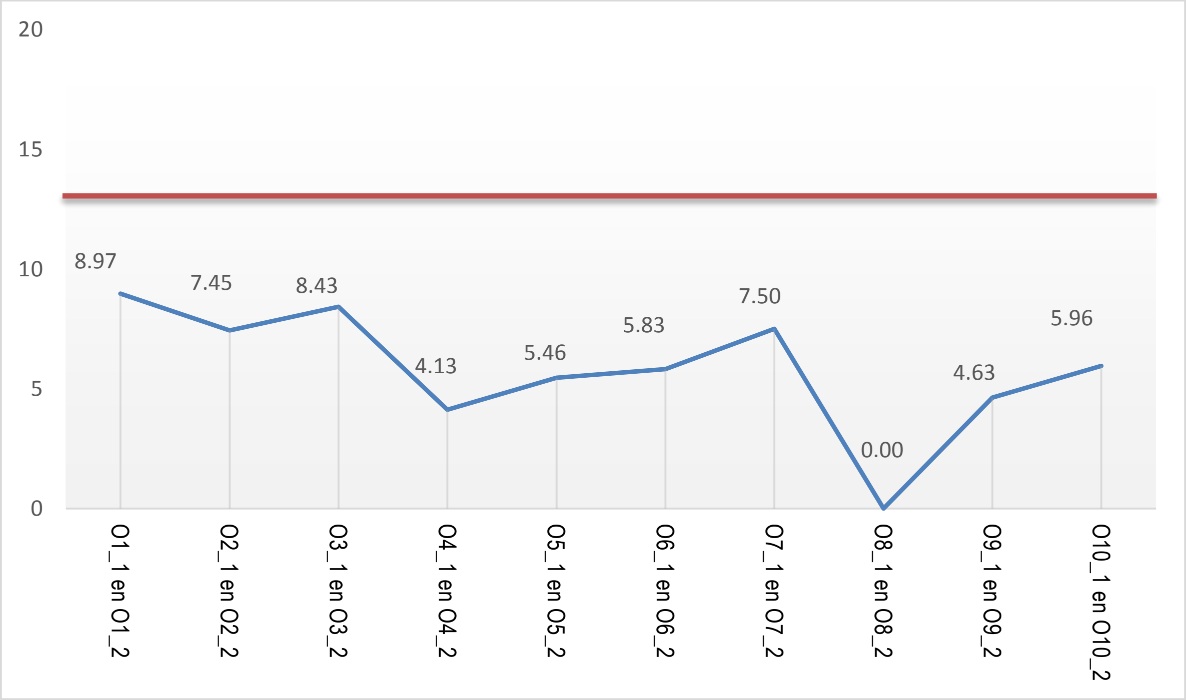
6.2.1 Gespesialiseerde korpus: sleutelwaardes van funksiewoorde

Die bevindinge van die analise van die sleutelwaardebepalings − die onderskeie datastelle is met mekaar vergelyk[170] – word in hierdie afdeling voorgehou[171]. Wanneer die sleutelwaardes van die onderskeie datastelle met mekaar vergelyk is, is die gemiddelde sleutelwaarde hoër as 15

[170] Vir die doel van die sleutelwaardebepalings is die onderskeie datastelle vergelyk soos in afdeling 5.12.3.2 Die gespesialiseerde korpus uiteengesit. Ten einde die sleutelwaardes te kon bepaal, is die grootste van die twee korpora wat gebruik is as die verwysingskorpus gebruik.

[171] Hierdie bevindinge word opsommend weergegee en ŉ opsomming van die tabelle word in Bylaag C gevind.

wat daarop dui dat die tekste deur verskillende outeurs geskryf is (Kotzé, 2007:393). Die eerste vergelyking wat onderneem is, was interne vergelykings om te bepaal of die twiets van ’n enkele outeur wel met mekaar ooreenstem. Figuur 6.1 is ’n opsomming van hierdie bevindinge:



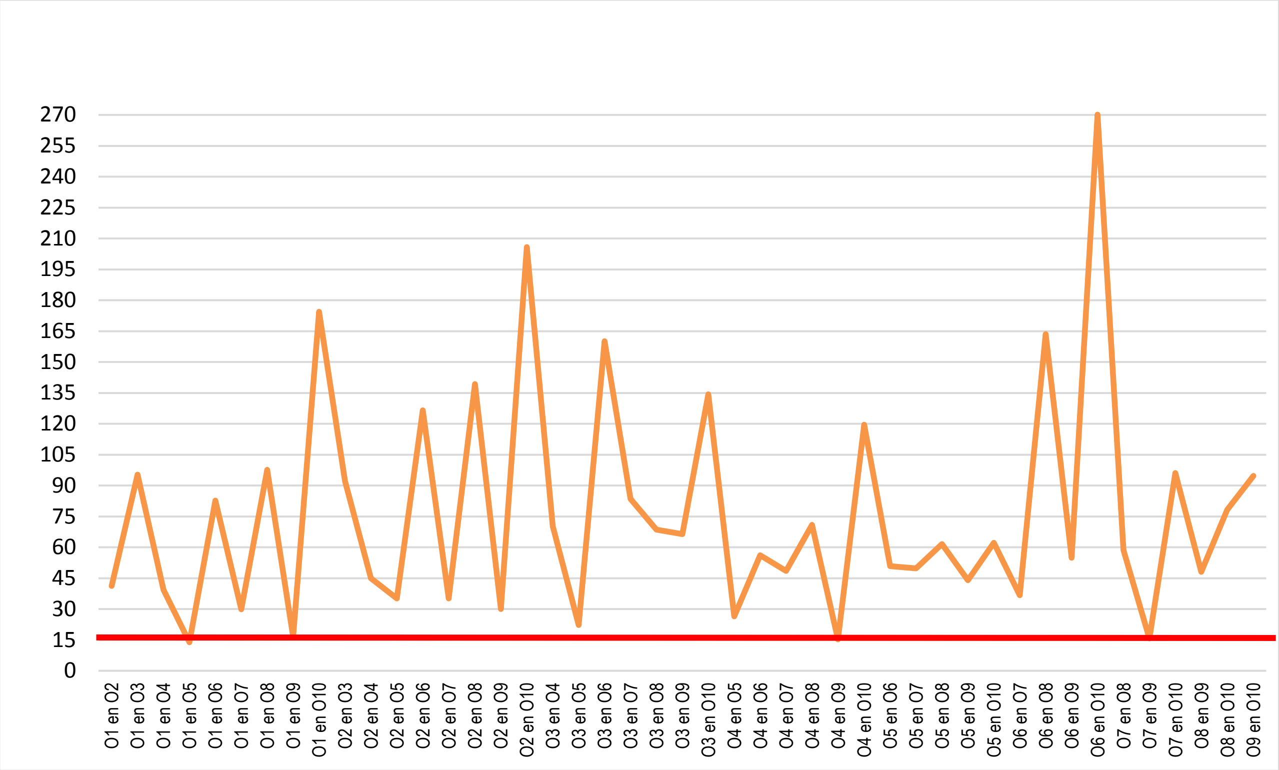
Figuur 6.1 Interne sleutelwaarde vergelykings van die grammatikale woorde

Vanuit Figuur 6.1 is dit duidelik dat die gemiddelde sleutelwaardes − wanneer die korpora intern met mekaar vergelyk was − laer as 15[172] is en daarom is daar deur die interne ondersoek bevestig dat die funksiewoorde wel kan aandui wanneer ’n teks deur dieselfde outeur geskryf is.

Vervolgens is die korpora (soos uiteengesit in Tabel 5.7) vergelyk ten einde die gemiddelde sleutelwaardes van die funksiewoorde te kon bepaal. Figuur 6.2 toon die sleutelwaarde vergelykings tussen die onderskeie korpora aan. Vanuit dié Figuur is dit duidelik dat sommige van die sleutelwaardes laer as 15[173] is (4 van die sleutelwaarde vergelykings is laer as 15), maar die meerderheid van die vergelykings het waardes van 15 en hoër wat daarop dui dat die outeurs van die tekste verskil.

[172] Die hoogste sleutelwaarde (wanneer die korpora intern met mekaar vergelyk is) was 8.97. Sien bylaag C vir ’n lys met die interne sleutelwaardevergelykings.

[173] Wanneer die sleutelwaardes van die onderskeie datastelle met mekaar vergelyk is, was die gemiddelde sleutelwaardes hoër as 15 wat daarop dui dat die tekste deur verskillende outeurs geskryf is (Kotzé, 2007:393).



Figuur 6.2 Sleutelwaarde vergelykings van die grammatikale woorde

Wanneer daar spesifiek na die tekste ondersoek ingestel is waar die gemiddelde sleutelwaarde laer as 15 is, blyk dit dat O1 en O5 (4,78), O1 en O9 (6.85), O3 en O4 (14,52) asook O4 en O7 (12,06) soveel ooreenstem dat dit nie moontlik is om sonder twyfel na gelang van die gemiddelde sleutelwaarde berekeninge te kon vasstel of die tekste deur dieselfde outeur geskryf is nie.

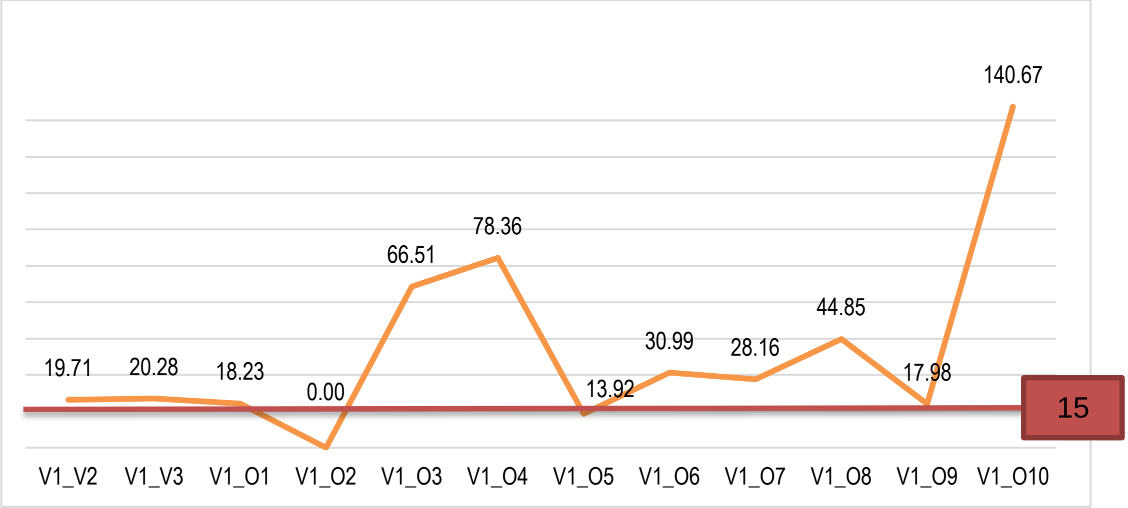
Met nadere ondersoek na die bogenoemde tekste − waarvan die gemiddelde sleutelwaardes laer as 15 is – blyk dit dat die onderskeie kategorieë[174] waarin Twitter verdeel is ŉ invloed op die taalgebruik het. O1, O5 en O9 is in die blogger-kategorie ingedeel terwyl O3 deel van die media-kategorie is. O4 en O7 is in die kategorie vir vermaaklikheidspersoonlikhede geplaas. Wanneer die korpora wat sleutelwaarde gemiddelde van minder as 15 het dus ondersoek is met betrekking tot die genoemde kategorieë, blyk dit dat korpora in dieselfde kategorie (hier kan dus ook na die genre van Twitter verwys word) ooreenstem. O3 en O4 val nie in dieselfde kategorie nie (O3 – media, O4 – vermaaklikheidspersoonlikhede), maar die gemiddelde sleutelwaarde is 14,52; dus baie naby aan die 15-afsnypunt om te kan aanvoer dat die tekste deur verskillende outeurs geskryf is.

[174] Sien hoofstuk 5 (afdeling 5.4 Data-insameling en -verwerking).

Vanuit die 30 vergelykings wat ten opsigte van gemiddelde sleutelwaardes van die funksiewoorde ondersoek is, is 26 sleutelwaardes dus hoër as 15 (die tekste is deur verskillende outeurs geskryf) terwyl daar slegs by 4 van die vergelykings nie sonder twyfel aangedui kon word dat die tekste nie deur dieselfde outeur geskryf is nie.

Alhoewel die gemiddelde sleutelwaardes van die funksiewoorde dus met 86,66% van die vergelykings verskil, dui dit daarop dat die tekste deur verskillende outeurs geskryf is (en in hierdie geval is dit wel bekend dat al die tekste deur verskillende individue geskryf is). Die gebruik van die grammatikale woorde se sleutelwaardes is dus nie alleen genoegsaam om sonder redelike twyfel te kan aantoon dat ʼn teks deur ʼn bepaalde individu geskryf is nie

6.2.2 Verdagte tekste: sleutelwaardes en inhoudswoorde

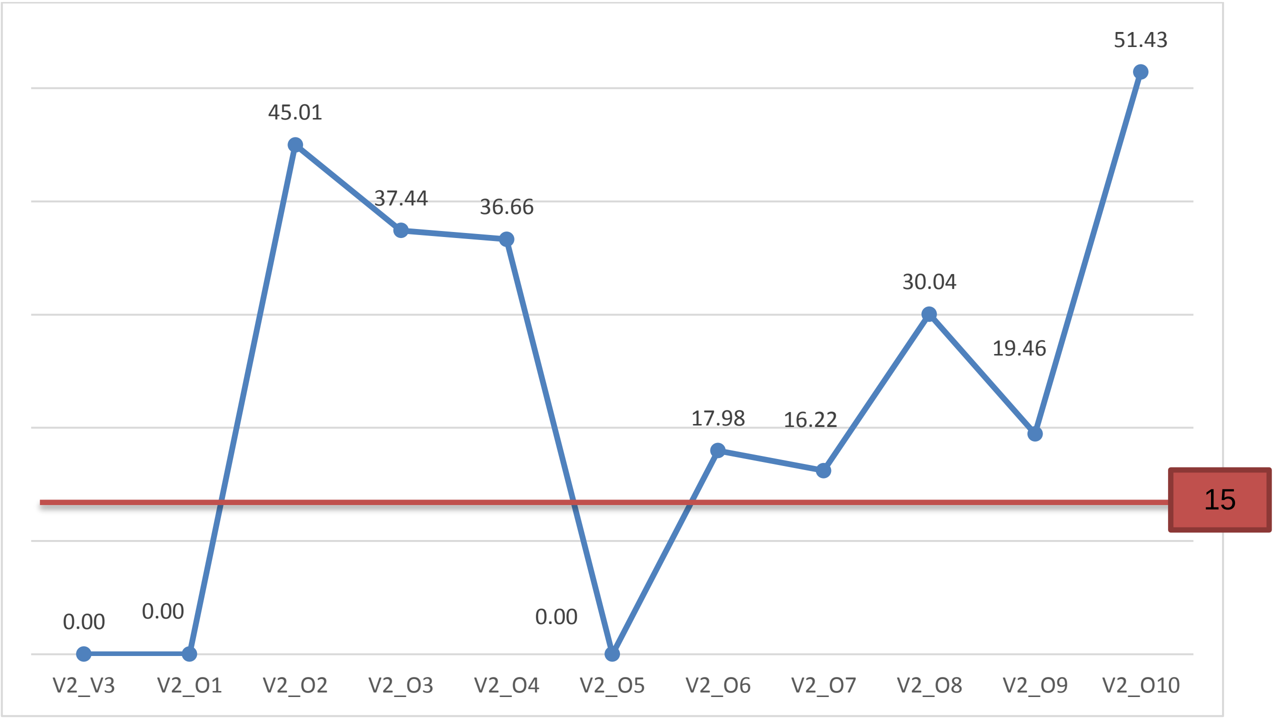


Figuur 6.3 Vergelyking van V1 se korpora met die gespesialiseerde korpora

Figuur 6.3 toon die gemiddelde sleutelwaardes aan van die vergelyking tussen die verdagte outeur (V1) en die tien outeurs vanuit die gespesialiseerde korpus. Soos reeds genoem, kan daar volgens Kotzé (2007:393) aangevoer word dat wanneer die sleutelwaardes van die onderskeie datastelle met mekaar vergelyk word, en die gemiddelde sleutelwaardes hoër is as 15, kan dit daarop dui dat die tekste deur verskillende outeurs geskryf is.

Vanuit Figuur 6.3 (die sleutelwaardevergelyking van die korpora van die eerste verdagte outeur (V1) met die korpora van die gespesialiseerde korpora) blyk dit dat die vergelyking van V1 en O2 se sleutelwaardevergelyking se gemiddeld 0 is terwyl die vergelyking van V1 en O5 se sleutelwaardes ʼn gemiddeld van 13,92 het. Daar kan dus vanuit die data in Figuur 6.3 aangevoer word dat die eerste verdagte teks (V1) nie deur O1, O3, O4, O6, O7, O8, O9 of O10 geskryf is nie. Aangesien die sleutelwaardes van die vergelyking tussen V1 en O5 13,93 (laer

as 15) is, kan daar nie met sekerheid gestel word dat O5 nie die outeur van V1 is nie, maar siende dat die vergelyking tussen O2 en V1 nog laer as dié van V1 en O5 is (0), kan daar aangevoer word dat O2 na gelang van hierdie vergelyking die moontlike outeur van V1 is.



Figuur 6.4 Sleutelwaardevergelyking van V2 se korpora met die gespesialiseerde korpora

Figuur 6.4 verskaf die data waar die korpus van die tweede verdagte teks (V2) met die gespesialiseerde korpora vergelyk is ten einde die gemiddelde sleutelwaardes te kon bepaal. Vanuit Figuur 6.4 blyk dit dat daar drie korpora is wat oor ’n gemiddelde sleutelwaarde van laer as 15 beskik. Daar kan dus nie sonder redelike twyfel aangevoer word dat die korpora nie deur verskillende outeurs geskryf is nie. Die drie korpora is V2 en V3 (0), V2 en O1 (0) asook V2 en O5 (0). Die oorblywende vergelykings (V2 in vergelyking met V1, O2, O3, O4, O6, O7, O8, O9 en O10) se gemiddelde sleutelwaardes is almal hoër as 15. Daar kan dus sonder twyfel aangevoer word dat die tweede verdagte teks (V2) nie deur ’n spesifieke outeur van die korpora geskryf is nie.

Aangesien daar drie vergelykings (sien Figuur 6.4) was waar die gemiddelde sleutelwaardes laer as 15 is, kan daar nie met sekerheid ’n enkele outeur geïdentifiseer word wat as die moontlike outeur van V2 beskou kan word nie. V2 en V3 (twee van die drie verdagte tekste) stem ook ooreen na gelang van die ondersoek na die sleutelwaardes en dus is dit duidelik vanuit hierdie analise dat V3 as ’n moontlike outeur van V2 beskou kan word. V2 stem verder

ook ooreen met die sleutelwaarde vergelyking van O1 en O5. Wanneer O1 en O5 verder geanaliseer is, blyk dit dat beide O1 en O5 in die persoonlike kategorie van Twitter geplaas is.

Hierdie ondersoek na die sleutelwaarde vergelyking van V2 toon dus dat V2 en V3 se gebruik van funksiewoorde na gelang van die sleutelwaarde vergelyking nie ooreenstem nie (dus kan V2 en V3 deur dieselfde outeur geskryf wees), maar ook dat daar twee outeurs vanuit die gespesialiseerde korpus is wat moontlik die outeur van V2 kan wees (O1 en O5). Vanuit die ondersoek na die sleutelwaardes van V2 (sien Figuur 6.5) kan daar vervolgens geargumenteer word dat die moontlike tien individue wat die outeur van V2 is, verminder kan word na twee individue (O1 en O5).

Figuur 6.5 Sleutelwaardevergelyking van V3 se korpora met die gespesialiseerde korpora

Figuur 6.5 toon die data aan in die ondersoek van die sleutelwaardewaarde vergelyking van V3 met die gespesialiseerde korpora. Vanuit Figuur 6.5 blyk dit dat daar vier vergelykings is waar die gemiddelde sleutelwaarde laer is as 15 en waarvoor daar dus nie aangevoer kan word dat die tekste deur verskillende outeurs geskryf is nie.

Die vier gemiddelde sleutelwaardes wat minder as 15 is, word gevind in die vergelykings van V3 en O1 (0), V3 en O5 (0), V3 en O6 (14,48) asook V3 en O9 (0) – hier moet ook in ag geneem word dat die gemiddelde sleutelwaarde van V2 en V3 (soos gesien in Figuur 6.4) ook laer is as 15. Dit kom dus voor asof V3 ooreenstem na gelang van die gemiddelde sleutelwaardeondersoek met vier bekende outeurs en een onbekende outeur. Die sleutelwaardevergelyking tussen V3 en O6 is so naby aan 15 (veral wanneer die p-waardes van

die ander korpora wat meer met dié van V3 ooreenstem in ag geneem word) dat daar aangevoer kan word dat O6 nie die outeur van die derde verdagte teks (V3) is nie.

Vanuit die analise om die gemiddelde sleutelwaardes te bepaal, kan daar vervolgens aangevoer word dat daar vier moontlike outeurs vir die derde verdagte teks is, naamlik V2, O1, O5 en O9. Hierdie vier outeurs is dus as die "hoofverdagtes" in die verdere ondersoek na die identifisering van die outeur van V3 beskou en verdere analises was nodig om die outeurs van die verdagte tekste te kon identifiseer.

Vervolgens is daar ondersoek ingestel na die ratiobepalings (gespesialiseerde korpus sowel as die tekste in die verdagte korpus) om te kon bepaal of hierdie stilometriese ondersoek toegepas kan word in die identifisering van mikroblogs se werklike outeurs.

6.3 Ratiobepalings

Soos in hoofstuk 5[175] bespreek, kan daar deur middel van ŉ ratioberekening bepaal word of die korpora wat ondersoek is prakties beduidend verskil of nie. Ten einde die ratio's te kon bepaal van die woorde wat deur die afsonderlike outeurs gebruik is, is die woordelysfunksie (*wordlist*) in WordSmith Tools gebruik. Die korpora (soos uitgebeeld in Tabel 5.7 en 5.8) is in WordSmith Tools ingevoer en volgens frekwensie gesorteer. Daarna is die woordelyste as Microsoft Excel-dokumente gestoor. Wanneer die ratio groter as 1,2 of kleiner as 0,8 is, kan die woord as prakties beduidend beskou word (vgl. Van Rooy & Terblanche, 2006:169) − daar kan gevolglik met redelike sekerheid gestel word dat die tekste deur verskillende outeurs geskryf is. Ten einde die prakties beduidende verskil te kon bepaal, het ek ondersoek ingestel na die funksiewoorde (lidwoorde, hulpwerkwoorde, voornaamwoorde, voegwoorde en voorsetsels [vgl. Biber *et al.,* 2003:26-33; Carney, 2015:91; Rocha *et al.,* 2017:8]).

Die ondersoek na die bogenoemde woordgroepe is uitgevoer aangesien funksiewoorde as woorde beskou word wat konteksonafhanklik is; daarom is hierdie woorde (vgl. Koppel & Schler, 2004) nie onderwerpspesifiek nie.

Die ratio van die onderskeie funskiewoorde is soos volg bereken: eerstens, is die korpora geannoteer ten einde die verskillende woordsoorte[176] te kon aandui[177]. Tweedens, is ŉ lys opgestel met die lidwoorde en die lys is as ŉ txt-dokument gestoor. Derdens, is die onderskeie outeurs se woordelyste in WordSmith Tools oopgemaak en die txt-dokument met die lidwoorde is as ŉ stoplys in die program ingevoer. Deur gebruik te maak van die stoplys-funksie, is alle

[175] Sien afdeling 5.6 Kwantitatiewe analise.
[176] Sien bylaag F, Woordsoort-etikette, soos beskryf deur Pilon (2005).
[177] Daar is gebruik gemaak van die woordsoortetiketeerder soos deur CTexT verskaf.

woorde wat nie op die lys verskyn nie (dus alle woorde wat nie lidwoorde is nie) van die woordelyste verwyder. Hierdie nuwe woordelyste is as Microsoft Excel-dokumente gestoor. Laastens, is die ratio's bereken[178] en daar is bepaal of die verskil as prakties beduidend beskou kan word.

Die ratio berekeninge kan as volg saamgevat word:

- Daar is vir elke datastel 'n woordelys in WordSmith Tools saamgestel.
- Sleutelwoordlyste is saamgestel. Die sleutelwoordlyste verwys na 'n lys waarin twee verskillende datastelle se woordelyste met mekaar vergelyk word. Ten einde die onderskeie funksiewoorde te identifiseer is daar van die rekenaarprogram NCHLT-Tagger gebruik gemaak ten einde die korpora te annoteer[179] en sodoende die verskillende woordsoorte te identifiseer. Daar is dan woordelyste saamgestel wat uit lidwoorde, voornaamwoorde, voorsetsels, hulpwerkwoorde en voegwoorde (die funksiewoorde) bestaan het. Hierdie woordelyste is telkens as stoplyste in WordSmith Tools ingevoer, en die opsie is gekies om alle woorde wat nie in die stoplyste voorkom nie vanuit die sleutelwoordlyste te verwyder. Sodoende is die onderskeie funksiewoorde wat in die vergelykings van die datastelle voorkom geïdentifiseer.
- Die onderskeie datastelle (soos uitgebeeld in Tabel 5.7 en 5.8) is dan volgens frekwensie gesorteer.
- Hierdie woordelyste is dan as Microsoft Excel-dokumente gestoor.
- Die frekwensie is vervolgens vir elke woord bereken:

$$\frac{Totale\ keer\ wat\ woord\ in\ korpus\ voorkom}{Totale\ woorde\ in\ korpus} \times 1000 = Frekwensie$$

Frekwensie van *af* (sien skermskoot 6.5) in O1

$$\frac{2}{8942} \times 1000 =$$ Frekwensie van *af* (sien skermskoot 6.5) in O1

∴ Frekwensie van *af* (sien skermskoot 6.5) in O1 = 0,22

[178] Sien afdeling 5.7 (Kwantitatiewe analise).

[179] Dit is belangrik om die leemte wat daar in die automatiese annoteerder voorkom te beklemtoon. In hierdie analise moes daar in ag geneem word dat die suksesratio van etikettering nie 100% is nie. Daarom het die moontlikheid bestaan dat van die woorde verkeerd geëtiketteer is. Die besluit is geneem om met die fout te werk ten einde te kon bepaal of díé soort analise wel 'n bydra tot die identifisering van outeurs van mikroblogtekste kan lewer.

- Volgende is die ratio vir die onderskeie datastelle bereken (die datastelle in die gespesialiseerde korpus is met mekaar vergelyk, en ten einde die bevindinge te verifieer is die tien sub-korpora van die gespesialiseerde korpus ook telkens met die verwysingskorpus vergelyk). Ratio's word soos volg bereken:

$$\frac{Frekwensie\ van\ woord\ in\ O1}{Frekwensie\ van\ woord\ in\ O2} \times 1000 = Ratio$$

Frekwensie van *af* (sien skermskoot 6.5) in O1 en O2

$$\frac{0{,}22}{0{,}00} \times 1000 = 0{,}00$$

∴ Ratio van *af* (wanneer O1 met O2 vergelyk word) = 0,00

∴ ratio van af = $< 0{,}8$

∴ ratio van *af* (O1 en O2) toon ʼn prakties beduidende verskil

∴ O1 en O2 is na gelang van die voorkoms van die woord *af* deur verskillende outeurs geskryf.

Aangesien daar by elk van die verskillende funksiewoordgroepe etlike woorde deur die NCHLT-Tagger geïdentifiseer is, is daar ʼn persentasie bereken waarby elk van die sub-korpora prakties beduidend van mekaar verskil, al dan nie. Hierdie persentasie is soos volg bereken:

Som van die funksiewoorde wat prakties beduidend verskil in elk van die vergelykings ten einde die prakties beduidende verskil te bereken gedeel deur die totale hoeveelheid funskiewoorde maal met 100.

Verder is die totale som van al die funksiewoorde wat prakties beduidend verskil in al tien die korpora is bereken en ʼn totaal is gevind. Daarna is die som van al die funksiewoorde wat nie prakties beduidend verskil nie bereken en ʼn totaal is gevind. Ten einde ʼn gemiddeld te kon bereken, is elk van die totale deur die totale aantal funskiewoorde (hetsy lidwoorde,

voornaamwoorde, voorsetsels, hulpwerkwoorde of voegwoorde) wat in al die vergelykings gevind is, gedeel en met 100 vermenigvuldig.

Neem as voorbeeld die berekeninge van die persentasies prakties beduidende verskille, al dan nie

X voornaamwoorde geïdentifiseer en geannoteer deur middel van die NCHLT-Tagger

45 vergelykings word uitgevoer

∴ *X* x 45 = Totale aantal funksiewoorde

Y hoeveelheid funksiewoorde met 'n prakties beduidende verskil in al 45 die vergelykings

∴ *Y* = (O1 en O2) + (O1 en O3) + (O1 en O4) + (O1 en O5) + (O1 en O6) + (O1 en O7) + (O1 en O8) + (O1 en O9) + (O1 en O10) + (O2 en O3) + (O2 en O4) + (O2 en O5) + (O2 en O6) + (O2 en O7) + (O2 en O8) + (O2 en O9) + (O2 en O10) + (O3 en O4) + (O3 en O5) + (O3 en O6) + (O3 en O7) + (O3 en O8) + (O3 en O9) + (O3 en O10) + (O4 en O5) + (O4 en O6) + (O4 en O7) + (O4 en O8) + (O4 en O9) + (O4 en O10) + (O5 en O6) + (O5 en O7) + (O5 en O8) + (O5 en O9) + (O5 en O10) + (O6 en O7) + (O6 en O8) + (O6 en O9) + (O6 en O10) + (O7 en O8) + (O7 en O9) + (O7 en O10) + (O8 en O9) + (O8 en O10) + (O9 en O10)

∴ *Y/X* x 100 = persentasie funksiewoorde wat prakties beduidend verskil (al dan nie).

Die berekening is dus vir alle funksiewoorde wat prakties beduidend verskil, sowel as die wat nie prakties beduidend verskil nie uitgevoer. Vervolgens gaan ek die metode wat gevolg is ten einde die ratio berekeninge uit te voer aan die hand van data vanuit die gespesialiseerde korpus voorhou. Hiedie voorbeeld word na gelang van die ratio berekeninge van die hulpwerkwoorde[180] aangebied.

Daar is 20 hulpwerkwoorde geannoteer deur die gebruik van die rekenaarprogram NCHLT-Tagger[181]. Die woorde is as 'n txt-dokument gestoor, en woordelyste is van elk van die datastelle

---

[180] Alle funksiewoorde se berekeninge is op dieselfde wyse uitgevoer. Aangesien daar slegs 21 hulpwerkwoorde in die gespesialiseerde korpus geïdentifiseer is, is daar ter wille van bondigheid besluit om hierdie funksiewoord-groep as voorbeeld te gebruik.

[181] Dit is belangrik om die leemte wat daar in die automatiese annoteerder voorkom te beklemtoon. In hierdie analise moes daar in ag geneem word dat die suksesratio van etikettering nie 100% is nie. Daarom het die moontlikheid bestaan dat van die woorde verkeerd geëtiketteer is. Die besluit is

opgestel in WordSmith Tools, waar die txt-dokumente met die hulpwerkwoorde as stoplyste ingevoer is. Die resultate van die woordelyste kan in tabel 6.2 gesien word.

Tabel 6.2 Voorstelling van die Tabelle waar die frekwensies van die voorkoms van die hulpwerkwoorde gedokumenteer is vanuit O1

Woord	Frekwensie	%	Tekste	%
1	10	0.110132158	1	100
2	3	0.033039648	1	100
3	9	0.099118941	1	100
4	73	0.803964734	1	100
5	123	1.354625583	1	100
6	5	0.055066079	1	100
7	244	2.687224627	1	100
8	80	0.881057262	1	100
9	12	0.132158592	1	100
10	7	0.077092513	1	100
11	10	0.110132158	1	100
12	7	0.077092513	1	100
13	7	0.077092513	1	100
14	4	0.044052862	1	100
15	46	0.50660795	1	100
16	2	0.022026431	1	100
17	20	0.220264316	1	100
18	12	0.132158592	1	100
19	36	0.396475762	1	100
20	22	0.24229075	1	100
21	5	0.055066079	1	100

Tabel 6.2 toon die frekwensies aan van die hulpwerkwoorde wat in O1 voorkom. Hierdie tabelle is vir elk van die korpora saamgestel, waarna ’n gesamentlike Excel dokument geskep is waarin die ratio’s bereken is (sien Tabel 6.3). Hierdie frekwensies is bereken soos volg bereken:

$$\frac{Totale\ keer\ wat\ woord\ in\ korpus\ voorkom}{Totale\ woorde\ in\ korpus} \times 1000 = Frekwensie$$

Tabel 6.3 Voorstelling van die Tabelle waar die frekwensies van die hulpwerkwoorde gedokumenteer is vanuit die gespesialiseerde korpus

Frekwensie O1	Frekwensie O2	Frekwensie O3	Frekwensie O4	Frekwensie O5
1.12	1.14	1.78	1.13	0.44
0.34	0.08	0.30	0.00	0.22
1.01	0.49	0.00	0.70	0.22
8.16	5.14	6.81	6.47	8.95
13.76	12.73	12.43	9.56	12.23

geneem om met die fout te werk ten einde te kon bepaal of dié soort analise wel ’n bydra tot die identifisering van outeurs van mikroblogtekste kan lewer.

Frekwensie O1	Frekwensie O2	Frekwensie O3	Frekwensie O4	Frekwensie O5
0.56	0.00	0.00	0.42	0.22
27.29	18.11	20.72	18.14	36.03
8.95	3.59	1.18	3.23	6.33
1.34	1.31	1.48	2.53	1.75
0.78	0.41	0.00	0.42	0.22
1.12	1.31	0.00	1.13	3.28
0.78	0.90	1.48	0.28	0.44
0.78	0.24	0.00	0.98	2.84
0.45	0.65	3.85	0.70	0.22
5.14	4.00	0.00	2.25	5.90
0.22	0.57	0.89	0.14	0.87
2.24	2.94	2.96	2.95	4.15
1.34	0.57	0.00	0.56	0.44
4.03	1.06	0.89	0.98	4.37
2.46	4.73	4.44	1.27	1.75
0.56	0.16	0.30	0.42	0.44

Ten einde die ratio's berekeninge uit te voer is die frekwensies soos in tabel 6.3 voorgestel gebruik in die onderstaande formule:

$$\frac{Frekwensie\ van\ woord\ in\ O1}{Frekwensie\ van\ woord\ in\ O2} \times 1000 = Ratio$$

Die onderskeie sub-korpora is met mekaar vergelyk, dit wil sê 'n totaal van 45 vergelykings is uitgevoer ten einde al die sub-korpora waaruit die gespesialiseerde korpus bestaan met mekaar te vergelyk. Hierdie bevindinge is ook in MS Excel uitgevoer (sien tabel 6.4).

Tabel 6.4 Voorstelling van die Tabelle waar die ratio's van die hulpwerkwoorde gedokumenteer is vanuit O1

Ratio_O1-O2	Praktiese beduidenheid	Ratio_O1-O3	Praktiese beduidenheid	Ratio_O1-O4	Praktiese beduidenheid	Ratio_O1-O5	Praktiese beduidenheid
0.98	Dieselfde	0.63	Verskillend	0.99	Dieselfde	2.56	Verskillend
4.11	Verskillend	1.13	Dieselfde	0.00	Verskillend	1.54	Verskillend
2.06	Verskillend	0.00	Verskillend	1.43	Verskillend	4.61	Verskillend
1.59	Verskillend	1.20	Dieselfde	1.26	Verskillend	0.91	Dieselfde
1.08	Dieselfde	1.11	Dieselfde	1.44	Verskillend	1.12	Dieselfde
0.00	Verskillend	0.00	Verskillend	1.33	Verskillend	2.56	Verskillend
1.51	Verskillend	1.32	Verskillend	1.50	Verskillend	0.76	Verskillend
2.49	Verskillend	7.56	Verskillend	2.77	Verskillend	1.41	Verskillend
1.03	Dieselfde	0.91	Dieselfde	0.53	Verskillend	0.77	Verskillend
1.92	Verskillend	0.00	Verskillend	1.86	Verskillend	3.58	Verskillend
0.86	Dieselfde	0.00	Verskillend	0.99	Dieselfde	0.34	Verskillend
0.87	Dieselfde	0.53	Verskillend	2.78	Verskillend	1.79	Verskillend
3.20	Verskillend	0.00	Verskillend	0.80	Verskillend	0.28	Verskillend
0.69	Verskillend	0.12	Verskillend	0.64	Verskillend	2.05	Verskillend
1.29	Verskillend	0.00	Verskillend	2.29	Verskillend	0.87	Dieselfde
0.39	Verskillend	0.25	Verskillend	1.59	Verskillend	0.26	Verskillend
0.76	Verskillend	0.76	Verskillend	0.76	Verskillend	0.54	Verskillend
2.35	Verskillend	0.00	Verskillend	2.39	Verskillend	3.07	Verskillend

Ratio_O1-O2	Praktiese beduidenheid	Ratio_O1-O3	Praktiese beduidenheid	Ratio_O1-O4	Praktiese beduidenheid	Ratio_O1-O5	Praktiese beduidenheid
3.80	Verskillend	4.53	Verskillend	4.09	Verskillend	0.92	Dieselfde
0.52	Verskillend	0.55	Verskillend	1.94	Verskillend	1.41	Verskillend
3.43	Verskillend	1.89	Verskillend	1.33	Verskillend	1.28	Verskillend

Die bevindinge van die ratio berekeninge van die hulpwerkwoorde van O1 en O2, O1 en O3, O1 en O4, sowel as O1 en O5 word in Tabel 6.4 uitgebeeld. Daar is ŉ kolom geskep waarin daar aangedui is of die bevindinge prakties beduidend verskil (verskil – verskillende outeurs) of nie prakties beduidend verskil nie (dieselfde - dieselfde outeur). Hierdie benaming is gegrond op die bepaling wat stel dat wanneer die ratio groter as 1,2 of kleiner as 0,8 is, kan die woord as prakties beduidend beskou word (vgl. Van Rooy & Terblanche, 2006:169) en daar kan gevolglik met redelike sekerheid gestel word dat die tekste deur verskillende outeurs geskryf is.

Die volgende stap in hierdie proses was om te tel hoeveel van die 21 hulpwerkwoorde in elk van die vergelykings beskik oor ŉ prakties beduidende verskil (verskil – verskillende outeurs). Aangesien ek weet dat die datastelle in die gespesialiseerde korpus deur verskillende outeures geskryf is, sou die meeste van die hulperkwoorde in elk van die vergelykings wel oor ŉ prakties beduidende verskil moes beskik ten einde hierdie ondersoek as geslaagd te kan beskou en dit dan op die verdagte tekste toe te pas.

Tabel 6.5 Voorstelling van die wyse waarop die prakties beduidende verskille (al dan nie) aangeteken is

	O1 en O2	O1 en O3	O1 en O4	O1 en O5
Verskillend	16	17	19	17
Dieselfde	5	4	2	4

Tabel 6.5 toon die wyse aan waarop die hulpwerkwoorde wat prakties beduidend verskil (verskil) of nie verskil nie (dieselfde) aangeteken is. Hierdie getalle is gevind deur die som van die aantal hulpwerkwoorde wat prakties beduidend verskil in die vergelyking van O1 en O2 te bereken, sowel as die som van die hulpwerkwoorde wat nie prakties beduidend verskil nie[182]. Vanuit Tabel 6.5 is dit duidelik dat 16 van die 21 hulpwerkwoorde vanuit die ratio bepaling tussen O1 en O2 wel prakties beduidend verskil – dit wil sê dat slegs vyf hulpwerkwoorde op so ŉ wyse deur beide O1 en O2 gebruik word dat daar nie sonder redelike twyfel aangevoer kan word dat die tekste deur verskillende outeurs geskryf is nie. Hierdie bevindinge kan vervolgens ook as persentasies weergegee word;

[182] Hierdie berekening is vir elke vergelyking uitgevoer.

$16 \div 21 \times 100 =$ 71,2 % van die hulpwerkwoorde in die vergelyking van O1 en O2 verskil prakties beduidend, en daarom kan daar aangevoer word dat daar in 71,2 % van die gebruik van hulpwerkwoorde deur O1 en O2 aangevoer kan word dat die tekste wel deur verskillende outeurs geskryf is. Daar kan ook aangevoer word dat daar slegs in 28,8 % van die kere wat hulpwerkwoorde deur O1 en O2 gebruik word nie met sekerheid aangevoer kan word dat die tekste wel deur verskillende outeurs geskryf is nie.

Deur die persentasies te bereken van die hoeveelheid kere wat daar 'n praktiese beduidende verskil, al dan nie, voorkom wanneer die ratioberekeninge uitgevoer word, kan die bevindinge gevolglik makliker voorgehou word.

Laastens is daar ook bepaal watter persentasie van die hulpwerkwoorde in al 45 vergelykings wel prakties beduidend verskil.

Totale aantal hulpwerkwoorde in al die vergelykings: $21 \times 45 = 945$

Totale aantal hulpwerkwoorde wat prakties beduidend verskil: $16 + 17 + 19 + 17$ ... (sien tabel 6.5 – die totale van al die hulpwerkwoorde wat prakties beduidend verskil is bymekaar getel) gedeel deur 945 (totale aantal hulpwerkwoorde in al 45 vergelykings) x 100.

$811 \div 945 \times 100 = 85{,}9\%$

Daar kan dus in 85,9 % van al die verglykings van die gebruik van lidwoorde ten opsigte van die ratiobepalings aanngevoer word dat die tekste wel deur verskillende outeurs geskryf is. In die hieropvolgende afdelings word die bevindinge rakende die ratiobepalings van die funsksiewoorde voorgehou aan die hand verduideliking in hierdie afdeling.

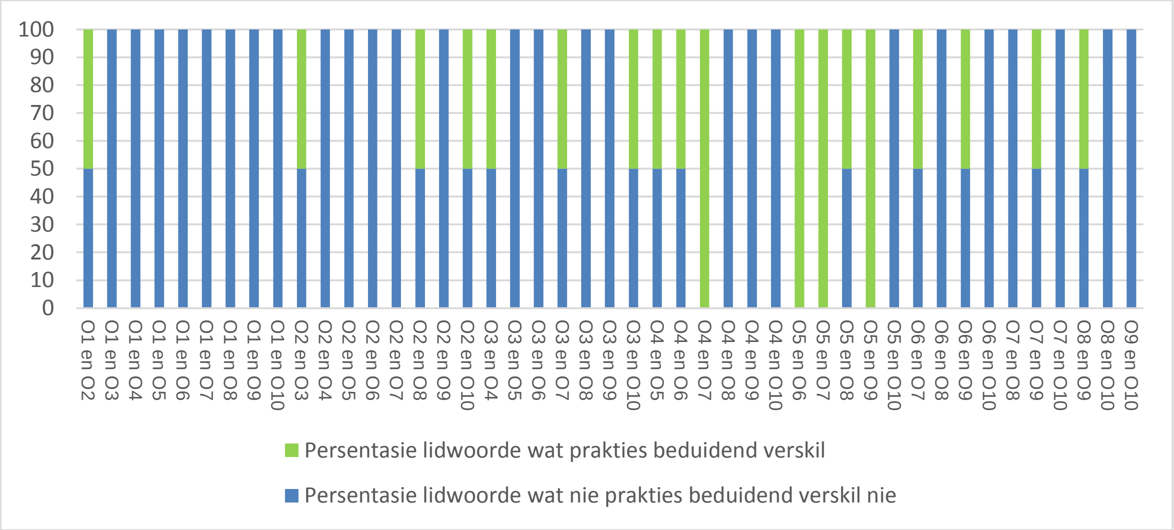
6.3.1 Ratiobepalings: lidwoorde

Die ratio van lidwoorde is soos volg bereken: eerstens, is die korpora geannoteer ten einde die verskillende woordsoorte[183] te kon aandui[184]. Tweedens, is 'n lys opgestel met die lidwoorde en die lys is as 'n txt-dokument gestoor. Derdens, is die onderskeie outeurs se woordelyste in WordSmith Tools oopgemaak en die txt-dokument met die lidwoorde is as 'n stoplys in die program ingevoer. Deur gebruik te maak van die stoplys-funksie, is alle woorde wat nie op die lys verskyn nie (dus alle woorde wat nie lidwoorde is nie) van die woordelyste verwyder. Hierdie

nuwe woordelyste is as Microsoft Excel-dokumente gestoor. Laastens, is die ratio’s bereken[185] en daar is bepaal of die verskil as prakties beduidend beskou kan word.

6.3.1.1 Gespesialiseerde korpus: lidwoorde

Figuur 6.6 toon die bevindinge van hierdie analise van die lidwoorde in die gespesialiseerde korpus aan:



Figuur 6.6 Voorstelling van die persentasie van die prakties beduidend verskil van die ratiobepalings van die lidwoorde

Vanuit Figuur 6.6[186] blyk dit dat met die samevatting van die onderskeie ondersoeke daar in 27 van die ratio-ondersoeke nie ’n prakties beduidende verskil waargeneem is nie en daar kan dus met sekerheid voorgehou word dat hierdie tekste (na aanleiding van die ondersoek na die lidwoordgebruik) deur verskillende outeurs geskryf is.

Die korpora waar daar wel ’n prakties beduidende verskil waargeneem is, is O1 en O2, O2 en O3, O2 en O8, O2 en O10, O3 en O7, O3 en O10, O4 en O5, O4 en O6, O4 en O7, O5 en O6, O5 en O7, O5 en O8, O5 en O9, O6 en O7, O6 en O9, O7 en O9 asook O8 en O9. In hierdie 18 korpora waar lidwoorde voorkom wat nie prakties beduidend verskil nie, is dit telkens die helfte van die lidwoorde wat nie prakties beduidend oor die korpora verskil nie.

Slegs in die ratiovergelykings van die frekwensies van lidwoorde tussen O4 en O7, O5 en O6, O5 en O7 asook O5 en O9 is daar geen lidwoorde wat prakties beduidend verskil nie. Daar kan

[186] Sien bylaag D vir ’n voorbeeld van die ratioberekeninge.

dus slegs in die geval van die bogenoemde vier korpora aangevoer word dat die tekste sonder twyfel deur verskillende outeurs geskryf is. Wanneer ondersoek ingestel is na die kategorieë[187] waarin hierdie korpora val, blyk dit dat die kategorieë waarin die outeurs geval het nie die gebruik van lidwoorde beïnvloed nie aangesien korpora vanuit verskillende kategorieë nie ŉ praktiese beduidende verskil toon nie.

Tabel 6.6 Uitbeelding van die ratioberekeninge van die lidwoorde in die korpora wat nie ŉ prakties beduidende verskil getoon het nie.

	O4 en O7	O5 en O6	O5 en O7	O5 en O9
Die	0,92	0,87	0,87	1,09
ŉ	084	0,84	0,84	0,85

Tabel 6.6 toon al die lidwoorde wat 100% nie prakties beduidend verskil nie. Soos genoem in afdeling 6.3 (Ratiobepalings), kan ŉ woord as prakties beduidend beskou word wanneer die ratio > 1,2 of < as 0,8 is (vgl. Van Rooy & Terblanche, 2006). Daar kan dus met redelike sekerheid gestel word dat die tekste deur verskillende outeurs geskryf is. In Tabel 6.6 toon al die voorbeelde > 1,2 of < 0,8 en daar kan dus met sekerheid aangevoer word dat die tekste deur verskillende outeurs geskryf is.

Ten einde te kon bepaal hoe suksesvol die ratio-analise van die lidwoorde is, is die gemiddeld van al die korpora se lidwoorde bereken wat wel ŉ prakties beduidende verskil toon. Hierdie gemiddeld is soos volg bereken: die totale som van al die lidwoorde wat prakties beduidend verskil in al tien die korpora is bereken en ŉ totaal van 68 is gevind. Daarna is die som van al die lidwoorde wat nie prakties beduidend verskil nie bereken en ŉ totaal van 22 is gevind. Ten einde ŉ gemiddeld te kon bereken, is elk van die totale deur die totale aantal lidwoorde (90) wat in al die vergelykings gevind is, gedeel en met 100 vermenigvuldig:

$$\therefore 68 \div 90 \times 100 = 75{,}5 \; (prakties\ beduidende\ verskil)$$

en

$$\therefore 22 \div 90 \times 100 = 24{,}5 \; (nie\ prakties\ beduidende\ verskil\ nie)$$

Wanneer die gemiddeld van die ratioberekeninge (oor al tien korpora) bereken is, blyk dit dat daar 75,5% gevalle is waar lidwoorde prakties beduidend verskil terwyl daar slegs in 24,5% van die gevalle aangevoer kan word dat daar nie ŉ prakties beduidende verskil voorkom nie. Dus kan daar in 75,5% van die ratioberekeninge ten opsigte van die praktiese beduidendheid van die lidwoorde aangevoer word dat die ondersoek suksesvol is aangesien die korpora wat ondersoek is wel deur verskillende outeurs geskryf is.

[187] Sien hoofstuk 5 (afdeling 5.4 – Data-insameling en -verwerking).

Ondersoek is ingestel na die individuele ratiovergelykings en die persentasie lidwoorde wat in elke vergelyking ’n prakties beduidende verskil aantoon. Soos reeds genoem, is daar in 18 van die 45 vergelykings bevind dat daar nie ’n prakties beduidende verskil voorkom nie. Dus kan daar in 27 van die lidwoorde se ratiovergelykings met redelike sekerheid aangevoer word dat die korpora wel deur verskillende outeurs geskryf is. Ook vir hierdie bevinding is daar ’n persentasie bereken. Hierdie persentasie is soos volg bereken:

$$Hoeveelheid \text{ korpora waarin die } lidwoorde\ prakties\ beduidend\ \text{verskil} \div totale\ hoeveelheid\ lidwoorde\ \times 100$$

$$\therefore 27 \div 45 \times 100 = 60\,\%$$

en

$$\text{H}oeveelheid \text{ korpora waar die } lidwoorde\ \ nie\ prakties\ beduidend\ \ verskil\ nie \div totale\ hoeveelheid\ lidwoorde\ \times 100$$

$$\therefore 18 \div 45 \times 100 = 40\%$$

Wanneer die persentasie bereken is om te bepaal of daar wel ’n prakties beduidende verskil is in die ratiobepalings van die lidwoorde in terme van al die korpora wat ondersoek is, kan daar aangevoer word dat daar in 60% van die ondersoeke wel sonder redelike twyfel gestel kan word dat die korpora deur verskillende outeurs geskryf is.

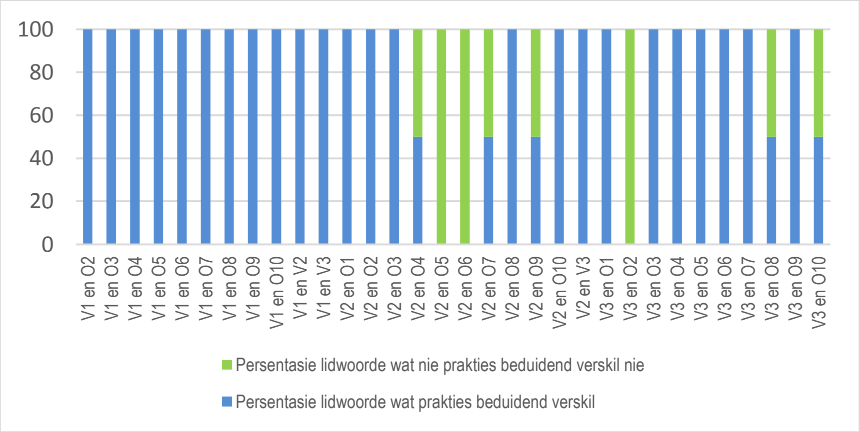
Laastens, is daar na elk van die individuele vergelykings ondersoek ingestel. In die geval van die lidwoorde is daar vasgestel dat die laagste persentasie van praktiese beduidendheid 50% is (O1 en O2, O2 en O3, O2 en O5, O2 en O8, O2 en O10, O3 en O4, O3 en O7, O3 en O10. O4 en O5, O4 en O6, O5 en O8, O6 en O7, O6 en O9, O7 en O9 asook O8 en O9). In die vergelykings van O4 en O7, O5 en O6, O5 en O7 asook O5 en O9 is al die lidwoorde wat ondersoek is prakties beduidend en in dié geval kan daar aangevoer word dat hierdie korpora deur verskillende outeurs geskryf is.

Vanuit die ratioberekeninge van die lidwoorde blyk dit dus dat daar in 60% van die vergelykings bewys kan word dat die korpora deur verskillende outeurs geskryf is. Vanuit die 40% gevalle waar die ratioberekeninge nie prakties beduidend is nie, en waar daar dus nie sonder twyfel aangevoer kan word dat die korpora deur verskillende outeurs geskryf is nie, is die laagste persentasie praktiese beduidendheid 50% – dus nie oorwegend verskillend of dieselfde nie oftewel 31,2%. In die geval van die vergelykings tussen O4 en O7, O5 en O6, O5 en O7 sowel as O5 en O9 is al die lidwoorde prakties beduidend. Daar kan dus slegs in 8,8% bewys word dat die lidwoordgebruik van die outeurs só ooreenstem dat daar nie bewys kan word dat die tekste deur verskillende outeurs geskryf is nie.

Verder blyk dit dat outeurs wel soms lidwoorde uitlaat ten einde by die karakterbeperking van die betrokke platform te kan hou. Tog is dit kenmerkend dat die outeurs, waar moontlik, wel van lidwoorde gebruik maak. Die gebruik van lidwoorde kan dus as ŉ eienskap van Afrikaans beskou word soos dit op Twitter gebruik word. Daar kan verder aangevoer word dat die ratiobepalings van lidwoorde as ŉ suksesvolle analise beskou kan word ten einde outeurs van mikroblogtekste te identifiseer wanneer die analise saam met verdere analises gebruik word.

6.3.1.2 Verdagte tekste: lidwoorde

Die analise van die gebruik van lidwoorde is uitgevoer soos in afdeling 6.3.1.1 voorgehou. Hierdie analise het ten doel gehad om te bepaal of die gebruik van lidwoorde moontlik gebruik kan word om outeurs van mikrobloginskrywings te identifiseer. In afdeling 6.3.3.1 is daar bepaal dat wanneer die ratiobepalings van lidwoorde in die gespesialiseerde korpus ondersoek word, dit wel genoegsame bewyse oor die outeurskap van mikroblogtekste bied en dat dit in die proses van outeurskapidentifikasie ondersoek moet word. Figuur 6.7 toon die bevindinge van hierdie analise aan:



Figuur 6.7 Voorstelling van die persentasie van prakties beduidende verskil van die ratiobepalings van die lidwoorde

Figuur 6.7 toon die bevindinge van die analise van die prakties beduidende verskil van die lidwoorde aan wat deur die drie verdagte tekste gebruik is. Vanuit hierdie analise kan daar etlike afleidings oor die gebruik van lidwoorde gemaak word.

In die vergelyking van die tien bekende outeurs se korpora met die eerste verdagte outeur (V1) blyk dit dat daar geen lidwoorde voorkom wat prakties beduidend verskil nie; daarom kan daar aangevoer word dat die korpora wat vergelyk is almal deur verskillende outeurs geskryf is. Vanuit Figuur 6.7 blyk dit dat die vergelykings tussen die onderskeie korpora met V1 100% prakties beduidend verskil. Dit wil sê dat geen van die bekende outeurs se gebruik van lidwoorde enigsins met die lidwoordgebruik van V1 ooreenstem nie.

Wanneer die tekste van V2 met die korpora van die bekende outeurs (sien Figuur 6.7) vergelyk is, is dit duidelik dat V2 se gebruik van lidwoorde baie meer ooreenstem met die gebruik van lidwoorde soos gebruik deur die bekende outeurs. Figuur 6.7 dui aan dat vanuit die 11 vergelykings ten opsigte van die gebruik van lidwoorde, ses vergelykings voorkom waar daar geen lidwoorde is wat nie prakties beduidend verskil nie – dus ses vergelykings waaruit dit duidelik is dat die tekste na gelang van die gebruik van lidwoorde moontlik deur verskillende outeurs geskryf is.

In die oorblywende vyf vergelykings is daar drie (V2 en O4, V2 en O7 asook V2 en O9) vergelykings waarin 50% van die lidwoorde wat deur die onderskeie outeurs gebruik is nie ’n prakties beduidend verskil nie – dus is daar ’n 50% kans dat die tekste deur dieselfde outeurs geskryf is. In die laaste twee (V2 en O5 asook V2 en O6) vergelykings is daar ’n 100% kans dat die tekste deur dieselfde outeurs geskryf is (geen prakties beduidende verskil nie). Vanuit die analise oor die praktiese beduidendheid van die gebruik van lidwoorde deur V2 blyk dit dat O5 en O6 die moontlike outeurs van die tweede verdagte teks is.

Die laaste verdagte teks wat geanaliseer is ten opsigte van die lidwoordgebruik was V3. Ook in hierdie vergelykings stem die bekende outeurs se lidwoordgebruik ooreen met die gebruik van lidwoorde soos gesien in V3 (sien Figuur 6.7). Alhoewel V2 se gebruik van lidwoorde met meer tekste ooreenstem, is daar drie tekste (V3 en O2 – 100% nie prakties beduidend nie, V3 en O8 – 50% prakties beduidend en V3 en O10 – 50% prakties beduidend) wat ooreenkomste met die lidwoordgebruik van V3 toon. Van hierdie drie tekste is die vergelyking van V3 en O2 die enigste vergelykings wat aantoon dat O2, na gelang van die gebruik van lidwoorde, die outeur van V3 kan wees terwyl daar ’n 50% kans is dat O8 en O10 die outeurs van V3 is.

Die gebruik van lidwoorde toon dus nie sonder enige twyfel aan wie die moontlike outeurs van die verdagte tekste is nie. Veral in die geval van V1 is daar geen aanduiding dat die verdagte teks met enige van die bekende outeurs se tekste ooreenstem nie. Verdere ondersoeke is dus uitgevoer om die moontlike outeurs van die verdagte tekste te kon identifiseer. Vervolgens word die ratio-analise van voornaamwoorde verskaf.

6.3.2 Ratiobepalings: voornaamwoorde

Net soos in die ondersoek na die lidwoorde, was die eerste stap in die ratiobepaling van die voornaamwoorde[188] om die geannoteerde korpora te ondersoek en die voornaamwoorde (soos geannoteer) te identifiseer. Die geïdentifiseerde voornaamwoorde is as ’n lys in txt-formaat

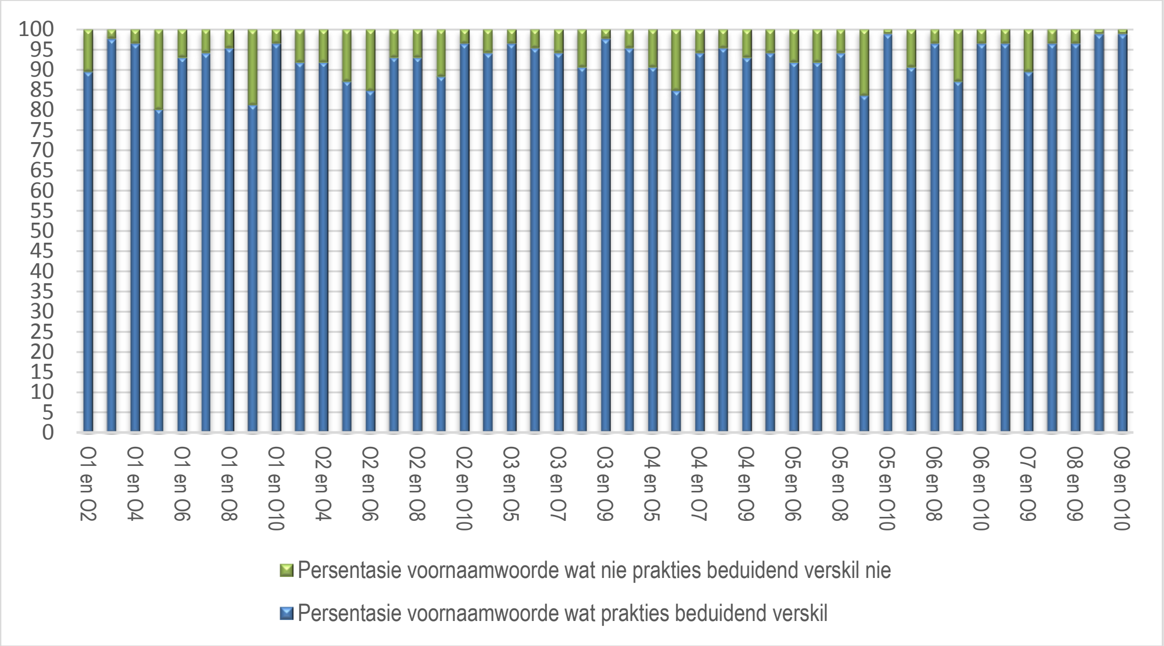
[188] Die voornaamwoorde is deur middel van die woordsoortetiketeerder, soos deur CTexT verskaf, geannoteer.

gestoor. Die korpora is in WordSmith Tools oopgemaak en die txt-dokument met die voornaamwoorde is as ŉ stoplys in WordSmith Tools ingevoer.

Alle woorde wat nie op hierdie stoplys voorgekom het nie is vanuit die onderskeie sleutelwoordlyste verwyder. Hierdie nuwe sleutelwoordlyste is as ŉ Microsoft Excel-dokument gestoor. 86 voornaamwoorde is in die onderskeie korpora geïdentifiseer en geanaliseer ten einde die ratio's te kon bepaal. Eerstens, word daar na die bevindinge vanuit die gespesialiseerde korpus verwys alvorens die verdagte tekste se bevindinge voorgehou word.

6.3.2.1 Gespesialiseerde korpus: voornaamwoorde

Figuur 6.8 beeld die bevindinge van die analise van die ratiobepalings van die voornaamwoorde[189] uit:



Figuur 6.8 Voorstelling van die prakties beduidende verskil van die gebruik van voornaamwoorde (persentasie)

Vanuit Figuur 6.8 blyk dit dat al die korpora wat ondersoek is, wel ŉ prakties beduidende verskil toon wanneer die ratio's van die voornaamwoorde ondersoek is. Alhoewel daar in elk van die 45 vergelykings wel bevind is dat daar voornaamwoorde voorkom waarvan die frekwensie van voorkoms nie prakties beduidend verskil nie, is hierdie voornaamwoorde in elk van die vergelykings 20% (oftewel 17 voornaamwoorde wat nie prakties beduidend verskil nie) of minder. O1 en O5 stem die meeste ooreen met betrekking tot die gebruik van voornaamwoorde. In hierdie vergelyking toon 17 van die 86 voornaamwoorde nie ŉ prakties beduidende verskil nie, en dit blyk dat die voornaamwoorde deur dieselfde outeur gebruik is. Die vergelykings tussen O5 en O10, O8 en O9 asook O9 en O10 toon die laagste voorkoms van

[189] Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

voornaamwoorde aan wat nie prakties beduidend verskil nie (slegs een voornaamwoord in elk van hierdie drie vergelykings is nie prakties beduidend nie).

Ten einde te kon bepaal hoe suksesvol die ratio-analise van die voornaamwoorde is, is die gemiddeld bereken van al die korpora se voornaamwoorde wat wel ’n prakties beduidende verskil aantoon. Hierdie gemiddeld is bereken soos verduidelik in afdeling 6.3.1.1:

$$\therefore 3591 \div 3870 \times 100 = 92{,}8\% \ (prakties\ beduidende\ verskil)$$

en

$$\therefore 279 \div 3870 \times 100 = 7{,}2\% \ (nie\ prakties\ beduidende\ verskil\ nie)$$

Na gelang van die persentasieberekeninge van die prakties beduidende verskil van die ratio’s van die voornaamwoorde wanneer die onderskeie korpora met mekaar vergelyk is, blyk dit dat daar in 92,8% van die geval sonder redelike twyfel aangevoer kan word dat die tekste deur verskillende outeurs geskryf is. Hierdie 92,8% verwys na al 45 van die vergelykings wat tussen die onderskeie korpora uitgevoer is.

Toe daar na die individuele resultate van die vergelykings ondersoek ingestel is, is daar bevind dat (soos gesien in Figuur 6.8) daar in al 45 van die vergelykings wel voornaamwoorde voorkom wat nie prakties beduidend verskil nie. Alhoewel daar in elk van die 45 vergelykings voornaamwoorde voorkom wat nie prakties beduidend verskil nie, dui die aantal voornaamwoorde daarop dat die tekste heel moontlik nie deur dieselfde outeur geskryf is nie.

Die vergelyking tussen O1 en O5 toon die grootste voorkoms van voornaamwoorde (17) wat nie prakties beduidend verskil nie. Die persentasie van voornaamwoorde wat dus daarop kan dui dat O1 en O5 dieselfde outeur is, is slegs 19,8% − die aantal voornaamwoorde wat prakties beduidend verskil. Dit kan dus daarop dui dat O1 en O5 wel verskillende outeurs is − 80,2%. Daar kan dus selfs in die geval van die ratiobepalings van die korpora wat die meeste ooreenstem ten opsigte van voornaamwoordgebruik, aangevoer word dat die tekste wel deur verskillende outeurs geskryf is.

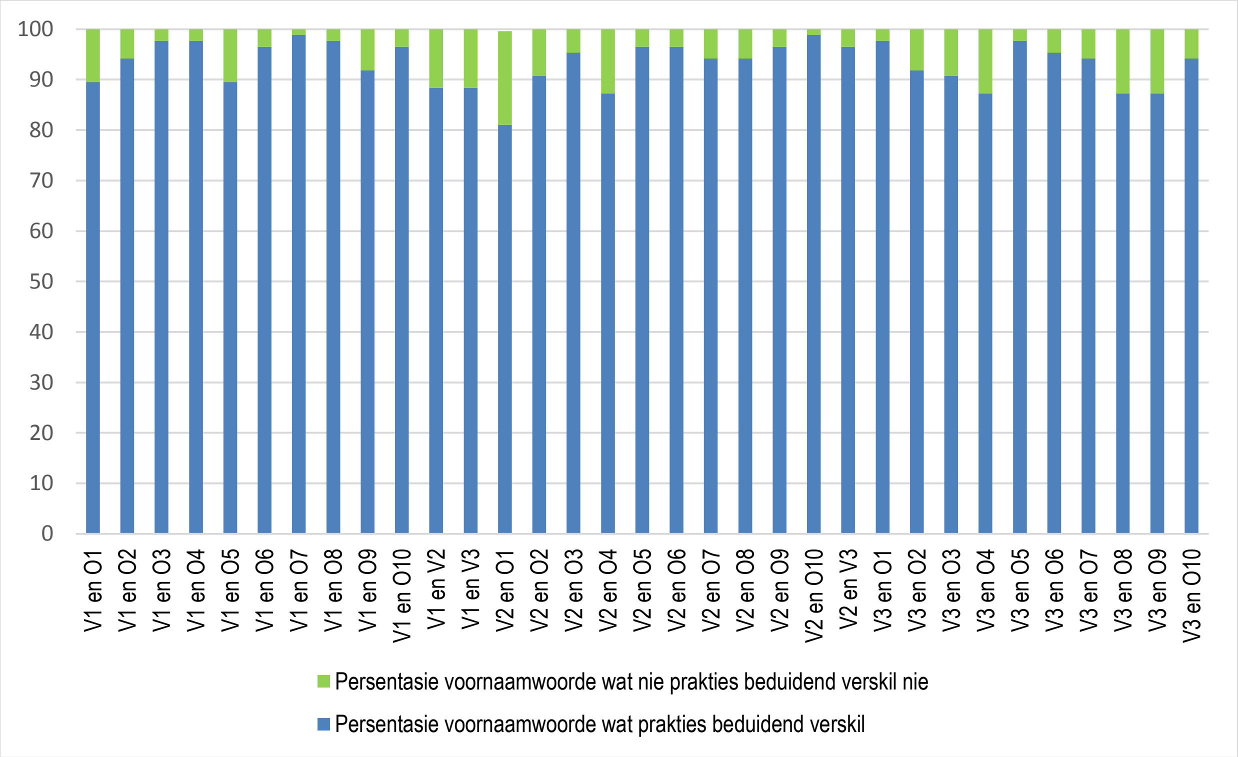
Wanneer die oorblywende 44 vergelykings verder geanaliseer is in terme van die onderskeie outeurs se voornaamwoordgebruik, blyk dit dat die korpora van O5 en O10, O8 en O9 asook O9 en O10 die minste ooreenstem in hulle gebruik van voornaamwoorde. 98,8% van voornaamwoorde wat in hierdie korpora gebruik word, verskil prakties beduidend (dus is die tekste deur twee verskillende outeurs geskryf) terwyl slegs 1,2% van die voornaamwoorde nie prakties beduidend verskil nie (dus deur dieselfde outeur geskryf).

Aangesien die praktiese beduidende verskil van die voornaamwoorde in die onderskeie korpora tussen 80,2% en 98,8% is, kan daar aangevoer word dat die ondersoek na die praktiese

beduidendheid van voornaamwoorde − wanneer outeurs van mikroblogs geïdentifiseer moet word − as suksesvol beskou kan word.

6.3.2.2 Verdagte tekste: voornaamwoorde

Die ratiobepalings van die voornaamwoorde − soos die woordsoorte deur die onderskeie outeurs in hulle korpora gebruik − is uitgevoer soos in afdeling 6.3 bespreek. Vanuit die data wat in hierdie analise verwerk is, kan die volgende afleidings gemaak word (sien Figuur 6.9):



Figuur 6.9 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die voornaamwoorde

In die ondersoek na die praktiese beduidendheid van die voornaamwoorde is daar soortgelyk aan die ondersoek na die gebruik van lidwoorde ŉ vergelykende analise uitgevoer ten einde die gebruik van voornaamwoorde deur die tien bekende outeurs te vergelyk met die gebruik van voornaamwoorde soos dit in die drie verdagte tekste voorkom.

Figuur 6.9 beeld die verwerkte data vanaf die ratiobepalings uit. Vanuit die vergelykings met die eerste verdagte teks (V1) word daar in elk van die vergelykings wel voornaamwoorde aangetref wat nie prakties beduidend verskil nie, maar daar is nooit meer as 12 voornaamwoorde nie (slegs in die geval van die vergelykings van V1 en V2 asook V1 en V3 – oftewel die verdagte tekste).

Wanneer die aantal voornaamwoorde wat nie prakties beduidend verskil tussen V1 en die tien bekende outeurs ondersoek is, blyk dit dat V1 en O9 asook V1 en O2 se gebruik van voornaamwoorde die meeste ooreenstem (V1 en O1 – 10% van die voornaamwoorde verskil nie prakties beduidend nie, V1 en O5 – 10% van die voornaamwoorde verskil nie prakties beduidend nie). Dit wil sê dat daar 10% van die voornaamwoorde wat deur V1, O1 en O5 gebruik word só ooreenstem dat daar aangevoer kan word dat die tekste deur dieselfde outeur geskryf is.

In die oorblywende sewe vergelykings is daar ŉ 94% tot 99% kans dat die tekste wel deur verskillende outeurs geskryf is. Dit kom voor asof die ratiobepalings van V1 se voornaamwoordgebruikweer eensas ŉ proses van eliminasie beskou kan word, en dat die moontlike outeurs van die teks wat die hoogste mate van ŉ prakties beduidende verskil aantoon ten opsigte van hulle voornaamwoordgebruik (wanneer dit met die eerste verdagte teks vergelyk word) nie as die moontlike outeurs van V1 beskou kan word nie.

Na gelang van die ondersoek na die gebruik van voornaamwoorde kan daar vervolgens aangevoer word dat O1, O2, O5 en O9 steeds as moontlike outeurs van V1 beskou kan word terwyl die ander bekende outeurs na gelang van die ondersoek na die gebruik van voornaamwoorde nie as moontlike outeurs beskou kan word nie.

Die ratiobepalings van die vergelyking tussen V2 en die tien bekende outeurs (sien Figuur 6.9) word vervolgens bespreek om te bepaal of daar prakties beduidende verskille in die voornaamwoordgebruik van die onderskeie outeurs plaasgevind het.

Vanuit die vergelykings wat uitgevoer is, is daar geen vergelyking in die voornaamwoordgebruik in die twee vergelykende korpora waar daar ŉ prakties beduidende verskil voorgekom het nie. In die vergelyking van die korpora van V2 en O10 het 1,1% van die voornaamwoorde wat deur die twee outeurs gebruik is nie ŉ prakties beduidende verskil getoon nie. 1,1% van die voornaamwoorde wat in die korpora van hierdie twee outeurs aangetref is, stem tot so ŉ mate ooreen dat daar aangevoer kan word dat die tekste deur dieselfde outeurs geskryf is.

V2 en O10 is die twee korpora wat die kleinste ooreenkoms ten opsigte van die gebruik van voornaamwoorde toon. Wanneer daar spesifiek ondersoek ingestel is na die korpora waar daar groter ooreenkomste in die gebruik van voornaamwoorde voorkom, is daar twee vergelykings (V2 en O1 – 18,6%, en V2 en O4 – 12,8%) wat ten opsigte van die ander vergelykings uitstaan in terme van die aantal voornaamwoorde wat as prakties beduidend beskou kan word – sien ook Figuur 6.9. Die aanname kan dus gemaak word dat O1 as moontlike outeur van die teks

van V2 beskou kan word na gelang van die ratiobepalings van die voornaamwoorde soos deur die onderskeie outeurs gebruik.

Die ratiobepalings van die derde verdagte teks (V3) word in Figuur 6.9 aangetoon − die voornaamwoordgebruik in hierdie korpora is vergelyk met die voornaamwoordgebruik in die tien bekende outeurs se korpora. Vanuit hierdie analise blyk dit dat voornaamwoorde in geen van die korpora tot so 'n mate gebruik word dat daar sonder twyfel aangevoer kan word dat die prakties beduidende verskil toon dat die tekste deur verskillende outeurs geskryf is nie.

Die vergelykings van V3 en O4, V3 en O8 asook V3 en O9 toon die grootste ooreenkomste. In al drie hierdie vergelykings (V3 en O4, V3 en O8 asook V3 en O9) verskil die voornaamwoordgebruik nie prakties beduidend nie (12,79%). In 12,79% van die keer wat voornaamwoorde gebruik is, kan daar aangevoer word dat dieselfde outeur die tekste geskryf het. In die oorblywende agt vergelykings het die persentasie voornaamwoorde wat nie prakties beduidend verskil nie tussen 2,32% en 9,3% gewissel. O4, O8 en O9 is dus as die beste kandidate vir die moontlike outeur van V3 beskou na gelang van die ratiobepalings van die voornaamwoorde wat deur die onderskeie outeurs gebruik is.

Die ondersoek na die ratiobepalings van die gebruik van voornaamwoorde, soos in die vorige ondersoeke, kan dus ook as 'n proses van eliminasie beskou word. Outeurs wat die minste ooreenkomste in terme van die gebruik van voornaamwoorde toon, is nie as moontlike outeurs van die verdagte tekste (V1, V2 en V3) beskou nie.

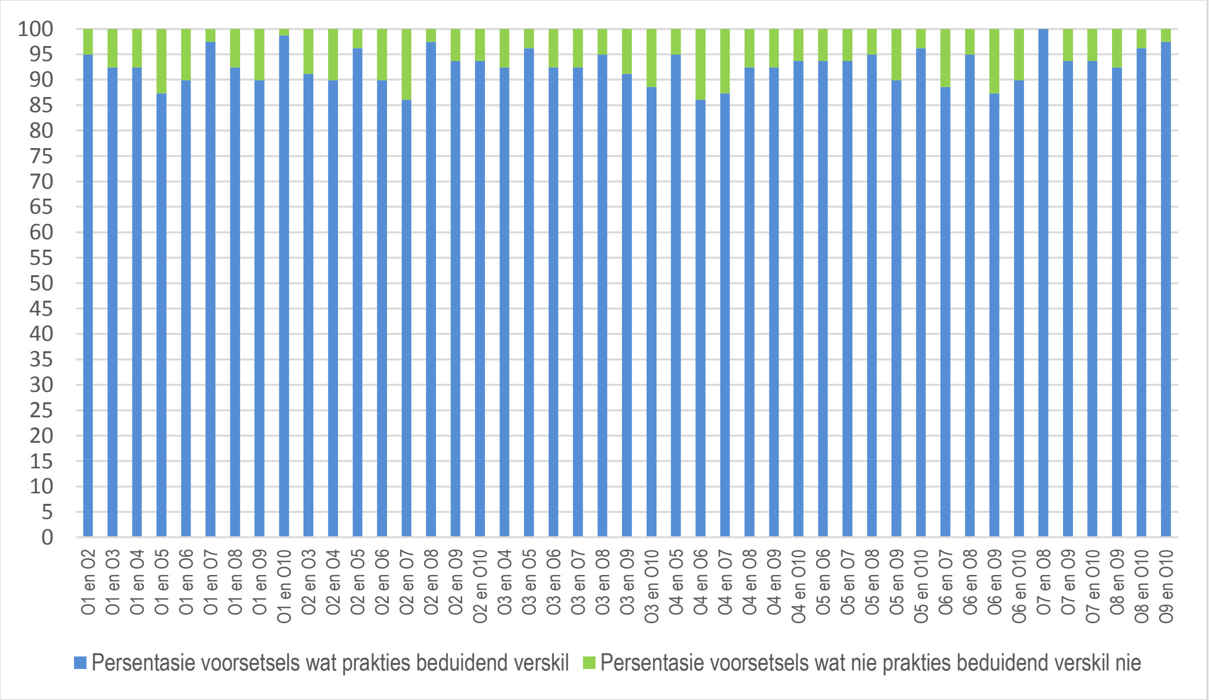
6.3.3 Ratiobepalings: voorsetsels

Ten einde die ratio van die voorsetsels[190] te kon bepaal, is daar gebruik gemaak van die tekste wat geannoteer is om die verskillende woordsoorte aan te dui. Soos in die geval van afdeling 6.3.1 en 6.3.2, is die voorsetsels as 'n txt-dokument gestoor en nadat die sleutelwoordlyste in WordSmith Tools oopgemaak is as 'n stoplys in die program ingevoer. Alle woorde wat nie op hierdie lys van voorsetsels (die stoplys) verskyn het nie, is outomaties vanuit die sleutelwoordlyste verwyder. Hierdie nuwe lyste is vanuit WordSmith Tools as Microsoft Excel-dokumente gestoor. 79 voornaamwoorde is in die onderskeie korpora geïdentifiseer en die ratio's van elk van hierdie vergelykings (sien afdeling 6.3.1) is bereken[191].

[190] Die voorsetsels is deur middel van die woordsoortetiketeerder, soos deur CTexT verskaf, geannoteer.
[191] Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

6.3.3.1 Gespesialiseerde korpus: voorsetsels

Aangesien daar 45 vergelykings uitgevoer is om die ratio's van die voorsetsels te kon bereken, verwys ek net opsommend na die bevindinge in die geval van lidwoorde en voornaamwoorde. Figuur 6.10 toon die voorsetsels aan wat volgens die ratiobepalings oor ŉ prakties beduidende verskil beskik het of nie:



Figuur 6.10 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die voorsetsels

Vanuit Figuur 6.10 blyk dit dat daar in slegs in die vergelyking tussen O7 en O8 geen voorsetsels deur die outeurs gebruik word wat nie prakties beduidend verskil nie. In die oorblywende 44 (van die 45 vergelykings) is daar wel voorsetsels nie prakties beduidend verskil nie. Dus kan daar sonder twyfel aangevoer word dat die tekste, soos ondersoek in die vergelyking van O7 en O8, deur verskillende outeurs geskryf is.

Ten einde te kon bepaal hoe suksesvol die ratio-analise van die voorsetsels was, is die gemiddeld van al die korpora se voorsetsels bereken wat wel ŉ prakties beduidende verskil aantoon. Hierdie gemiddeld is soos volg bereken (verduidelik in afdeling 6.3.1):

$$\therefore 3295 \div 3555 \times 100 = 92{,}7\% \ (\mathit{prakties\ beduidende\ verskil})$$

en

$$\therefore 260 \div 3555 \times 100 = 7{,}3\% \ (\mathit{nie\ prakties\ beduidende\ verskil\ nie})$$

Na gelang van die persentasieberekeninge van die prakties beduidende verskil van die ratio's van die voorsetsels wanneer die onderskeie korpora met mekaar vergelyk is, blyk dit dat daar in 92,7% van die geval sonder redelike twyfel aangevoer kan word dat die tekste deur verskillende outeurs geskryf is. Hierdie 92,7% verwys na al 45 vergelykings wat tussen die onderskeie korpora uitgevoer is. Die oorblywende 7,3% van die voorsetsels wat wel nie prakties beduidend verskil nie, is in 44 van die 45 vergelykings van die korpora aangetref (slegs in die vergelyking van O7 en O8 is daar geen voorsetsels gevind wat prakties beduidend verskil nie).

Ten einde te kon bepaal hoe die individuele resultate van die voorsetselondersoek (ratiobepalings) daaruitsien, is daar na die resultate van die ratiobepalings ondersoek ingestel soos uiteengesit in Figuur 6.10. Soos reeds genoem, het daar slegs in die vergelyking van O7 en O8 geen voornaamwoorde voorgekom wat prakties beduidend verskil nie. Daarom is daar geen twyfel dat hierdie twee tekste nie deur verskillende outeurs geskryf is nie.

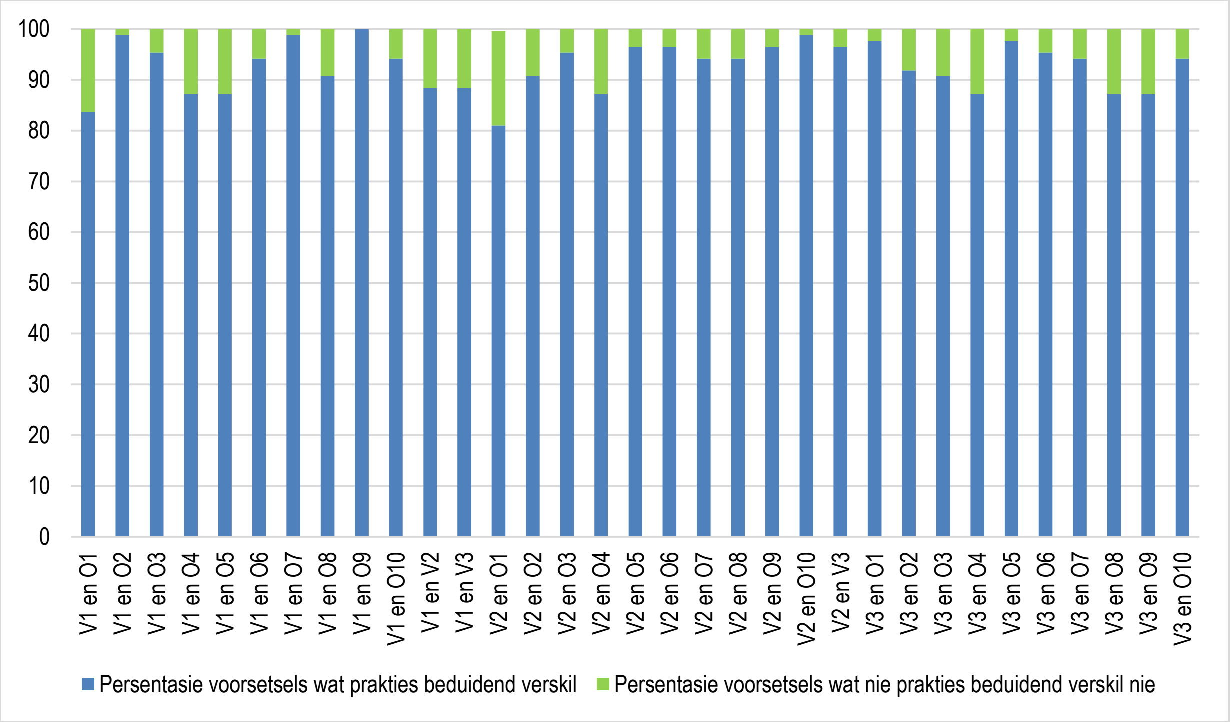
In die oorblywende 44 vergelykings is daar bevind dat die voorsetselgebruik van die onderskeie outeurs wel insiggewend is. Die laagste persentasie voorsetsels wat prakties beduidend is in die onderskeie vergelykings word in die vergelyking van O2 en O7 asook O4 en O6 aangetref. Vanuit hierdie ratiobepaling tussen O2 en O7 en O4 en O6 is daar bepaal dat slegs 86,1% van die ratiobepalings prakties beduidend verskil. Dit wil sê dat die laagste persentasie waar die ratiobepalings van die gebruik van voorsetsels in die onderskeie vergelykings van die korpora aantoon dat daar wel 'n prakties beduidende verskil voorkom (dat die tekste wel deur verskillende outeurs geskryf is) 86,1% is en dat in hierdie vergelyking van die korpora daar 'n 13,9% kans is dat die tekste deur dieselfde outeur geskryf is[192]. In die oorblywende 43 tekste wissel die persentasies en daar 'n praktiese beduidende verskil is waargeneem − tussen 88,6% en 100%. Dit dui daarop dat die gebruik van voorsetsels − wanneer die ratio-analises uitgevoer is − wel beskou kan word as 'n aanduider van outeurskap.

Hierdie bevinding van die persentasies dui daarop dat daar sonder redelike twyfel aangevoer kan word dat die voorsetselgebruik van verskillende outeurs op Twitter verskil. Om die praktiese beduidende verskil te kon bereken, is al 45 van die vergelykings ondersoek: hoeveel van die voorsetsels deur die onderskeie outeurs gebruik is wanneer die ratiobepalings uitgevoer is tesame met die feit dat die gemiddelde voorkoms van voornaamwoorde soos bereken deur die ratiobepalings 92,7% is. Daarom kan daar aangevoer word dat die ratiobepalings, wanneer voorsetsels ondersoek word, as 'n suksesvolle middel tot die identifisering van outeurs op mikroblogs beskou kan word.

[192] Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

6.3.3.2 Verdagte tekste: voorsetsels

Ten einde die ondersoek na die ratiobepalings van die voorsetsels te kon uitvoer, is die onderskeie korpora (soos in afdeling 6.3.3 voorgehou) vergelyk. Die bevindinge is as ŉ Microsoft Excel-dokument gestoor en Figuur 6.11 bied ŉ opsomming van hierdie bevindinge:



Figuur 6.11 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die voorsetsels

Die ondersoek na die prakties beduidende verskil van die voorsetsels, soos opgesom in Figuur 6.11, toon aan dat die voorsetsels op só ŉ manier deur die onderskeie outeurs gebruik word dat daar aangevoer kan word dat die tekste waarin die voorsetsels voorkom deur dieselfde outeur geskryf is.

Wanneer daar spesifiek na die vergelyking van die voorsetselgebruik van die eerste verdagte outeur (V1) en die bekende outeurs se tekste ondersoek ingestel is, blyk dit dat die korpora van O1 (16,27%), O4 (12,79%), O5 (12,79%), V2 (11,62%) en V3 (11,62%) die grootste ooreenkomste ten opsigte van die gebruik van voorsetsels toon. Die voorsetsels in die oorblywende ses vergelykings waar daar nie ŉ prakties beduidende verskil voorkom nie wissel tussen 9,30% en 0% (die vergelykings tussen die tekste van V1 en O9 toon geen voorsetsels wat prakties beduidend verskil nie en dus kan daar sonder enige twyfel aangevoer word dat O9 nie die outeur van die eerste verdagte teks – V1 – is nie).

Vanuit die ratio-ondersoek na die tweede verdagte teks (V2) ten opsigte van die gebruik van voorsetsels (sien Figuur 6.11) blyk dit dat die tekste van V2 en O1 oor die meeste voorsetsels beskik wat nie prakties beduidend verskil nie − 18,6% van die voorsetsels wat deur die twee outeurs gebruik word, stem só ooreen dat dit voorkom asof die voorsetsels deur dieselfde outeur gebruik word. In die oorblywende tien vergelykings wissel die voorsetselgebruik tussen 12,7% en 1,16% met geen prakties beduidende verskil nie.

Hierdie ratiobepalings toon dus aan dat daar in elk van die 11 vergelykings van die korpora wel voorsetsels voorkom wat deur die onderskeie outeurs gebruik word wat nie prakties beduidend verskil nie. Daar is dus voorsetsels deur elk van die outeurs gebruik wat só ooreenstem dat daar aangevoer kan word dat daar moontlike ooreenkomste in outeurskap kan wees. Daar is in geen van die vergelykings meer as 18,6% voorsetsels wat só ooreenstem dat daar aangevoer kon word dat die tekste wat ooreengestem deur dieselfde outeur geskryf is nie.

O1 toon na gelang van die ratiobepalings van voorsetsels die grootste ooreenkoms met V2 en daarom kan die aanname gemaak word dat O1 as die beste kandidaat vir outeurskap van V2 beskou kan word.

Die derde verdagte teks (V3) se ratiobepalings (sien Figuur 6.11) van die voorsetsels soos dit deur die onderskeie outeurs gebruik word, toon ’n maksimum van 12,79% voorsetselgebruik wat nie prakties beduidend verskil nie. In die twee vergelykings (V3 en O8 asook V3 en O9) verskil 12,79% van die voorsetsels wat in die twee korpora aangetref word nie prakties beduidend nie en dit blyk dus asof dit deur dieselfde outeur geskryf is.

In die oorblywende agt vergelykings wissel die persentasie voorsetsels tussen 2,32% en 9,30% wat nie prakties beduidend verskil nie. Die aantal voorsetsels wat só ooreengestem dat dit blyk dat die tekste deur dieselfde outeurs geskryf is, is tot die minimum beperk. Na gelang van die ondersoek na ’n praktiese beduidende verskil, of nie, van die voorsetsels in die derde verdagte teks kan daar aangevoer word dat O4 en O9 die moontlike outeurs van die derde verdagte teks (V3) kan wees.

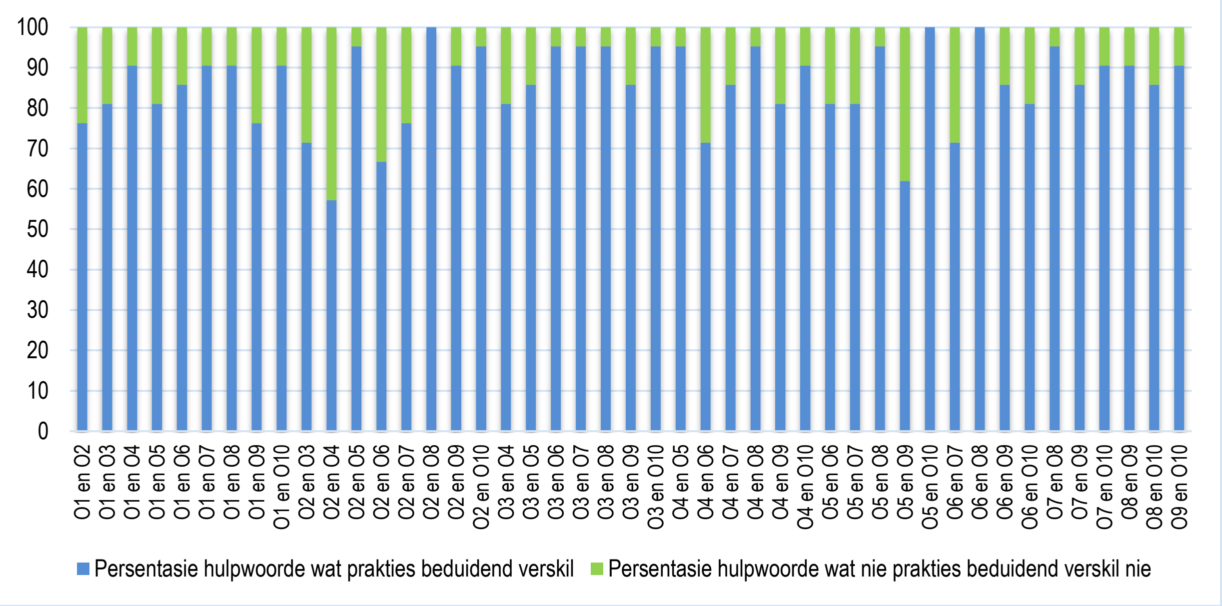
Die ondersoek na die voorkoms van voorsetsels en die ratiobepalings dui gevolglik aan dat V1 en O1, V2 en O1 en V3 en O4 en O9 oor die grootste aantal voorsetsels beskik wat prakties beduidend verskil en wel deur dieselfde outeur geskryf is na gelang van die ratiobepalings. In die hieropvolgende afdeling word daar ondersoek ingestel na die voorkoms van hulpwerkwoorde.

6.3.4 Ratiobepalings: hulpwerkwoorde

Die ratio's van die hulpwerkwoorde[193] is op dieselfde manier as die ratio's van die lidwoorde, voornaamwoorde en voorsetsels bepaal. Die geannoteerde tekste is geanaliseer en ŉ lys is saamgestel van al die hulpwerkwoorde wat in elk van die korpora voorgekom het. Hierdie lys met hulpwerkwoorde is as ŉ txt-dokument gestoor. Die sleutelwoordlyste van die onderskeie vergelykings van die korpora is in WordSmith Tools oopgemaak en die txt-dokument met die hulpwerkwoorde is as ŉ stoplys in die program ingevoer. Al die woorde wat nie op die lys voorgekom het nie is outomaties verwyder. Hierdie nuwe WordSmith Tools lyste het nou uit slegs die hulpwerkwoorde van die twee korpora wat met mekaar vergelyk is, bestaan en is as Microsoft Excel-dokumente gestoor. 21 hulpwerkwoorde is in die onderskeie korpora geïdentifiseer. In die hieropvolgende afdelings (6.3.4.1 en 6.3.4.2) word die ratiobepalings van die gespesialiseerde korpus en die verdagte tekste bespreek.

6.3.4.1 Gespesialiseerde korpus: hulpwerkwoorde

Soos in afdeling 6.3.4 voorgehou, is die ratio's van die 21 hulpwerkwoorde in die vergelykings van die gespesialiseerde korpus bereken. Aangesien daar 45 vergelykings uitgevoer is, is daar ŉ persentasie vir die praktiese beduidendheid van elke vergelyking bereken (sien Figuur 6.12[194]).



Figuur 6.12 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die hulpwerkwoorde

[193] Die hulpwerkwoorde is deur middel van die woordsoortetiketeerder, soos deur CTexT verskaf, geannoteer.
[194] Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

Figuur 6.12 toon die prakties beduidende verskil van die gebruik van hulpwerkwoorde aan wanneer die ratiobepalings tussen die verskillende korpora bereken is. Vanuit Figuur 6.12 is dit duidelik dat daar wel ŉ prakties beduidende verskil in al 45 die vergelykings van die korpora waargeneem kan word toe die gebruik van hulpwerkwoorde ondersoek is. Figuur 6.12 dui verder ook aan dat daar in 2 van die 45 vergelykings van die onderskeie korpora (O2 en O8 asook O5 en O10) geen hulpwerkwoorde gevind is wat nie oor ŉ prakties beduidende verskil beskik nie. Daar kan vervolgens aangevoer word dat daar in elk van die oorblywende 43 vergelykings wat uitgevoer is wel hulpwerkwoorde gebruik is.

Die ondersoek na die praktiese beduidendheid van die woorde toon aan dat die korpora deur dieselfde outeurs geskryf is. Ten einde te kon bepaal hoe suksesvol die ratio-analise van die hulpwerkwoorde was, is die korpora wat met mekaar vergelyk is se hulpwerkwoorde wat wel ŉ prakties beduidende verskil aantoon, bereken. Hierdie persentasie is bereken soos verduidelik in afdeling 6.3.1:

$$\therefore 811 \div 945 \times 100 = 85{,}8\% \ (prakties\ beduidende\ verskil)$$

en

$$\therefore 134 \div 945 \times 100 = 14{,}2\% \ (nie\ prakties\ beduidende\ verskil\ nie)$$

Na gelang van die persentasieberekeninge van die prakties beduidende verskil van die ratio's van die hulpwerkwoorde wanneer die onderskeie korpora met mekaar vergelyk is, blyk dit dat daar in 85,8% van die geval sonder redelike twyfel aangevoer kan word dat die tekste deur verskillende outeurs geskryf is. Hierdie 85,8% verwys na al 45 vergelykings wat tussen die onderskeie korpora uitgevoer is ten einde die ratio van die onderskeie outeurs se gebruik van hulpwerkwoorde te bepaal.

Die bogenoemde berekening van die persentasie hulpwerkwoorde wat deur die onderskeie outeurs gebruik is, was ŉ vergelyking van al die korpora. Wanneer daar spesifiek na die individuele resultate van hierdie ratiobepalings van die gebruik van hulpwerkwoorde ondersoek ingestel is, is dit duidelik dat daar in slegs 2 van die 45 vergelykings (O2 en O8 asook O5 en O10) geen hulpwerkwoorde voorkom wat oor ŉ praktiese beduidende verskil beskik nie. Hierdie twee tekste is dus, soos reeds genoem, sonder twyfel deur verskillende outeurs geskryf. Wanneer die oorblywende 43 vergelykings ondersoek is, is daar gevind dat daar wel steeds ŉ definitiewe individuele voorkeur by elk van die outeurs is ten opsigte van die gebruik van hulpwerkwoorde.

Die laagste persentasie hulpwerkwoorde wat prakties beduidend is wanneer die ratiobepalings bereken is, word in die vergelyking van O5 en O9 gevind. Vanuit die vergelyking van hierdie twee outeurs se korpora ten opsigte van die gebruik van hulpwerkwoorde verskil slegs 61,9% van die hulpwerkwoorde wat deur die outeurs gebruik word prakties beduidend (deur verskillende outeurs geskryf). 38,1% van die hulpwerkwoorde wat deur O5 en O9 gebruik word, stem ooreen − in 38,1% van die geval waar die outeurs van hulpwerkwoorde gebruik gemaak, kan daar nie sonder redelike twyfel aangevoer word dat die tekste deur verskillende outeurs geskryf is nie.

Aangesien hierdie twee outeurs se gebruik van hulpwerkwoorde soveel ooreenstem, is daar verder ondersoek ingestel na die outeurs se korpora. Beide O5 en O9 is in die bloggerkategorie van Twitter-gebruikers geplaas. Met nadere ondersoek blyk dit dat beide die outeurs geneig is om Twitter as ’n platform te gebruik waar hulle ’n redelik negatiewe wêreldbeskouing met hulle volgelinge deel.

Aangesien hierdie twee outeurs se tekste (met verwysing na die gebruik van hulpwerkwoorde) soveel ooreenstem en aangesien daar in geen van die ander ratiobepalings (voornaamwoorde, voorsetsels en lidwoorde) outeurs gevind is waar die tekste soveel ooreenstem nie, is die gebruik van hulpwerkwoorde verder geanaliseer deur na die gebruik van afsonderlike hulpwerkwoorde ondersoek in te stel (sien Tabel 6.7):

Woord	Voorkoms in O5	Voorkoms in O9	Frekwensie van O5	Frekwensie van O9	Ratio vir O5 en O9	Prakties beduidende verskil
begin	2	2	6,01	7,27	0,83	Ja
behoort	1	0	3,00	0,00	0,00	Nee
bly	1	1	3,00	3,64	0,83	Ja
gaan	41	16	123,12	58,18	2,12	Nee
het	56	67	168,17	243,64	0,69	Nee
hoef	1	2	3,00	7,27	0,41	Nee
is	165	127	495,50	461,82	1,07	Ja
kan	29	25	87,09	90,91	0,96	Ja
kom	8	8	24,02	29,09	0,83	Ja
kon	1	1	3,00	3,64	0,83	Ja
laat	15	2	45,05	7,27	6,19	Nee
mag	2	1	6,01	3,64	1,65	Nee
moenie	13	0	39,04	0,00	0,00	Nee
moes	1	1	3,00	3,64	0,83	Ja
moet	27	13	81,08	47,27	1,72	Nee
probeer	4	2	12,01	7,27	1,65	Nee

Woord	Voorkoms in O5	Voorkoms in O9	Frekwensie van O5	Frekwensie van O9	Ratio vir O5 en O9	Prakties beduidende verskil
sal	19	7	57,06	25,45	2,24	Nee
sou	2	6	6,01	21,82	0,28	Nee
wil	20	20	60,06	72,73	0,83	Ja
word	8	6	24,02	21,82	1,10	Ja
wou	2	2	6,01	7,27	0,83	Ja

Tabel 6.7 Hulpwerkwoorde soos gebruik deur O5 en O9

Vanuit Tabel 6.7 is dit duidelik dat vanuit die 21 hulpwerkwoorde wat geannoteer en in die ratiobepalings gebruik is, O5 en O9 se gebruik van hulpwerkwoorde in 10 van die gevalle ooreenstem en in 11 van die gevalle verskil. Daar is dus slegs ŉ prakties beduidende verskil waargeneem wanneer die twee outeurs die woorde ***begin***, ***bly***, ***is***, ***kan***, ***kom***, ***kon***, ***moes***, ***wil wou*** en ***word*** gebruik wat daarop dui dat die tekste deur verskillende outeurs geskryf is.

Daar was geen prakties beduidende verskil wanneer hierdie twee outeurs (O5 en O9) die hulpwerkwoorde ***behoort***, ***gaan***, ***het***, ***hoef***, ***laat***, ***mag***, ***moenie***, ***moet***, ***probeer***, ***sal*** en ***sou*** gebruik nie en dus kan daar nie sonder redelike twyfel aangevoer word dat hierdie tekste deur verskillende outeurs geskryf is nie. O5 en O9 kan vervolgens nie net ten opsigte van die ratiobepalings van die hulpwerkwoorde wat deur die twee outeurs gebruik is, geïdentifiseer word nie. Die hulpwerkwoorde wat wel prakties beduidend is (***begin***, ***bly***, ***is***, ***kan***, ***kom***, ***kon***, ***moes***, ***wil***, ***wou*** en ***word***) kan wel in die outeurskapidentifikasie gebruik word indien albei hierdie outeurs (O5 en O9) moontlike verdagtes is in ŉ situasie waar die outeur van ŉ mikroblog-teks geïdentifiseer moet word.

Die oorblywende 42 ratioberekeninge van die gebruik van hulpwerkwoorde is ondersoek en dit blyk die laagste persentasie hulpwerkwoorde wat prakties beduidend is, is in die vergelyking van O2 en O6 gevind. In die ratiobepaling[195] van O2 en O6 is die gebruik van die hulpwerkwoorde deur hierdie twee outeurs 66,7% (prakties beduidend) en 33,3% van die keer nie prakties beduidend nie. Dit wil sê dat daar 66,7% keer van hulpwerkwoorde deur O2 of O6 gebruik gemaak word en daar dus aangevoer kan word dat die tekste wel deur verskillende outeurs geskryf is. Ander vergelykings waar die gebruik van hulpwerkwoorde se prakties beduidende verskil nie so hoog is nie sluit in:

- **O2** en **O3 – 71,4%** van die keer wat hulpwerkwoorde deur dié twee outeurs gebruik is, is daar wel ŉ prakties beduidende verskil waargeneem.

[195] Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

- **O4** en **O6 – 71,4%** van die keer wat hulpwerkwoorde deur dié twee outeurs gebruik is, is daar wel ’n prakties beduidende verskil waargeneem.
- **O6** en **O7 – 71,4%** van die keer wat hulpwerkwoorde deur dié outeurs gebruik is, is daar wel ’n prakties beduidende verskil waargeneem.

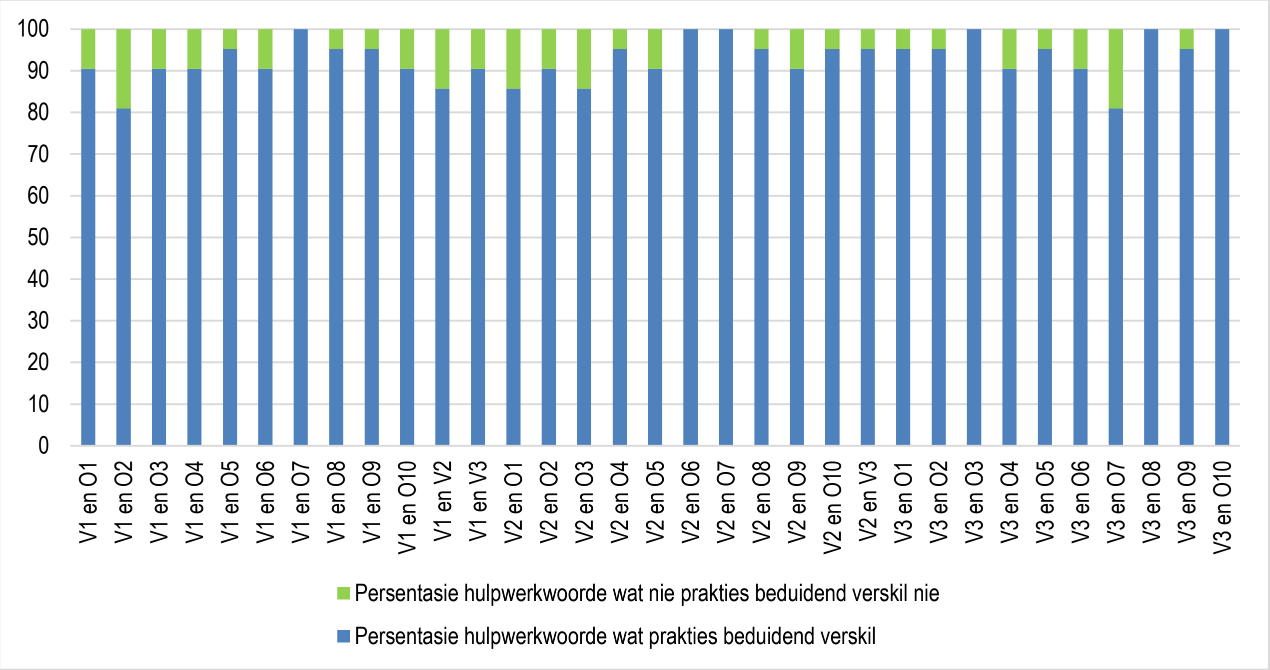
In die oorblywende 33 vergelykings is daar bevind dat die hulpwerkwoordgebruik van die onderskeie outeurs wel insiggewend is. Hierdie bevindings is soortgelyk aan dié wat in die ratiobepalings van die voorsetsels en voornaamwoorde bevind is waar die persentasie waartydens daar wel ’n prakties beduidende verskil in die gebruik van hulpwerkwoorde waargeneem is, wissel tussen 75% en 100%.

Die ondersoek na die gebruik van hulpwerkwoorde met betrekking tot die ratiobepalings om die prakties beduidende verskil te kon bereken, was nie altyd so suksesvol soos met die ander ratiobepalings nie. Ten spyte daarvan is daar etlike vergelykings waar die gebruik van hulpwerkwoorde nie ’n baie hoë persentasie van praktiese beduidendheid getoon het nie en kan hierdie ondersoek steeds gebruik word ten einde moontlike verdagte outeurs van ’n onbekende mikroblog-teks te elimineer. Die persentasie hulpwerkwoorde wat wel oor ’n prakties beduidende verskil beskik wanneer die resultate van al 45 vergelykings bereken is, is 85,8%.

Daar kan dus aangevoer word dat die gebruik van hulpwerkwoorde wel as ’n hulpmiddel in die identifikasie van mikroblogouteurs beskou kan word, maar aangesien die ondersoek na die gebruik van hulpwerkwoorde se resultate nie so goed is soos die resultate by die gebruik van lidwoorde, voornaamwoorde en voorsetsels nie, kan hierdie soort ondersoek nie alleen gebruik word ten einde outeurskapidentifikasie uit te voer nie.

6.3.4.2 Verdagte tekste: hulpwerkwoorde

Figuur 6.13 bied ’n opsomming van die bevindinge van die ratiobepalings van die voorkoms van hulpwerkwoorde in die verdagte tekste:



Figuur 6.13 Voorstelling van die persentasie van ŉ prakties beduidende verskil van die ratiobepalings van die hulpwerkwoorde

Figuur 6.13 dui die persentasies hulpwerkwoorde aan wat prakties beduidend verskil in die verdagte tekste. In die ondersoek na die prakties beduidende verskil van die gebruik van hulpwerkwoorde − wanneer die eerste verdagte teks (V1) met die gespesialiseerde korpus en V2 en V3 vergelyk is – blyk dit dat daar in die vergelyking van V1 en O2 die meeste hulpwerkwoorde voorkom wat nie prakties beduidend verskil nie (19,04%). In die vergelyking van V1 en V2 is 14,42% van die hulpwerkwoorde nie prakties beduidend verskillend nie. Die oorblywende nege vergelykings se persentasies was tussen 9,54% en 4,76% hulpwerkwoorde wat nie prakties beduidend verskil nie. In die vergelyking van V1 en O7 is daar geen hulpwerkwoorde wat prakties beduidend verskil nie en daar kan dus gestel word dat V1 na gelang van die gebruik van hulpwerkwoorde sonder enige twyfel nie deur dieselfde outeur geskryf is nie.

Tydens die analisering van die tweede verdagte teks (V2) om die prakties beduidende verskil van die hulpwerkwoorde te kon ondersoek, blyk dit dat daar twee vergelykings is waar al die hulpwerkwoorde prakties beduidend verskil – dit wil sê daar kan sonder enige twyfel aangevoer word dat die twee tekste in die vergelyking (V2 en O6 asook V2 en O7) deur verskillende outeurs geskryf is. In vier van die oorblywende nege vergelykings (V2 en O4, V2 en O8, V2 en O10 asook V2 en V3) verskil vyf hulpwerkwoorde nie prakties beduidend nie. 5% van die hulpwerkwoorde word tot só ŉ mate deur die onderskeie outeurs gebruik dat daar dus gestel kan word dat dieselfde outeurs die tekste geskryf het.

Drie van die laaste vyf vergelykings ten opsigte van die ratiobepalings van die hulpwerkwoorde toon 10% hulpwerkwoorde wat nie prakties beduidend verskil nie (slegs 10% van die hulpwerkwoorde wat in die tekste voorkom, wat met mekaar vergelyk is – V2 en O2, V2 en O5 asook V2 en O9 – is deur dieselfde outeur geskryf).

Dit is die vergelykings van V2 en O1 (14%) sowel as V2 en O3 (14%) wat die grootste ooreenkoms ten opsigte van die ratiobepalings van die hulpwerkwoorde toon. 14% van die hulpwerkwoorde verskil nie prakties beduidend nie en dui daarom aan dat 14% van die hulpwerkwoorde se gebruik in die onderskeie korpora tot só ʼn mate ooreenstem dat daar aangevoer kan word dat dieselfde outeurs die tekste geskryf het. Die mate waartoe die gebruik van hulpwerkwoorde ooreenstem, is nie sodanig dat daar sonder redelike twyfel aangevoer kan word dat O1 of O3 die outeur van die tweede verdagte teks is nie. Daar kan aangevoer word dat O1 en O3 die beste moontlike outeurskandidate is na gelang van die ondersoek na die gebruik van hulpwerkwoorde.

Ten opsigte van die derde verdagte teks blyk dit dat daar drie vergelykings is waar daar geen hulpwerkwoorde voorkom wat prakties beduidend verskil nie (sien Figuur 6.13) en daar kan vervolgens aangevoer word dat O3, O8 en O10 nie die outeurs van V3 kan wees na gelang van die ratiobepalings nie. In vier van die elf vergelykings het 5% van die hulpwerkwoordgebruik nie prakties beduidend verskil nie. Hierdie vier outeurs (O1, O2, O5 en O9) gebruik dus 5% van hulle hulpwerkwoorde só dat dit voorkom asof hulle die outeur van V3 kan wees. 5% is egter ʼn té lae persentasie om te kan aanvoer dat O1, O2, O5 of O9 wel die outeur van V3 is.

Die hulpwerkwoordgebruik van O4 en O6 stem met 10% van die hulpwerkwoordgebruik van V3 ooreen en dus is die kans groter dat O4 of O6 die outeur van die derde verdagte teks is na gelang van die ratiobepalings van die hulpwerkwoorde. 19% van die hulpwerkwoorde wat deur V3 en O7 gebruik word, toon nie ʼn prakties beduidende verskil nie en dus is O7 die mees waarskynlike outeur van die derde verdagte teks na gelang van die ondersoek na die gebruik van hulpwerkwoorde. Weer eens kan daar dus aangevoer word dat die ratiobepalings eerder beskou moet word as ʼn manier waarop moontlike verdagte outeurs van tekste geëlimineer kan word as wat dit sonder twyfel die outeurs van tekste aandui.

6.3.5 Ratiobepalings: voegwoorde

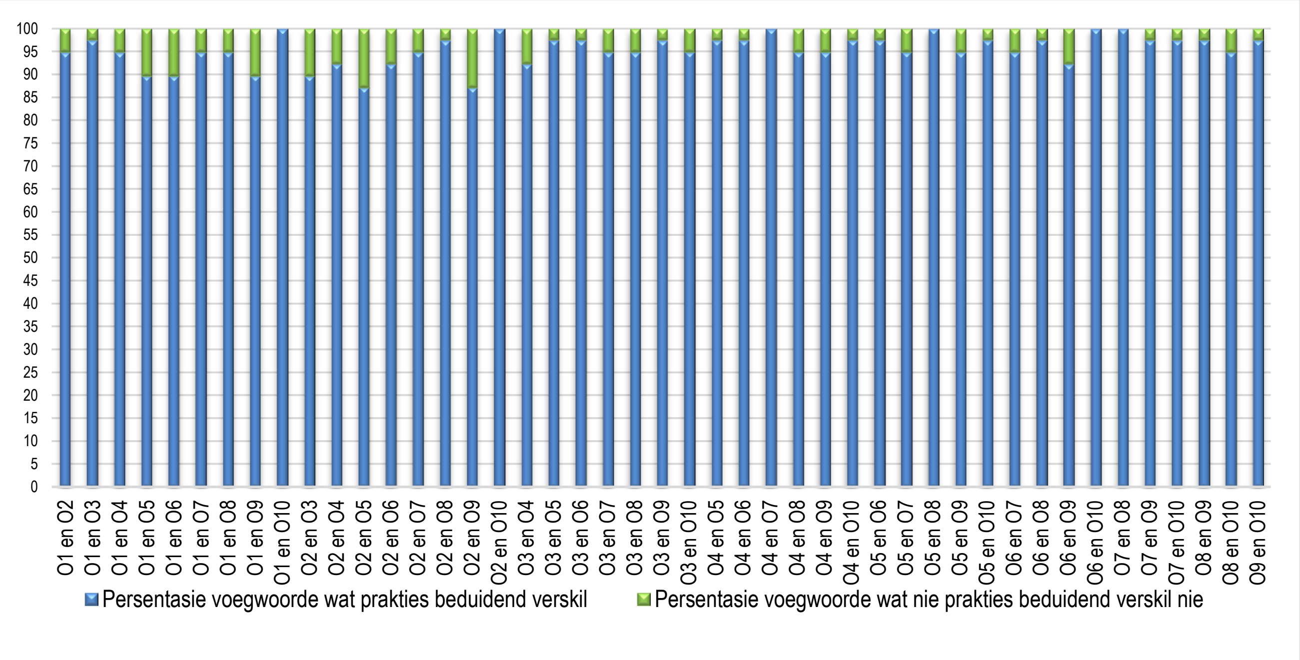
Die laaste ratiobepaling wat uitgevoer is ten einde die praktiese beduidendheid van die grammatikale woorde te kon bepaal, was die ratiobepalings van die voegwoorde[196]. Soos in die

[196] Die voegwoorde is deur middel van die woordsoortetiketeerder, soos deur CTexT verskaf, geannoteer.

vorige analises is die geannoteerde korpora geanaliseer, alle voegwoorde is geïdentifiseer en as 'n txt-dokument gestoor. Die onderskeie korpora is in WordSmith Tools oopgemaak en die txt-dokument met die voegwoorde is as 'n stoplys in WordSmith Tools ingevoer. Al die woorde wat nie op hierdie stoplys voorgekom het nie is outomaties vanuit die sleutelwoordlyste verwyder. Die nuwe lyste wat net uit die voegwoorde bestaan het wat in die onderskeie korpora voorgekom het, is as Microsoft Excel-dokumente gestoor. 39 voegwoorde is geïdentifiseer en die ratio's vir elk van die onderskeie korpora vergelykings[197] is bereken.

6.3.5.1 Gespesialiseerde korpus: voegwoorde

Soos in afdeling 6.3.5 voorgehou, is die ratio's van die 39 voegwoorde in die vergelykings van die onderskeie korpora bereken. Aangesien daar 45 vergelykings uitgevoer is, is daar 'n persentasie vir die praktiese beduidendheid van elke vergelyking bereken (sien Figuur 6.14[198]).



Figuur 6.14 Voorstelling van die praktiese beduidende verskil van die ratiobepalings van die hulpwerkwoorde

Figuur 6.14 toon aan dat daar in 6 van die 45 vergelykings geen voegwoorde gebruik word met 'n praktiese beduidend nie. Dus kan daar sonder enige twyfel aangevoer word dat wanneer die tekste van O1 en O10, O2 en O10, O4 en O7, O5 en O8, O6 en O10 asook O7 en O8 vergelyk word deur middel van die ratiobepalings van die gebruik van voegwoorde, ten einde die praktiese beduidendheid te bepaal, die outeurs wat die betrokke tekste geskryf het wel verskil.

197 Sien afdeling 5.12 (Samestelling van die korpora).
198 Sien bylaag D vir 'n voorbeeld van die ratioberekeninge.

Ten einde te kon bepaal hoe suksesvol die ratio-analise van die voegwoorde was, is die gemiddeld van al die korpora se voegwoorde wat wel ’n prakties beduidende verskil aantoon, bereken. Hierdie gemiddeld is bereken soos verduidelik in afdeling 6.3.1:

$$\therefore 1674 \div 1755 \times 100 = 95{,}3\% \; (prakties\ beduidende\ verskil)$$

en

$$\therefore 81 \div 1755 \times 100 = 4{,}7\% \; (nie\ prakties\ beduidende\ verskil\ nie)$$

Na gelang van die persentasieberekeninge van die prakties beduidende verskil van die ratio’s van die voegwoorde wanneer die onderskeie korpora met mekaar vergelyk is, blyk dit dat daar in 95,3% van die geval sonder redelike twyfel aangevoer kan word dat die tekste deur verskillende outeurs geskryf is. Hierdie 95,3% verwys na al die 45 vergelykings wat tussen die onderskeie korpora uitgevoer is.

Ten einde verder te kon bepaal hoe die individuele resultate van die ondersoek na die gebruik van voorsetsels (ratiobepalings) daaruitsien, is daar na die resultate van die ratiobepalings ondersoek ingestel soos uiteengesit in Figuur 6.14. Soos genoem, het die ratiobepalings van die voegwoorde in die vergelykings van O1 en O10, O2 en O10, O4 en O7, O5 en O8, O6 en O10 asook O7 en O8 aangedui dat daar geen prakties beduidende verskil in die gebruik van voegwoorde voorkom nie. Na gelang van die bogenoemde is daar dus geen twyfel dat hierdie ses tekste nie deur verskillende outeurs geskryf is nie.

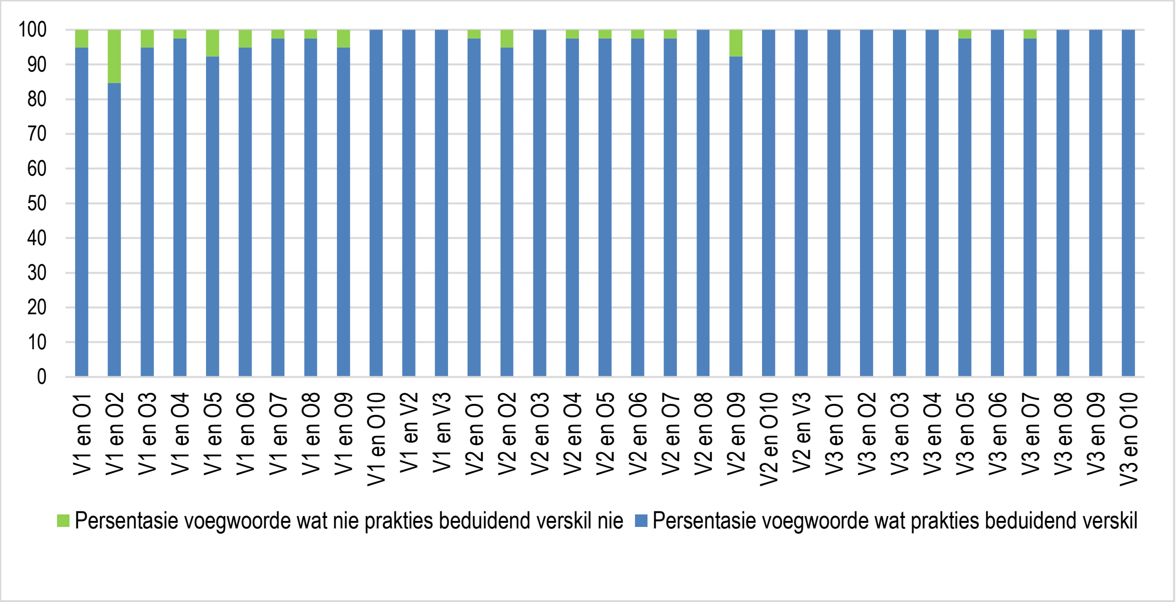
In die oorblywende 39 vergelykings is daar bevind dat die gebruik van voegwoorde deur die onderskeie outeurs wel insiggewend is. Die laagste persentasie voegwoorde wat prakties beduidend in die onderskeie vergelykings is, is in die vergelyking van O2 en O9 aangetref. Vanuit hierdie ratiobepaling tussen O2 en O9 is daar bepaal dat slegs 87,2% van die ratiobepalings van die voegwoorde prakties beduidend verskil. Dit wil sê dat die laagste persentasie waar die ratiobepalings van die gebruik van voegwoorde in die onderskeie vergelykings van die korpora aangetoon het dat daar wel ’n prakties beduidende verskil voorkom (dat die tekste wel deur verskillende outeurs geskryf is) is 87,2% en in hierdie vergelyking van die korpora is daar dus ’n 12,8% kans dat die tekste deur dieselfde outeur geskryf is[199]. In die oorblywende 38 tekste het die persentasies gewissel en ’n praktiese beduidende verskil is wel waargeneem tussen 89,7% en 99%. Dit dui daarop dat die gebruik van voegwoorde wanneer die ratio-analises uitgevoer is wel beskou kan word as ’n aanduider van outeurskap.

[199] Sien bylaag D vir ’n voorbeeld van die ratioberekeninge.

Die ratiobepalings van die voegwoorde om die praktiese beduidendheid van die gebruik van hierdie woordsoorte te kon bepaal wanneer die onderskeie outeurs se korpora met mekaar vergelyk is, voltrek die ratiobepalings van die grammatikale woorde. In hierdie afdeling (afdeling 6.3 – ratiobepalings) is ondersoek ingestel na frekwente woorde, leksikale woorde en ook grammatikale woorde (lidwoorde, voornaamwoorde, voorsetsels, hulpwerkwoorde en voegwoorde). Vanuit hierdie afdeling van die kwantitatiewe analise blyk dit dus dat outeurs van mikroblogtekste wel geïdentifiseer kan word, maar ten einde die stelling te kon verifieer, is die verdagte tekste vervolgens ondersoek.

6.3.5.2 Verdagte tekste: voegwoorde

Soos in afdelings 6.3.5 bespreek, is die onderskeie korpora in WordSmith Tools ingevoer en woordelyste is as Microsoft Excel-dokumente gestoor nadat stoplyste van die voegwoorde in WordSmith ingevoer is en alle woorde wat nie voegwoorde was nie vanuit die lyste verwyder is. Die bevindinge van die ratiobepalings word in Figuur 6.15 voorgehou:



Figuur 6.15 Voorstelling van die persentasie van 'n prakties beduidende verskil van die ratiobepalings van die voegwoorde

In Figuur 6.15 word die bevindinge van die ratiobepalings van die voegwoorde voorgehou soos ondersoek is in die vergelykings van die drie verdagte tekste (V1, V2 en V3) met die korpora van die bekende outeurs. In die vergelyking van die gebruik van voegwoorde soos gebruik in die eerste verdagte teks (V1) en die korpora van die bekende outeurs is daar drie vergelykings waarin daar geen voegwoorde voorkom wat prakties beduidend verskil nie en kan daar aangevoer word dat O10, V2 en V3 sonder twyfel nie die outeurs van V1 is nie.

3% van die voegwoorde wat deur V1 gebruik word, word ook tot só ’n mate deur O4, O7 en O8 gebruik dat die voegwoorde nie prakties beduidend verskil nie, en daar dus aangevoer kan word dat die tekste deur dieselfde outeurs geskryf is. 5% van die voegwoorde wat in die vergelykings van V1 met O1, O3, O6 en O9 gebruik is, verskil nie prakties beduidend nie. In die vergelyking van V1 en O5 is 8% van die voegwoorde wat deur die twee outeurs gebruik is nie prakties beduidend verskillend nie, maar O2 toon die grootste ooreenkoms met die eerste verdagte teks.

In hierdie vergelyking is 15% van die voegwoorde wat deur O2 en V1 gebruik word nie prakties beduidend verskillend nie (15% van die voegwoorde se gebruik kom voor asof dit deur dieselfde outeur gebruik word). 15% is ’n baie klein persentasie, en daar kan weer eens aangevoer word dat die ratiobepalings van die voegwoorde gebruik kan word om die moontlike outeurs te elimineer eerder as om die outeur van V1 te identifiseer.

Die voegwoordgebruik van V2 het relatiewe lae persentasies voegwoorde opgelewer wat nie prakties beduidend verskil nie (tussen 0% en 5%). Die grootste ooreenkoms kom voor in die vergelyking van V2 en O9 waar 8% van die voegwoorde nie prakties beduidend verskil nie. Die ratiobepalings van V3 toon nog minder ooreenkomste as dié van V2 en in die vergelykings was dit slegs in die vergelyking van V3 en O5 asook V3 en O7 waar daar enigsins voegwoorde voorkom wat nie prakties beduidend verskil nie. Dit blyk dus dat O5 (3%) en O7 (3%) die enigste outeurs is wat moontlik die outeurs van V3 kan wees. In die oorblywende agt vergelykings is daar geen voegwoorde wat nie prakties beduidend verskil nie en dus is O1, O2, O3, O4, O6, O8, O9 en O10 definitief nie die outeurs van V3 nie.

Die ondersoek na die ratiobepalings kan vanuit die ondersoek na die gebruik van grammatikale woorde (lidwoorde, voornaamwoorde, voorsetsels, hulpwerkwoorde en voegwoorde) soos voorgehou in afdelings 6.3.1 tot 6.3.5 as ’n suksesvolle ondersoek na die identifikasie van outeurs van mikrobloginskrywings beskou word. Alhoewel die ratiobepalings nie alleen gebruik kan word om ’n outeurskapidentifikasie uit te voer nie, is dit sonder twyfel ’n belangrike stap in die proses en moontlike outeurs kan dus geëlimineer word wat die poel van verdagtes verminder. Woorde en karakters per twiet word in die volgende afdeling in hierdie stilometriese analise ondersoek.

6.4 Sinne, woorde en karakters per twiet

Twitter beskik oor ’n maksimum karakterbeperking van 140 karakters per twiet en juis daarom is daar ondersoek ingestel na die aantal woorde en die gemiddelde aantal karakters wat die

onderskeie outeurs per twiet gebruik[200]. Hierdie analise is in Microsoft Excel uitgevoer. Elke outeur se korpus is in Microsoft Excel ingevoer (sien Tabel 6.8) en die woorde en karakters per twiet is sodoende bereken.

Tabel 6.8 Voorstelling van ŉ korpus in Microsoft Excel waardeur die woorde en karakters per twiets bepaal is

Twiets	Karakters [201]per twiet	Woorde per twiet
Ek dink mense wat die Rubix cube oplos, doen dit perongeluk.	60	11
As ek nog een bio lees met "papercut survivor" in, gaan ek death by papercuts op myself gaan.	93	18
Ek het die afgelope paar maande net junk food geëet. Toe ek gesond begin eet toe word ek siek. Moral of the story: ek gaan nou 'n burger kry	140	29
Begin net êrens. You can't steer a parked car.	46	9
Iemand, motiveer my.	20	3
Dis die tyd van die jaar waar ek nie meer omgee as daar 'n stukkie papier aan my fizzer vasplak nie. Ek hoop maar my maag kan dit verteer.	138	29

Hierdie analise, soos gesien in Tabel 6.8, is vir elk van die korpora saamgestel. Verder is die woorde en karakters per twiet ook in die interne korpora bereken sowel as in die verdagte tekste. Die onderskeie datastelle in die gespesialiseerde korpus is dus nie net met mekaar vergelyk nie, maar ook intern (dit wil sê die sub-korpus van elke deelnemer) vergelyk om sodoende te kon bepaal of daar ŉ ooreenkoms of verskil waargeneem kan word in die woorde en karakters wat per twiet gebruik is. Die sinne[202] is in elke twiet getel en aangeteken en die totaal van die sinne in die korpora is bereken. Daarna is die totale aantal sinne per korpus deur die totale aantal twiets per korpus gedeel en ŉ gemiddeld is sodoende bereken vir elke korpus se sinne.

Soortgelyk aan die sinne per korpus, is die aantal woorde per twiet getel en ŉ somtotaal van die aantal woorde per korpus is bereken. Hierdie totale aantal woorde per korpus is deur die aantal twiets per korpus gedeel om ŉ gemiddeld van die aantal woorde per twiet te bereken. Laastens, is die aantal karakters per twiet bereken in Microsoft Excel en weer eens is daar ŉ gemiddelde aantal karakters per twiet vir elke korpus bereken. Ten einde die bogenoemde te kon uitvoer, verwys ek eerstens na die bevindinge van die analise van die gespesialiseerde korpus alvorens die verdagte tekste se bevindinge voorgehou word.

---

[200] Alhoewel daar na gelang van die ondersoek bevind is dat die hoeveelheid sinne, woorde en karakters ook sonder moeite elk as 'n afsonderlike ondersoek geanaliseer kan word, is daar teen hierdie werkswyse besluit aangesiend die doel van hierdie studie is om eksploratief ondersoek in te stel na die aspekte wat wel gebruik kan word in die identifisering van die outeur van Twitter inskrywings in Afrikaans.

[201] Die karaktertelling sluit spasies in.

[202] 'n Sin verwys in hierdie studie na enige opeenvolging van woorde (grammatikale eenheid) wat 'n leesteken – punt, uitroepteken of vraagteken – voorafgaan, of waar 'n hutswoord of emotikon as die einde van 'n gedagte beskou kan word. Daar is dus nie onderskeid tussen hoofsinne of afhanklike sinne getref nie.

6.4.1 Gespesialiseerde korpus: sinne, woorde en karakters per twiet

Tabel 6.9 toon die interne vergelykings van die woorde per twiet, sinne per twiet en karakters per twiet aan:

Tabel 6.9 Sinne, woorde, karakters en sinne in die gespesialiseerde korpus

Korpus	Totale aantal sinne in korpus	Gemiddelde aantal sinne per twiet	Gemiddelde aantal woorde per twiet	Gemiddelde aantal karakters per twiet
O1	944	1,41	13,17	68,23
O1_1	453	1,35	13,44	68,99
O1_2	491	1,46	12,83	66,64
O2	1577	2,67	20,22	123,17
O2_1	743	2,45	20,25	124,58
O2_1	834	2,91	20,31	121,67
O3	266	1,29	15,31	99,02
O3_1	134	1,25	14,64	95,32
O3_2	132	1,33	16,06	102,77
O4	978	2,02	14,14	82,95
O4_1	467	1,92	13,6	76,23
O4_2	511	2,12	14,73	89,73
O5	556	1,67	13,9	67,78
O5_1	259	1,58	13,57	66,13
O5_2	297	1,76	14,02	69
O6	1657	2,17	18,5	116,3
O6_1	858	2,22	18,55	116,42
O6_2	799	2,12	18,41	115,84
O7	790	1,73	12,1	62,29
O7_2	402	1,73	12,27	62,78
O7_2	388	1,73	12,03	62,07
O8	443	1,48	11,18	78,65
O8_1	224	1,48	10,78	75,09
O8_2	219	1,47	11,31	79,91
O9	467	1,7	17,43	92,21
O9_1	244	1,66	16,65	89,19
O9_2	223	1,76	18,62	95,8
O10	1190	2,38	14,54	96,5
O10_1	579	2,46	15,5	103,23
O10_2	611	2,31	13,81	90,77

Vanuit Tabel 6.9 blyk dit dat die gemiddelde aantal sinne per twiet verskil van een gebruiker na ŉ ander. Wanneer daar ondersoek ingestel is na die interne vergelykings van die korpora, is dit duidelik dat die korpora ooreenstem met betrekking tot die gemiddelde sinslengte per twiet. In die interne vergelyking is die grootste verskil tussen die gemiddelde sinslengtes per twiet in die korpora van O2 en O10 gevind. O2 en O2_1 verskil met 0,12 en O2 en O2_2 verskil met 0,46 terwyl O2_1 en O2_2 met 0,22 verskil. O10 en O10_1 verskil met 0,12 en O10 en O10_2 verkil met 0,07. O10_1 en O10_2 verskil met 0,15. In die oorblywende interne vergelykings het die

korpora telkens met minder as 0,10 verskil en daarom kan daar aangevoer word dat die gemiddelde lengte van die sinne wat outeurs op Twitter gebruik as uniek aan elke outeur beskou kan word en dus gebruik kan word om outeurskap saam met ander veranderlikes te bepaal.

Ten einde die bogenoemde stelling te kon toets, is die gemiddelde sinslengtes van O1 tot O10 ook met mekaar vergelyk. Vanuit hierdie vergelyking stem O1 en O8 (0,07), O5 en O9 (0,03), O5 en O7 (0,06) asook O7 en O9 (0,03) se gemiddelde sinslengtes met mekaar ooreen. Korpora wat die meeste verskil ten opsigte van die gemiddelde lengte van die sinne per twiet is O1 en O2 (1,26), O2 en O3 (1,38), O2 en O8 (1,19) asook O3 en O10 (1,09). Aangesien dit slegs 4 van die 50 vergelykings is waar daar nie met sekerheid aangevoer kan word dat die outeurs se gemiddelde sinslengte per twiet verskil nie, kan die gemiddelde lengte van sinne per twiet ook as ŉ moontlike aanduiding van outeurskap beskou word wanneer outeurs van mikroblogtekste geïdentifiseer moet word.

Dit kom voor asof ŉ gemiddeld van 1,85 sinne per twiet as ’n gewilde sinslengte vir sinne in Afrikaans op Twitter beskou kan word[203]. Wanneer die interne vergelyking tussen die outeurs uitgevoer is, blyk dit dat die aantal sinne per twiet as uniek aan elke outeur beskou kan word en dus gebruik kan word om outeurskap te bepaal. Hierdie stelling is verder getoets wanneer die verdagte tekste ondersoek is.

Vanuit Tabel 6.9 blyk dit dat die gemiddelde aantal woorde per twiet tussen outeurs verskil. Wanneer daar na die interne vergelykings van die korpora ondersoek ingestel is, is dit duidelik dat die korpora ooreenstem met betrekking tot die gemiddelde aantal woorde per twiet. Wanneer die korpora intern met mekaar vergelyk is, word die grootste verskil tussen die gemiddelde aantal woorde per twiet in die korpora van O9 aangetref. O9 en O9_1 verskil met 0,78. O9 en O9_2 verskil met 1,19 terwyl O9_1 en O9_2 met 1,97 verskil. Die interne vergelykings van O2 (O2 en O2_1 – 0,03; O2 en O2_2 – 0,09; O2_1 en O2_2 – 0,06) en O6 (O6 en O6_1 – 0,05; O6 en O6_2 – 0,09; en O6_1 en O6_2 – 0,14) verskil baie min van mekaar.

Wanneer daar na die vergelykings tussen O1 tot O10 ondersoek ingestel is, blyk dit dat die gemiddelde aantal woorde per vergelyking met meer as 1 verskil wanneer O1 en O4 (0,97), O1 en O5 (0,73), O3 en O10 (0,77), O4 en O5 (0,24), O4 en O10 (0,4), O5 en O10 (0,64) asook O7 en O8 (0,92) met mekaar vergelyk is. Sommige van die vergelykings het ook verskille van meer

[203] Die gemiddeld is bepaal na gelang van die ondersoek na die gespesialiseerde en verwysingskorpus.

as vyf opgelewer (O2 en O4 – 6,08; O2 en O5 – 6,32; O2 en O7 – 8,12; O2 en O8 – 9,04; O2 en O10 – 5,68; O6 en O8 – 7,32; O7 en O9 – 5,33; en O8 en O9 – 6,25).

Daar kan dus aangevoer word dat die ondersoek na die gemiddelde aantal woorde wat ŉ outeur per twiet gebruik wel beskou kan word as ŉ manier waarop outeurs van mikroblogtekste geïdentifiseer kan word. Hierdie ondersoek na die gemiddelde aantal woorde per twiet is nie op sigself genoeg om outeurs te identifiseer nie, dit moet saam met ander analises gebruik word ten einde outeurskapidentifikasie op mikroblogs uit te voer.

Die laaste ondersoek in hierdie afdeling verwys na die gemiddelde aantal karakters per twiet. Die bevindinge van hierdie ondersoek word in Tabel 6.5 weergegee en is soos volg geïnterpreteer. Wanneer die interne vergelykings van die korpora geanaliseer is, toon die vergelyking van O10 en O10_1 (6,73), O10 en O10_2 (5,73) asook O10 en O10_2 (12,46) die grootste verskil in die gemiddelde karakters per twiet. O6 en O6_1 (0,12), O6 en O6_2 (0,46) en die interne vergelyking van O6_1 en O6_2 (0,58) stem die meeste ooreen met verwysing na die gemiddelde aantal karakters per twiet.

Die vergelykings van O1 tot O10 toon oor die algemeen baie groot verskille. Die twee outeurs se gemiddelde aantal karakters per twiet wat die meeste ooreenstem, is O1 en O5 (0,45). Ook O1 en O7 (5,94), O2 en O6 (6,87), O3 en O9 (6,81), O3 en O10 (2,52), O4 en O8 (4,3), O4 en O9 (9,26), O5 en O7 (5,49) asook O9 en O10 (4,29) kan as relatief soortgelyk beskou word (alhoewel die gemiddelde aantal karakters per twiet aansienlik verskil). Die oorblywende 41 vergelykings toon verskille van tussen 10, 87 (O5 en O8) en 60,88 (O2 en O7).

Vanuit die ondersoek na die aantal woorde en karakters per twiet kan die aanname gemaak word dat twiets in Afrikaans in hierdie versameling van korpora uit ŉ gemiddeld van ongeveer 88,71 karakters[204] en 15,05 woorde per twiet bestaan. Alhoewel sommige van die outeurs vanuit die gespesialiseerde korpus wel oor ŉ baie hoër of laer gemiddeld beskik wanneer die woorde en karakters per twiet geanaliseer is, kan die bogenoemde getalle as ŉ aanvaarbare gemiddeld beskou word. Hierdie bevinding rakende die gemiddelde aantal woorde en karakters per twiet sluit ook aan by navorsing gedoen deur Go *et al* (2009) wat bevind het dat 14 woorde en 78 karakters per twiet as die norm in Engels beskou kan word en Crystal (2011:47) se bevinding dat 14,7 woorde per twiet in Engels as die norm beskou kan word.

Die resultate van hierdie analise kan dus gebruik word om outeurs van mikrobloginskrywings te identifiseer mits daar bykomende ondersoeke (soos die ratiobepalings en sleutelwaardes) ook

[204] Vanuit die verwysingskorpus is die gemiddelde hoeveelheid karakters 87,96, terwyl die gemiddelde hoeveelheid woorde per twiet 13,01 is.

uitgevoer word. Dit kom voor asof elk van die analises wat tot dusver op die onderskeie korpora (beide intern en ekstern) uitgevoer is, gebruik kan word in die identifisering van outeurs van mikrobloginskrywings. Ten einde die bogenoemde stelling te kon verifieer, is die verdagte tekste vervolgens geanaliseer ten opsigte van die sinne, woorde en karakters per twiet.

6.4.2 Verdagte tekste: sinne, woorde en karakters per twiet

In afdeling 6.4.1 is die sinne, woorde en karakters per twiet van die tien bekende outeurs bespreek (sien Tabel 6.9). Aangesien daar ʼn karakterbeperking van toepassing is op Twitter (maksimum van 140 karakters per twiet, spasies ingesluit), is daar ondersoek ingestel na die aantal sinne, woorde en die gemiddelde aantal karakters wat die onderskeie outeurs per twiet gebruik. Hierdie analise is in Microsoft Excel uitgevoer. Elke outeur se korpus is in Microsoft Excel ingevoer en die woorde en karakters per twiet is sodoende bereken.

Aangesien die analise in afdeling 6.4.1 aandui dat daar wel onderskeid getref kan word tussen die onderskeie outeurs se gemiddelde sinne, woorde en karakters per twiet, is die data met dié van die drie verdagte tekste vergelyk (sien Tabel 6.10):

Tabel 6.10 Sinne, woorde, karakters en sinne in die verdagte tekste

Korpus	Totale aantal sinne in korpus	Gemiddelde aantal sinne per twiet	Gemiddelde aantal woorde per twiet	Gemiddelde aantal karakters per twiet
V1	90	3	20,07	121,21
V2	147	1,63	15,66	81,71
V3	44	1,38	12,88	61,28

Vanuit Tabel 6.10 is dit duidelik dat al drie die verdagte tekste se gemiddelde aantal sinne, woorde en karakters per twiet verskil. Ten einde die vraag te kon beantwoord of outeurs van mikrobloginskrywings geïdentifiseer kan word, is die karakters, sinne en woorde van die drie outeurs (V1, V2 en V3) met dié van die tien bekende outeurs vergelyk (sien Tabel 6.10). Die bevindinge word in Tabel 611 voorgehou:

Tabel 6.11 Gemiddelde aantal sinne, woorde en karakters vanuit die gespesialiseerde korpus wat met die verdagte tekste ooreenstem

Korpus	Totale aantal sinne in korpus	Gemiddelde aantal sinne per twiet	Gemiddelde aantal woorde per twiet	Gemiddelde aantal karakters per twiet
V1	90	3	20,07	121,21
O2	1 577	2,67	20,22	123,17
V2	147	1,63	15,66	81,71
O3	266	1,29	15,31	99,02
O4	978	2,02	14,14	82,95
O5	556	1,67	13,90	67,76
V3	44	1,38	12,88	61,28
O1	944	1,41	13,17	68,23
O7	790	1,73	12,10	62,29

Tabel 6.11 toon aan met watter van die bekende outeurs se tekste die gemiddelde aantal sinne, woorde en karakters per twiet van die verdagte tekste ooreenstem. O2 is in elk van die vergelykings die mees waarskynlike outeur van die V1-teks. Drie outeurs stem met V2 ooreen. O3 en V2 se gemiddelde aantal woorde per twiet stem ooreen, maar die sinne en karakters per twiet verskil aansienlik van mekaar. Die gemiddelde aantal karakters per twiet van O4 stem ooreen met dié van V2 terwyl O5 en V2 ooreenkomste toon ten opsigte van die gemiddelde aantal sinne per twiet. Al drie hierdie outeurs (O3, O4 en O5) kan dus as moontlike outeurs van die tweede verdagte outeur beskou word.

V3 se gemiddelde aantal sinne, woorde en karakters per twiets stem ooreen met beide O1 en O7. In die geval van O1 stem die gemiddelde aantal sinne per twiet met dié van V3 ooreen terwyl die gemiddelde aantal karakters en woorde per twiet van V3 en O7 as soortgelyk beskou kan word. O7 se gemiddelde aantal sinne per twiet is slegs 0,32 hoër as dié van O1 en daarom (tesame met die feit dat V3 en O7 se gemiddelde aantal karakters en woorde per twiet ŉ sterk ooreenkoms toon) kan O7 as die moontlike outeur van V3 beskou word.

Vanuit hierdie ondersoek kan dieselfde gevolgtrekking gemaak word as vanuit afdeling 6.4.1 (waar die bekende outeurs se korpora vergelyk is ten einde te kon bepaal of die gemiddelde aantal sinne, woorde en karakters per twiet wel as ŉ middel tot outeurskapidentifikasie op mikroblogs gebruik kan word). Dit blyk dus dat hierdie ondersoek, soos in die ratiobepalings, die aantal moontlike outeurs van ŉ verdagte mikroblog-teks kan verminder en na gelang van die ondersoek is O1 (V1), O3, O4 en O5 (V2) en O7 (V3) die moontlike outeurs van die verdagte tekste.

Daar was nog nie ŉ enkele ondersoek wat alleenlik genoegsame bewys gelewer het dat ŉ spesifieke outeur ŉ teks geskryf het nie. Aangesien outeurs nie gebind word aan enige spel- of taalkonvensies nie en aangesien Twitter ŉ sosialemediaplatform is wat wêreldwyd gebruik word, was dit verder nodig om ondersoek na die verhouding tussen die gebruik van Afrikaans en ander tale (spesifiek Engels in hierdie geval) in te stel om te kon bepaal of die taalverhouding in die twiets aanleiding kan gee tot die identifikasie van outeurs van mikrobloginskrywings.

6.5 Taalverhouding in die twiets

Aangesien daar nie ŉ definitiewe reël op Twitter voorkom oor spel- of taalkonvensies nie, kan Twitter-gebruikers self kies in watter taal of hoe puristies hulle die taal van hulle keuse wil gebruik. Ten einde ondersoek te kon instel na die gebruikerskeuse op Twitter (Afrikaans of

Engels), is die verhouding tussen die aantal Afrikaanse en Engelse twiets asook die aantal hertwiets bereken.

Bogenoemde is bereken deur die aanvanklike onverwerkte data te analiseer. Die data, soos dit aanvanklik vanaf Twitter gekopieer is, is ondersoek en die aantal eie Afrikaanse twiets, Afrikaanse hertwiets, eie Engelse twiets en Engelse hertwiets is getel (sien Skermskoot 6.5 vir 'n voorbeeld van die onverwerkte data):

Skermskoot 6.5 Uitbeelding van die onverwerkte data wat in die ondersoek na die taalverhouding gebruik is

Soos dit duidelik blyk vanuit Skermskoot 6.5, word die hertwiets duidelik aangedui. Vervolgens is daar vier Microsoft Word-dokumente geskep (eie Afrikaans, eie Engels, RT_Engels en RT_Afrikaans) en die hertwiets, eie Afrikaanse, eie Engelse en hertwiets in Engels is na die gepaste dokumente gekopieer. Die bevindinge van hierdie analise word vervolgens voorgehou. Daar sal eerstens na die gespesialiseerde korpus verwys word alvorens die verdagte tekste se taalverhoudings voorgehou word.

6.5.1 Gespesialiseerde korpus: taalverhouding

Nadat die stappe, soos in afdeling 6.5 voorgehou, uitgevoer is, is daar tabelle saamgestel ten einde die data voor te hou:

Tabel 6.12 Voorbeeld van die taalverhouding (gespesialiseerde korpus)

Totale twiets in onverwerkte data O1	896	100%
Totale eie twiets	721	80,5%
Afrikaanse twiets	669	74,7%
Engelse twiets	53	5,9%
Totale hertwiets	175	19,5%

Afrikaanse hertwiets	120	13,4%
Engelse hertwiets	55	6,1%

In Tabel 6.12 word daar ’n getal en ’n persentasie weergegee. Die getal verwys na die aantal twiets wat in elke geval voorkom (dus 896 onverwerkte twiets, 699 eie Afrikaanse twiets en 175 twiets wat deur O1 hertwiet is). Die persentasie is bereken ten einde die getalle te normaliseer[205] sodat daar ’n vergelyking tussen die onderskeie outeurs se korpora kon plaasvind. Die persentasies wat elk van die aspekte wat in die taalverhouding ondersoek is van elk van die afsonderlike korpora, is soos volg bereken:

- Persentasie vir totale eie twiets = 721 (totale eie twiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 80,5%.
- Persentasie vir eie Afrikaanse twiets = 699 (totale eie Afrikaanse twiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 74,7%.
- Persentasie vir eie Engelse twiets = 53 (totale eie Engelse twiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 5,9%.
- Persentasie vir totale hertwiets = 175 (totale eie twiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 19,5%.
- Persentasie vir Afrikaanse hertwiets = 120 (totale Afrikaanse hertwiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 13,4%.
- Persentasie vir Engelse hertwiets = 55 (totale Engelse hertwiets) ÷ 896 (totale twiets in onverwerkte data) × 100 = 6,1%.

Die bostaande berekeninge is vir elk van die korpora uitgevoer (sien Tabel 6.13) om vas te stel of die verhouding tussen Afrikaanse en Engelse twiets sowel as die gebruik van hertwiets as ’n aanduiding van outeurskap beskou kon word[206]:

Tabel 6.13 Verhouding tussen Afrikaanse en Engelse twiets en hertwiets in die gespesialiseerde korpus

Korpus	Totale twiets in onverwerkte data	Totale Eie twiets		Afrikaanse twiets		Engelse twiets		Totale hertwiets		Afrikaanse hertwiets		Engelse hertwiets	
O1	896	721	80,5%	669	74,7%	53	5,9%	175	19,5%	120	13,4%	55	6,1%
O2	667	627	94,0%	589	88,3%	38	5,7%	40	6,0%	4	0,6%	36	5,4%
O3	834	264	31,7%	206	24,7%	58	6,9%	570	68,3%	101	12,1%	469	56,2%

205 Die getalle is genormaliseer deur ’n persentasie uit te werk angesien die somtotaal van al die twiets (soos uiteengesit in Tabel 6.12 en 6.13) 100% moes vorm.

206 Twiets wat in beide Afrikaans en Engels geskryf is, is na gelang van die taal wat as die primêre taal beskou kan word geklassifiseer en verder as kodewisseling ondersoek.

Korpus	Totale twiets in onverwerkte data	Totale Eie twiets		Afrikaanse twiets		Engelse twiets		Totale hertwiets		Afrikaanse hertwiets		Engelse hertwiets	
O4	900	506	56,2%	484	53,8%	22	2,4%	394	43,8%	233	25,9%	161	17,9%
O5	573	346	60,4%	333	58,1%	13	2,3%	227	39,6%	185	32,3%	42	7,3%
O6	872	821	94,2%	768	88,1%	54	6,2%	51	5,8%	44	5,0%	7	0,8%
O7	1 058	461	43,6%	457	43,2%	4	0,4%	597	56,4%	432	40,8%	165	15,6%
O8	995	362	36,4%	308	31,0%	54	5,4%	633	63,6%	483	48,5%	150	15,1%
O9	559	403	72,1%	275	49,2%	126	22,5%	156	27,9%	26	4,7%	130	23,3%
O10	858	637	74,2%	501	58,4%	136	15,9%	221	25,8%	92	10,7%	129	15,0%

Vanuit Tabel 6.13 is die verhouding tussen die onderskeie outeurs wat in die studie ondersoek is se eie Afrikaanse en Engelse sowel as Afrikaanse en Engelse hertwiets sigbaar. Soos aangedui, is 'n persentasie vir elk van die verhoudings uitgewerk om 'n vergelyking te kon maak. Die afleiding wat vervolgens vanuit Tabel 6.13 gemaak kan word, is dat daar wel 'n verskil waargeneem kan word in die verhouding tussen die Afrikaanse en Engelse twiets en in die aantal keer wat die onderskeie outeurs ander Twittergebruikers se twiets hertwiet. Met nadere ondersoek blyk dit dat individue wat in die nuuskategorie geplaas is, geneig is om meer te hertwiet (O3 en O8). Hierdie outeurs skroom nie om in Engels te twiet of hertwiet nie, maar hierdie gebeurtenis kan toegeskryf word aan die feit dat alhoewel outeur O3 en O8 steeds meer in Afrikaans twiet wanneer hul hulle eie twiets saamstel, die outeurs probeer om 'n meer internasionale gehoor te bereik met hulle twiets.

O2 en O10 is beide in die nuus- en die bloggerkategorie geplaas, maar neig meer na die bloggerskategorie. Dit blyk duidelik dat die twee outeurs meer van eie twiets gebruik maak as hertwiets. Wanneer daar spesifiek na die soort twiets ondersoek ingestel is wat O2 en O10 hertwiet, sluit dit aan by nuuswaardige gebeurtenisse wat by die oorkoepelende tema van hulle profiel aansluit (en die twiets is nuuswaardig met betrekking tot die hooftemas wat na vore kom in O2 en O10 se eie twiets). Die groot verskil was dat O2 en O10 oorwegend in Afrikaans twiet en dat die hertwiets oorwegend in Engels plaasvind (sien Tabel 6.9).

O1, O9 en O5 is ook in die bloggerkategorie geplaas en dit is duidelik dat hulle eie twiets baie meer (O1 – 80,5% teenoor 19,5%; O5 – 60,4% teenoor 39,6%; en O9 – 72,1% teenoor 27,9%) as die hertwiets is. Hierdie outeurs se eie Afrikaanse twiets is ook meer as hulle eie Engelse twiets en met betrekking tot die hertwiets is daar nie 'n definitiewe patroon sigbaar wat aan die kategorie gekoppel kon word nie aangesien O1 en O5 se hertwiets hoofsaaklik in Afrikaans is terwyl O9 se hertwiets[207] meer in Engels is.

[207] Aangesien hierdie korpus slegs uit tien outeurs bestaan het en daar slegs drie outeurs per kategorie was, is daar spesifiek navorsing oor die aard van Twitter in Afrikaans met 'n baie groter korpus gedoen alvorens so 'n afleiding gemaak kon word.

Vanuit die ondersoek na die drie outeurs wat in die vermaaklikheidskategorie geplaas is, is dit duidelik dat hierdie outeurs (O4, O6 en O7) nie noodwendig voorkeur aan Afrikaanse twiets gee nie en ook dat hulle graag van hertwiets gebruik maak. 56,2% van O4 se twiets is deur die outeur self geformuleer en 43,8% is hertwiets, maar die twiets en hertwiets is oorwegend Afrikaans. O6, daarenteen, se korpus bestaan uit 94,2% eie twiets (waarvan 88,1% in Afrikaans is) terwyl slegs 5,8% van die outeur se twiets hertwiets is (ook hier is die meerderheid van die twiets in Afrikaans). Die verhouding tussen eie twiets en hertwiets in O7 is redelik naby aan mekaar, alhoewel die outeur van meer hertwiets as eie twiets gebruik maak (43,6% eie twiets teenoor 56,4% hertwiets). Tog blyk dit verder dat O7 se twiets en hertwiets oorwegend in Afrikaans is.

Hierdie ondersoek na die verhouding tussen die onderskeie outeurs se twiets, hertwiets en Afrikaanse en Engelse twiets toon dus aan dat daar sekere afleidings is wat wel vanuit hierdie ondersoek gemaak kan word na gelang van die kategorieë waarin die bepaalde outeurs geplaas is. Dit is duidelik dat elk van die onderskeie outeurs steeds hulle persoonlike voorkeur ten opsigte van Engelse en Afrikaanse twiets en hertwiets het. Hierdie patroonmatigheid sowel as ʼn persoonlike voorkeur is verder ondersoek wanneer die verdagte tekste ondersoek is ten einde te bepaal hoe hierdie aspekte bydra tot die identifisering van outeurs van mikroblogs.

6.5.2 Verdagte tekste: taalverhouding in twiets

In afdeling 6.5.1 is daar bevind dat daar wel ʼn verskil waargeneem kan word in die verhouding tussen die Afrikaanse en Engelse twiets sowel as in die aantal keer wat die onderskeie outeurs ander Twittergebruikers se twiets hertwiet. Met nadere ondersoek blyk dit dat die kategorie waarin bepaalde outeurs geplaas is ook ʼn invloed op die gebruik van hertwiets en dies meer het.

Vervolgens word die verdagte tekste se aantal oorspronklike Afrikaanse twiets, Afrikaanse hertwiets, eie Engelse twiets en Engelse hertwiets voorgehou (sien Tabel 6.14) alvorens dit met die aantal eie Afrikaanse twiets, Afrikaanse hertwiets, eie Engelse twiets en Engelse hertwiets van die bekende outeurs vergelyk is ten einde te bepaal of daar patroonmatigheid in die gebruik van die bogenoemde waargeneem kon word.

Tabel 6.14 Taalverhouding in die verdagte tekste

Korpus	**Totale twiets in onverwerkte data**	**Totale Eie twiets**		**Afrikaanse twiets**		**Engelse twiets**		**Totale hertwiets**		**Afrikaanse hertwiets**		**Engelse hertwiets**	
V1	61	43	70,49%	42	68,85%	1	1,64%	18	29,51%	6	9,84%	12	19,67 %

Korpus	Totale twiets in onverwerkte data	Totale Eie twiets		Afrikaanse twiets		Engelse twiets		Totale hertwiets		Afrikaanse hertwiets		Engelse hertwiets	
V2	125	101	80,80%	96	76,80%	5	4,00%	24	19,20%	10	8,00%	14	11,20 %
V3	60	35	58,33%	32	53,33%	3	5,00%	25	41,67%	22	36,67 %	3	5,00%

Vanuit Tabel 6.14 blyk dit dat die taalverhoudings in die drie verdagte tekste wel verskil. Hierdie bevindinge is met die bevindinge in die analise van die taalverhouding van elk van die tien bekende outeurs se korpora vergelyk (sien Tabel 6.13 vir die gespesialiseerde korpus se onderskeie taalverhoudings). Vanuit hierdie vergelyking is dit duidelik dat daar nie ŉ enkele bekende outeur is wie se taalverhoudings sonder twyfel met dié van die verdagte tekste ooreenstem nie (sien Tabel 6.15):

Tabel 6.15 Ooreenkomste tussen die verdagte tekste en tekste vanuit die gespesialiseerde korpus ten opsigte van taalverhouding

Korpus	Totale twiets in onverwerkte data	Totale eie twiets		Afrikaanse twiets		Engelse twiets		Totale hertwiets		Afrikaanse hertwiets		Engelse hertwiets	
V1	61	43	**70,49%**	42	**68,85%**	1	**1,64%**	18	**29,51%**	6	**9,84%**	12	**19,67%**
O1	896	721	80,50%	669	**74,70%**	53	5,90%	175	19,50%	120	13,40%	55	6,10%
O4	900	506	56,20%	484	53,80%	22	**2,40%**	394	43,80%	233	25,90%	161	**17,90%**
O9	559	403	**72,10%**	275	49,20%	126	22,50%	156	**27,90%**	26	4,70%	130	23,30%
O10	858	637	74,20%	501	58,40%	136	15,90%	221	25,80%	92	**10,70%**	129	15,00%
V2	125	101	**80,80%**	96	**76,80%**	5	**4,00%**	24	**19,20%**	10	**8,00%**	14	**11,20%**
O1	896	721	**80,50%**	669	**74,70%**	53	5,90%	175	**19,50%**	120	13,40%	55	6,10%
O6	872	821	94,20%	768	88,10%	54	6,20%	51	5,80%	44	**5,00%**	7	0,80%
O8	995	362	36,40%	308	31,00%	54	**5,40%**	633	63,60%	483	48,50%	150	**15,10%**
V3	60	35	**58,33%**	32	**53,33%**	3	**5,00%**	25	**41,67%**	22	**36,67%**	3	**5,00%**
O2	667	627	94,00%	589	88,30%	38	5,70%	40	6,00%	4	0,60%	36	**5,40%**
O5	573	346	**60,40%**	333	**58,10%**	13	2,30%	227	**39,60%**	185	**32,30%**	42	7,30%
O8	995	362	36,40%	308	31,00%	54	**5,40%**	633	63,60%	483	48,50%	150	15,10%

Tabel 6.15 toon aan met watter korpora in die taalverhoudings ondersoek is (vanuit die gespesialiseerde korpus) en elk van die verdagte tekste wat ooreenstem na gelang van die aspekte (totale eie twiets, Afrikaanse twiets, Engelse twiets, totale hertwiets, Afrikaanse hertwiets en Engelse hertwiets).

Dit is duidelik vanuit Tabel 6.15 dat daar nie ŉ enkele korpus is wat met al die aspekte van ŉ verdagte teks ooreenstem nie, maar V1 en O9, V2 en O1 asook V3 en O5 beskik oor die meeste aspekte wat as ooreenstemmend beskou kan word. Soos in die vorige ondersoeke kan die taalverhoudings van ŉ verdagte teks gebruik word as ŉ middel om moontlike outeurs te identifiseer, maar dié ondersoek kan nie sonder twyfel outeurs van mikroblogs aandui nie. Die

laaste aspek wat vervolgens in die kwantitatiewe analise ondersoek is leesbaarheidstoetse is die laaste ondersoek in hierdie afdeling.

6.6 Leesbaarheidstoetse

Leesbaarheidstoetse verwys na die toetse wat ŉ leesbaarheidsindeks bepaal het van elk van die korpora wat in hierdie studie gebruik is. Daar is talle aanlyn programmatuur beskikbaar om die leesbaarheidsindeks van ŉ teks te bepaal, maar hierdie programme is nie vir Afrikaanse tekste beskikbaar nie. Combrink (1992:i) voer aan dat die Misindeks gebruik kan word om die leesbaarheid van ŉ Afrikaanse prosateks te bepaal. Combrink (1992:i) het hierdie toets in 1989 saamgestel na aanleiding van die *Gunning Fog Index*. Die Misindeks is gebaseer op sinslengte en die gebruik van meerlettergrepige woorde. Combrink (1992:i-ii) het die onderstaande metode saamgestel ten einde die Misindeks te kan bereken:

i. Die teks moet oor 100 woorde of meer beskik – as 100 nie op ŉ sinseinde val nie, sluit die hele sin by die teks in.
ii. Werk die gemiddelde sinslengte uit:

$$Gemiddelde\ woordlengte\ van \sin = \frac{Aantal\ woorde}{Aantal\ sinne^{208}}$$

iii. Tel die aantal woorde met drie of meer lettergrepe in die teks en deel die getal deur twee.
iv. Neem die antwoord vanuit ii en iii verkry en vermenigvuldig met 0,4[209].

Die antwoord wat by stap iv (4) gekry word, is die Misindeks van die teks. Die resultaat is die Misindeks van ŉ teks − ŉ waarde wat die moeilikheidsgraad van ŉ teks aandui. In hierdie studie is daar verwag dat die Misindeks relatief laag sou wees aangesien die tekste wat gebruik is uit slegs ŉ beperkte aantal karakters bestaan het waarmee ŉ boodskap oorgedra is. Volgens Combrink (1992:ii) moet die Misindeks soos volg geïnterpreteer word:

5 en 6 − taamlik maklik.

7 en 8 – standaard.

9 tot 11 – taamlik moeilik.

12 tot 15 – moeilik.

---

[208] Dubbelpunte, kommapunte en aandagstrepe dien as ŉ punt, en sinsdele word as volle sinne getel mits die dele voor en ná die genoemde leestekens volsinne is.

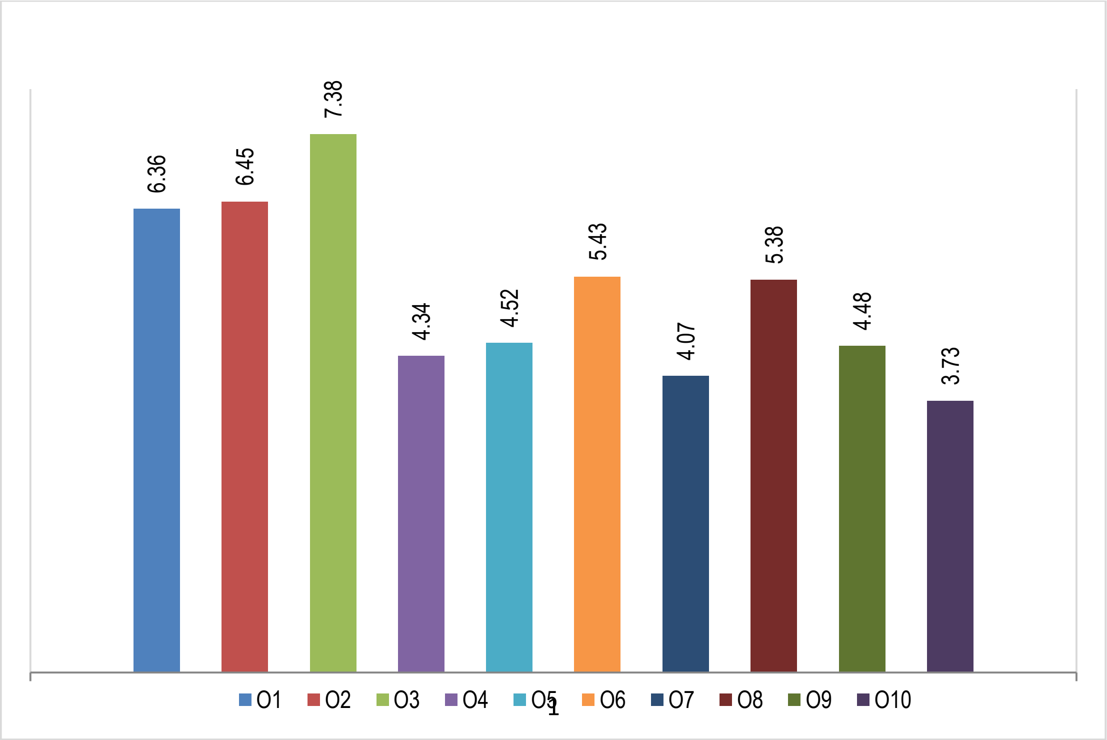
[209] Deur met 0,4 te vermenigvuldig, is die resultaat van die berekining ŉ aanduiding van hoeveel jaar skoolopleiding ŉ leser benodig om die kommunikasielading in ŉ teks te kan verstaan.

16 en hoër – baie moeilik.

Combrink (1992:ii) voer verder aan dat in wetenskaplike en tegniese prosa sekere woorde en woordgroepe nie verkort kan word nie; daarom moet daar voorsiening hiervoor gemaak word, en kan ŉ maksimum van ŉ ekstra twee by elk van die afdelings gevoeg word om meer akademiese tekste te kan akkommodeer.

6.6.1 Gespesialiseerde korpus: Misindeks

Vanuit die misindeksberekeninge[210] is dit duidelik dat daar ooreenkomste tussen die leesbaarheidsvlak van die interne korpusvergelykings voorkom. Die tien outeurs se misindekse is met mekaar vergelyk en daar is telkens ŉ verskil in die indekse geïdentifiseer – sien Figuur 6.16:



Figuur 6.16 Misindeksreslutate vir die gespesialiseerde korpus

Figuur 6.16 verskaf die Misindeks van die gespesialiseerde korpus. Vanuit die analise blyk dit dat daar ŉ waarneembare verskil in die leesbaarheid van die gespesialiseerde korpus sigbaar is en dus kan dié analise ook gebruik word in die ondersoek na die identifisering van mikroblogouteurs.

210 Sien Bylaag M vir die Misindeks-bevindinge.

Alle waardes (sien Figuur 6.16) wat dus laer as ses aangedui is, kan as taamlik maklik beskou word[211], tekste wat dus as "taamlik maklik" beskou is, sluit O4, O5, O6, O7, O8, O9 en O10 in. Ses van hierdie tekste het deel gevorm van die blogger- en vermaaklikheidskategorie terwyl die sewende teks (O8) in die nuuskategorie geplaas is. Tekste met 'n Misindeks tussen 7 en 8 word as standaard beskou, en in hierdie ondersoek is die oorblywende drie korpora (O1, O2 en O3) se leesbaarheid as standaard beskou. O2 en O3 vorm deel van die nuuskategorie en O1 is in die bloggerkategorie geplaas.

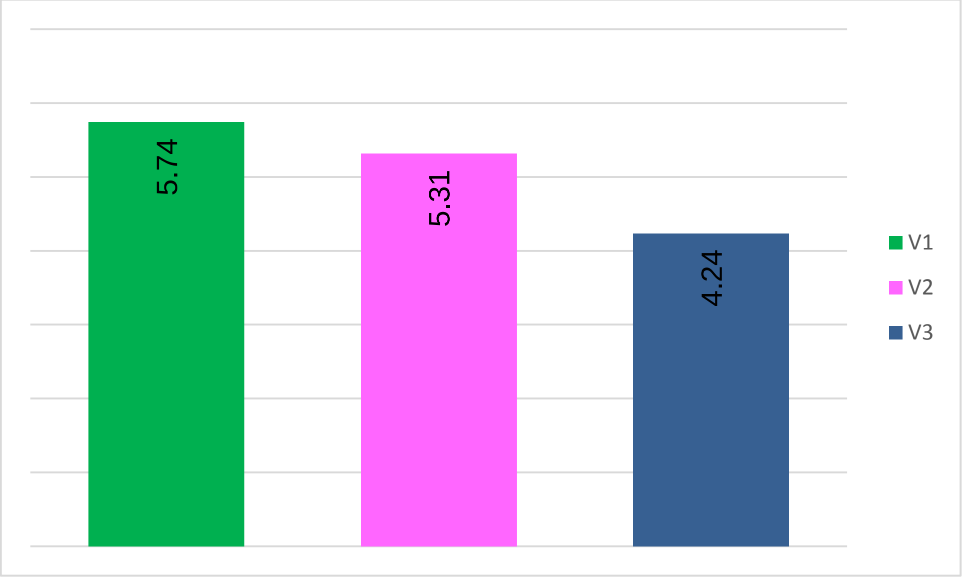
Vanuit die bepaling van die onderskeie korpora se misindekse blyk dit dat daar nie 'n enkele korpus voorkom wat moeiliker as die "standaard" is wanneer die Twitter-tekste geanaliseer is nie. Soos reeds genoem, is daar 'n karakterbeperking op Twitter aanwesig wat wel outeurs "dwing" om hulle tekste so eenvoudig moontlik te hou ten einde die boodskap aan die teikengehoor oor te dra.

Dit is opvallend dat die tekste wat in die nuuskategorie geplaas is, deurlopend oor 'n hoër Misindeks beskik as die tekste vanuit die persoonlike (bloggers-) en vermaaklikheidskategorie. Hierdie tendens kan toegeskryf word aan die feit dat die nuuskategorie oor meer feitlike inligting beskik en omdat die teikengehoor anders daaruit sien as die volgelinge van byvoorbeeld bloggers- en vermaaklikheidsprofiele. Ten slotte kan daar aangevoer word dat na gelang van die analise van die gespesialiseerde korpus dat die Afrikaans wat op Twitter gebruik word 'n gemiddelde Misindeks van 5,21 het of 'n baie maklike leesbaarheidvlak.

6.6.2 Verdagte tekste: Misindeks

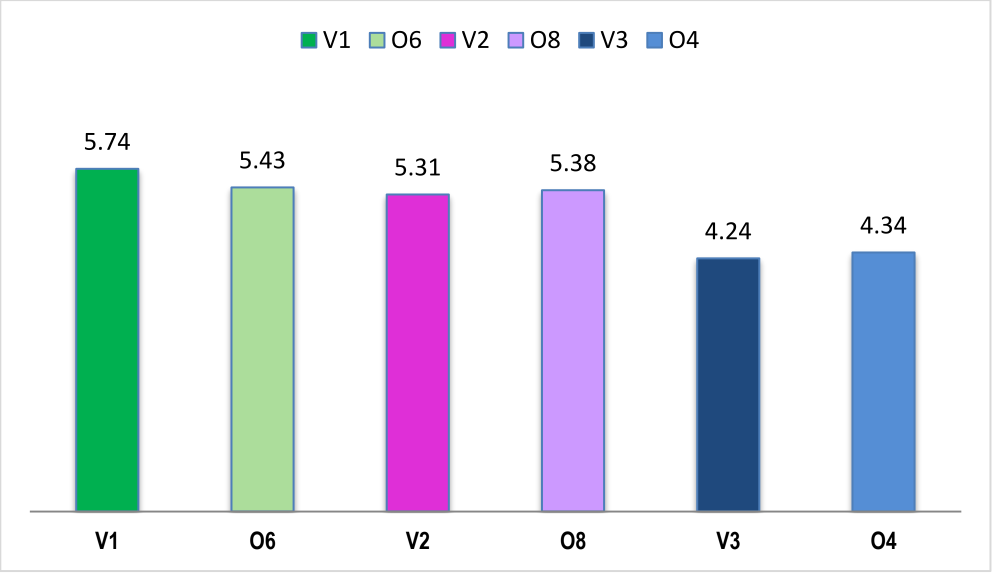
Die leesbaarheid van die gespesialiseerde korpus toon aan dat daar wel onderskeid getref kan word tussen die onderskeie outeurs se tekste. Daar is bevind dat die leesbaarheidsindeks van tekste wat deel uitmaak van die nuuskategorie oor die algemeen moeiliker is as die tekste wat in die bloggers- of vermaaklikheidskategorie geval[212] het en dat die Afrikaans wat op Twitter gebruik word as "baie maklik" ten opsigte van die leesbaarheidsvlak beskou kan word. Die drie verdagte tekste (V1, V2 en V3) is op dieselfde manier as die gespesialiseerde korpora geanaliseer (sien afdeling 6.6).

[211] Sien Combrink (1992:ii) sowel as afdeling 6.7 vir die uiteensetting van die waardes van die leesbaarheidsvlakke van die misindeks.
[212] Die leesbaarheidsindeks van die moeilikste tekste val steeds in die "standaard" groep, dit wil sê die misindeks is minder as agt.



Figuur 6.17 Die misindeks van die verdagte tekste

Figuur 6.17 dui die Misindeks van die drie verdagte tekste aan. Ook in die ondersoek na die leesbaarheid blyk dit dat die drie tekste verskil ten opsigte van hulle leesbaarheid – ten spyte daarvan dat al drie die tekste as "baie maklik" beskou kan word na gelang van Combrink (1992:ii) se definisie.



Figuur 6.18 Misindeks-vergelyking tussen die verdagte tekste en die gespesialiseerde korpus

Wanneer die verdagte tekste met die gespesialiseerde korpus se bevindinge vergelyk is, stem V1 met O6 en V2 met O8 ooreen terwyl V3 en O4 as ooreenstemmend beskou kan word (sien Figuur 6.18) – dit wil sê die korpora toon die grootste ooreenkoms met mekaar. Na gelang van die bevindinge oor die kategorieë waarin die outeurs geval het, kan daar aangevoer word dat

V1, V2 en V3 in óf die bloggers-, óf die vermaaklikheidskategorie moet val. Soortgelyk aan die ander kwantitatiewe analises kan daar dus aangevoer word dat ʼn analise van die Misindeks ook as ʼn manier beskou kan word om verdagte outeurs te identifiseer.

6.7 Sintese van die stilometriese analise

In hierdie hoofstuk is daar ondersoek ingestel na sleutelwoorde; ratiobepalings (lidwoorde, voornaamwoorde, voorsetsels, hulpwerkwoord en voegwoorde); die sinne, woorde en karakters per twiet; die taalverhouding in die twiets en laastens, die leesbaarheidsindeks van die onderskeie korpora. In elk van die afdelings is daar outeurs vanuit die gespesialiseerde korpus geïdentifiseer wat na gelang van die ondersoek as moontlike outeurs van die verdagte tekste beskou is. In hierdie afdeling word daar ʼn oorsig van die bevindinge verskaf ten einde te kan bepaal watter outeurs ná afloop van die stilometriese analise as moontlike outeurs van die verdagte tekste beskou kon word.

Die interne sleutelwaarde-analise toon aan dat die sleutelwaardes van die grammatikale woorde deurgaans minder as 15[213] is en daarom bewys dit dat hierdie korpora deur dieselfde outeurs geskryf is. Vanuit die interne ondersoek na die sleutelwaardes, is dit die vergelyking tussen O1_1 en O1_2 wat die grootste verskil aantoon (8,97), maar aangesien die bevinding steeds laer as 15 is, bewys dit selfs wanneer daar na die hoogste sleutelwaarde van die interne vergelykings verwys word dat die tekste deur dieselfde outeur geskryf is. Hierdie interne vergelyking het dus aangetoon dat sleutelwaardes wel outeurs van mikrobloginskrywings kan aantoon aangesien die interne vergelykings wel moet aantoon dat die tekste deur dieselfde outeurs geskryf is.

Die vier korpora wat ooreenstem ten opsigte van die ondersoek na die sleutelwaardes, is O1 en O5 (4,78), O1 en O9 (6,85), O3 en O4 (14,52) asook O4 en O7 (13,06). Die sleutelwaardes van O1 en O5 asook O1 en O9 is so laag dat daar nie vanuit hierdie analise aangevoer kan word dat die tekste deur verskillende outeurs geskryf is nie. Die sleutelwaardes van O3 en O4 asook O4 en O7 is nader aan die 15-afsnypunt, maar steeds kan daar nie sonder redelike twyfel aangevoer word dat hierdie tekste deur verskillende outeurs geskryf is nie. Aangesien slegs 4 uit 45 van die sleutelwaarde-ondersoeke nie kon aandui dat die tekste deur verskillende outeurs geskryf is nie, kan die ondersoek na die sleutelwaardes as ʼn suksesvolle ondersoek in die identifisering van outeurs van mikrobloginskrywings beskou word. Die ondersoek na die sleutelwaardes is ook op die drie verdagte tekste uitgevoer en die korpora is geïdentifiseer wat na gelang van die sleutelwaarde-analise met die verdagte tekste ooreenstem – sien Tabel 6.16:

[213] Wanneer die sleutelwaardes hoër as 15, dui dit daarop dat die tekste deur verskillende outeurs geskryf is (Kotzé, 2007:393).

Tabel 6.16 Moontlike outeurs van die drie verdagte tekste na gelang van die sleutelwaarde-analise

Verdagte teks	Gespesialiseerde korpus	P-waarde
V1	O2	0,00
	O5	13,92
V2	O1	0,00
	O5	0,00
V3	O1	0,00
	O5	0,00

Tabel 6.16 dui aan dat V1 na alle waarskynlikheid deur O2 geskryf is terwyl beide V2 en V3 deur O1 en O5 geskryf is. In die ondersoek na die sleutelwaardes (sien afdeling 6.2.2) is daar bepaal dat V2 en V3 soveel ooreenkomste toon dat hierdie twee tekste moontlik deur dieselfde outeur geskryf is − 'n bevinding wat dus deur die sleutelwaarde-analise van 'n vergelyking tussen die verdagte tekste en die gespesialiseerde korpus ondersteun word. Die afleiding wat dus op hierdie stadium van die ondersoek gemaak kan word, is dat indien V2 en V3 nie deur dieselfde outeur geskryf is nie, sal die twee tekste minstens in dieselfde kategorie (sien afdeling 5.4) geplaas kon word. Die sleutelwaarde-analise kan as 'n suksesvolle analise beskou word in die proses van outeurskapidentifikasie.

Die analise van die ratio's het gevolg op die sleutelwaarde-analise. Soos in afdeling 6.3 (Ratiobepalings) bespreek, is daar ondersoek ingestel na die gebruik van grammatikale woorde. Vanuit die ratiobepalings kan die afleiding gemaak word dat wanneer tekste vanaf 'n sosialemediaplatform met mekaar vergelyk word (in dié deel van die kwantitatiewe analise is elk van die verdagte tekste met die korpora van die tien bekende outeurs vergelyk), die ratiobepalings nie noodwendig die moontlike outeur van 'n verdagte teks aandui nie, maar die analise van die grammatikale woorde (lidwoorde, voornaamwoorde, hulpwerkwoorde en voegwoorde) kan wel gebruik word om die moontlike outeurs van verdagte tekste te elimineer of die lys van moontlike verdagtes te verminder.

In geen van die analises ten opsigte van die gebruik van grammatikale woorde (hulpwerkwoorde, voornaamwoorde, voegwoorde en voorsetsels) is daar 'n enkele outeur waar die praktiese beduidende verskil kleiner as 81% is nie. Aangesien die persentasie grammatikale woorde dus nie prakties beduidend verskil nie (wat daarop dui dat die tekste deur dieselfde outeurs geskryf is), so laag is, kan daar nie sonder redelike twyfel aangevoer word dat 'n bepaalde outeur V1, V2 of V3 geskryf het nie. In elk van die vergelykings wat uitgevoer is, is

daar tien of meer moontlike outeurs[214] en na gelang van die bevindinge vanuit afdeling 6.3 kan die moontlike outeurs van die drie verdagte tekste verminder word (sien Tabel 6.17):

Tabel 6.17 Moontlike outeurs van die drie verdagte tekste na gelang van die ratiobepalings

Verdagte teks	Grammatikale woord	Moontlike outeurs	Persentasie grammatikale woorde wat nie prakties beduidend verskil het nie
V1	Hulpwerkwoorde	**O2**	19%
	Voegwoorde	**O2**	15%
	Voornaamwoorde	**V2**	12%
		V3	12%
		O1	10%
		O5	10%
	Voorsetsels	**O1**	16%
V2	Hulpwerkwoorde	**O1**	14%
	Hulpwerkwoorde	**O3**	14%
	Voegwoorde	**O9**	8%
	Voornaamwoorde	**O1**	19%
	Voorsetsels	**O1**	19%
V3	Hulpwerkwoorde	**O7**	19%
	Voegwoorde	**O5**	3%
		O7	3%
	Voornaamwoorde	**O4**	13%
		O8	13%
	Voorsetsels	**O9**	13%
		O4	13%
		O8	13%
		O9	13%

Tabel 6.17 bied ŉ opsomming van die outeurs wat die grootste ooreenkomste met die verdagte tekste toon toe die gebruik van hulpwerkwoorde, voegwoorde, voornaamwoorde en voorsetsels ondersoek is. Vanuit die Tabel is dit duidelik dat daar nie ŉ enkele outeur is wat in al die verdagte tekste se ratiobepalings uitstaan as ŉ moontlike outeur nie. In die geval van V1 kom O2 en O1 gereeld voor as outeurs. V2 en O1 toon baie meer ooreenkomste terwyl V3 en O4, O7 en O8 ooreenkomste toon. Soos reeds genoem, kan die ratiobepalings vervolgens gebruik word om die aantal moontlike outeurs van die verdagte tekste te verminder. Die outeurs wat in hierdie analise gereeld ooreenkomste toon, is verder met behulp van die verdagte tekste ondersoek.

In die analise van die lidwoorde is daar wel van die lidwoorde wat ŉ 50% prakties beduidende verskil aantoon, maar aangesien daar slegs twee lidwoorde is (bepaald en onbepaald), is hierdie bevinding verwag. In die ratiobepalings van die eerste verdagte teks (V1) is daar geen lidwoorde wat nie prakties beduidend verskil nie en dus is die lidwoordgebruik van V1 uniek in

[214] Ten einde die ondersoek so volledig moontlik te kon uitvoer, is die verdagte tekste ook met mekaar vergelyk.

dié sin dat geen van die bekende outeurs waarmee die korpus van V1 vergelyk is ten opsigte van die gebruik van lidwoorde moontlike outeurs van die teks kan wees nie.

In die ondersoek na die praktiese beduidendheid van die lidwoordgebruik van V2 in vergelyking met die bekende outeurs verskil 50% van die lidwoorde in die vergelyking met O4, O7 en O10 nie prakties beduidend nie. Dus kan die aanname gemaak word dat 50% van die lidwoorde wat deur V2 gebruik word in só ŉ mate met die gebruik van lidwoorde deur O4, O7 en O10 ooreenstem dat hierdie drie outeurs moontlik die tweede verdagte teks geskryf het.

Die derde verdagte teks (V3) se ratiobepalings oor die gebruik van lidwoorde toon aan dat 50% van die lidwoorde wat deur die onderskeie outeurs gebruik word (O10 en O8) nie prakties beduidend verskil nie. Die vergelyking tussen V3 en O2 toon aan dat 100% van die lidwoorde nie prakties beduidend verskil nie, en dus is O2 as ŉ moontlike outeur van die derde verdagte teks beskou.

Die ratiobepalings moet in die identifisering van mikroblogouteurs ondersoek word al kan die analises nie sonder redelike twyfel ’n enkele outeur van ŉ verdagte tekste aantoon nie. Die analises elimineer steeds van die outeurs wat as moontlike verdagtes van tekste beskou word. Sinne, woorde en karakters per twiet is daarna ondersoek en vanuit dié ondersoek is daar ook bewyse van outeurskap geïdentifiseer (sien Tabel 6.18).

Soos in die analises wat reeds in die sintese bespreek is, het die ondersoek in die eerste plek intern op die tekste van die tien outeurs van die gespesialiseerde korpus gefokus. In die ondersoek – sien afdeling 6.4.1 – is daar bevind dat die korpora van ’n enkele outeur wél tot so ’n mate ooreenstem dat daar nie redelike twyfel bestaan dat die tekste deur ’n enkele outeur geskryf is nie. Die tekste van die tien outeurs van die gespesialiseerde korpus is daarna met mekaar vergelyk om te bepaal of daar wel verskille in die gemiddelde gebruik van die aantal woorde, sinne en karakters per twiet voorkom.

Soortgelyk aan die verwagte bevinding van die interne ondersoek, is daar ook hier bevind dat ’n beduidende verskil in die gemiddelde aantal sinne, woorde en karakters per twiet wat per outeur gebruik is, voorkom. Daar kan dus aangevoer word dat die tekste wel deur verskillende outeurs geskryf is. Ten einde die bogenoemde bevindinge te kon verifieer, is die verdagte tekste en die gespesialiseerde korpus met mekaar vergelyk (sien Tabel 6.18):

Tabel 6.18 Moontlike outeurs van die drie verdagte tekste na gelang van die ondersoek na die sinne, woorde en karakters per twiet

Korpus	Gespesialiseerde korpus	Totale aantal sinne in korpus	Gemiddelde aantal sinne per twiet	Gemiddelde aantal woorde per twiet	Gemiddelde aantal karakters per twiet
V1		90	3,00	20,07	121,21
	O2	1 577	2,67	20,22	123,17
V2		147	**1,63**	**15,66**	**81,71**
	O3	266	1,29	**15,31**	99,02
	O4	978	2,02	14,14	**82,95**
	O5	556	**1,67**	13,9	67,76
V3		44	**1,38**	**12,88**	**61,28**
	O1	944	**1,41**	13,17	68,23
	O7	790	1,73	**12,1**	**62,29**

Vanuit die bevindinge van die ondersoek na die sinne, woorde en karakters per twiet soos in Tabel 6.18 voorgehou, kan O2 as die moontlike outeur van V1 beskou word. V2 toon ooreenkomste met O3, O4 en O5. As een van die outeurs as ’n moontlike verdagte outeur van V2 geïdentifiseer kan word na gelang van die ooreenkomste in die gemiddelde aantal sinne, woorde en karakters per twiet, kan O4 as sodanig beskou word. V3 stem met O1 en O7 ooreen, maar aangesien O7 se gemiddelde aantal sinne per twiet slegs 0,32 hoër as dié van O1 is, kan O7 as die moontlike outeur van V3 beskou word.

Die ondersoek na die gemiddelde aantal sinne, woorde en karakters per twiet kan as ’n suksesvolle ondersoek in die identifiseringsproses van mikroblogouteurs beskou word. Die taalverhouding in die twiets van die onderskeie outeurs is daarna ondersoek. In hierdie analise is die verhouding tussen die aantal Afrikaanse en Engelse twiets sowel as die aantal hertwiets bereken. Soos in die ondersoek na die sinne, woorde en karakters per twiet is die ondersoek eers as ’n interne vergelyking op die gespesialiseerde korpus uitgevoer alvorens die tien outeurs van die gespesialiseerde korpus se tekste met mekaar vergelyk is. Vanuit die ondersoek is daar bevind dat die interne vergelyking grotendeels ooreenstem.

Hierdie bevinding is verwag aangesien die tekste wel deur dieselfde outeurs geskryf is. Die vergelyking van die gespesialiseerde korpus het aangedui dat daar wel ’n waarneembare verskil in die taalverhouding van die twiets sigbaar is (weer eens soos verwag aangesien hierdie tekste deur verskillende outeurs geskryf is). Ten einde hierdie bevindinge te kon verifieer, is die verdagte tekste se taalverhouding met dié van die gespesialiseerde korpus vergelyk (sien Tabel 6.19):

Tabel 6.19 Moontlike outeurs van die drie verdagte tekste na gelang van die ondersoek na die taalverhouding in die twiets

Korpus	Gespesialiseerde korpus	Totale eie twiets	Afrikaanse twiets	Engelse twiets	Totale hertwiets	Afrikaanse hertwiets	Engelse hertwiets
V1		**70,49%**	**68,85%**	**1,64%**	**29,51%**	**9,84%**	**19,67%**
V2		**80,80%**	**76,80%**	**4,00%**	**19,20%**	**8,00%**	**11,20%**
	O1	**80,50%**	**74,70%**	5,90%	**19,50%**	13,40%	6,10%
V3		**58,33%**	**53,33%**	**5,00%**	**41,67%**	**36,67%**	**5,00%**
	O5	**60,40%**	**58,10%**	2,30%	**39,60%**	**32,30%**	7,30%

Vanuit Tabel 6.19 is daar ’n enkele outeur wat as die moontlike outeur van elk van die drie verdagte tekste voorgehou kan word. Alhoewel daar meer as een outeur is wat telkens met die verdagte tekste ooreenstem (sien Tabel 6.18), word die outeur wat die grootste ooreenkoms toon hier aangedui. In hierdie ondersoek is dit slegs V1 wat nie ’n duidelike ooreenkoms met ’n enkele outeur toon nie (O1, O4, O9 en O10 stem ooreen met elke aspek wat in die ondersoek na die taalverhouding geanaliseer is).

In die geval van V2 en V3 is daar ’n enkele outeur (onderskeidelik O1 en O5) wat as moontlike outeur van die teks beskou kan word. Ten spyte daarvan dat die analise van die taalverhouding in V1 nie die gewenste resultate gelewer het nie, is die bevindinge van V2 en V3 genoegsaam om te kan aanvoer dat die ondersoek na die taalverhouding in twiets wel ondersoek moet word in die ondersoek na die analise van mikroblogouteurs.

Die laaste ondersoek in die stilometriese analise was die ondersoek na die leesbaarheid van beide die gespesialiseerde korpus en die verdagte tekste. Hierdie analise is met behulp van die Misindeks (Combrink, 1992:i) uitgevoer.

’n Waarneembare verskil is vasgestel in die leesbaarheid van die gespesialiseerde korpus (sien afdeling 6.7.1) en dus kan dié analise ook gebruik word in die ondersoek na die identifisering van mikroblogouteurs. In die analise van die verdagte tekste is daar bepaal dat V1 met O6 en V2 met O8 ooreenstem terwyl V3 en O4 as ooreenstemmend beskou kan word. Die korpora toon dus die grootste ooreenkoms met mekaar.

Na gelang van die bevindinge oor die kategorieë waarin die outeurs geval het, kan daar aangevoer word dat V1, V2 en V3 in of die bloggers-, of die vermaaklikheidskategorieë moes val. Dit kom dus voor of die ondersoeke wat in die stilometriese analise uitgevoer is telkens suksesvol was ten einde die outeurs van die verdagte tekste te kon identifiseer. Tabel 6.20 is ’n opsomming van die bevindinge van die stilometriese analise. Vanuit hierdie opsomming kan die gevolgtrekking dus gemaak word dat O2 in vyf van die aspekte wat in hierdie afdeling

ondersoek is as die moontlike outeur van V1 geïdentifiseer kan word. O1 is in vyf van die aspekte wat ondersoek is as die outeur van V2 geïdentifiseer terwyl O5 drie keer as die moontlike outeur van V3 geïdentifiseer is.

Tabel 6.20 Opsomming van die moontlike outeurs van die verdagte tekste na gelang van die stilometriese analise

Korpus	V1	V2	V3
Sleutelwaarde	O2	O1 & O5	O1 & O5
Ratiobepaling: lidwoorde	-	O4, O7 & O10	O2
Ratiobepaling: voornaamwoorde	O1 & O5	O1	O4, O8 & O9
Ratiobepaling: voorsetsels	O1	O1	O4, O8 & O9
Ratiobepaling: hulpwerkwoorde	O2	O1 & O3	O7
Ratiobepaling: voegwoorde	O2	O9	O5 & O7
Sinne, woorde en karakters per twiet	O2	O4	O7
Taalverhouding in twiets	-	O1	O5
Leesbaarheidstoetse	O6	O8	O4

Alhoewel daar nie sonder twyfel aangevoer kan word dat V1 deur O2, V2 deur O1 en V3 deur O5 geskryf is nie, is hierdie outeurs (O1, O2 en O5) geïdentifiseer as die sterkste kandidate vir outeurskap na gelang van die stilometriese analise. Die outeurs van die verdagte tekste word in afdeling 7.10.2 en 8.3.2.2 aangedui.

6.8 Samevatting

In hierdie stilometriese analise is daar ondersoek ingestel na die sleutelwaardes van grammatikale woorde. Verder het ek na die ratio’s van frekwente woorde, die leksikale woorde, lidwoorde, voorsetsels, voornaamwoorde en hulpwerkwoorde ondersoek ingestel. Daar is ondersoek ingestel na die sinne, woorde en karakters per twiet, die taalverhouding van die twiets, en daar is ook leesbaarheidstoetse uitgevoer. In elk van die analises is outeurs geïdentifiseer as moontlike outeurs van die verdagte tekste (sien afdeling 6.7).

Outeurskapidentifikasie kan dus wel op Twitter uitgevoer word na gelang van al die bogenoemde aspekte wat deel uitgemaak het van die stilometriese analise. Rocha *et al.* (2017:6) beweer dat korter tekste verskeie uitdagings voorhou, en dat langer tekste meer stilometriese eienskappe openbaar as korter tekste. Soos in afdeling 5.2 (navorsingsontwerp) bespreek is, het ek van ’n gemengde metode gebruik gemaak om te verseker dat die bevindinge vanuit hierdie afdeling (kwantitatiewe analise) geverifieer kon word. Vanuit hierdie afdeling (stilometriese analise) kan die veronderstelling gemaak word dat die stilometriese analise reeds ’n moontlike aanduiding van outeurskap kan gee.

Die ondersoek na die verdagte tekste toon aan dat daar nie sonder redelike twyfel 'n enkele outeur geïdentifiseer kan word vir elk van die verdagte tekste nie, maar dat die moontlike verdagte outeurs verminder kan word. Na gelang van die kwantitatiewe analise van die verdagte tekste het dit vervolgens geblyk dat O1, O2, O5 en O6 as moontlike outeurs van V1 beskou kan word, maar dat O2 as die gewildste verdagte outeur vir V1 beskou word.

In die geval van V2 is dit slegs O2 wat glad nie as 'n moontlike outeur beskou word nie, maar O1 is die outeur wat die meeste ooreenkomste met V2 toon en kan dus as die beste moontlikheid vir outeurskap na gelang van die stilometriese analise beskou word. O10 is die enigste outeur vanuit die gespesialiseerde korpus wat nie as 'n moontlike outeur van V3 beskou word nie, maar soortgelyk aan V2 kan daar na gelang van die stilometriese analise aangevoer word dat O5 as die sterkste kandidaat vir outeurskap van V3 beskou word. Vervolgens is 'n kwalitatiewe analise uitgevoer en daar is veral op ooreenkomste tussen V1 en O2, V2 en O5 asook V3 en O5 gefokus om te bepaal of hierdie outeurs wat deur die kwantitatiewe analise as moontlike outeurs van die verdagte tekste geïdentifiseer is wel V1, V2 en V3 geskryf het.

Hoofstuk 7

EMPIRIESE ANALISE: STILISTIESE EN TEKSANALISE

For last year’s words belong to last year’s language
And next year’s words await another voice.

T.S. Eliot *Four Quartets* (1943)

7.1 Inleiding

Hierdie hoofstuk fokus op die beskrywende, kwalitatiewe analise (stilistiese en teksanalise) wat gedoen is. Die stilistiese analise, soos in hoofstuk 5 bespreek, verwys volgens McMenamin (2002:163) na die beginsel dat geen twee outeurs se taalstyl dieselfde is nie, en dat daar selfs variasie binne ’n enkele outeur se skryfstyl sigbaar is[215]. Hierdie stylkenmerke kan deur middel van ’n stilistiese analise geïdentifiseer word waar daar volgens McMenamin (2002:163) na patrone van variasie in geskrewe taal ondersoek ingestel kan word. Wybenga (1987:2) argumenteer dat stilistiek ook beskou kan word as ’n “studie van die afwyking van die normale” en dat daar in só ’n ondersoek vasgestel word wat vir ’n sekere gemeenskap op ’n bepaalde tyd vir ’n spesifieke genre en inhoud as die norm beskou kan word. Dít wat dan van die norm afwyk, word as stilistiek beskou.

Stilistiek kan na gelang van die bogenoemde argument van Wybenga (1987:2) as die wetenskap van die onverwagte en styl as die uitbuiting van die norm beskou word. Daar kan dus geargumenteer word dat ’n stilistiese analise betrekking het op die identifisering van stylmerkers in ’n bepaalde teks, en die moontlikheid dat hierdie stylmerkers kan afwyk van dít wat as die norm beskou word.

Ten einde die stilistiese en teksanalise te kon uitvoer, het ek elk van die datastelle geanaliseer ten opsigte van agt verskillende kategorieë. Die kategorieë is gekies na gelang van navorsing wat reeds oor outeurskapidentifikasie gedoen is soos bespreek in hoofstuk 5[216]. In die bespreking is daar na McMenamin (2002), Coulthard en Johnson (2007), Michell (2013), Thiart (2014) en Kotzé en Oosthuizen (2015) se metodes van outeurskapidentifikasie verwys. Vanuit die bespreking is die agt kategorieë saamgestel. Hierdie kategorieë sluit in[217]:

[215] Sien afdeling 2.3.2.3 (Idiolek).
[216] Sien afdeling 5.6.1 (Stilistiese analise).
[217] Sien hoofstuk 5 vir die uiteensetting van die verskillende kenmerke wat stilisties ondersoek kan word.

1. **Algemeen**: die algemene kategorie het die ondersoek na algemene kenmerke ingesluit:
    - **Twitter-spesifieke tekselemente**: Die Twitter-spesifieke tekselemente verwys na elemente wat kenmerkend aan sosialemediaplatforms is, naamlik hiperskakels, die gebruik van die hutsteken en die @-gebruikerskonvensie − wat steeds van gebruiker tot gebruiker kan verskil. McMenamin (2002) verwys spesifiek na die analise van ŉ teksformaat, maar aangesien die teksformaat grotendeels deur die platform bepaal word (sien Skermskoot 2.7), is daar in die ondersoek byvoorbeeld nie na die manier waarop die datum geskryf word of lynspasiëring ondersoek ingestel nie.
    - **Leestekengebruik**: ondersoek is ingestel na die gebruik van punte aan die einde van sinne, die gebruik van ŉ ellips asook die gebruik van meer as een leesteken (twee of meer uitroeptekens, vraagtekens en dies meer).
    - **Hoofletters**: die ondersoek na hoofletters het die gebruik van hoofletters ingesluit aan die begin van ŉ sin, hooflettergebruik ten einde ŉ woord te beklemtoon en ook die gebrek aan hoofletters in eie- en selfstandige naamwoorde.
2. **Sintaksis**: die sintaktiese kenmerke van die twiets:
    - **Sinstipe**: verwys na die gebruik van stel-, vraag-, uitroep- of bevelsinne.
    - **Woordsoorte**: Daar is ondersoek ingestel na die sinstipes en die gebruik van woordsoorte.
    - **Die invloed van Engels**: verwys na sinne wat ŉ invloed vanaf Engels toon − hierdie invloed kan ook as anglisisties beskryf word.
3. **Morfologie**: morfologiese afwykings in die onderskeie korpora is ondersoek en het ingesluit:
    - **Afkortings**: Verwys na die proses waardeur ŉ nuwe woord gevorm word "op grond van die beginletters van ŉ woord of woordgroep". Daar is onderskeid getref tussen "ware afkortings" en akronieme volgens Van Huyssteen (2014:198). Vir die doel van hierdie studie is daar na foutlose en foutiewe afkortings, akronieme en SMS-afkortings ondersoek ingestel.
    - **Inkorting**: "Ter wille van ekonomie in taalgebruik maak taalgebruikers dikwels in geskrewe taal gebruik van inkorting om iets kort en kragtig te stel" (Van Huyssteen, 2014:198). Daar word onderskeid getref tussen twee verskillende

soorte inkorting, naamlik *knipsels sonder die inagneming van die morfologiese struktuur*, en *knipsels met die inagneming van die morfologiese struktuur*[218].

- **Weglating**: die proses waartydens ŉ letter of ŉ klank van ŉ woord weggelaat word − soortgelyk aan sinkopee in formele taalkunde (Thiart, 2014:80).

4. **Semantiek**: Semantiek kan beskou word as tekens in taal, waar tekens in ŉ abstrakte manier na alles wat inligting oordra verwys. Ten einde abstrakte tekens in Twitter-taal te ondersoek word die gebruik van emotikons geanaliseer, ten einde die semantiese waarde en uniekheid van hierdie tekens in twiets te bepaal.

5. **Kodewisseling**: Namba (2004:2) verwys na die matriks-taal (ML) en die ingebedde taal (EL), waar ML na sprekers se dominante taal verwys. Myers-Scotton (2001:25) verwys na die kontakfenomeen waar verskillende tale bydra tot ŉ tweetalige uiting, en waar nie net die ML bydra tot die grammatikale struktuur van ŉ woord nie. Vir die doel van die studie verwys **dubbelmorfologie** na Engelse woorde wat ŉ Afrikaanse voorvoegsel bevat. Verder word daar ook ondersoek ingestel na die gebruik van Engelse woorde in die twiets. Ek verwys ook na die gebruik van Engelse woorde in hutswoorde.

6. **Tussenwerpsels**: Rijhwani *et al. (*2017:1973) verwys na tussenwerpsels as ŉ "certain string of alphabetic characters representing generic interjections or sounds." Lockyer (2014:69) voer aan dat die belangrikste doel van tussenwerpsels is om emosie (positief of negatief) oor te dra en "woorde" word hoofsaaklik in informele diskoers-situasies gebruik. Ook Biber *et al. (*2003:450) definieer tussenwerpsels as woorde wat as ŉ uitroep dien en emosie oordra. Meinard (2014:156) stel dat tussenwerpsels "context dependent" is "… and interjections can sum up a whole sentence in a non-elliptical manner and depend on the context they appear in."

7. **Kraswoorde**: Ondersoek is ingestel na die frekwensie waarteen sommige outeurs kraswoorde gebruik.

8. **Foute-analise**: Die ondersoek na spelling verwys na woorde wat verkeerd gespel is, en herhaaldelik só verkeerd gespel word en dus as ŉ kenmerk van ŉ bepaalde outeur beskou kan word. Verder is daar ondersoek ingestel na die weglating van die dubbel nie

[218] Vir die doel van hierdie studie is daar nie onderskeid getref tussen knipsels met, en knipsels sonder die inagneming van die morfologiese struktuur nie.

in ontkenning, die skryf van woorde soos dit in gesproke taal gebruik word[219], die verkeerde spelling van die ŉ en ook die gebruik van voegwoorde aan die begin van ŉ sin.

9. **N-gram analise**: verwys in rekenaarlinguistiek na ŉ reeks items (*n*) wat in opeenvolging in gesproke of geskrewe taal voorkom.

Die bogenoemde nege kategorieë, met die onderskeie onderafdelings, is in die stilistiese ondersoek geanaliseer. Die analise is op elk van die tien datastelle uitgevoer, en ek het ook intern ondersoek ingestel na die voorkoms van die genoemde kenmerke ten einde te kon bepaal of daar verskille in die individue se stylkenmerke voorgekom het. Elk van die deelnemers se waarnemings word eers individueel voorgehou alvorens die vergelykende analise bespreek word. Daar word ook telkens na die verwysingskorpus verwys ten einde te kon bepaal of kenmerke eie aan ŉ deelnemer is en of dit ook op Twitter in die algemeen gevind kon word. Ná afloop van hierdie kwalitatiewe analise, is daar ondersoek ingestel na die verdagte tekste ten einde te kon bepaal of outeurs van mikrobloginskrywings suksesvol geïdentifiseer kan word. Ten einde die analises te kon uitvoer, het ek van *ATLAS.ti* gebruik gemaak. Vervolgens word die analise van die algemene kenmerke voorgehou.

7.2 Algemeen

Die algemene kenmerke verwys na Twitter-spesifieke tekselemente, leestekengebruik en die gebruik van hoofletters. Ten einde hierdie analise te kon uitvoer, is daar van *ATLAS.ti* en WordSmith Tools gebruik gemaak. Ten einde *ATLAS.ti* te kon gebruik, is die onderskeie korpora (verwerk[220] om in WordSmith Tools gebruik te kon word) as txt-dokumente[221] in die program ingevoer. Kodes is vir elk van die aspekte wat ondersoek is, geskep en die tekste is dienooreenkomstig geanaliseer. Vervolgens word die ondersoek na die Twitter-spesifieke tekselemente, leestekengebruik en hoofletters bespreek.

7.2.1 Twitter-spesifieke tekselemente

McMenamin (2002:216) verwys na tekselemente wat die volgende insluit: verskille in die uitleg van dokumente, die manier waarop ŉ datum gebruik word, leestekengebruik met betrekking tot aanspreekvorms en lynspasiëring − die uitleg van twiets word grotendeels deur ŉ betrokke

[219] Alhoewel die skryf van woord soos dit in gesproke taal gebruik word as ‘n emotiewe kenmerk van RBK beskou kan word, het ek besluit om die gebruik van woorde as deel van die foute analise te analiseer.
[220] Sien afdeling 5.12.2 (Dataverwerking).
[221] Sien Tabel 5.5 vir die uiteensetting van die korpora soos geanaliseer in die kwalitatiewe analise.

platform bepaal[222]. In die analise oor Twitter-spesifieke[223] tekselemente, het ek na die gebruik van die hutsteken, die verwysing na gebruikers en hiperskakels ondersoek ingestel aangesien bogenoemde aspekte nie deur ’n die platform bepaal word nie, maar deur gebruikers.

Hong *et al. (*2011) het ’n studie gedoen waarin daar ondersoek ingestel is na die manier waarop verskillende taalgebruikers verskeie aspekte[224] van Twitter gebruik. Hierdie outeurs (2017:519) verwys na die navorsing van Honeycutt en Herring (2009), boyd *et al.*, F(2010) en Kwak *et al.* (2010) oor die funksies van die @-gebruikerskonvensie asook hertwiets en volgelinge en argumenteer dat hierdie navorsing slegs betrekking het op twiets in Engels. Hulle (Hong *et al.*, 2011:519) het daarna navorsing gedoen oor die gebruik van hiperskakels, hutswoorde, die @-gebruikerskonvensie, antwoorde op twiets en ook hertwiets in 104 verskillende tale. Ten einde die bevindinge voor te hou, het Hong *et al.* (2011) die persentasie van elk van die bogenoemde aspekte bereken wat in elk van die 104 tale voorkom. Hierdie bevindinge van Hong *et al.* (2011) is dus in die ondersoek gebruik om te bepaal of Afrikaans soos dit op Twitter gebruik word, ooreenkomste met internasionale bevindinge toon.

Die Twitter-spesifieke kenmerke is eerstens geanaliseer ten opsigte van die frekwensie waarteen elk van die outeurs wat deel uitgemaak het van die studie die spesifieke kenmerk gebruik – ’n genormaliseerde waarde tot 1 000 woorde[225] is bereken. Daar is bepaal of die interne vergelyking van elk van die outeurs in die gespesialiseerde korpus tot só ’n mate ooreengestem het dat daar aangevoer kon word dat die frekwensie van die voorkoms van die bepaalde aspek wel kenmerkend aan ’n outeur is[226]. Die verdagte tekste is op dieselfde manier geanaliseer, en die genormaliseerde waarde is met die bevindinge vanuit die ondersoek na die gebruik van algemene kenmerke in die gespesialiseerde korpus vergelyk. Vir elk van die Twitter-spesifieke kenmerke is die persentasie van gebruik ook bereken[227] en met Hong *et al.* (2017:519) se bevindinge vergelyk. Die analise van die gebruik van hutswoorde word eerstens voorgehou.

7.2.1.1 Hutswoorde

Zappavigna (2015:274-275) argumenteer dat hutswoorde op mikroblogplatforms (soos Twitter) ontstaan het en daarna na ander platforms (televisie en bemarking) versprei het. Hutswoorde is volgens Bruns en Burgess (2011) en Zappavigna (2015:275) gebruikersgegenereerde

---

[222] Sien hoofstuk 2 (afdeling 2.6.2.2 – Twitter).

[223] Alhoewel baie van hierdie kenmerke ook op ander sosialemediaplatforms aangetref word, word daar vir die doel van hierdie studie daarna verwys as Twitter-spesifieke elemente.

[224] Die aspekte sluit hiperskakels, hutswoorde, verwysings na gebruiker, hertwiets en antwoorde op twiets in.

[225] Alhoewel sommige gebruikers wel van meer as een woord na die hutsteken gebruik maak (dit wil sê daar is soms ’n sinsnede ná die hutsteken) kan die hutsteken en die daaropvolgende woord(e) as ’n enkele entiteit beskou word; daarom normaliseer die waardes ná 1 000 woorde. Net só is die @ gebruikerskonvensie en die hiperskakel as ’n enkele entiteit (woord) beskou.

[226] Sien Bylaag E.

[227] Sien afdeling 7.2.1.1 vir die uiteensetting van die persentasie berekeninge vir die gebruik van die aspekte soos genoem.

sleutelwoorde (sleutelwoorde kan na 'n enkele woord, afkorting, frase of selfs 'n sin verwys) met 'n hutsteken (#) as voorvoegsel.

Die hutsteken word algemeen op Twitter gebruik en boyd *et al.* (2010) sowel as Tsur en Rappoport (2012:643-644) voer aan dat die gebruik van 'n hutsteken gebruikers toelaat om 'n tema aan hulle twiet te koppel (om die reikwydte van twiets te vergroot of om twiets in 'n internasionale konteks te plaas) wat ander Twitter-gebruikers toelaat om twiets te vind wat oor 'n bepaalde onderwerp handel. Zappivigna (2015:275) sluit aan by die gebruik van hutswoorde deur te beklemtoon dat hutswoorde verskillende betekenisse kan oordra soos die semantiese domein van twiets om twiets te verbind met bestaande hutswoorde (#FF – sien voorbeeld 283) of om as emosioneel gelaaide metakommentaar te dien wat verskillende taalkundige funksies kan verrig.

Ten einde te kon onderskei tussen die hutsteken (#) en die hutswoord(e) wat saam met die hutsteken[228] gebruik is, het ek vervolgens na 'n hutsteken verwys: net na die # self en hutswoorde wat na die # en die woord(e) (afkorting, frase of sin) wat daarop volg verwys. Aangesien Twitter-gebruikers gereeld van 'n hutsteken gebruik maak, is dit moontlik dat sekere hutswoorde, die hutsteken-frekwensie of selfs die lengte van hutswoorde kenmerkend van 'n bepaalde individu se idiolek kon wees.

Die eerste ondersoek wat uitgevoer is ten einde die hutstekengebruik te ondersoek, was die voorkoms van hutswoorde in elk van die datastelle. Die interne vergelyking[229] het aangetoon dat daar in drie van die interne vergelykings van die gespesialiseerde korpus 'n definitiewe ooreenkoms in die gebruik van hutswoorde was. Die interne vergelyking van die oorblywende sewe korpora het aangetoon dat die aantal hutstekens wat per korpus gebruik is nie konstant was nie en daar nie 'n definitiewe patroon gevolg is nie. Die gebruik van hutswoorde is dus verder geanaliseer. Die vergelyking van die gespesialiseerde korpus se gebruik van hutswoorde het 'n definitiewe verskil aangetoon met betrekking tot die frekwensie van die gebruik van hutswoorde (sien Tabel 7.1). Wanneer die verdagte tekste se genormaliseerde waarde (tot 1 000 woorde) met die gespesialiseerde korpus vergelyk is, het dit geblyk dat daar tog ooreenkomste in die frekwensie van die gebruik van die hutswoorde waargeneem kon word (sien Tabel 7.1):

Tabel 7.1 Die gebruik van hutstekens in die onderskeie korpora

Korpus	**Totale aantal Twiets**	**Totale aantal sinne**	**Totale aantal woorde**	**Totale hutswoorde in korpus**	**Maksimum # per twiet**	**Hutswoorde genormaliseer tot 1 000 woorde**	**Gemiddelde hutswoorde per twiet**
O1	669	944	8 942	24	4	2,68	0,04

228 Sien Skermskoot 2.6 (Voorbeeld van 'n twiet met 'n hutswoord).

229 Sien Bylaag E.

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale hutswoorde in korpus	Maksimum # per twiet	Hutswoorde genormaliseer tot 1 000 woorde	Gemiddelde hutswoorde per twiet
O2	589	1577	12 259	397	3	32,38	0,67
O3	206	266	3 379	37	2	10,95	0,18
O4	484	978	7 110	121	3	17,02	0,25
O5	333	556	4 579	14	2	3,06	0,04
O6	768	1657	16 340	47	2	2,88	0,06
O7	457	790	5 655	4	1	0,71	0,01
O8	308	443	3 690	283	5	76,69	0,92
O9	275	467	4 918	16	4	3,25	0,06
O10	499	1190	7 744	363	4	46,88	0,73
V1	42	90	844	0	0	0,00	0,00
V2	96	147	1409	1	1	0,71	0,01
V3	32	44	412	1	1	2,43	0,03

Vanuit Tabel 7.1 is dit duidelik dat wanneer hierdie bevindinge (die gebruik van hutswoorde in die verdagte tekste) met die analise van die gebruik van hutswoorde vanuit die gespesialiseerde korpus vergelyk is, daar ’n aansienlike verskil in die frekwensie van die gebruik van hutstekens was en daar kon in die geval van V2 (0,71) en V3 (2,43) aangevoer word dat die genormaliseerde waarde van die hutstekengebruik ooreenkomste getoon het met onderskeidelik O7 (0,71) en O1 (2,68). Die frekwensie waarteen elk van die outeurs hutswoorde gebruik, kan dus reeds in ŉ mate as ŉ aanduiding van outeurskap dien.

Wanneer die maksimum hutswoorde ondersoek is wat elk van die outeurs per twiet gebruik, het dit geblyk dat daar vanuit die data afgelei kon word dat daar gemiddeld nie meer as vier hutswoorde per twiet gebruik word nie, en dat ’n gemiddeld van twee hutswoorde per twiet as die aanvaarde gebruik in Afrikaans op Twitter beskou kan word. Ten einde die maksimum hutswoorde per twiet sowel as die gemiddelde aantal hutstekens per twiet te kon verifieer, is daar ook na die verwysingskorpus ondersoek ingestel. In die gespesialiseerde korpus was nege hutswoorde per twiet die maksimum wat aangetref is (die nege hutswoorde per twiet het slegs eenmalig voorgekom). Die gemiddelde aantal hutswoorde wat per twiet gebruik is, (sien Tabel 7.1) was tussen 0,01 en 0,96 per twiet.

Ten einde ŉ vergelyking tussen die aantal hutswoorde wat in Afrikaans gebruik is te kon tref met die bevindinge vanuit Hong *et al.* (2017:519) se studie, is die aantal twiets in die gespesialiseerde korpus getel waarin daar hutswoorde voorkom. Dit wil sê dat as een van die outeurs vier hutswoorde in ŉ twiet gebruik het, is daar net een hutswoord getel aangesien daar bepaal moes word in hoeveel persent van die totale twiets vanuit die gespesialiseerde korpus daar hutswoorde gebruik is.

Daar is 942 twiets in die gespesialiseerde korpus wat hutswoorde bevat. Die totale aantal twiets in die gespesialiseerde korpus is 4 588. Die persentasie is dus soos volg bereken:

$$Persentasie\ twiets\ in\ Afrikaans\ waarin\ hutswoorde\ gebruik\ word = \frac{Hoeveelheid\ twiets\ waarin\ hutswoorde\ voorkom}{\text{Totale aantal twiets in die gespesialiseerde korpus}} \times 100$$

$$\therefore\ Persentasie\ twiets\ in\ Afrikaans\ met\ hutswoorde = \frac{942}{4588} \times 100$$

$$\therefore Persentasie\ twiets\ in\ Afrikaans\ met\ hutswoorde = 20{,}5$$

In Hong *et al.* (2011:519) se ondersoek is daar bevind dat Duits die hoogste gebruik van die hutswoorde (18%) toon, gevolg deur Engels (14%) en Nederlands (13%). Die gebruik van hutswoorde in Afrikaans is dus effens hoër as die taal waarin die meeste hutswoorde voorkom, maar tog plaas dit die gebruik van hutswoorde in Afrikaans in konteks met die gebruik daarvan in ander tale.

Die ondersoek na die frekwensie van die gebruik van hutswoorde het insiggewende data gelewer, maar ten einde die kenmerke wat as uniek tot die gebruik van hutswoorde deur elk van die onderskeie outeurs beskou word te kon identifiseer, word daar in die hieropvolgende onderafdelings (afdeling 7.2.1.1.1 en 7.2.1.1.2) verwys na die manier waarop hutswoorde deur die outeurs gebruik is.

7.2.1.1.1 Gespesialiseerde korpus: hutswoorde

Soos in Tabel 7.1 aangedui, maak elk van die outeurs vanuit die gespesialiseerde korpus wel van hutswoorde gebruik. Voorbeelde vanuit die gespesialiseerde korpus word vervolgens voorgehou alvorens hierdie voorbeelde sowel as kenmerke van die onderskeie outeurs bespreek word:

Voorbeeld 1: **O1** – x My sussie: "Noem dit face of the devil spawn." #ekhetditdesign #prettynifty #selfportrait… https://www.instagram.com/p/BFEfI4Ppk

Voorbeeld 2: **O1** – Die dosent dink ek het my naam van Karen Zoid gekry, minwetend dat die account al 3 jaar bestaan. #syleerditbymy #KCOM112

Voorbeeld 3: **O2** – Vir joernaliste: ""@GEBRUIKER[230]: How to avoid 3 common mistakes with #statistics - our latest guide: https://goo.gl/eUiYc2 #factchecking"

Voorbeeld 4: **O2** – Verkragtingstorie 6.50 in die oggend op #RSG se #Monitor. Kan aan beter tipes nuus dink om die oggend mee te begin. #goeieNuusAsb!

Voorbeeld 5: **O3** – Uitspraak voorbehou in die #OscarPistorius saak

Voorbeeld 6: **O3** – #jesuischarlie

Voorbeeld 7: **O4** – Hierdie babies is beautiful!! #lekkerpaartekke

Voorbeeld 8: **O4** – the wingless moth vs The Lion #nog'nkakfliekidee

[230] Ten einde die identiteit van die gebruikers waarna die onderskeie outeurs in hulle twiets verwys te beskerm, is die gebruiker se naam (byvoorbeeld @mariskanel) telkens met die woord "GEBRUIKER" vervang (@GEBRUIKER).

Voorbeeld 9: **O4** – Vroeg op vir die boks, Lekke chillie so vroeg in die oggend, dis ma MAYwether vir jou! #moerhompacquiao!

Voorbeeld 10: **O5** – Elke 2de meisie is deesdae 'n #kaalvoetkind, maar hulle kan skaars yskas toe loop sonder skoentjies.

Voorbeeld 11: **O6** – Vanaand, 17:30, is ek op #LekkerTV (kanaal 146) met 'n onderhoud op die bank. Ons gesels lekker oor allerlei... http://fb.me/1TE3U50hp

Voorbeeld 12: **O6** - Of #ek #verstaan #iets #nie, #maar #hierdie #manier #van #skryf #irreteer #die #hel #uit #my #uit. #geirreteerd.... http://fb. me/7FpTVRxWm

Voorbeeld 13: **O7** – #Wildveiling: En nou.... Horison, die buffelbul van die aand, is onder die hamer. @GEBRUIKER

Voorbeeld 14: **O7** – "Hy het vir die boer gesê hy gaan hom doodmaak." #Rouxville #plaasmoord http://www.netwerk24.com/Nuus/Hof/rouxville-moord-2-in-hof-20160202 …

Voorbeeld 15: **O8** – #Dans in #OggendOpRSG @GEBRUIKER "Rock daai lyfie"... in jou badkamer, in jou motor, op jou kantoortafel... @GEBRUIKER @GEBRUIKER

Voorbeeld 16: **O8** – Go #Bulls ! Van 'n Stormer, maar mede-Suid-Afrikaner A201#bullsvsrebels #SuperRugby

Voorbeeld 17: **O9** – WTAF. Hoe sterk kan iemand se lighter game wees? Is 'n lighter 'n valid rede om verlief te raak op iemand? #FMJFOK.

Voorbeeld 18: **O9** – Elke dag wat ek nie hierdie stuk kak mens se skedel vergruismet my stapler nie. #ThatIsStrengthThatIsGrace

Voorbeeld 19: **O10** – #ClubChamps16 Na 10 boulbeurte is die @GEBRUIKER 54-1 op die telbord.

Voorbeeld 20: **O10** – Skare sing voort "#FeesMustFall." na die uitstelling van die vergadering. @GEBRUIKER @GEBRUIKER

Voorbeeld 1 tot 20 toon die verskillende maniere aan waarop hutswoorde deur die onderskeie outeurs in die gespesialiseerde korpus gebruik is. Hierdie voorbeelde dui reeds aan dat die doel van hutswoorde nie net verskil in die twiets nie, maar dat daar verskillende maniere is waarop die hutswoorde gebruik kan word.

Wanneer die posisie ondersoek is waar die onderskeie gebruikers kies om hutswoorde te plaas, is daar bepaal dat die meeste hutswoorde aan die einde van ŉ twiet geplaas word (voorbeelde 18, 17, 13, 8, 7 en 2) of dat hutswoorde in die sin geplaas word sonder om die leesbaarheid of koherensie van ŉ sin te beïnvloed (sien voorbeelde 4, 10, 11, 12 en 16). Hutswoorde word ook aan die begin van ŉ sin/twiet aangetref (sien voorbeelde 13, 15 en 19). Voorbeeld 13 kan as ŉ hapax legomena[231] beskou word aangesien die gebruik van ŉ hutswoord gevolg deur ŉ

[231] "Hapax legomena" verwys na woorde wat slegs een maal in ŉ teks voorkom (Coulthard & Johnson, 2010:526) – sien ook afdeling 3.5.1 (Benaderings tot outeurskapidentifikasie).

dubbelpunt slegs een maal in die gespesialiseerde korpus aangetref is. O7 het die hutswoord in dié geval as 'n opsomming van die tema van die twiet gebruik.

'n Verdere bevinding in die analise van die plasing van hutswoorde was dat O1, O2, O5, O6, O7, O9 en O10 hutswoorde altyd ná die leesteken aan die einde van die voorafgaande sin plaas. O4 plaas hutswoorde voor en ná die leesteken terwyl O3 (voorbeeld 5 en 6) nie van leestekens gebruik maak in 'n sin/twiet waar Twitter-spesifieke tekselemente voorkom nie[232]. O2 en O3 plaas hutswoorde altyd voor die verwysing na 'n gebruiker wanneer hierdie twee Twitter-spesifieke tekselemente aan die einde van 'n sin gebruik word.

Daar kan ook tussen twee kenmerkende gebruike onderskei word wanneer hutswoorde en hiperskakels aan die einde van 'n sin gebruik word. Terwyl daar nie 'n patroonmatigheid in die korpora van O4, O3, O6, O8, O9 en O10 geïdentifiseer kon word nie, plaas O2 hutswoorde slegs ná die hiperskakel terwyl O1, O5 en O7 hutswoorde telkens voor die hiperskakel plaas.

Alhoewel daar slegs in die geval van O7 en O3 kenmerke oor die plasing van hutswoorde geïdentifiseer is wat in geeneen van die ander outeurs in die gespesialiseerde korpus aangetref is nie, kan die bevindinge oor die plasing van hutswoorde steeds gebruik word ten einde die outeurs van mikrobloginskrywings te identifiseer.

'n Verdere gebruik wat ondersoek is, was die gebruik van leestekens en hoofletters in hutswoorde. Dit is slegs in die korpora van O2 (voorbeeld 4) en O4 (voorbeeld 9) wat die outeurs leestekens as deel van hutswoorde gebruik. Dit is opvallend dat O2 en O4 in beide die twee voorbeelde 'n sin ná die hutswoorde gebruik.

Ten opsigte van die gebruik van hoofletters in hutswoorde is daar vanuit die gespesialiseerde korpus waargeneem dat O3, O9 en O10 hutswoorde met hoofletters skryf. Só ook sal hierdie outeurs (O3, O9 en O10) − sien voorbeelde 20, 19, 18, 17 en 5 − elke woord wat in hutswoorde gebruik word met 'n hoofletter skryf. Wanneer 'n sin of frase ná die hutsteken volg, word elk van die woorde ook met 'n hoofletter geskryf (#LagMaar #DisSeer #BroekeSalSyDra).

'n Opvallende kenmerk in die korpus van O10 is dat hierdie outeur aanvanklik die hutswoord #MyNWU sonder enige hoofletters gebruik (#mynwu), maar dan later wel van hoofletters gebruik maak (#MyNWU). In teenstelling met O10 wat hoofletters begin gebruik het, het O1 aanvanklik hoofletters (#MyNWU en #KCOM121) gebruik, maar met verloop van tyd het O1 eenvoudig nie meer van hoofletters gebruik gemaak nie (#mynwu en #kcom121).

O2 en O7 (voorbeelde 4 en 14) pas deurlopend taalreëls toe oor die gebruik van hoofletters en die outeurs gebruik deurlopend hutswoorde met hoofletters behalwe wanneer hutswoorde in die

[232] O3 gebruik nie leestekens aan die einde van 'n sin as daar hutswoorde, verwysings na gebruikers of hiperskakels aan die einde van die sin gebruik word nie.

middel van die sin geplaas is en dit nie ŉ eienaam was nie. Indien hutswoorde (wat nie eiename is nie) aan die einde van ŉ sin/twiet geplaas word, word dié woord steeds met ŉ hoofletter geskryf. O2 maak ook van hoofletters gebruik om te onderskei tussen die begin van ŉ nuwe woord wanneer ŉ sin of frase ná hutstekens gebruik word (#KykTVenWordSlim)[233].

O4 het van 121 hutswoorde (sien Tabel 7.1) gebruik gemaak. 51 van hierdie hutswoorde (of 42% van die hutswoorde) is met hoofletters geskryf. Geen patroonmatigheid is gevind ten opsigte van die gebruik van hoofletters deur O4 nie. O8 se gebruik van hoofletters in hutswoorde het ook geen definitiewe patroonmatigheid geïdentifiseer nie, die outeur het byvoorbeeld in ŉ twiet na *#Tanzania* verwys en enkele twiets later *#tanzania* gebruik. O5 het slegs twee keer van hoofletters in hutswoorde gebruik gemaak en in beide gevalle is die woord *Afrikaans* met ŉ hoofletter geskryf. O6 is die enigste outeur wat deurlopend nie hoofletters in hutswoorde gebruik nie – slegs ŉ enkele voorbeeld (sien voorbeeld 11) is gevind waar hoofletters in ŉ hutswoord gebruik is en dié gebruik kan toegeskryf word aan die amptelike[234] hutsterm wat deur kanaal 146 gebruik word.

Sommige twiets het slegs uit ŉ hutswoord (voorbeeld 6) bestaan en wanneer die twiet slegs uit ŉ hutswoord bestaan het, kan die leser telkens self die konteks en die bedoelde boodskap aflei. Alhoewel die gebruik van slegs ŉ hutswoord as ŉ twiet nie as ŉ afwyking van die norm beskou kan word nie[235], is dit slegs O5 en O3 vanuit die gespesialiseerde korpus wat net van ŉ hutswoord in ŉ twiet gebruik maak en indien daar in die verdagte tekste ŉ voorbeeld is waar ŉ twiet net uit ŉ hutswoord bestaan, sal O5 of O3 as die moontlike outeurs van die teks oorweeg kan word. In hierdie ondersoek is daar, soos bespreek, bepaal dat daar woorde, frases en sinne ná die hutsteken gebruik word (sien voorbeeld 1 tot 20) − ŉ gebruik wat deur al die outeurs in die gespesialiseerde korpus gevolg is en wat by Bruns en Burgess (2011) en Zappavigna (2015:275) se navorsing oor die gebruik van hutswoorde in Engels aansluit.

Vanuit die ondersoek oor die gebruik van hutswoorde is daar ook bevind dat O1, O2, O3, O8 en O10 van afkortings[236] en akronieme[237] in hutswoorde gebruik maak terwyl O8 en O9 ook SMS-afkortings[238] in hutswoorde gebruik (sien voorbeeld 17). In die korpus van O2, O4 en O10 is daar morfologiese afwykings in hutswoorde aangetref. O2 maak van die inkorting vir die woord Potchefstroom (*#By***Potch***Tronk*) gebruik – die voorkoms van dié inkorting vir Potchefstroom

[233] Daar is ŉ enkele sin wat op die hutstekens volg waar daar nie deur middel van die gebruik van hoofletters tussen die aanvang van nuwe woorde onderskei kan word nie (sien voorbeeld 4),

[234] Daar is ŉ soektog op Twitter na die hutswoord *#LekkerTv* uitgevoer (27 September 2017), en vanuit die soektog blyk dit dat dit as die norm beskou kan word om hoofletters in die hutswoord te gebruik.

[235] Na gelang van ŉ soektog op Twitter (20 September 2017) sowel as ŉ ondersoek in die verwysingskorpus is daar wel twiets wat slegs uit ŉ enkele hutswoorde bestaan.

[236] Sien afdeling 7.4 (Morfologie).

[237] Sien afdeling 7.4 (Morfologie).

[238] Sien afdeling 7.4 (Morfologie).

word in verskeie hutswoorde aangetref indien ŉ soektog op Twitter en in die verwysingskorpus uitgevoer is.

O4 gebruik die woord *tekke* in ŉ hutswoord (*#lekkerpaar***tekke**). In ŉ soektog in die verwysingskorpus sowel as op Twitter is daar bevind dat *tekke* nie in ander hutswoorde in Afrikaanse twiets gevind kon word nie. Daarom kan die gebruik van *tekke* in ŉ hutswoord (sowel as ŉ twiet[239]) as kenmerkend tot die taalgebruik van O4 beskou word. *#wiele***oppi***pad* is nog ŉ voorbeeld van ŉ morfologiese afwyking (weglating) wat in O4 se hutswoorde aangetref word. Hierdie hutswoord kan as ŉ hapax legomena[240] beskou word aangesien die gebruik van hierdie hutswoord slegs een maal in die korpus van O4 aangetref is, en ŉ soektog in die verwysingskorpus en Twitter na die woord *oppi* het slegs verwysings na OppiKoppi gelewer. In die korpus van O10 kom die morfologiese afwyking *leka* (weglating – lekker) in ŉ hutswoord voor (*#FinalejaarIsLeka*). Alhoewel die woord *leka* telkens op Twitter gebruik word en in die verwysingskorpus voorkom, is die morfologiese afwyking slegs een keer in die gespesialiseerde korpus geïdentifiseer. Daarom kan die gebruik van hierdie woord, veral in hutswoorde, as kenmerkend tot die taalstyl van O10 beskou word.

ŉ Kenmerk wat ook in die gespesialiseerde korpus se taalgebruik, ten opsigte van hutswoorde geïdentifiseer is, is die gebruik van tussenwerpsels[241] in hutswoorde. O1 (*#ohwell*), O4 (*#eksêwhoooohaoo*) en O2 (*#Sjoe!*) gebruik ŉ enkele voorbeeld van tussenwerpsels in hulle hutswoorde, en daar kan aangevoer word dat indien ŉ verdagte teks tussenwerpsels in hutswoorde gebruik, O1, O2 en O4 as moontlike outeurs van die verdagte teks oorweeg kan word.

ŉ Verdere ondersoek is uitgevoer ten einde die taalgebruik in hutswoorde te ondersoek. In die korpora van O1, O4, O8 en O9 kom kodewisseling voor ten opsigte van die hutswoorde en twiets (die twiets is in Afrikaans en die hutswoorde in Engels) en in die hutswoorde self (sien voorbeeld 1). In teenstelling met O1, O4 en O8 gebruik O6 slegs kodewisseling tussen die taal van die twiets en die hutswoorde – in hierdie studie kan dit dus as kenmerkend tot die taalstyl van O6 beskou word. In die twiets van O2, O3 en O7 is hutswoorde deurlopend in dieselfde taal as die twiet gebruik. Daar is ŉ enkele voorbeeld in die korpus van O2 (*#DisSad*) en O10 (*#GooiSomeGees*) waar kodewisseling in die hutswoorde voorkom. In die korpus van O5 kom kodewisseling slegs voor wanneer die outeur ŉ kraswoord gebruik. Alhoewel O5 nie die enigste outeur is wat kraswoorde in hutswoorde gebruik nie, is dit kenmerkend dat hierdie outeur se hutswoorde altyd in Afrikaans is behalwe wanneer die hutswoord ŉ kraswoord is –

[239] Sien afdeling 7.4 (Morfologie) – die morfologiese afwyking *tekke* (tekkies – skoene) word sewe keer in die korpus van O4 aangetref.

[240] "Hapax legomena" verwys na woorde wat slegs een maal in ŉ teks voorkom (Coulthard & Johnson, 2010:526) – sien ook afdeling 3.5.1 (Benaderinge tot outeurskapidentifikasie).

[241] Sien afdeling 7.6 (Tussenwerpsels).

hierdie kenmerk is nie in enige van die ander korpora aangetref nie, en kan vervolgens as kenmerkend tot O5 se taalstyl beskou word.

Vanuit die ondersoek na die gebruik van hutswoorde, kan die gevolgtrekking gemaak word dat die gebruik van kraswoorde[242] as deel van hutswoorde ŉ kenmerk van Afrikaans kan wees soos dit op Twitter gebruik word. Alhoewel voorbeeld 21 tot 24 slegs in die korpora van O4 en O5 voorkom, is daar etlike voorbeelde van kraswoorde wat deel uitmaak van die hutswoorde in die verwysingskorpus − en ook op Twitter self gevind word – en daarom kan die voorbeelde nie as kenmerkend van die bepaalde outeurs beskou word nie. Wanneer daar met ŉ beperkte aantal verdagte tekste vanaf Twitter gewerk word (soos in die geval van my studie), kan daar wel aangevoer word dat die gebruik van kraswoorde in hutswoorde as ŉ aanduiding van outeurskap beskou kan word. Voorbeeld 21 tot 24 beeld die gebruik van kraswoorde as hutswoorde uit:

Voorbeeld 21: **O4** – Vroeg op vir die boks, Lekke chillie so vroeg in die oggend, dis ma MAYwether vir jou! #**moer**hompacquiao!

Voorbeeld 22: **O4** − As jou kinders nog nie by die sirkus was nie, wys hulle more die State of the Nation Address.. #hierkom**kak**

Voorbeeld 23: **O4** – Boys, vandag gaan sunblock nie werk nie, verf jou met 4 blikke plascon! Drink julle bier vandag vinnig anders word dit tee! #hitte**fokken**golf

Voorbeeld 24: **O5** – #dingewatmy**afpis** : ouens wat in elke message "hehehehe" tik.

In die geval van O9 het dit geblyk asof hierdie outeur nie die hutsteken voor ŉ sinsnede plaas nie (sien voorbeeld 25 en 26), maar eerder net ŉ frase met die kraswoord[243] aan die einde van die twiet:

Voorbeeld 25: **O9** – Vergeet van 'n afspraak waarvoor ek R400 upfront neer gesit het. "R200 boete as jy herskeduleer, of jy verloor..." **Hahahahfokjousienjoumore.**

Voorbeeld 26: **O9** – Wanneer iemand julle planne kanselleer vir die Boeremark: Uhm, okay? Dis nie asof ek al 'n week beplan wat ek gaan eet nie. **Dankiefokjoubye.**

O9 gebruik dus nie die hutsteken nie, maar die stylkenmerke wat in hutswoorde gevind word, is wel sigbaar. Hierdie kenmerk kan as kenmerkend van die outeur beskou word aangesien dit nie in enige van die tekste van die oorblywende nege outeurs in die gespesialiseerde korpus gesien is nie[244].

Vanuit die ondersoek na die gebruik van hutswoorde in die gespesialiseerde korpus is daar bevind dat die tien outeurs hutswoorde oorwegend gebruik om die primêre boodskap van ŉ twiet te beklemtoon alhoewel die hutswoorde nie altyd as ŉ uitbreiding van die onderwerp van ŉ

[242] Sien ook afdeling 7.7. (Kraswoorde).
[243] Die gebruik van kraswoorde word in afdeling 7.7 verder bespreek.
[244] Die verwysingskorpus is ook gebruik, en daar is ook nie in hierdie korpus voorbeelde gevind waar ŉ Twitter-gebruiker die styl van ŉ hutswoord gebruik, maar nie ŉ hutsteken aan die begin van die frase plaas nie.

twiet beskou kan word nie (sien voorbeeld 15 en 16). O2, O3, O7 en O10 gebruik hutswoorde veral om ŉ twiet met ŉ spesifieke nuuswaardige gebeurtenis te skakel, en in enkele gevalle om met ŉ internasionale gehoor te skakel (sien voorbeeld 5 en 6). In die geval van O5 en O9 se gebruik van hutswoorde, word hutswoorde gebruik om ŉ spesifieke emosie aan die gehoor oor te dra of om ŉ opinie oor ŉ spesifieke gebeurtenis oor te dra (kwaad – *#dingewatmyafpiss).* Hierdie bevindinge oor die doel van hutswoorde in Afrikaans sluit aan by die navorsing van boyd *et al.* (2010), Tsur en Rappoport (2012:643-644) en Zappavigna (2015:275)

Vanuit afdeling 7.2.1.1.1 is dit duidelik dat die ondersoek na die gebruik van hutstekens nie baie betroubaar is om outeurs te identifiseer nie. Wanneer daar 'n beperkte aantal verdagte outeurs is, kan hierdie kenmerk wel gebruik word as ŉ stilistiese kenmerk om die outeurs van tekste te identifiseer of om die aantal moontlike verdagte outeurs te verminder.

Vervolgens is die gebruik van hutswoorde in die verdagte tekste ondersoek om sodoende te bepaal of die kenmerke oor die gebruik van hutswoorde (soos in afdeling 7.2.1.1.1) wel ŉ rol kan speel in die identifisering van die outeurs van mikrobloginskrywings.

7.2.1.1.2 Verdagte tekste: hutswoorde

In afdeling 7.2.1.1 word daar verwys na die frekwensie van die gebruik van hutswoorde in die gespesialiseerde korpus en die verdagte tekste asook die ooreenkoms in die frekwensie van die gebruik van hutswoorde (sien Tabel 7.1). Afdeling 7.2.1.1.1 hou die stylkenmerke van die gebruik van hutswoorde in die gespesialiseerde korpus voor. In hierdie afdeling is daar ondersoek ingestel na die manier waarop hutswoorde in die verdagte tekste gebruik is om sodoende te kan bepaal of daar outeurs vanuit die gespesialiseerde korpus is wie se gebruik van hutswoorde met die verdagte tekste ooreenstem. Soos in Tabel 7.1 aangetoon, kom daar slegs in die korpus van die tweede (V2) en derde (V3) verdagte tekste hutswoorde voor en dan ook net een hutsteken per korpus – sien voorbeeld 27 en 28:

Voorbeeld 27: **V2** – Ek weet wie môre in die hoof se kantoor gaan wees. **#GEBRUIKER**

Voorbeeld 28: **V3** – Dis goed om te droom. Maak net seker dat jy nie nét droom nie, maar beplan ook. **#fortunecookie**

In hierdie twee voorbeelde van die gebruik van hutswoorde wat in die verdagte tekste voorkom, is die hutstekens in beide die voorbeelde aan die einde van die twiet geplaas. In die geval van V2 (voorbeeld 27) is die hutsteken 'n eienaam wat die onderwerp van die twiet voorhou – dus versterk die hutswoord die primêre boodskap van die twiet. Voorbeeld 28 (V3) kan as 'n voorbeeld van kodewisseling beskou word waar die twiet in Afrikaans is, maar die hutswoord in Engels – die hutswoord dui aan waar die inligting vir die twiet gevind is.

Wanneer die bogenoemde aspekte van die twiets vergelyk is met die bevindinge vanuit afdeling 7.2.1.1.1, blyk dit dat O1 hutswoorde gebruik om die primêre boodskap van die twiet te beklemtoon. O7 en O8 maak ook van hutswoorde gebruik om die onderwerp van die twiet te beklemtoon. Daar kan dus aangevoer word dat ten spyte van die beperkte aantal hutswoorde wat in die korpus van V2 voorkom, O1, O7 en O8 as moontlike outeurs van die teks geïdentifiseer kan word na gelang van ooreenkomste in die gebruik van hutswoorde

In die geval van hutswoorde wat in V3 voorkom, was dit moeilik om moontlike outeurs op grond van hutswoorde te identifiseer aangesien die manier waarop V3 hutswoorde gebruik as ŉ algemene gebruik op Twitter beskou kan word. In die korpus van V1 is daar geen hutswoorde nie en daarom kan daar nie ŉ vergelyking getref word nie.

Die ondersoek na die gebruik van die hutswoorde was dus nie suksesvol ten einde outeurs van mikrobloginskrywings te identifiseer nie, maar dit bly steeds ŉ kenmerk wat in die ondersoek na outeurskapidentifikasie op mikroblogs uitgevoer kan word aangesien die gebruik van hutswoorde telkens kan verander.

Die afleiding kan vervolgens gemaak word dat die gebruik van hutswoorde nie net na gelang van die frekwensie van gebruik ondersoek moet word nie, maar ook ten opsigte van die manier waarop hutswoorde gebruik word. Alhoewel die gebruik van hutswoorde nie ŉ algemene verskynsel in die verdagte tekste is nie, is daar steeds stylooreenkomste sigbaar tussen die hutswoorde soos in die verdagte tekste gebruik en in die gespesialiseerde korpora.

V2 se gebruik van ŉ eienaam as ŉ hutswoord om sodoende die onderwerp van die twiet te beklemtoon sowel as die gebruik van kodewisseling in die hutswoord omweer eensdie boodskap van die twiet te versterk, stem ooreen met die gebruik van hutswoorde deur O1, O7 en O8. Hierdie drie outeurs is dus na gelang van die ondersoek na die gebruik van hutswoorde moontlike outeurs van V2. V1 en V3 se gebruik van hutswoorde kan nie as sodanig beskou word en moontlike outeurs van die tekste na gelang van die ondersoek na die aspek van die teksformaat kan nie geïdentifiseer word nie. V3 se gebruik van hutswoorde kan wel as ooreenstemmend met dít wat as algemene kenmerke van die gebruik van hutswoorde op Twitter (in Afrikaans) geïdentifiseer is, beskou word.

7.2.1.2 Verwysing na gebruikers

Die verwysing na gebruikers verwys na die @-teken wat gebruik word voor ŉ spesifieke gebruiker se naam om sodoende die gebruiker by die gesprek in te sluit[245]. Hierdie @-gebruikerskonvensie word volgens Honeycutt en Herring (2009) gebruik om twiets aan spesifieke gebruikers te rig en kan dus as ŉ aanspreekvorm beskou word. Werry (1996:52) sluit by Honeycutt en Herring aan deur te stel dat hierdie aanspreekvorm gebruikers se aandag daarop vestig dat die boodskap spesifiek vir hulle bedoel is − ŉ ietwat problematiese aspek op sosiale media. Huberman *et al*. (2008:3) onderskei in dié verband tussen direkte en indirekte boodskappe. ŉ Direkte boodskap sluit die @-gebruikerverwysing in terwyl ŉ indirekte boodskap nie aan ŉ spesifieke individu gerig word nie. Deur ŉ direkte verwysing na ŉ gebruiker in te sluit, ontvang die spesifieke gebruiker (waarna daar verwys is) ŉ boodskap/kennisgewing dat daar na hom/haar op Twitter verwys word.

Honeycutt en Herring (2009:2) verwys na die gebruik van die @-gebruikerkonvensie as ŉ manier om koherensie op aanlyn platforms te skep en om een-tot-een (persoonlike) kommunikasie na te boots – dus word daar met ŉ spesifieke gebruiker in gesprek getree. Ten spyte van die bogenoemde, argumenteer Honeycutt en Herring (2009:5-6) dat die funksie van die verwysing na gebruikers ook gebruik word om na ŉ spesifieke gebruiker te verwys, om ŉ spesifieke ligging aan te dui (@ Restaurant) of die gebruik van die @-teken in ŉ e-posadres.

Aangesien daar nie spesifieke riglyne oor die plasing van die @-gebruikerskonvensie bestaan nie, kan outeurs kies waar in die twiet hulle dit wil plaas. Vervolgens is die verwysing na gebruikers ondersoek ten einde te kon bepaal of hierdie konvensie as ŉ kenmerkende eienskap van outeurs op Twitter beskou kan word.

Soos in die ondersoek na die gebruik van hutswoorde, is daar ondersoek ingestel na die aantal verwysings na gebruikers wat in elk van die datastelle gebruik is, die aantal @-tekens wat per twiet gebruik is sowel as die gemiddelde aantal verwysings na gebruikers wat per datastel voorgekom het (hierdie waarde is tot 1 000 woorde genormaliseer – sien Tabel 7.2).

Tabel 7.2 @-gebruikerskonvensies in die onderskeie datastelle

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale verwysing na gebruikers in korpus	Maksimum verwysing na gebruikers per twiet	Verwysing na gebruikers per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal verwysings na gebruikers per twiet
O1	669	944	8 942	24	**1**	2,68	0,04
O2	589	1 577	12 259	28	**1**	2,28	0,05
O3	206	266	3 379	180	**3**	53,27	0,87
O4	484	978	7 110	233	**6**	32,77	0,48
O5	333	556	4 579	3	**1**	0,66	0,01

[245] Sien ook afdeling 2.6.2.2 (Twitter).

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale verwysing na gebruikers in korpus	Maksimum verwysing na gebruikers per twiet	Verwysing na gebruikers per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal verwysings na gebruikers per twiet
O6	768	1 657	16 340	30	**2**	1,84	0,04
O7	457	790	5 655	40	**5**	7,07	0,09
O8	308	443	3 690	299	**7**	81,03	0,97
O9	275	467	4 918	47	**8**	9,56	0,17
O10	499	1 190	7 744	904	**5**	116,74	1,81
V1	42	90	844	1	1	1,18	0,02
V2	96	147	1409	5	2	3,55	0,05
V3	32	44	412	1	1	2,43	0,03

Tabel 7.2 dui die totaal aan wat die @-gebruikerskonvensie (verwysing na gebruikers) in die onderskeie datastelle gebruik is, die maksimum verwysings na gebruikers per twiet, die gemiddelde aantal verwysings na gebruikers per datastel sowel as die genormaliseerde waarde (per 1 000 woorde) van die verwysings na gebruikers per datastel. Vanuit Tabel 7.2 kan daar talle gevolgtrekkings oor die gebruik van die @-gebruikerskonvensie gemaak word.

Eerstens, blyk dit dat die onderskeie outeurs se frekwensie van die gebruik van die @-gebruikerskonvensie per korpus verskil. Die interne vergelyking het aangetoon dat die interne korpora ooreenstem met die gebruik van die verwysings na gebruikers (sien bylaag E – gemiddelde aantal verwysings na gebruikers per twiet).

ŉ Opvallende aspek van die @-gebruikerskonvensie is die feit dat terwyl die meeste gebruikers minder as 178,8 (ŉ gemiddeld van minder as 0,45[246] of ŉ genormaliseerde waarde van minder as 30,79)[247] verwysings na gebruikers per korpus gebruik , blyk dit dat O3, O4, O8 en O10 se gebruik van die @-gebruikerskonvensie baie hoër as die waarneembare gemiddeld in die betrokke korpora is. Hierdie vier outeurs is in die persoonlike nuus- en vermaaklikheids-kategorie geplaas. Daar kan dus nie aangevoer word dat die kategorie (of die doel waarvoor outeurs Twitter gebruik) ŉ invloed het op die aantal verwysings na gebruikers nie.

Wanneer die gemiddelde aantal verwysings na gebruikers (Tabel 7.2) van die verdagte tekste vergelyk is met die gemiddelde aantal verwysings na gebruikers van die outeurs vanuit die gespesialiseerde korpus, blyk dit dat V1 (1,18) en O5 (0,66), V2 (3,55) en O1 (2,68) asook V3 (2,43) en O2 (2,28) sterk ooreenkomste toon ten opsigte van die genormaliseerde aantal verwysings na gebruikers per korpora.

Na gelang van Hong *et al.* (2011:519) se studie is die persentasie twiets waarin daar verwysings na gebruikers voorkom ook bereken:

$$Persentasie\ twiets\ in\ Afrikaans\ met\ verwysings\ na\ gebruikers = \frac{1\,028}{4\,588} \times 100$$

[246] Die verdagte tekste (sien Tabel 7.2) se gemiddelde hoeveelheid verwysings na gebruikers stem ook met die bevinding ooreen.
[247] Sien Tabel 7.2.

$$\therefore\ Persentasie\ twiets\ in\ Afrikaans\ met\ verwysings\ na\ gebruikers = 22{,}4\%$$

22,4% van die twiets wat in die studie gebruik is, bevat verwysings na gebruikers. Wanneer hierdie bevinding met Hong *et al.* (2011:519) se bevindinge vergelyk word, is dit duidelik dat Twitter-gebruikers aansienlik minder verwysings na gebruikers gebruik as gebruikers van die top tien tale soos voorgehou in Hong *et al.* (2011) se studie. Die persentasie twiets wat in die Afrikaanse korpus voorkom (22,4%), toon die grootste ooreenkoms met Duits (36%).

Vanuit hierdie ondersoek na die gebruik van die @-gebruikerskonvensie kan daar aangevoer word dat die aantal verwysings na gebruikers per twiets wel as ŉ aanduiding van outeurskap beskou kan word. Ten einde die analise te kon voltooi, is daar vervolgens ondersoek ingestel na die manier waarop die @-gebruikerskonvensie deur elk van die outeurs gebruik word.

7.2.1.2.1 Gespesialiseerde korpus: verwysings na gebruikers

Ten einde kenmerkende gebruike van die verwysings na gebruikers te kon voorhou soos geïdentifiseer vanuit die gespesialiseerde korpus, word voorbeelde vanuit elk van die korpora verskaf voordat die voorbeelde bespreek word:

Voorbeeld 29: **O1** – Damn, **GEBRUIKER** Waar het jy my rooster gekry, stalker much? **@GEBRUIKER**

Voorbeeld 30: **O2** – **"@GEBRUIKER**: Prof. GEBRUIKER namens #NWU-raad:ontgroening gebeur weer, want nuwe studente kom elke jaar, hulle is jonk, hormoongedrewe". Oi!

Voorbeeld 31: **O3** – **.@GEBRUIKER** "Die ANC werf nie stemme nie, hulle steel dit," Malema

Voorbeeld 32: **O4** – Mag ons vandag die Aussies knak soos 'n glow stick **@GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER**

Voorbeeld 33: **O5** – Want **@GEBRUIKER** maak alles beter.

Voorbeeld 34: **O5** – Mag ek net sê dat ek stukkend is oor vanaand Die Boekklub se laaste episode was? **@GEBRUIKER**

Voorbeeld 35: **O6** – Ek deel 'n verjaardag met **@GEBRUIKER** Boogskutters is oulike mense! ❤

Voorbeeld 36: **O7** – Laaste aand van vier vol shows by die **@GEBRUIKER** in die kaap. Môre vat ons hom bu **@GEBRUIKER** op kknk

Voorbeeld 37: **O8** – :) x **@GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER**

Voorbeeld 38: **O8** – Vanaand op my tafel. Baie dankie **@GEBRUIKER** #oliveoil **@GEBRUIKER**

Voorbeeld 39: **O9** – As 'n hoër krag my net op 'n Sondagmiddag #7delaanomnibus tyd by **@GEBRUIKER** kan uitbring, sal dit bedoos wees. Dankie by voorbaat.

Voorbeeld 40: **O9** – YAAASSSSSSSSSSSSSSSSSSS FOK FOK FOK. **@GEBRUIKER**

Voorbeeld 41: **O10** – TELLING: UJ is tans 120/3 na 15 boulbeurte op die telbord. @**GEBRUIKER** @**GEBRUIKER** #VarsityCricket

Voorbeeld 42: **O10** – Die @**GEBRUIKER** is vierde op die telbord teenoor UWC wat sesde lê. Verrigtinge skop om 18:00 af. @**GEBRUIKER** saai dit ook uit. @**GEBRUIKER**

Voorbeeld 29 tot 42 dui die gebruik van die @-gebruikerskonvensie van O1 tot O10 aan. Vanuit hierdie voorbeelde is dit duidelik dat daar geen vaste reël bestaan wat gebruikers aanraai hoe om die verwysings na gebruikers op ʼn bepaalde manier te gebruik nie. Daar kan dus ook vanuit die gegewe voorbeelde gesien word dat die verwysings na gebruikers aan die begin van ʼn twiet geplaas word, in ʼn twiet self gebruik word of aan die einde van ʼn twiet geplaas word. Die verwysing na gebruikers word − soos deur Honeycutt en Hering (2009) voorgehou − as ʼn aanspreekvorm en ʼn verwysing na ʼn spesifieke plek gebruik.

In die korpus van O1 (sien voorbeeld 29) verwys die outeur eers na ʼn spesifieke individu en O1 gebruik eers aan die einde van die twiet die @-gebruikerskonvensie. Dus word die spesifieke individu wat aangespreek word se naam (eienaam) in die teks gebruik en die gebruikersnaam word bygevoeg om seker te maak dat die individu wat in die twiet aangespreek word wel die boodskap sien. O1 is die enigste outeur in die gespesialiseerde korpus wat die verwysings na gebruikers só gebruik. Die manier waarop O1 beide die individu se eienaam en gebruikersnaam in die korpus gebruik, stem baie ooreen met informele gesprekvoering: die individu word eers aangespreek en dan word die spesifieke Twitter-element bygevoeg.

O2 en O3 plaas meestal[248] die verwysing na ʼn gebruiker aan die begin van ʼn twiet. O3 plaas eers ʼn punt en dan die verwysing na die gebruiker (sien voorbeeld 31). O2 het die verwysing 16 keer aan die begin van ʼn twiet geplaas (in geen van hierdie 16 verwysings plaas O2 ʼn punt voor die @-gebruikerskonvensie nie), drie maal aan die einde en agt maal in die middel van ʼn twiet. Die verwysings na gebruikers, soos deur O2 gebruik, verrig verskillende funksies. Wanneer die verwysing na die gebruiker deur O2 in die middel van die twiet geplaas word, spreek die outeur die gebruiker aan (dit dien as ʼn eienaam) of die outeur verwys na ʼn aanhaling van die gebruiker waarna verwys word (aanduiding van outeurskap). Wanneer O2 die verwysing na die gebruiker aan die einde van ʼn twiet plaas, wil O2 ʼn spesifieke persoon (of maatskappy) van ʼn nuuswaardige gebeurtenis inlig.

O3 maak gebruik van ʼn punt voor die @-gebruikerskonvensie wanneer die verwysing na die gebruiker aan die begin van ʼn twiet geplaas word. In die korpus van O3 is daar nie ʼn enkele verwysing na ʼn gebruiker aan die begin van ʼn twiet wat nie deur ʼn punt voorafgegaan is nie. Daar kan dus aangevoer word dat (veral wanneer daar ʼn beperkte aantal verdagte outeurs is en

[248] Beide O2 en O3 gebruik ook die @-gebruikerskonvensie in die middel van ʼn twiet, maar dan word die @-gebruikerskonvensie as ʼn aanspreekvorm, soos in gesprekvoering, gebruik, en word die twiet se leesbarheid nie beïnvloed nie.

O3 een van die verdagte persone is) die gebruik van ŉ punt voor die @-gebruikerskonvensie aan die begin van ŉ sin ŉ aanduiding kan wees dat O3 die outeur[249] van die teks is.

Die @-gebruikerskonvensie wat aan die begin van die twiet aangetref word, kan nie as ŉ manier beskou word waarop O3 met ŉ spesifieke gebruiker in gesprek wil tree nie, aangesien O3 hierdie @-gebruikerskonvensie gebruik om na instansies eerder as individue te verwys. O3 gebruik Twitter hoofsaaklik as ŉ joernalis en die @-gebruikerskonvensie word in die outeur se geval gebruik as ŉ aanduiding van eienaarskap – vir wie ŉ spesifieke storie geskryf is (sien voorbeeld 31). Indien die twiet steeds ŉ aanhaling is, sal die individu wat aangehaal word aan die einde van die twiet genoem word. Wanneer die verwysing na die gebruiker in die middel van die twiet geplaas word, doen die verwysing nie afbreek aan die leesbaarheid van die twiet nie. Die outeurs (O1, O3, O4, O5, O6, O7, O8, O9 en O10) gebruik die @-gebruikerskonvensie as ŉ aanspreekvorm (ŉ eienaam) en dus lees die twiet soos ŉ gesprek waarby die gebruiker waarna daar verwys word, ingesluit word.

Voorbeeld 37 wys hoe O8 die @-gebruikerskonvensie gebruik om na etlike (in die geval van die voorbeeld, vier verskillende gebruikers) gebruikers te verwys. In hierdie voorbeeld blyk dit dat die outeur slegs ŉ emosie (geluk) wil oordra en met die gebruikers waarna daar verwys is, wil deel. Die @-gebruikerskonvensie word dus gebruik om ŉ direkte boodskap aan ŉ individu oor te dra en nie net as ŉ aanspreekvorm nie.

Wanneer die manier ondersoek is waarop die @-gebruikerskonvensie deur die verskillende outeurs gebruik is in terme van die kategorieë (soos in afdeling 5.4 bespreek), is die gevolgtrekking gemaak dat die rede waarom die verskillende outeurs Twitter gebruik ŉ invloed het op die manier waarop die @-gebruikerskonvensie gebruik word. Die outeurs wat in die nuuskategorie geval het (O2, O3, O8 en O10), verwys deurlopend na gebruikers/profiele wat by "nuuswaardige" twiets betrokke is of wat deel van die gebeurtenis is waaroor daar getwiet word.

Outeurs wat in die vermaaklikheidskategorie (O7, O4 en O6) geval het, verwys deurlopend na ander kunstenaars, teaters en bedank spesifieke organisasies en individue en maak dus meer gebruik van die @-liggingfunksie (soos deur Honeycutt & Henning [2009] voorgehou). Die persoonlike kategorie (O1, O2, O5, O9 en O10) verwys na ander Twitter-gebruikers en die twiets word dikwels aan ŉ spesifieke individu gerig. Daar kan dan ook in Afrikaans − soos dit op Twitter gebruik word − gesien word dat die @-gebruikerskonvensie as ŉ aanspreekvorm beskou kan word (soos bespreek deur Honeycutt & Herring, 2009).

Vanuit afdeling 7.2.1.2.1 kan daar dus voorgehou word dat die @-gebruikerskonvensie soos deur die betrokke outeurs gebruik nie noodwendig as uniek beskou kan word nie. Die punt wat

[249] Die gebruik van ŉ punt voor die verwysing na ŉ gebruiker aan die begin van ŉ sin kon nie in enige ander korpus (verwysingskorpus of die gespesialiseerde korpus) gevind word nie, en ŉ soektog op Twitter het ook nie enige ooreenstemmende resultate gelewer nie.

O3 herhaaldelik voor die verwysing na die gebruiker plaas (aan die begin van ŉ twiet) kan as die enigste voorkoms van ŉ kenmerkende gebruik van die @-gebruikerskonvensie beskou word. Daar kan wel verder ondersoek ingestel word na gebruikers waarna daar frekwent verwys word, indien ŉ verdagte individu ŉ misdryf begaan[250]. Dit is ook duidelik dat die kategorie waarin ŉ bepaalde outeur geplaas word, ŉ invloed het op die manier waarop die verwysing na gebruikers gebruik word in twiets. Dit is verder duidelik dat die outeurs van Afrikaanse twiets ook die funksie van verwysings na gebruikers gebruik soos voorgehou deur Honeycutt en Herring (2009). Die bevindinge vanuit hierdie afdeling (afdeling 7.2.1.2.1) het as basis gedien vir die ondersoek na die gebruik van die verwysings na gebruikers in die verdagte tekste.

7.2.1.2.2 Verdagte tekste: verwysings na gebruikers

In afdeling 7.2.1.2.1 is daar ondersoek ingestel na die manier waarop die @-gebruikerkonvensie in Afrikaanse twiets gebruik word ten einde die gebruik van die konvensie in die verdagte tekste met die gespesialiseerde korpus te kon vergelyk. Voorbeelde (43-48) van die verwysings na gebruikers in die verdagte tekste word eerstens voorgehou:

Voorbeeld 43: **V1** – Kyk gerus na my artikel vandag in Beeld en op Netwerk24: SAUK nie so onredelik oor beeldmateriaal http://www.netwerk24.com/Stemme/Aktueel/sauk-nie-so-onredelik-oor-beeldmateriaal-20160602 … via @**GEBRUIKER**

Voorbeeld 44: **V2** – Geluk my mooiste. Jy verdien elke bietjie geluk wat na jou kant toe kom. Jy't 'n spesiale plek in my hart. @**GEBRUIKER**

Voorbeeld 45: **V2** – Ek voel hierdie is op jou van toepassing. @**GEBRUIKER**

Voorbeeld 46: **V2** – Ek's verlief op @**GEBRUIKER** se stem. Dankie @**GEBRUIKER** dat jy my introduce het aan sy voice notes.

Voorbeeld 47: **V2** – Ek en my een vriendin hap mekaar die hele dag. @**GEBRUIEKR**: "Is julle ooit nice met mekaar." Vriendin: "Net op Woensdae."

Voorbeeld 48: **V3** – Okay so @**GEBRUIKER** het my klaar oortuig Dat goeie mense nog bestaan That bio tho 😱👌👏

Wanneer voorbeeld 43 tot 48 met die voorbeelde vanuit die gespesialiseerde korpus vergelyk is (voorbeeld 29 tot 42), is daar sekere ooreenkomste geïdentifiseer in die gebruik van die @-gebruikerskonvensie. Dit blyk dat voorbeeld 43 tot 48 nie net ooreenstem met dít wat as die norm vir die verwysing na gebruikers in Afrikaans soos dit op Twitter gebruik word, identifiseer nie, maar dat daar ook ooreenskomste tussen die verdagte tekste en van die outeurs vanuit die gespesialiseerde korpora waargeneem is.

[250] Vir die doel van hierdie studie is die gebruikers waarna die outeurs verwys nie voorgehou nie. Dit is gedoen om die identiteit van die deelnemers te beskerm.

Die manier waarop V2 die @-gebruikerskonvensie in die middel van ŉ twiet plaas sonder dat dit die leesbaarheid/verstaanbaarheid van die boodskap steur sowel as die wisseling in die plasing van verwysings na gebruikers (hetsy in die middel van ŉ twiet of aan die einde van ŉ twiet) stem ooreen met die manier waarop O1 (sien voorbeeld 29) verwysings na gebruikers gebruik – sien voorbeeld 46 en 47. In hierdie voorbeelde (29, 46 en 47) word die @-gebruikerskonvensie as ŉ direkte aanspreekvorm (in die plek van ŉ eienaam) gebruik.

V3 plaas die @-gebruikerskonvensie (sien voorbeeld 48) in die middel van ŉ twiet en gebruik dit as ŉ aanspreekvorm − ŉ kenmerk wat in die korpora van O1, O3, O4, O5, O6, O7, O8, O9 en O10 gevind word. Tog kan O2, O3, O4, O6, O7, O8 en O10 nie as die outeurs van die derde verdagte teks beskou word nie aangesien die @-gebruikerskonvensie slegs gebruik word om ŉ direkte boodskap aan ŉ spesifieke gebruiker oor te dra (ŉ aanspreekvorm in die plek van ŉ eienaam) en daar is geen ander funksie soos die @-ligging gebruik nie en ook nie die manier waarop daar na ŉ gebruiker verwys is om erkenning aan die individu se opinie te gee nie. Die gebruik van die verwysing na gebruikers is nie in die korpus van V3 sigbaar nie. O1, O5 en O9 kan na gelang van die funksie wat die @-gebruikerskonvensie in die korpus vervul as outeurs van V3 beskou word.

Soos in die geval van V3, verwys V1 slegs een keer na ŉ gebruiker in die teks wat vir die ondersoek gebruik is. In die voorbeeld (sien voorbeeld 43) word die @-gebruikerskonvensie aan die einde van die twiet geplaas en ŉ *via* word voor die verwysing gebruik. Hierdie plasing van *via* dui daarop dat die inligting waarna daar in die twiet verwys word op die profiel van die gebruiker gevind kan word. Hierdie eienskap is egter nie uniek tot V1 nie, maar kan as ŉ kenmerk beskou word wat veral in twiets van individue wat nuusgebeure deel, gevind word. Vervolgens kan die aanname vanuit hierdie @-gebruikerskonvensie en die inhoud van V1 se twiets gemaak word dat daar ŉ groot moontlikheid bestaan dat die outeur van V1 een van die outeurs is wat Twitter ter wille van nuus (die nuuskategorie – O2, O3 en O8) gebruik.

Alhoewel daar baie min van die @-gebruikerskonvensie in die verdagte tekste gebruik gemaak word – hierdie is ŉ tendens wat as een van die moontlike beperkings in so ŉ klein korpus beskou kan word – blyk dit dat daar steeds genoegsaam patroonmatighede waargeneem kon word om moontlike outeurs van die verdagte tekste te kan identifiseer. V3 is na gelang van die analise van die genormaliseerde aantal verwysings na gebruikers moontlik deur O1, O5 of O9 geskryf. V2 en O1 stem tot so ŉ mate ooreen dat O1 die moontlike outeur van die teks kan wees, en na gelang van die analise na die manier waarop die @-gebruikerskonvensie gebruik word deur V1, is O2, O3 en O8 moontlike outeurs van die korpus.

7.2.1.3 Hiperskakels

Die gebruik van hiperskakels is die laaste kenmerk wat in die stilistiese analise van Twitter-spesifieke tekselemente ondersoek is. Yi *et* al. (2016:1808) voer aan dat hiperskakels veral in twiets gebruik word om ʼn manier te vind waarop boodskappe betekenisvol oorgedra kan word ten spyte van die karakterbeperking op hierdie platform (Twitter). Hiperskakels verwys na ʼn internetskakel wat gebruikers na ʼn ander webblad neem wanneer daar op die skakel gekliek word en hiperskakels vul dus die boodskap in twiets aan (Yi *et al.*, 2016:1808). Die kommunikatiewe doelwit van Twitter-gebruikers sluit daaglikse gesprekvoering, die deel van inligting, hiperskakels en nuuswaardige gebeure in (vgl. Dunlap & Lowenthal, 2009; Java *et al.*, 2007; Starbird *et al.*, 2010).

Soos in afdeling 7.2.1.1 en 7.2.1.2 is daar eerstens ondersoek ingestel na die frekwensie van die gebruik van hiperskakels deur die onderskeie outeurs (ʼn interne vergelyking van die gespesialiseerde korpus, die onderskeie outeurs in die gespesialiseerde korpus met mekaar, en ook die verdagte tekste). Sien Tabel 7.3:

Tabel 7.3 Die gebruik van hiperskakels in die onderskeie datastelle

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van hiperskakels in die korpus	Hiperskakels per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hiperskakels per twiet
O1	669	944	8 942	13	1,45	0,02
O2	589	1 577	12 259	37	3,02	0,06
O3	206	266	3 379	7	2,07	0,03
O4	484	978	7 110	34	4,78	0,07
O5	333	556	4 579	0	0,00	0,00
O6	768	1 657	16 340	628	38,43	0,82
O7	457	790	5 655	16	2,83	0,04
O8	308	443	3 690	86	23,31	0,28
O9	275	467	4 918	3	0,61	0,01
O10	499	1 190	7 744	6	0,77	0,01
V1	42	90	844	1	1,18	0,02
V2	96	147	1409	1	0,71	0,01
V3	32	44	412	0	0,00	0,00

Tabel 7.3 dui die totale aantal keer aan wat hiperskakels in die onderskeie datastelle gebruik is, die maksimum hiperskakels per twiet, die gemiddelde aantal hiperskakels per datastel sowel as die genormaliseerde waarde (per 1 000 woorde[251]) van die hiperskakels per datastel. Vanuit Tabel 7.3 kan daar etlike gevolgtrekkings oor die gebruik van hiperskakels gemaak word.

[251] Net soos in die geval van die hutswoorde is die hiperskakels se waarde na 1 000 woorde genormaliseer aangesien hiperskakels (ten spyte daarvan dat hiperskakels uit meer as een woord bestaan) ook as ʼn entiteit beskou kon word.

Dit is duidelik vanuit Tabel 7.3 dat elk van die onderskeie outeurs hiperskakels teen ŉ unieke frekwensie gebruik[252]. O6 maak die meeste van hiperskakels gebruik (628 hiperskakels in ŉ korpus wat uit 768 twiets bestaan – basies ŉ hiperskakel per twiet: 0,82 hiperskakels per twiet). Daar kan nie aangevoer word dat die gebruik van hiperskakels ŉ definitiewe bewys is dat O6 die outeur van ŉ teks is nie, maar wanneer die hiperskakels so frekwent (ŉ hiperskakel per teks) gebruik word, moet O6 as ŉ verdagte persoon (of ŉ moontlike outeur) in die identifisering van mikrobloginskrywings oorweeg word.

In teenstelling met O6 maal O5 van geen hiperskakels gebruik nie. Aangesien elk van die ander outeurs in die gespesialiseerde korpus wel hiperskakels in hulle twiets gebruik, kan die afwesigheid van hiperskakels − net soos ŉ hoë frekwensie van hiperskakels (O6) − as ŉ aanduiding beskou word dat die individu as ŉ moontlike outeur beskou kan word.

Die oorblywende agt outeurs het deurlopend van hiperskakels gebruik gemaak. Dit is duidelik vanuit die interne vergelyking tussen die korpora[253] dat daar sterk ooreenkomste in die gebruik van hiperskakels waargeneem is – hierdie aspek word verder deur die genormaliseerde gemiddeld van die gebruik van hiperskakels beklemtoon.

Wanneer die verdagte tekste se gebruik van hiperskakels met die gespesialiseerde korpus vergelyk is (sien Tabel 7.3) stem V1 (1,18) met O1 (1,45) ooreen, V2 (0,71) en O9 (0,61) stem ooreen en V3 (0,00) en O5 (0,00) se genormaliseerde waarde per 1 000 woorde vir die gebruik van hiperskakels is identies.

Na gelang van Hong *et al.* (2011:519) se navorsing is die totale persentasie twiets bereken wat hiperskakels bevat:

$$Persentasie\ twiets\ in\ Afrikaans\ met\ verwysings\ na\ gebruikers = \frac{830}{4588} \times 100$$

$$\therefore\ Persentasie\ twiets\ in\ Afrikaans\ met\ verwysings\ na\ gebruikers = 18\%$$

In Hong *et al.* (2011:519) se ondersoek is daar bevind dat Duits die hoogste gebruik van hutswoorde (39%) toon, gevolg deur Frans (37%) en Engels (25%). Die gebruik van hiperskakels in Afrikaans is dus effens laer as die drie tale waarin die meeste hutswoorde voorkom, maar tog plaas dit die gebruik van hutswoorde in Afrikaans in konteks met die gebruik van hiperskakels in ander tale. Die 18% Afrikaanse twiets waarin daar wel hiperskakels gebruik word, stem ooreen met Nederlands, Koreaans en Maleis (17% – hierdie is dus die tale waarin die vierde meeste hiperskakels voorkom).

[252] Sien Bylaag E vir die interne vergelykings van die gebruik van hiperskakels.
[253] Sien Bylaag E vir die interne vergelykings van die gebruik van hiperskakels.

Die frekwensie waarteen die onderskeie outeurs van hiperskakels gebruik gemaak, kan vervolgens as ʼn aanduiding van outeurskap beskou word. Ten einde die stilistiese ondersoek na hiperskakels te kon voltooi, verwys ek vervolgens na voorbeelde van die hiperskakels soos deur die outeurs gebruik.

7.2.1.3.1 Gespesialiseerde korpus: hiperskakels

Daar is reeds verwys na die manier waarop hiperskakels deur Twitter-gebruikers gebruik word om die kommunikatiewe waarde van (kort) boodskappe uit te brei en om meer inligting oor te dra. Voorbeeld 49 tot 60 is voorbeelde van hoe die outeurs in die gespesialiseerde korpus hiperskakels in hulle twiets gebruik:

Voorbeeld 49: **O1** – My lewe in 'n prentjie-storie. **https://instagram.com/p/BEQZyePJkIZ/**

Voorbeeld 50: **O1** – Ek as kind, probably. **https://vine.co/v/in3t2zuYXxP**

Voorbeeld 51: **O2** – "@GEBRUIKER: Essex bus collides with lorry carrying steam engine **http://trib.al/amEYPXP** ". Gewonder wat is nuus? Hier's 'n voorbeeld.

Voorbeeld 52: **O3** – Hierdie stuk deur @GEBRUIKER is baie interessant. Ek sien uit daarna om te sien hoe die party gaan reageer **http://www.bdlive.co.za/opinion/2016/01/21/ maimane -race-plamust-be-incorporated-in-da …**

Voorbeeld 53: **O4** – Ek en @GEBRUIKER geniet n stukkie banjokapperjol oppad Potties toe! **https://instagram.com/p/51otDvPcpd/**

Voorbeeld 54: **O6** – Ek en GEBRUIKER probeer 'n 'selfie'. Ons weeklikse kuier-meeting aand. "Maak werk jou plesier, dan sal jy... **http://fb.me/2LJo51o6c**

Voorbeeld 55: **O7** – John Deere het laat weet hulle skenk R1 vir elke share vd video to R50000 share die rook uit hom uit **https://www.facebook.com/permalink.php?story_fbid= 10153802718262936&id=76283842935 …**

Voorbeeld 56: **O7** – Ek het vir jou n website. **http://www.caringdaisies.co.za** dankie hulle ontvang die skenkings

Voorbeeld 57: **O8** – Paarlberg #sonsondergang #sundown #mountains @GEBRUIKER **https://www.instagram.com/p/BDqgjw3GQkk/**

Voorbeeld 58: **O8** – #Dans in #OggendOpRSG The Kinks - All Day and All of the Night **https://youtu.be/mMWNwHof0kc** via @YouTube @GEBRUIKER

Voorbeeld 59: **O9** – Lekker vibes. :0 @GEBRUIKER **https://youtu.be/oD4I7XdTEM** 9Thrown like a star in my vast sleep, I opened my eyes to take a peek.

Voorbeeld 60: **O10** – NWU-Pukke veg tot in halfeindstryd #VarsityCricket #ShowYourColours #GooiSomeGees @GEBRUIKER @GEBRUIKER **https://m.facebook.com/story. php?story_fbid=95292691506294&substory_index=0&id=128062830595955 …**

Hiperskakels word soos vanuit voorbeeld 49 tot 60 gesien − met die uitsondering van O2, O7 en O8 − altyd aan die einde van ŉ twiet geplaas. O2, O7 en O8 plaas hiperskakels soms in die middel van ŉ twiet (sien voorbeelde 51, 56 en 59). Wanneer hiperskakels deur hierdie outeurs in die middel van ŉ twiet gebruik word, doen die hiperskakel nie afbreuk aan die leesbaarheid van die twiet nie.

O7 verwys lesers van die twiet na ŉ spesifieke webblad en ná die plasing van ŉ hiperskakel sluit die outeur die twiet af. O2 plaas hiperskakels na ŉ nuusberig en verskaf ná ŉ hiperskakel ŉ opinie oor die spesifieke berig. O8 maak meestal van hiperskakels gebruik wat na YouTube verwys, gevolg deur ŉ verwysing na YouTube en/of ŉ spesifieke gebruiker.

Wanneer die soort hiperskakels wat deur die outeurs gebruik word, ondersoek is, kan die gevolgtrekking gemaak word dat die outeurs na ander sosialemediaplatforms (Facebook, Instagram, Vine en in sommige gevalle – O1 – Twitter self) en nuuswebwerwe verwys.

O1 verwys na al die bogenoemde sosialemediaplatforms. Daar word meestal na Instagram (ses skakels) en Vine (ses skakels) verwys terwyl O2 slegs na webwerwe verwys waar nuus gedeel word (Media24 ingesluit). O2 verwys ook na mediamense, ŉ hiperskakel wat deur geen van die ander outeurs gebruik word nie. O9 en O10 maak gebruik van hiperskakels wat na sosialemediaplatforms verwys, en geen bepaalde patroonmatigheid (of ŉ voorkeur aan die sosialemediaplatform waarna verwys word) is waargeneem nie.

O3 gebruik hiperskakels om lesers na webwerwe te neem wat handel oor nuuswaardige gebeurtenisse waaroor ŉ spesifieke twiet gehandel het. Daar was ŉ enkele hiperskakel wat lesers van ŉ twiet na ŉ webblad neem ter viering van "wêreld Arabiese-taaldag". Daarom kan die aanname nie gemaak word dat die outeur slegs van hiperskakels gebruik maak om lesers na "harde nuusstories" te lei nie.

Die hiperskakels wat deur O4 gebruik word, is hiperskakels na YouTube of na webblaaie waar bemarking vir optredes en dies meer gedoen word. O7 gebruik hiperskakels soortgelyk aan O4 alhoewel gebruikers ook hiperskakels hertwiet soos deur ander Twitter-gebruikers geplaas. Hierdie gebruik van hiperskakels kan as kenmerkend van die vermaaklikheidskategorie beskou word alhoewel die webblaaie waarna die outeurs in die twiets verwys wel inligting oor ŉ spesifieke outeur kan voorhou. O6 is ook in die vermaaklikheidskategorie geplaas, maar soos reeds genoem, gebruik O6 aansienlik meer hiperskakels as enige van die ander outeurs.

Die soort hiperskakels wat O6 gebruik, is ondersoek en dit is duidelik dat hierdie outeur ŉ unieke manier het waarop hiperskakels gebruik word. Voorbeeld 54 wys hoe O6 hiperskakels gebruik. O6 gebruik oorwegend hiperskakels na Facebook. Die outeur plaas ŉ twiet, maar telkens is dit slegs ŉ segment van die volle storie − soos dit op Facebook geplaas is.

Hiperskakels word dus deur ’n ellips voorgegaan. Vanuit die 628 hiperskakels wat deur die outeur gebruik is, is slegs een hiperskakel nie ’n skakel na Facebook nie (sien voorbeeld 61):

Voorbeeld 61: **O6** – Vanaand doen ek 'n uur lange onderhoud tussen 21:00 - 22:00 op VCR. Luister 'online' by **http://www.906fmstereo.com ... http://fb.me/3AEgAEdCt**

Vanuit voorbeeld 61 is dit duidelik dat selfs wanneer O6 van ’n ander hiperskakel gebruik maak as dié na Facebook, is die hiperskakel deur nóg ’n hiperskakel na Facebook gevolg. Wanneer die korpus ondersoek is waar die outeur slegs van hiperskakels na Facebook gebruik maak, kan daar aangeneem word dat O6 moontlik die outeur van die tekste (twiets) kan wees.

O8 plaas hiperskakels op geen bepaalde plek in twiets nie, maar 33 van die 86 hiperskakels (38,4%) is na YouTube, 47 na Instagram (54,7%) terwyl slegs 6 (oftewel 6,9%) na ander webwerwe verwys. Alhoewel verwysings na Instagram en YouTube nie so kenmerkend is soos O6 se verwysing na Facebook nie, kan die gebruik van hiperskakels wel as ’n aanduiding van outeurskap beskou word veral wanneer die frekwensie van die gebruik van hiperskakels geanaliseer is.

Die analise van die gebruik van hiperskakels het dus aangetoon dat daar geen vaste plek is waar hiperskakels in twiets gebruik word nie, maar dat gebruikers wat in Afrikaans twiet wel van hiperskakels gebruik maak. Vanuit die analise van O6 se gebruik van hiperskakels, kan daar voorgehou word dat dit wel moontlik is om outeurs te identifiseer na gelang van die gebruik van hiperskakels na Facebook.

Ook O8 se gebruik van hiperskakels kan as kenmerkende beskou word. Wanneer die frekwensie van die gebruik van hiperskakels saam met die soort hiperskakel ondersoek is, kan die teksformaat van twiets as ’n aanduiding van outeurskap beskou word mits dit saam met ander aspekte ondersoek word. Ten einde te kon bepaal of die ondersoek na die gebruik van hiperskakels wel insiggewend is ten einde die outeurs van mikrobloginskrywings te identifiseer, is die gebruik van hiperskakels in die verdagte tekste ondersoek.

7.2.1.3.2 Verdagte tekste: hiperskakels

Soos in afdeling 7.2.1.3 voorgehou, verwys hiperskakels na internetskakels wat gebruikers na ’n ander webblad neem. Aangesien die kommunikatiewe doelwit van Twitter-gebruikers daaglikse gesprekvoering en die deel van inligting insluit, is hiperskakels ’n nuttige hulpmiddel om hierdie nuuswaardige gebeure te deel (vgl. Dunlap & Lowenthal, 2009; Java *et al.*, 2007; Starbird *et al.*, 2010). In afdeling 7.2.1.3.1 is daar bepaal dat die soort hiperskakel sowel as die posisie van hiperskakels wel ’n aanduiding van outeurskap beskou kan word. In hierdie afdeling is voorbeelde van die verdagte tekste gebruik waarin hiperskakels voorkom en ooreenkomste met die gespesialiseerde korpus word bespreek.

Voorbeeld 62: **V1** – Kyk gerus na my artikel vandag in Beeld en op Netwerk24: SAUK nie so onredelik oor beeldmateriaal **http://www.netwerk24.com** via @GEBRUIKER

Voorbeeld 63: **V2** – Ek mag dalk 'n slegte mens wees, maar tenminste het ek nie 'n Instagram account vir my hond nie. **https://instagram.com/p/BJk056cBzn3/**

Die soort hiperskakels wat deur V1 gebruik word (sien voorbeeld 62) toon ŉ groot ooreenkoms met die soort hiperskakels wat in die korpora van O2 en O3 gevind word. Hierdie outeurs – O2 en O3 – is in die nuuskategorie Twitter geplaas, en dus word die bewyse al hoe sterker dat V1 een van die outeurs vanuit die nuuskategorie moet wees (sien ook voorbeeld 51 en 52).

In die geval van V2 stem die soort hiperskakels wat in die verdagte teks voorkom met die hiperskakels wat deur O1 gebruik word (sien voorbeeld 49, 50 en 62). Beide die outeurs verwys na ander sosialemediaplatforms ten einde die boodskap wat oorgedra word te versterk. Alhoewel O8 ook na sosialemediaplatforms verwys (sien voorbeeld 57 en 58) – is 38,4% van die hiperskakels na YouTube en 54,7% na Instagram – die ooreenkoms tussen O8 en V2 is dus nie so beduidend nie.

Die verskil tussen O8 en O1 se gebruik van hiperskakels na Instagram is die meegaande boodskap wat oorgedra word aan lesers. O1 verwys na "ek" soos in voorbeeld 49 (**My** lewe in 'n prentjie-storie**.** https://instagram.com/p/BEQZyePJkIZ/) en die hiperskakels brei uit oor die outeur self. ŉ Verdere kenmerk wat in beide die korpus van O1 en V2 gevind word, is die gebruik van ŉ punt aan die einde van ŉ twiet, maar voor hiperskakels. O8 daarenteen gebruik hiperskakels na Instagram om soos in voorbeeld 57 meer inligting aan lesers oor die onderwerp van die twiet te verskaf (**Paarlberg** #sonsondergang #sundown #mountains @GEBRUIKER https://www.instagram.com/p/BDqgjw3GQkk/). Daarom kan daar aangevoer word dat O1 ŉ sterker kandidaat vir die outeurskap van V2 is as O8.

Geen hiperskakels is in die korpus van V3 gevind nie, en ook nie in die korpus van O5 nie. Aangesien hierdie twee korpora (V3 en O5) die enigste korpora in die studie is waar geen hiperskakels voorkom nie, kan die afwesigheid van dié kenmerk juis as ŉ aanduiding van outeurskap beskou word. Vanuit die ondersoek na die gebruik van hiperskakels blyk dit dat daar sonder twyfel wel patroonmatighede geïdentifiseer kan word wat aanleiding kan gee tot die identifikasie van outeurs van mikrobloginskrywings.

Die ondersoek na die gebruik van hutstekens, hutswoorde, verwysings na gebruikers en hiperskakels kan wel as ŉ bydraende faktor in die identifisering van outeurs van mikrobloginskrywing beskou word. Daar is enkele kenmerke geïdentifiseer wat as kenmerke van Afrikaans soos dit op Twitter gebruik word, beskou kan word. Vervolgens is ondersoek

ingestel na die gebruik van leestekens ten einde die ondersoek na die algemene kenmerke van Twitter te voltooi.

7.2.2 Leestekengebruik

In die ondersoek na leestekengebruik is daar spesifiek na die gebruik van leestekens aan die einde van sinne, die manier waarop (indien wel) ’n ellips gebruik is sowel as die gebruik van meer as een leesteken ondersoek ingestel. Soos in afdeling 7.2.1 word daar eerstens na die frekwensie van die gebruik van die verskillende aspekte verwys ten opsigte van die gebruik van leestekens alvorens kenmerke in die styl van die gespesialiseerde korpus en verdagte tekste voorgehou word. Daar is reeds verwys na die maniere waarop die sosiale aard van platforms aanleiding daartoe gee dat taalkonvensies nie noodwendig gevolg word nie. Die vraag is dus gestel of Twitter-gebruikers wel leestekens aan die einde van sinne gebruik.

Volgens Gleason (2016:44) poog die meerderheid Twitter-gebruikers (Engels) om goed gestruktureerde en koherente twiets wat nie net die basiese taalbeginsels volg nie, maar ook oor die nodige leestekens beskik, te plaas. In teenstelling met Gleason (2016) voer Al-Salman (2017:178) aan dat sosialemediagebruikers nie die taalreëls oor die gebruik van hoofletters[254] of leestekens volg nie. In hierdie afdeling is daar spesifiek na die gebruik van leestekens in sinne ondersoek ingestel ten einde te kan bepaal of outeurs wel die gepaste leestekens in die sinne waaruit ’n twiet bestaan, gebruik en of leestekens moontlik uitgelaat word ten einde die karakterbeperking te oorkom wat op Twitter voorkom. In Tabel 7.4 word die aantal keer wat leestekens wel aan die einde van sinne gebruik word, aangedui sowel as die aantal keer wat daar nie leestekens gebruik word nie:

Tabel 7.4 Die gebruik van leestekens in twiets

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Leestekens gebruik aan die einde van ’n sin	Frekwensie van gebruik van leesteken aan die einde van 'n sin	Leestekens uitgelaat aan die einde van ’n sin	Frekwensie van die uitlating van leestekens aan die einde van 'n sin
O1	669	944	8 942	917	971,4	27	28,6
O2	589	1 577	12 259	1499	950,54	78	49,46
O3	206	266	3 379	75	281,95	191	718,05
O4	484	978	7 110	759	776,07	219	223,93
O5	333	556	4 579	527	947,84	39	70,14
O6	768	1 657	16 340	1217	734,46	440	265,54
O7	457	790	5 655	443	560,76	347	439,24
O8	308	443	3 690	198	446,95	245	553,05
O9	275	467	4 918	461	987,15	6	12,85
O10	499	1 190	7 744	1021	857,98	169	142,02

[254] Sien afdeling 7.2.3 (Hoofletters).

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Leestekens gebruik aan die einde van 'n sin	Frekwensie van gebruik van leesteken aan die einde van 'n sin	Leestekens uitgelaat aan die einde van 'n sin	Frekwensie van die uitlating van leestekens aan die einde van 'n sin
V1	42	90	844	88	977,78	2	22,22
V2	96	147	1409	140	952,38	7	47,62
V3	32	44	412	36	818,18	8	181,82

Tabel 7.4 toon die aantal keer wat leestekens aan die einde van sinne gebruik word sowel as die aantal keer wat dit uitgelaat word[255]. O3 laat meer leestekens aan die einde van 'n sin uit as die aantal keer wat leestekens wel aan die einde van sinne gebruik word. Die genormaliseerde waarde tot 1 000 sinne vir die uitlaat van leestekens aan die einde van 'n sin os hoër as die gebruik van leestekens. O3 en O8 is die enigste twee outeurs wat nie deurlopend meer van leestekens aan die einde van 'n twiet gebruik maak nie. Die vergelyking tussen die outeurs vanuit die gespesialiseerde korpus het aangetoon dat elk van die outeurs (wanneer die genormaliseerde waarde tot 1 000 sinne vergelyk is) tot só 'n mate verskil dat dit wel as 'n hulpmiddel in die identifisering van outeurs van mikrobloginskrywings gebruik kan word.

Wanneer die bevindinge wat in Tabel 7.4 voorgehou word, vergelyk is met die frekwensie van die gebruik en weglating van leestekens aan die einde van 'n sin soos deur die drie verdagte outeurs gebruik, is ooreenkomste gevind tussen V1 (977,78 – leestekens gebruik; 22,22 – leestekens ontbreek) en O1 (971,4 – leestekens gebruik; 28,60 – leestekens ontbreek), V2 (952,38 – leestekens gebruik; 47,62 leestekens ontbreek) en O2 (950,54 – leestekens gebruik; 49,46 – leestekens ontbreek) en tussen V3 (818,18 – leestekens gebruik; 181,82 – leestekens ontbreek) en O10 (857,98 – leestekens gebruik; 142,02 – leestekens ontbreek). Dus kan moontlike outeurs van twiets deur middel van die ondersoek na die frekwensie van die gebruik van leestekens geïdentifiseer word.

Soortgelyk aan die gebruik van leestekens is daar verder ook ondersoek ingestel na die gebruik van meer as een leesteken. Tabel 7.5 dui die gebruik van meer as een leesteken in die gespesialiseerde korpus en die verdagte tekste aan:

Tabel 7.5 Die gebruik van meer as een leesteken in twiets

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van oortollige leestekens in die korpus	Oortollige leestekens per korpus genormaliseer tot 1 000 sinne	Gemiddelde aantal oortollige leestekens per twiet
O1	669	944	8 942	7	7,42	0,01
O2	589	1 577	12 259	1	0,63	0,00
O3	206	266	3 379	82	308,27	0,40
O4	484	978	7 110	326	333,33	0,67

[255] Sien Bylaag E vir die interne vergelyking van die gespesialiseerde korpus se gebruik van leestekens.

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van oortollige leestekens in die korpus	Oortollige leestekens per korpus genormaliseer tot 1 000 sinne	Gemiddelde aantal oortollige leestekens per twiet
O5	333	556	4 579	64	115,11	0,19
O6	768	1 657	16 340	77	46,47	0,10
O7	457	790	5 655	5	6,33	0,01
O8	308	443	3 690	7	15,80	0,02
O9	275	467	4 918	1	2,14	0,00
O10	499	1 190	7 744	35	29,41	0,07
V1	42	90	844	0	0,00	0,00
V2	96	147	1409	3	2,13	0,03
V3	32	44	412	2	4,85	0,06

Vanuit Tabel 7.5 is dit duidelik dat O1, O2, O7, O8 en O9 minder as tien keer (genormaliseer tot 15,8) meer as een leesteken gebruik. O5, O6 en O10 maak onderskeidelik 64, 77 en 35 (115,11, 46,47 en 29,41) maal meer van net een leesteken gebruik. O3 en O4 maak die meeste gebruik van meer as een leesteken (O3 -308,27 en O4 - 333,33). Die aantal keer wat elk van die onderskeie outeurs dus van meer as een leesteken gebruik maak, verskil só dat dit wel moontlik is om tussen die verskeie outeurs te kan onderskei na gelang van die frekwensie van die gebruik van oortollige leestekens[256]. Die interne vergelyking het aangedui dat elk van die outeurs se gebruik van meer as een leesteken ooreenstem.

Wanneer die gebruik van meer as een leesteken in die verdagte tekste vergelyk is met die gebruik van meer as een leesteken deur die outeurs van die gespesialiseerde korpus, is ooreenkomste waargeneem in die frekwensie waarteen die verdagte outeurs van meer as een leesteken gebruik maak. Aangesien al drie die verdagte tekste oor ʼn genormaliseerde waarde (tot 1 000 sinne) van minder as vyf beskik, blyk dit dat die verdagte tekste se outeurs, na gelang van die frekwensie, O2 (0,63) of O9 (2,14) kan wees.

Vanuit die ondersoek na die frekwensie waarteen elk van die outeurs in die studie leestekens aan die einde van sinne gebruik (Tabel 7.4 en 7.5) en ook die voorkoms van meer as een leesteken, blyk dit dat Twitter-gebruikers in Afrikaans telkens meer as een uitroep- of vraagteken aan die einde van ʼn twiet plaas. Hierdie "oormatige" gebruik van leestekens word gebruik om emosies uit te druk en sluit aan by McMenamin (2002:216-231) se studie oor hoe die manier waarop leestekens in ʼn teks gebruik word (hetsy deur af te wyk van die norm of om dit juis te volg), aangewend kan word om die outeurs van tekste te identifiseer.

Verder het die bevindinge Verhoef (2001:8) se stelling dat leestekengebruik op internetplatforms spaarsamig is, gekontrasteer aangesien die outeurs wat in hierdie studie gebruik is nie net deurlopend wel van leestekens aan die einde van sinne en/of twiets gebruik maak nie, maar ook van meer as een leesteken gebruik maak ten einde emosie oor te dra – ʼn

[256] Sien Bylaag E – interne vergelyking van die gespesialiseerde korpus se gebruik van meer as een leesteken.

aspek wat die sosiale aard van hierdie platform beklemtoon en karakterbeperking ondermyn. In die volgende onderafdelings is daar na kenmerkende eienskappe in die gebruik van leestekens in die gespesialiseerde korpus en die verdagte tekste ondersoek ingestel.

7.2.2.1 Gespesialiseerde korpus: die gebruik van leestekens

Die eerste aspek van die gebruik van leestekens wat voorgehou word, is die gebruik (al dan nie) van leestekens aan die einde van sinne. Aangesien die afwesigheid van ŉ leesteken nie die leesbaarheid van twiets beïnvloed nie, word slegs enkele voorbeelde vanuit die gespesialiseerde korpus voorgehou waar leestekens ontbreek:

Voorbeeld 64: **O3** – Maharaj sê joernaliste moet aanhou druk op die regering plaas, hulle moet nie nee vir 'n antwoord of leuns aanvaar nie

Voorbeeld 65: **O8** – Mens raak dit eintlik nooit gewoond nie 📼

Voorbeeld 66: **O9** – https://m.youtube.com/watch?v=SK6FoQ05EHk … If only you could see the facial expressions I make when I sing along to this version. Ek voel baie wit en Daspoort

Voorbeeld 67: **O4** – Ek hoop Father Christmas het die bokke se olie en water gecheck en die filters vervang!! Dis amper tyd

Voorbeeld 68: **O7** – Gaan gou so bietjie gesels op Groot fm se drive show

Wanneer daar ondersoek ingestel is na voorbeelde van twiets waar leestekens nie aan die einde van sinne gebruik word nie, blyk dit eerstens dat daar talle faktore is wat ŉ invloed op die gebruik van leestekens aan die einde van sinne kan uitoefen. Wanneer outeurs emotikons, hutswoorde, hiperskakels en/of die @-gebruikerskonvensie ná sinne of woorde gebruik (of aan die einde van twiets), blyk dit ŉ algemene gebruik te wees. Dit is slegs met uitsondering (sien voorbeeld 64, 66, 67 en 68) waar die outeurs nie van leestekens aan die einde van twiets gebruik maak nie. Soos reeds genoem, is O3 en O8 die enigste twee outeurs wat nie deurlopend van leestekens aan die einde van twiets gebruik maak nie.

Die afwesigheid van leestekens aan die einde van sinne (of twiets) verminder nie die leesbaarheid van twiets nie. Die bedoelde boodskap word steeds oorgedra. Aangesien Twitter ŉ sosialemediaplatform is, kan die afwesigheid van die konsekwente gebruik van leestekens aan die einde van sinne as ŉ voorbeeld van die informele aard van Afrikaans beskou word soos dit op Twitter gebruik word.

Daar is verder ondersoek ingestel na die gebruik van meer as een leesteken. Vervolgens word voorbeelde voorgehou van oortollige leestekengebruik deur die onderskeie outeurs:

Voorbeeld 69: **O1** – Yeay! Nou kan ons Potch plat ry. Geluk liefste**!!!** ❤

Voorbeeld 70: **O1 –** WAT IS DIT MET THE 100 WAT AL DIE HOOFKARAKTERS DOODMAAK???

Voorbeeld 71: **O4** – “Ek hoop die sywurms sneak enetjie**!!!** Come on Pumas**!!!**”, “Speel hulle NFL**???** Daai was forward**!!**”

Voorbeeld 72: **O5** – Ek HAAT dit om 'n flippen eier te wees**!!** Maar om die dood kry ek nie die simpel prentjie verander nie :|

Voorbeeld 73: **O5** – Geluk jy is in 'n privaat skool. Moet ek vir jou hande klap??

Voorbeeld 74: **O6** – Dis SO spesiaal**!!!** Een van my 'Facebook friends' het vir my 'n pakkie gestuur! Een vir elke dag van die week vir… http://fb.me/23TQ5Knyu

Voorbeeld 75: **O6** – Ek doen dit nog my hele lewe? Gedink almal weet van daardie woord?? 😉

Voorbeeld 76: **O7** – 14 man**!!!** Jy verloor teen 14 man. Ek gaan nou drink en show. Drink want daai game gaan homself nie vergeet nie

Voorbeeld 77: **O7** – Vreemd ekt self die kreef geduik en die wyn persent gekry. Maar jy sit daar vêr en oordeel. Wat doen jy ?????

Voorbeeld 78: **O8** – Dis #Pumba **!!** #Serengeti #LiveTheJourney #tanzania #africa https://www.instagram.com/p/BFjfKPrGQhk/

Voorbeeld 79: **O10** – “as die beste**!!**”

Voorbeeld 69 tot 79 dui die gebruik van meer as een leesteken aan soos gevind in die gespesialiseerde korpus. Die gewildste leestekens wat herhaal word, is die vraagteken en die uitroepteken. Die gebruik van hierdie twee leestekens sluit by Rao *et al.* (2010:41) se navorsing aan wat aanvoer dat die karakterbeperking op Twitter aanleiding daartoe gee dat die manier waarop emosies uitgedruk word redelik beperk is en dat meer as een leesteken (veral meer as een uitroep- of vraagteken) gebruik word om die intonasie van sprekers se stem in ŉ gesproke boodskap na te boots.

Die manier waarop O1, O4, O5, O6, O7, O8 en O10 meer as een vraag- of uitroepteken gebruik, is telkens om meer emosie aan lesers oor te dra en aangesien daar nie ŉ patroonmatigheid in die aantal uitroep- of vraagtekens is wat deur die outeurs gebruik word nie, kan daar nie aangevoer word dat die eienskap kenmerkend aan slegs ŉ enkele outeur is nie. Vanuit voorbeeld 69 tot 79 blyk dit egter asof daar ŉ verband tussen die emosie en die aantal leestekens is en hoe meer emosie (woede, dankbaarheid, liefde, verbasing) die outeurs wil oordra, hoe meer uitroep- en vraagtekens word gebruik.

Dit is ook opmerklik dat slegs O1, O5 en O8 van emotikons gebruik maak wanneer daar meer as een uitroepteken in ŉ sin gebruik word en dus kan daar gestel word dat indien daar twee of meer uitroeptekens en emotikons in ŉ twiet voorkom, O1, O5 of O8 as moontlike outeur van ŉ twiet oorweeg kan word.

Die enigste voorbeeld van oortollige leestekengebruik was ŉ ekstra punt in O2 se korpus (sien voorbeeld 80)[257] wat ná die aanhalingsteken gebruik is. Aangesien hierdie fout slegs een maal voorgekom het, is dit moontlik dat dit slegs ŉ tikfout was:

Voorbeeld 80: **O2** – "@GEBRUIKER: Witness 1: says she and her husband was woken up by a woman screaming at about 3am**.".** Volg haar gerus.

Voorbeeld 81: **O3** – @Beeld_Nuus is by die perskonferensie waar ministers Nkoana-Mashabane, Chabane en Peters die media sal toespreek oor begrafnisreelings

@ is in die algemeen deur O3 gebruik (sien voorbeeld 81). Die gebruik van 'n punt voor die verwysing na 'n gebruiker aan die begin van 'n sin kon dus as 'n algemene kenmerk van O3 se korpora beskou word[258]. Wanneer die verwysingskorpus geanaliseer is ten opsigte van die voorkoms van @ het dit geblyk dat dit slegs een maal in die hele korpus voorgekom het, en dan ook nie aan die begin van 'n sin nie. Buiten die gebruik van @ is daar slegs ŉ enkele ander voorkoms van oortollige leestekengebruik, en daar kon soos by O2 aangevoer word dat hierdie oortollige leesteken ŉ tikfout is.

ŉ Verdere aspek van die gebruik van leestekens wat ondersoek is, is die gebruik van ŉ verkeerde ellips (.. of ....). Die Afrikaanse Woordelys en Spelreëls (2017:113) stel dat die ellipsteken (ook genoem die beletselteken) uit drie stippels (…) bestaan en dit maak lesers van ŉ teks daarop bedag dat ŉ lang pouse volg, dat die spreker onderbreek is of dat letters, woorde of selfs stukke inligting weggelaat is. Aangesien O4 en O5 die enigste outeurs in die gespesialiseerde korpora is wat verkeerdelik van die ellips gebruik gemaak, kan daar aangevoer word dat wanneer O4 en O5 ŉ verdagte in ŉ ondersoek na die identifisering van ŉ onbekende outeur is, ŉ teks met voorbeelde van **..** of **....** as ŉ aanduiding van moontlike outeurskap beskou kan word. Dit is belangrik om daarom te let dat hier nie verwys word na die verkeerde gebruik van die ellips ten opsigte van reël 13.21 tot 13.25 van die Afrikaanse Woordelys en Spelreëls (2017:113-114) nie, maar slegs na die skryf van die ellips as .. of ....:

Voorbeeld 82: **O9** – Populêre woorde op Twitter: Adventure. Tee. Vegan. Seer. God. Kaalvoet. Prinses. Woorde.. Als goed en wel, WAT DIE FOK is 'n Hartsvriendin?

O9 maak oor die algemeen (daar is slegs een voorbeeld in die korpus en dit kan dalk net ŉ tikfout wees) nie van oortollige leestekens gebruik nie, en pas die aanvaarde taalreëls oor lees- en skryftekens toe. Dit op sigself is kenmerkend van hierdie outeur, maar hierdie patroonmatigheid kan nie as ŉ kenmerkende eienskap beskou word nie aangesien daar op Twitter self ook ander individue is wat soortgelyke eienskappe toon.

[257] Volgens Müller (2003:412), reël 9.16 – "By 'n aanhaling in die direkte rede (anders as waar bloot 'n woord of 'n frase aangehaal word) word 'n leesteken (punt, komma, uitroep- of vraagteken) aan die einde binne die aanhalingsteken geplaas."
[258] Sien ook afdeling 7.2.1.2 (Verwysing na gebruikers).

Ten opsigte van die gebruik van ander leestekens is daar bevind dat kommas konsekwent deur al die outeurs gebruik word en daar word nie afgewyk van die taalreëls ten opsigte van die gebruik van kommas nie − soos voorgehou in die Afrikaanse Woordelys en Spelreëls (2017:118-130). Aandagstrepe word nie deur O3 en O7 gebruik nie terwyl O10, O8, O6 en O2 die aandagstreep gebruik soos voorgeskryf in reël 13.2[259] van die Afrikaanse Woordelys en Spelreëls (2017:104). Die gebruik van die aandagstreep soos voorgehou in reël 13.3[260] (2017:104) word in die korpora van O10, O8, O6 en O2 gevind terwyl voorbeelde van reël 13.1[261] (2017:103) in die korpus van O2 en O10 voorkom. Aangesien daar nie van die taalnorm vir die gebruik van aandagstrepe afgewyk word nie, kan die gebruik daarvan nie as uniek tot die onderskeie outeurs beskou word nie, maar daar kan wel aangevoer word dat wanneer daar ŉ beperkte aantal verdagte outeurs is, die gebruik van die aandagstreep, soos bespreek, wel gebruik kan word om moontlike outeurs te identifiseer.

In die korpus van O1, O6 en O8 is daar wel afwykings in die gebruik van die aandagstreep:

Voorbeeld 83:	**O1** – Mens kry twee tipes studente: - dié wie vet word a.g.v. drank - dié wie maer word omdat hulle nie weet hoe om kos te maak nie.
Voorbeeld 84:	**O4** – Leeus leeeus leeeus!!!! Lions - Single by Radio Raps
Voorbeeld 85:	**O6** – Oe, vanaand is my 'date' Mej. Suid Afrika by die Glamour awards by die 5ster gastehuis, Morrels - Fairlands. http://fb.me/2oVoycCxy
Voorbeeld 86:	**O8** – #JOUNE in #OggendOpRSG @GEBRUIKER Daylight - Maroon 5 Dark Horse - Katy Perry My heart will go on - Celine Dion

O1, O4 en O8 gebruik die aandagstreep om na musiek te verwys. Al drie die outeurs verskaf die liedjie se naam gevolg deur ŉ aandagstreep en dan die sanger se naam (sien voorbeeld 84 en 86). Hierdie gebruik van die aandagstreep word ook in die verwysingskorpus gevind. O1 gebruik die aandagstreep wat met reël 13.2 (Afrikaanse Woordelys & Spelreëls, 2017:104) ooreenstem, maar ŉ dubbelpunt word steeds geplaas. Hierdie gebruik van die aandagstreep kom slegs een keer in die korpus van O1 voor. O6 (voorbeeld 85)gebruik die aandagstreep in die plek van die @-teken. Die @-teken – soos deur Honeycutt en Herring (2009:5-6) voorgehou – word gebruik om ŉ spesifieke ligging aan te dui. O6 is die enigste outeur wat die aandagstreep vir hierdie doel gebruik en dus kan dit as kenmerkend tot die outeurskap van O6 beskou word. ŉ Verdere gebruik van ŉ leesteken wat slegs in die korpus van O6 aangetref word, is die gebruik van die skuinsstreep – ook genoem die solidus of virgule (Afrikaanse Woordelys en Spelreëls, 2017:133). O6 gebruik agt keer die leesteken en daar kan tussen drie verskillende funksies onderskei word (sien voorbeeld 87 tot 89):

[259] Reël 13.2 van die AWS (2017:104) stel, "die aandagstreep word in die plek van 'n ellipsteken, 'n dubbelpunt, of woorde soos **naamlik**, **te wete**, **byvoorbeeld** of **soos** om 'n verduideliking, 'n reeks items of 'n wending vooraf te gaan."

[260] Reël 13.3 van die AWS (2017:104) stel "die aandagstreep word na 'n aanhaling en voor die outeur se naam gebruik."

[261] Reël 13.1 van die AWS (2017:103) stel, "die aandagstreep word gebruik om 'n parentese aan te dui."

Voorbeeld 87: **O6** – Dankie @GEBRUIKER, @GEBRUIKER vir fantastiese verblyf in die Kaap! Moet asb nie twee keer dink om by hulle te gaan bly nie! **9,5 / 10!**

Voorbeeld 88: **O6** – By **Thai spa's / restaurante**. Dischem het ook 'n Coconut water. (Maar weet jy hoe sleg smaak daai goed, Ronel?) lol

Voorbeeld 89: **O6** – Moenie vanaand 20:00 JAM / (kyknet) misloop nie! Ek en André gaan lekker gesels oor gedeelde passies. http://fb.me/vGz55pCM

In voorbeeld 87 gebruik O6 die solidus soos voorgehou in reël 13.57 van die Afrikaanse Woordelys en Spelreëls (2017:136) om "'n deel uit 'n totaal aan te dui." Voorbeeld 88 dui op die gebruik van die solidus volgens reël 13.51 (Afrikaanse Woordelys en Spelreëls, 2017:134) waar 'n alternatiewe keuse tussen twee of meer items voorgehou word. Die gebruik van die solidus in voorbeeld 89 is 'n afwyking van die taalreëls en in die konteks van die twiet kan lesers aflei dat die outeur die solidus in die plek van die woord *op* gebruik. Die twiet se karaktertelling is 124, dus was die vervanging van die woord met die solidus nie as gevolg van die karakterbeperking nie, en die gebruik kan as kenmerkend aan O6 beskou word.

Daar is slegs 27 voorbeelde van die gebruik van kommapunte in die gespesialiseerde korpus (O9 – 5; O6 – 19; O5 – 1; en O2 – 2). O9, O6 en O5 gebruik telkens die kommapunt in die plek van 'n dubbelpunt – reël 13.15[262] en reël 13.17[263] van die Afrikaanse Woordelys en Spelreëls (2017:110-111). O2 is die enigste outeur wat die kommapunt volgens die taalreëls gebruik – reël 13.45[264] van die Afrikaanse Woordelys en Spelreëls (2017:130). Wanneer 'n kommapunt dus soos in reël 13.45 (2017:130) in 'n mikroblogteks gebruik word, kan O2 as die moontlike outeur van die teks oorweeg word.

Die laaste leesteken wat ondersoek is, is die gebruik van hakies. Al tien die outeurs in die gespesialiseerde korpus maak gebruik van hakies en in elk van die outeurs se korpora is hakies gebruik soos voorgehou in reël 13.26[265] (Afrikaanse Woordelys & Spelreëls, 2017:116). O6 maak die meeste van hakies gebruik (62 keer; in 8% van O6 se twiets is hakies gebruik). O9, O6, O5, O2 en O1 (voorbeeld 90 tot 93) gebruik hakies ten einde alternatiewe of opsionele getalle, letters of morfeme aan te dui – reël 13.27 (Afrikaanse Woordelys & Spelreëls, 2017:117):

Voorbeeld 90: **O1** – Mens kry dosente en dan kry mens **do(o)sente**. Ken die verskil

[262] Reël 13.15 stel dat die dubbelpunt ná 'n frase of sin geplaas word wat 'n uitbreidende frase of sin inlei (Afrikaanse Woordelys & Spelreëls, 2017:110).
[263] Reël 13.17 verwys na die gebruik van die dubbelpunt wat volg ná 'n "woord, frase of sin wat 'n woord, frase of sin in die direkte rede of 'n aanhaling inlei wat in die sin geïntegreer is" (Afrikaanse Woordelys & Spelreëls, 2017:111).
[264] Volgens die Afrikaanse Woordelys en Spelreëls (2017:130) verwys reël 13.45 na die gebruik van 'n kommapunt tussen twee onafhanklike hoofsinne in die plek van 'n neweskikkende voegwoord.
[265] "Hakies word gebruik om 'n parentese in die sin aan te dui" (Afrikaanse Woordelys & Spelreëls, 2017:116).

Voorbeeld 91: **O2** – Hoop almal praat en skryf met **groot(er)** versigtigheid en respek oor regter in #OscarPistorius saak. Sy verdien nie beledigings nie.

Voorbeeld 92: **O5** – Nog 'n flippen Maandag. Nog 'n flippen week. Skool meng by my **vry(e)** tyd in.

Voorbeeld 93: **O6** – Kyk **Nou(ks)** ! Ek en Lydia braai vir die gaste Karoo skaaptjoppies met báíe vet aan! En (met permissie gesê) ons... http://fb.me/3WIk2JlGP

In die korpus van O7 word daar ŉ afwyking van die taalreëls aangetref met betrekking tot die gebruik van hakies. O7 (voorbeeld 94) voeg ŉ ekstra spasie tussen die linkerhakie en die teks wat daarop volg en ná die teks en die regterhakie (Afrikaanse Woordelys & Spelreëls, 2017:115):

Voorbeeld 94: **O7** – Die vars nuwe Valiant plaat **(nagrit)** maak n koue nag mooi. Die waarheid en mooi akoorde. Ek skiet my deamon dood met dooswyn 2

Aangesien O7 die enigste outeur in die gespesialiseerde korpus is wat hierdie fout begaan wanneer hakies gebruik word, kan dit as kenmerkend aan die taalstyl van O7 beskou word. Hierdie fout is deurlopend in al die twiets van O7 aangetref wanneer hakies gebruik word.

Na gelang van die ondersoek na die gebruik van leestekens deur die outeurs van die gespesialiseerde korpus is dit duidelik dat daar etlike gebruike in die onderskeie korpora aangetref word wat gebruik kan word om die outeurs van mikrobloginskrywings te identifiseer. In die volgende afdeling is die gebruik van leestekens in die verdagte tekste ondersoek om sodoende te bepaal wie die outeurs van hierdie drie tekste kan wees na gelang van die gebruik van leestekens.

7.2.2.2 Verdagte tekste: die gebruik van leestekens

In afdeling 7.2.2 is daar na die frekwensie van die gebruik van leestekens (sien Tabel 7.4 en 7.5) verwys. Daar is bepaal dat daar reeds na gelang van die frekwensie, ooreenkomste tussen outeurs vanuit die gespesialiseerde korpus en die verdagte tekste waarneembaar is. In afdeling 7.2.2.1 is daar spesifiek ondersoek ingestel na die onderskeie outeurs se gebruik van leestekens. Vervolgens word voorbeelde van die gebruik van leestekens vanuit die drie verdagte tekste voorgehou:

Voorbeeld 95: **V1** – Kyk gerus na my artikel vandag in Beeld en op Netwerk24: SAUK nie so onredelik oor beeldmateriaal http://www.netwerk24.com/Stemme/Aktueel/sauk-nie-so-onredelik-oor-beeldmateriaal-20160602 … via @GEBRUIKER

Voorbeeld 96: **V1** – Hoop van harte Ivor Price is deur kykNET opgeraap. Hy is 'n seldsame talent. En ek hoop ander SAUK-mense wat afgedank is, is gou weer op lug

Voorbeeld 97: **V2** – Ek weet wie môre in die hoof se kantoor gaan wees. #margaretvanwyk

Voorbeeld 98: **V2** – Ek mag dalk 'n slegte mens wees, maar tenminste het ek nie 'n Instagram account vir my hond nie. https://instagram.com/p/BJk056cBzn3/

Voorbeeld 99: **V3** – Hierdie prinses sal haarself red 👑

Voorbeeld 100: **V3** – Dis goed om te droom. Maak net seker dat jy nie nét droom nie, maar beplan ook. #fortunecookie

Soos aangetoon in afdeling 7.2.2.1 is daar talle faktore wat ŉ invloed op die gebruik van ŉ leesteken aan die einde van ŉ sin kan uitoefen soos die gebruik van emotikons, hutswoorde, hiperskakels en/of die @-gebruikerskonvensie ná ŉ sin. Wanneer die stylkenmerke van die gebruik van leestekens (al dan nie) aan die einde van ŉ sin ondersoek is, blyk dit dat V1 (sien voorbeeld 95) nie ná ŉ verwysing na ŉ gebruiker ŉ punt plaas nie en dan ook (soos in voorbeeld 96) die punt aan die einde van ŉ sin bloot uitlaat. Aangesien V1 slegs een maal die punt aan die einde van ŉ sin uitlaat (98,9% van V1 se sinne beskik wel oor die gebruiklike leesteken aan die einde van die sin) kan dit nie as ŉ kenmerk van V1 se outeurskap beskou word nie.

Voorbeeld 95 (geen leesteken ná ŉ verwysing na ŉ gebruiker nie) is ook die enigste verwysing na ŉ gebruiker wat aan die einde van ŉ twiet voorkom en dus kan dit weer eens nie as ŉ kenmerk van die outeurskap van V1 beskou word nie. Vanuit die gespesialiseerde korpus is dit veral die korpus van O2 wat ooreenkomste met V1 toon ten opsigte van die gebruik van leestekens aan die einde van ŉ sin:

Voorbeeld 101: **O2** – Ingeval jy hierdie gemis het in Beeld: Drukmedia weer in politieke soeklig http://www.netwerk24.com/stemme/2015-05-05-boodskappers-met-die-slegte-nuus-drukmedia-weer-in-politieke-soeklig … via @GEBRUIKER

Wanneer voorbeeld 95 met voorbeeld 101 vergelyk word, is die ooreenkoms tussen die twee outeurs se styl onbetwisbaar. V2 en V3 gebruik (al dan nie) leestekens aan die einde van ŉ sin en stem ooreen met ŉ algemene kenmerk van Afrikaans soos dit op Twitter gebruik word. Daar word selde punte gebruik ná emotikons, hutswoorde, hiperskakels en verwysings na gebruikers. Dit blyk dus dat outeurs (ook die verdagte outeurs) oorwegend wel van leestekens aan die einde van sinne gebruik maak.

Vervolgens verskaf ek spesifieke voorbeelde vanuit elk van die verdagte tekste ten einde te bepaal of die manier waarop meer as een leesteken gebruik word ook ooreenstem met die gebruik van meer as een leesteken vanuit die gespesialiseerde korpora:

Voorbeeld 102: **V3** – MOENIE VIR MY LIEG NIE EK SAL JOU VERWURG**!!**

Voorbeeld 103: **V3** – 🙋‍♀️🙌SELFDE**!!**

Voorbeeld 104: **V2** – Kan Apple net vir my rustig raak met al hierdie iOS Updates**??**

Voorbeeld 105: **V2** – Dié**!!** Veral 28, 29 & 50.

Vanuit voorbeeld 102 tot 105 is dit duidelik dat die outeurs van die verdagte tekste − soortgelyk aan die bevindinge vanuit afdeling 7.2.2.1 – ook van meer as een uitroep- of vraagteken gebruik maak wanneer die oormatige gebruik van leestekens aan die einde van sinne/twiets ondersoek is. Aangesien dit blyk om ’n algemene tendens op Twitter te wees, kan daar nie na gelang van die stilistiese analise van die gebruik van meer as een vraag- of uitroepteken moontlike outeurs vir V2 en V3 geïdentifiseer word nie. V1 maak van geen oortollige leestekens gebruik nie − ’n kenmerk wat in die korpus van O2 aangetref word − en O2 kan dus as moontlike outeur van V1 geïdentifiseer word. Die ondersoek na die verkeerdelike gebruik van ellipse[266] in die verdagte tekste het aangetoon dat dit slegs in die korpus van V3 is wat hierdie afwyking aangetref word (voorbeeld 106):

Voorbeeld 106: **V3** – Ek het omgegee.. nou is ek net van fok jou ✌

Vanuit die gespesialiseerde korpus is dit slegs O4 en O5 wat die ellips soortgelyk aan V3 gebruik; daarom kan hierdie twee outeurs by die moontlike outeurs van V3 gevoeg word. Ten opsigte van die gebruik van ander leestekens is daar bevind dat kommas ook konsekwent[267] deur al die outeurs gebruik word en dat daar nie afgewyk word van die taalreëls ten opsigte van die gebruik van kommas nie − soos voorgehou in die Afrikaanse Woordelys en Spelreëls (2017:118-130).

Aandagstrepe word deur V2 en V3 gebruik, maar in albei die outeurs se korpora word die aandagstreep volgens die taalreëls (2017:104 in reël 13.2[268]) gebruik. Aangesien daar nie van die taalnorm vir die gebruik van aandagstrepe afgewyk word nie, kan die gebruik nie as uniek tot die onderskeie outeurs beskou word nie, maar daar kan wel aangevoer word − na gelang van die analises soos voorgehou in afdeling 7.2.2.1 − dat O10, O8, O6 en O2 ook aandagstrepe op dieselfde manier gebruik en as moontlike outeurs van V2 en V3 beskou kan word. Daar is geen skuinsstrepe of hakies in die verdagte tekste aangetref nie, en V1 is die enigste outeur wat van ’n kommapunt gebruik maak. Die kommapunt word gebruik soos voorgehou in reël 13.15[269] (Afrikaanse Woordelys & Spelreëls, 2017:110). O2 is die enigste outeur vanuit die

[266] Die Afrikaanse Woordelys en Spelreëls (2017:113) stel dat die ellipsteken (ook genoem die beletselteken) uit drie stippels (…) bestaan, en dat dit die lesers van die teks daarop bedag maak dat ’n lang pouse volg, dat die spreker onderbreek is, en dat letters, woorde of selfs stukke inligting weggelaat is.

[267] Daar is wel twiets waarin kommas ontbreek, maar dit blyk ’n algemene verskynsel op die platform te wees wat eerder betrekking op swak taalkennis het.

[268] Reël 13.2 van die Afrikaanse Woordelys en Spelreëls (2017:104) stel: “die aandagstreep word in die plek van ’n ellipsteken, ’n dubbelpunt, of woorde soos **naamlik**, **te wete**, **byvoorbeeld** of **soos** om ’n verduideliking, ’n reeks items of ’n wending vooraf te gaan.”

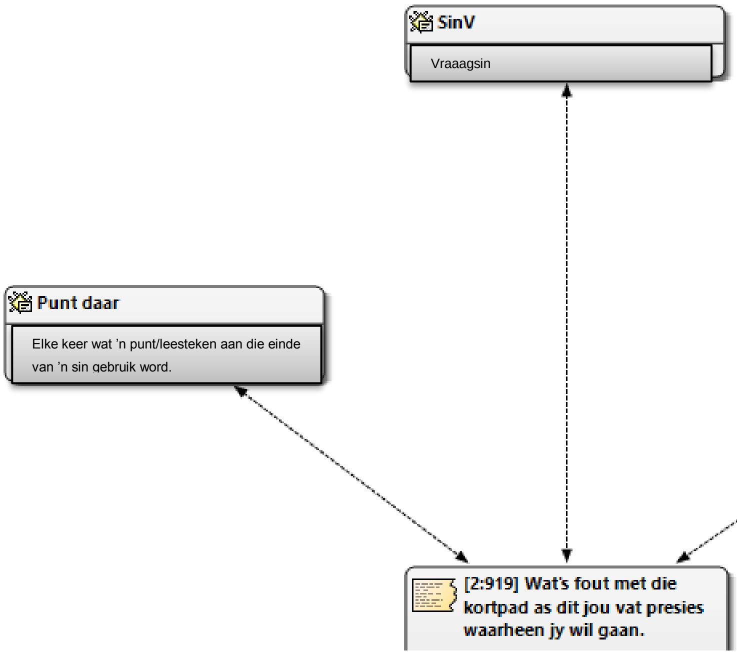
[269] Reël 13.15 stel dat die dubbelpunt ná ’n frase of sin geplaas word wat ’n uitbreidende frase of sin inlei (Afrikaanse Woordelys & Spelreëls, 2017:110).

gespesialiseerde korpus wat die kommapunt ook só gebruik en dus kan O2 na gelang van die gebruik van die kommapunt die mees waarskynlike kandidaat vir die outeurskap van V1 wees.

Alhoewel daar baie minder leestekens in die verdagte tekste gevind is, blyk dit dat die leestekens wat in die verdagte tekste gebruik word wel ŉ bydra kan lewer in die identifisering van outeurs van mikrobloginskrywings. In die hieropvolgende afdeling is die gebruik van hoofletters in die gespesialiseerde korpus en die verdagte tekste geanaliseer.

7.2.3 Hoofletters

Die ondersoek na hoofletters is uitgevoer na aanleiding van Barry en Luna (2012:2), Bhargava *et al.* (2013) en McMenamin (2002:229) wat aanvoer dat die gebruik van hoofletters op sosiale media kan afwyk van die normale taalkonvensies. Verhoef (2001:8) het in haar studie oor Internetafrikaans aangevoer dat "hooflettergebruik spaarsamig" is wanneer internettekste ondersoek word. Vervolgens is die gebruik van hoofletters aan die begin van ŉ sin, hooflettergebruik ten einde ŉ woord te beklemtoon en ook die gebrek aan hoofletters in eiename ondersoek. Ten einde die ondersoek na die gebruik van hoofletters aan die begin van ŉ sin uit te voer, is *ATLAS.ti* gebruik. Elk van die korpora is in die program ingevoer en ŉ kode is geskep wat aan elke sin toegeken is wat wel met ŉ hoofletter begin (sien Skermskoot 7.2) en wanneer hoofletters ontbreek.



Skermskoot 7.1 Voorstelling van ŉ twiet waarin die hoofletter aan die begin van ŉ sin met *ATLAS.ti* gekodeer is

Soos gesien in Skermskoot 7.1, is daar ŉ kode (byvoorbeeld, *hoofletter sin daar*) aan elke sin toegeken waar die hoofletter wel aan die begin van ŉ sin gebruik is. Waar daar nie hoofletters

aan die begin gebruik is nie, is die sin soortgelyk as in Skermskoot 7.1 gekodeer. Nadat al die sinne in die onderskeie korpora gekodeer is, is die data as ’n Microsoft Excel-dokument gestoor, en die frekwensie waarteen elk van die onderskeie outeurs hoofletters gebruik het of nie is bereken. Die bevindinge van die bogenoemde word in Tabel 7.6 voorgehou:

Tabel 7.6 Die gebruik van hoofletters in die onderskeie korpora

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Hoofletters aanwesig aan die begin van 'n sin	Hoofletters aan die begin van ’n sin per korpus genormaliseer tot 1 000 sinne	Hoofletters afwesig aan die begin van 'n sin	Hoofletters afwesig aan die begin van ’n sin per korpus genormaliseer tot 1 000 sinne
O1	669	944	8 942	924	978,81	20	21,19
O2	589	1 577	12 259	1 521	964,49	56	35,51
O3	206	266	3 379	266	1000,00	0	0,00
O4	484	978	7 110	870	889,57	108	110,43
O5	333	556	4 579	509	915,47	47	84,53
O6	768	1 657	16 340	1 477	891,37	180	108,63
O7	457	790	5 655	767	970,89	23	29,11
O8	308	443	3 690	437	986,46	6	13,54
O9	275	467	4 918	456	976,45	11	23,55
O10	499	1 190	7 744	1 085	911,76	105	88,24
V1	42	90	844	90	1000,00	0	0,00
V2	96	147	1409	144	979,59	3	20,41
V3	32	44	412	40	909,09	4	90,91

In Tabel 7.6 word die gebruik van hoofletters voorgehou. Vanuit hierdie Tabel is dit duidelik dat die gebruik van hoofletters aan die begin van sinne (of die weglaat daarvan) tussen die outeurs vanuit die gespesialiseerde korpus verskil. Wanneer die frekwensie waarteen hoofletters aan die begin van sinne gebruik word (of nie) ondersoek is, blyk dit dat V1 (hooflettergebruik – 1 000,00; hoofletters uitgelaat – 0,00) en O3 (hooflettergebruik – 1 000,00; hoofletters uitgelaat – 0,00) ooreenstem terwyl V2 (hooflettergebruik – 979,59; en hoofletters uitgelaat – 20,41) en O1 (hooflettergebruik – 978,81; en hoofletters uitgelaat – 21,19) sowel as V3 (hooflettergebruik – 909,09; en hoofletters uitgelaat – 90,91) en O10 (hooflettergebruik – 911,76; en hoofletters uitgelaat – 88,24) ooreenstem in terme van die genormaliseerde waarde van die gebruik en die uitlaat van hoofletters. Hoofletters wat vir beklemtoning gebruik word, is ook in die onderskeie korpora in *ATLAS.ti* gekodeer ten einde te bepaal hoeveel – en indien wel – die outeurs van hoofletters gebruik maak om woorde of sinne te beklemtoon. Tabel 7.7 toon die bevindinge aan:

Tabel 7.7 Gebruik van hoofletters ten einde woorde te beklemtoon

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters vir beklemtoning in korpus	Totale aantal hoofletters vir beklemtoning in korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters vir beklemtoning per twiet
O1	669	944	8 942	39	4,36	0,06
O2	589	1 577	12 259	24	1,96	0,04
O3	206	266	3 379	0	0,00	0,00
O4	484	978	7 110	105	14,77	0,22
O5	333	556	4 579	23	5,02	0,07

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters vir beklemtoning in korpus	Totale aantal hoofletters vir beklemtoning in korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters vir beklemtoning per twiet
O6	768	1 657	16 340	172	10,53	0,22
O7	457	790	5 655	13	2,30	0,03
O8	308	443	3 690	11	2,98	0,04
O9	275	467	4 918	23	4,68	0,08
O10	499	1 190	7 744	130	16,79	0,26
V1	42	90	844	0	0,00	0,00
V2	96	147	1409	3	2,13	0,03
V3	32	44	412	5	12,14	0,16

Wanneer die vergelyking van die frekwensie (genormaliseer tot 1 000 woorde) waarteen die onderskeie outeurs hoofletters gebruik om woorde te beklemtoon intern ondersoek is, is dit slegs O3, O5, O7 en O9 wat deurlopend konsekwent is met die frekwensie (genormaliseer tot 1 000 woorde) waarteen hoofletters vir beklemtoning gebruik word. Die gebruik van hoofletters deur O1, O2, O4, O6, O8 en O10 verskil intern so baie[270] dat daar nie aangevoer kan word dat dit dieselfde outeur is nie.

ŉ Verdere ondersoek na die gebruik van hoofletters vir beklemtoning (sien Tabel 7.7) dui daarop dat O10, O4 en O6 die meeste hiervan gebruik maak (O10 – 130 of genormaliseer tot 16,79; O4 – 105 of genormaliseer tot 14,77; O6 – 172 of genormaliseer tot 10,53). O2 gebruik die minste hoofletters vir die beklemtoning van woorde (23 of genormaliseer tot 1,96). Die oorblywende vyf outeurs (O1, O2, O5, O7, O8 en O9) gebruik hoofletters ter wille van beklemtoning − ook elk teen ŉ ander genormaliseerde waarde. O3 gebruik glad nie hoofletters om woorde te beklemtoon nie en op sigself is die afwesigheid van hoofletters ter wille van beklemtoning in dié geval ook ŉ aanduiding van outeurskap.

Wanneer die genormaliseerde waarde vir die gebruik van hoofletters ter wille van beklemtoning − soos uitgebeeld in Tabel 7.7 − vergelyk is, is daar bevind dat V1 (0,00) en O3 (0,00), V2 (2,13) en O2 (1,96) en O7 (2,30) sowel as V3 (12,14) en O6 (10,53) ooreenstem. Vanuit die ondersoek na die gebruik van hoofletters ter wille van beklemtoning van woorde blyk dit dus dat daar wel tussen die onderskeie outeurs onderskei kan word indien ŉ genormaliseerde waarde ondersoek word.

Tabel 7.8 Hoofletters wat nie volgens die taalreëls gebruik is nie

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters nie volgens die taalreëls in korpus gebruik nie	Verwysing na hoofletters nie volgens die taalreëls gebruik nie per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters nie volgens die taalreëls gebruik nie per twiet
O1	669	944	8 942	4	0,45	0,01
O2	589	1 577	12 259	1	0,08	0,00
O3	206	266	3 379	7	2,07	0,03
O4	484	978	7 110	8	1,13	0,02

[270] Sien Bylaag E.

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters nie volgens die taalreëls in korpus gebruik nie	Verwysing na hoofletters nie volgens die taalreëls gebruik nie per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters nie volgens die taalreëls gebruik nie per twiet
O5	333	556	4 579	0	0,00	0,00
O6	768	1 657	16 340	13	0,80	0,02
O7	457	790	5 655	81	14,32	0,18
O8	308	443	3 690	9	2,44	0,03
O9	275	467	4 918	0	0,00	0,00
O10	499	1 190	7 744	10	1,29	0,02
V1	42	90	844	0	0,00	0,00
V2	96	147	1409	0	0,00	0,00
V3	32	44	412	0	0,00	0,00

Vanuit Tabel 7.8 is dit duidelik dat O9 en O5 nie een keer verkeerdelik hoofletters gebruik nie terwyl die oorblywende agt outeurs nie altyd die taalkonvensies volg wanneer hoofletters gebruik word nie. Die meeste hoofletters word in die korpus van O7 aangetref en word nie volgens die taalkonvensies gebruik nie (81, genormaliseer tot 1 000 woorde – 14,32). In die oorblywende sewe outeurs (O1, O2, O3, O4, O6, O8 en O10) se korpora is die hoogste voorkoms van onkonvensionele hooflettergebruik in O8 gevind (’n genormaliseerde waarde van 2,44).

Die vergelyking van die onkonvensionele gebruik van hoofletters (wanneer die frekwensie van voorkoms tot 1 000 woorde genormaliseer word) toon dat elk van die outeurs hoofletters op ’n ander manier gebruik. Vanuit Tabel 7.8 is dit duidelik dat daar nie ’n enkele voorbeeld in die verdagte tekste is waar die outeurs nie van hoofletters gebruik maak soos deur die Afrikaanse spellreëls voorgeskryf nie. In Tabel 7.8 is daar twee korpora (O5 en O9) wat ook nie hoofletters uitlaat nie terwyl O1 (0,45) en O2 (0,08) minimaal nie die taalreëls volg ten opsigte van die gebruik van hoofletters. O5, O9, O1 en O2 kan vervolgens by die lys van moontlike outeurs vir V1, V2 en V3 gevoeg word.

Vervolgens is die stylkenmerke van die gebruik van hoofletters in die gespesialiseerde korpus sowel as die verdagte tekste ondersoek.

7.2.3.1 Gespesialiseerde korpus: hooflettergebruik in twiets

Vanuit hierdie analise kom dit voor asof die outeurs meestal aan die normale taalkonvensies voldoen en hoofletters gebruik. Voorbeelde van die afwesigheid van hoofletters aan die begin van sinne sluit in:

Voorbeeld 107: **O4** – Ek moet sê daai Ierse slot is so **lank Hulle** meet hom in altitude!

Voorbeeld 108: **O6** – **groot** asseblief **kersvader**... http://fb.me/7e4yYS0Jo

Voorbeeld 109: **O10** – Die @GEBRUIK se Wilmar Arnoldi voeg nog vyf punte tot die telbord **by. @gebruiker** verdoel suksesvol. 24-7 vir Pukke.

Voorbeeld 107 tot 109 toon aan hoe die outeurs hoofletters uitlaat aan die begin van ŉ sin. Aangesien dit nie ŉ kenmerkende eienskap van enige van die outeurs is nie, is daar ondersoek ingestel na die uitlaat van hoofletters aan die begin van ŉ sin om te bepaal of dit as ŉ eienskap van Afrikaans op Twitter beskou kan word. Dit kom voor asof daar nie ŉ bepaalde patroonmatigheid in die afwesigheid van hoofletters is nie.

Vanuit die ondersoek na die gebruik van hoofletters kan daar dus aangevoer word dat hoofletters oor die algemeen wel gebruik word en met die uitsondering van hoofletters ná die @-gebruikerkonvensie, daar so veel moontlik probeer word om die taalreëls ten opsigte van hoofletters aan die begin van ŉ sin te volg.

7.2.3.2 Verdagte tekste: hooflettergebruik in twiets

In afdeling 7.2.3.1 is daar bevind dat daar in teenstelling met Verhoef (2001:8) se bevinding dat “hooflettergebruik spaarsamig is” wanneer internettekste ondersoek word asook Barry en Luna (2012:2), Bhargava *et al.* (2013) en McMenamin (2002:229) wat aanvoer dat die gebruik van hoofletters op sosiale media kan afwyk van die normale taalkonvensies, so ver moontlik deur die outeurs wat deel van die gespesialiseerde korpus uitmaak, probeer om hoofletters te gebruik soos bepaal deur die normale taalkonvensies van Afrikaans.

Die stylkenmerke wat in die drie verdagte tekste sigbaar is, word in voorbeeld 110 tot 113 voorgehou:

Voorbeeld 110: **V1** – En as SAUK sy stasie privatiseer, wat bly oor? En gaan Multichoice die arm massas bedien? Gratis? Komaan, Dawie, ons het goeie SAUK nodig

Voorbeeld 111: **V2** – Ek's nie wie ek was 'n jaar terug nie **Ek's** nie wie ek was 'n maand terug nie **Ek's** nie wie ek was 'n week terug nie **Ek's** ’n verkleurmannetjie

Voorbeeld 112: **V3** − Ek het omgegee**..** **nou** is ek net van fok jou ✌️

Voorbeeld 113: **V3** – **ek** weet dis moeilik en ek weet dis seer maar ek belowe dit word beter as jy kies om hieruit te leer

Voorbeeld 110 is ’n aanduiding van hoe V1 deurlopend hoofletters gebruik soos voorgehou deur die taalreëls van Afrikaans. Vanuit voorbeeld 111 tot 113 is dit duidelik dat daar nie ’n bepaalde patroonmatigheid waargeneem kan word wanneer hoofletters nie gebruik word soos deur die norm daar gestel word nie. Wanneer die voorbeelde van V2 en V3 (voorbeeld 111, 112 en 113) geanaliseer is, kan die vraag gestel word of die gebrek aan hoofletters aan die begin van ’n sin doelbewus was.

Aangesien die meerderheid Twitter-gebruikers van slimfone gebruik maak om te twiet, kan die aanname gemaak word dat die uitlaat van punte aan die einde van ’n sin asook die lukrake

gebruik van hoofletters – soos in die geval van voorbeeld 111 waar die *ek's* herhaaldelik met ŉ hoofletter geskryf word – doelbewus is aangesien die meeste slimfone outomaties 'n hoofletter plaas nadat 'n leesteken gebruik is of aan die begin van 'n sin (sien voorbeeld 112 en 113). Selfs al word die twiets vanaf 'n gewone sleutelbord getwiet, moet die outeurs doelbewus die taalreëls ignoreer om die hoofletters uit te laat.

Soos in afdeling 7.2.3.1 kan daar na gelang van die ondersoek na die gebruik of die uitlaat van hoofletters aan die begin van 'n sin aanvoer dat die aantal keer wat outeurs wel hoofletters gebruik kan bydra tot die identifisering van outeurs van mikrobloginskrywings, maar dat die stylkenmerke as 'n algemene aspek van Twitter beskou kan word.

7.2.3.3 Gespesialiseerde korpus: hoofletters vir beklemtoning

Ritter *et al.* (2011:1527) argumenteer dat hoofletters in twiets gebruik word om belangrike inligting te beklemtoon of om ŉ bepaalde emosie oor te dra. Só kan outeurs ŉ enkele woord in hoofletters gebruik om uit te beeld hoe kwaad, hartseer of verbaas hulle is. Vervolgens word voorbeelde van die gebruik van hoofletters om woorde te beklemtoon, voorgehou:

Voorbeeld 114: **O1** – **EK WIL NETBAL SPEEL**!

Voorbeeld 115: **O2** – Oukei, dis nie n mediakwessie nie, maar ek sien dit op TV: grensregters by rugby met broeke wat so **hOOg** sit. Enige verduideliking? #rugby

Voorbeeld 116: **O4** – boys as jy babalas het, en jy **MOET** een van dié doen sal jy:

Voorbeeld 117: **O5** – Hou op inmeng. Hou. Op. Inmeng. **HOU. FOKKEN. OP. INMENG**.

Voorbeeld 118: **O6** – Om 'n fout te begaan in die lewe, maak **VERLOORDERS** bang om te probeer. Om 'n fout te begaan in die lewe, inspireer **WENNERS** om te probeer.

Voorbeeld 119: **O7** – **AI EK WENS ELK HET NOU NAGRIT KON HOOR EN 10 TON ROOIWYN KON SAAM DRINK**

Voorbeeld 120: **O9** – Current mood: Gee my alkohol en jou ma se kredietkaart **OF GEE MY JOU FOKKEN LEWE.**

Voorbeeld 114 tot 120 toon aan hoe die outeurs vanuit die gespesialiseerde korpus hoofletters gebruik om woorde te beklemtoon. Dit is duidelik dat daar nie ŉ definitiewe patroon waargeneem kan word wanneer O1, O2, O4, O5, O6, O7 en O8 se korpora ondersoek is nie en dat verskeie woorde (voorbeeld 115, 116 en 118) frases (voorbeeld 117 en 120) en soms ook sinne (voorbeeld 114 en 119) met hoofletters geskryf word.

Hierdie kenmerk is ook in die verwysingskorpus nagegaan en ook in hierdie korpus is daar menige voorbeelde van woorde, frases en sinne wat in hoofletters geskryf word. Hierdie woorde, sinne en frases verleen ekstra emosie aan die twiets, beklemtoon die emosie en die boodskap wat die outeurs probeer om aan die lesers oor te dra. Die gebruik van hoofletters ten

einde woorde te beklemtoon kan dus as ŉ algemene kenmerk van Afrikaans beskou word soos dit op Twitter gebruik word en dit word spesifiek gebruik om emosie oor te dra of om die boodskap van twiets te beklemtoon.

In die korpus van O9 is daar ŉ patroonmatigheid geïdentifiseer (sien voorbeelde 120). Hierdie outeur gebruik oorwegend kraswoorde in hoofletters. Alhoewel O5, O4 en O1 ook kraswoorde in hoofletters gebruik, gebruik hierdie outeurs hoofletters om kraswoorde te beklemtoon, onderskeidelik 1 (0.02%), 4 (0,05%) en 1 (0,01%) keer. O9 daarenteen gebruik 12 keer kraswoorde in hoofletters – dit wil sê 0,24% van O9 se korpus is kraswoorde wat in hoofletters geskryf word.

Verder kan daar aangevoer word dat 52,1% van alle woorde wat deur O9 in hoofletters geskryf word, kraswoorde is terwyl die persentasie kraswoorde wat in hoofletters geskryf word slegs 4,3% (O5), 3,8% (O4) en 2,6% (O1) van die totale hoofletters uitmaak wat in die onderskeie korpora gebruik word om woorde, sinne en frases te beklemtoon. Die afleiding wat hieruit gemaak is, is dat wanneer kraswoorde frekwent in hoofletters geskryf word, kan O9 wel moontlik die outeur van die teks wees.

Soos genoem, is O4, O6 en O10 die outeurs wat die meeste van hoofletters ter wille van beklemtoning gebruik maak. Geen patroonmatigheid is by O6 en O4 waargeneem nie. Wanneer O10 se gebruik van beklemtoonde hoofletters egter ondersoek is, word twee woorde deurlopend in hoofletters geskryf. Hierdie twee woorde, *TELLING (*die woord word tien keer gebruik*)* en *PAALTJIE (*die woord word 64 keer gebruik), vorm 57% van die 130 hoofletters wat vir beklemtoning in hierdie outeur se korpus gebruik word (sien voorbeeld 121 en 122):

Voorbeeld 121: **PAALTJIE**: @GEBRUIKER se Ruan Haasbroek verloor sy paaltjie op 19 lopies. 182/3 na 16.5 boulbeurte. @GEBRUIKER #VarsityCricket

Voorbeeld 122: **TELLING**: 125/8 na 17 boulbeurte vir NMMU. @GEBRUIKER @GEBRUIKER #VarsityCricket

Hierdie woorde toon aan waaroor die twiet gehandel het en soos gesien in die voorbeelde is daar steeds verwysings na gebruikers en hutswoorde in die twiets. Dit kom dus voor asof die outeur doelbewus vir lesers wou aankondig waaroor die twiet gehandel het. Hierdie eienskap is kenmerkend aan O10 en kan as ŉ identifiserende kenmerk beskou word.

Die ondersoek na die gebruik van hoofletters om woorde te beklemtoon kan dus wel as ŉ maatstaf van outeurskap gebruik word wanneer outeurs van mikrobloginskrywings geïdentifiseer moet word. Hoofletters word gebruik om emosie oor te dra of om die onderwerp in twiets te versterk en kan as ŉ kenmerk van Afrikaans, soos dit op Twitter gebruik word, beskou word.

7.2.3.4 Verdagte tekste: hoofletters vir beklemtoning

Voorbeeld 123 tot 126 dui aan hoe (en waar) die onderskeie outeurs van die verdagte tekste hoofletters gebruik ten einde ’n woord, frase of sin te beklemtoon:

Voorbeeld 123: **V2** – DIS 'N QWERTY KEYPAD. TIK JOU WOORDE VOLLEDIG UIT!

Voorbeeld 124: **V2** – Hy: "Ek wens ek kon teken." HOEKOM? Hy: "Dan sou ek jou oor en oor geteken het, want jy's beeldskoon."

Voorbeeld 125: **V3** – MOENIE VIR MY LIEG NIE EK SAL JOU VERWURG!!

Voorbeeld 126: **V3** – Jannie jammergat? TOTSIENS!!

Wanneer die bostaande voorbeelde met die voorbeelde in afdeling 7.2.3.3 vergelyk word, kan daar aangevoer word dat V2 (voorbeeld 123 en 124) sterk ooreenkomste toon met die korpus van O1 (voorbeeld 114). In beide die korpora gebruik die outeurs hoofletters vir beklemtoning as hulle ’n sterk boodskap aan lesers wil oordra of wanneer ’n woord gebruik word wat beklemtoon moet word ten einde die belangrikheid van die spesifieke woord oor te dra.

Die korpora van V3 (voorbeeld 125 en 126) stem ooreen met die korpus van O5 (sien voorbeeld 117) aangesien albei die outeurs geneig is om negatiewe boodskappe deur middel van die gebruik van hoofletters te beklemtoon. Na gelang van hierdie analise kan daar vervolgens aangevoer word dat O2 en O7 as moontlike outeurs van V2 beskou kan word na gelang van die frekwensie waarteen hoofletters vir beklemtoning gebruik word. Ten opsigte van die stylkenmerke wat waargeneem is wanneer hoofletters ter beklemtoning gebruik is, blyk dit dat V2 en O1 ’n groter ooreenskoms toon as die ander outeurs vanuit die gespesialiseerde korpus. V3 kan na gelang van die analise van die frekwensie waarteen hoofletters vir beklemtoning gebruik is moontlik deur O6 geskryf wees, maar wanneer daar spesifiek ondersoek ingestel is na die manier waarop die hoofletters vir beklemtoning gebruik is, blyk dit dat V3 en O5 hoofletters ter wille van beklemtoning op dieselfde manier gebruik.

Vervolgens is die uitlaat van hoofletters ondersoek om te bepaal of hierdie aspek van leestekengebruik wel gebruik kan word wanneer die outeurs van mikrobloginskrywings geïdentifiseer moet word.

7.2.3.5 Gespesialiseerde korpus: verkeerde gebruik van hoofletters

Ten einde die stylkenmerke in die stilometriese ondersoek te kon analiseer, verskaf ek vervolgens voorbeelde vanuit die onderskeie korpora waar outeurs nie hoofletters gebruik soos in die taalreëls van Afrikaans vervat nie:

Voorbeeld 127: **O1** – As jy in n koshuis op die PUK bly is sondag die nuwe **vrydag**

Voorbeeld 128: **O3** – Gordhan sê die hoof van **Sars** (in hierdie geval Tom Moyane) is nie op dieselfde vlak as 'n minister nie @GEBRUIKER

Voorbeeld 129: **O4** – Sien uit na die **sharks** game vanaand!

Voorbeeld 130: **O6** – Waar kry Cash and **carry** my selfoonnommer?

Voorbeeld 131: **O7** – Ons Rock n Roll die naweek in Tzaneen en **saterdag** en **sondag** doen ek akoestiese opvoerings by Rubys in die mooi **vaal rivier** dorpie Parys

O1 gebruik vier keer hoofletters wat afwyk van die taalkonvensies. Drie van die voorbeelde is dae van die week (sien voorbeeld 127) wat sonder 'n hoofletter geskryf word (sien Afrikaanse Woordelys & Spelreëls, 2017:47; reël 9.3). Voorbeeld 128 toon aan hoe O3 nie hoofletters in akronieme gebruik nie. Die akroniem *SANRAL* (O3 gebruik dit as *sanral*) kom vyf keer in die korpus van O3 voor.

Die oorblywende twee voorbeelde in die korpus is ook akronieme wat sonder hoofletters geskryf word (*ABSA* en *SARS*). SANRAL, ABSA en SARS word egter deur van die ander outeurs ook gebruik en met hoofletters geskryf. Ook in die verwysingskorpus word daar hoofletters gebruik wanneer SANRAL, ABSA en SARS in twiets genoem word[271].

Ten einde te bepaal of hierdie kenmerk van O3 se hooflettergebruik slegs op O3 van toepassing is, is 'n soektog op Twitter geloods. Daar is bevind dat daar wel ander gebruikers is wat soortgelyk aan O3 hoofletters in akronieme gebruik. Vir die doel van hierdie studie is die afwesigheid van die hoofletters in die akronieme SANRAL, SARS en ABSA 'n kenmerk wat slegs in die korpus van O3 voorkom en wanneer hierdie akronieme dus in twiets gebruik word (as sanral, absa en sars) kan O3 as 'n moontlike verdagte outeur ondersoek word[272].

O7 (die outeur in wie se korpus daar die meeste voorbeelde voorkom van hoofletters wat nie volgens die taalreëls gebruik word nie) gebruik oorwegend nie hoofletters in eiename nie (sien Afrikaanse Woordelys & Spelreëls, 2017:47; reël 9.3) en O7 verwys na die KKNK as kknk. O4, O6, O8 en O10 begaan deurlopend die fout om eiename nie met hoofletters te skryf nie, maar nie teen dieselfde hoë frekwensie as O7 nie.

In die korpora van O4, O6, O7, O8 en O10 word sommige eiename wel met hoofletters geskryf en dus is dit nie kenmerkend van een van die outeurs nie, maar dit kan eerder as 'n kenmerk van die informele aard van Afrikaans beskou word soos dit op Twitter gebruik word (dit kan ook toegeskryf word aan die gebruik van slimfone).

---

[271] Sien ook afdeling 7.4 (Morfologie).

[272] Volgens Müller (2007:270), reël 4.11 sowel as die Afrikaanse Woordelys en Spelreëls (2017:18), reël 3.10 en 3.11 kan akronieme met of sonder hoofletters geskryf word. Müller (2007:270) gaan voort deur aan te voer dat Media24 gewoonlik slegs die eerste letter van die akroniem met 'n hoofletter skryf. Dus is die afwesigheid van hoofletters in die korpus van O3 nie 'n fout nie. Verder is die korpus van O3 in die nuuskategorie geplaas, dus is díé eienskap kenmerkend van die kategorie.

In die korpus van O2 is daar slegs een voorbeeld waar die gebruik van ’n hoofletter nie die taalreëls van Afrikaans volg nie. Die fout is in ’n hutswoord aangetref (*#New zeeland#*) en hierdie kon dalk net ’n tikfout gewees het[273].

Vanuit die ondersoek na die manier waarop die outeurs nie die taalreëls volg nie wanneer hoofletters gebruik word, blyk dit dat eiename gereeld sonder hoofletters geskryf word, maar daar kan nie ’n patroonmatigheid by die onderskeie outeurs waargeneem word nie, en dus is dit eerder as ’n algemene kenmerk van Afrikaans beskou soos dit op Twitter gebruik word. Ook akronieme en afkortings word sonder die gepaste hoofletters geskryf, maar daar kan nie ’n patroonmatigheid aan die voorkoms van hierdie kenmerke gekoppel word nie. O3 is die enigste outeur in wie se korpus die taalreëls met betrekking tot hoofletters só verbreek word dat die voorkoms van die foute (sars, absa en sanral) as ’n kenmerk van outeurskap beskou kan word.

7.2.3.6 Verdagte tekste: verkeerde gebruik van hoofletters

Vanuit die ondersoek is dit duidelik dat daar nie ’n enkele voorbeeld in die verdagte tekste gevind kon word waar die outeurs nie van hoofletters gebruik maak soos deur die Afrikaanse taalreëls voorgeskryf nie. Die feit dat daar geen voorbeelde in die verdagte tekste voorkom waar die outeurs hoofletters uitlaat nie, kan dalk toegeskryf word aan die feit dat die verdagte korpora so klein is. Die ondersoek na die gebruik van hoofletters kan dus na gelang van die bevindinge in hierdie analise wel as ’n manier beskou word waarop die outeurs van mikrobloginskrywings geïdentifiseer kan word. Dit wil voorkom dat die algemene kenmerke van Twitter (soos in afdeling 7.2 bespreek) ’n waardevolle bydra lewer tot die beantwoording van die navorsingsvrae soos gestel in hoofstuk 1 (afdeling 1.3). Die bespreking van die sintaktiese ondersoek word vervolgens voorgehou.

7.3 Sintaksis

Du Plessis (1988:11) beskou sintaksis as die onderlinge verhouding tussen woorde in sinne. In die geval van mikroblogs ontstaan die vraag: hoe lyk die sinne wat gebruikers op hierdie platform gebruik? Hoe beïnvloed die karakterbeperking die sinne wat gebruikers gebruik en hoe figureer sintaksis in outeurskapidentifikasie?

Wat kan as die basiese sintaksis van Afrikaans soos dit op mikroblogs gebruik word, beskou word? Carstens (2011:40) verwys na die argument dat sinne nie na willekeur gevorm kan word nie, en dat daar altyd ’n onderliggende patroon is wat gevolg moet word. Sinne word saamgestel (en soos reeds genoem moet sinne sinvol – in ’n bepaalde konteks – wees ten

[273] Dit is interessant dat O2 nie eerder van “New Seeland” gebruik gemaak het nie aangesien hierdie outeur deurlopend die taalreëls volg. Dit kan wees dat “#New zeeland” as ’n hutswoord ’n groter reikwydte aan die twiet gegee het.

einde kommunikasie te verseker) uit verskillende woorde. Hierdie woorde word egter nie net lukraak gekies nie, maar moet aan die basiese patroonmatige norm voldoen om sodoende semantiese betekenis oor te dra. Woorde word op grond van vorm en funksie in verskillende groepe gedeel.

Die data wat vir die korpusanalise gebruik is, is gemerk met betrekking tot woordsoorte (woordsoortetiketteerder[274]) om te kon bepaal of die sinskonstruksie van mikroblog-Afrikaans verskil van die konstruksie wat as die basiese sintaktiese beginsels van Afrikaans aanvaar word en ook om te kon bepaal hoe sintaksis in outeurskapidentifikasie figureer [275]. Wat wel van belang in die studie was, is die orde waarin hierdie verskillende woordgroepe veronderstel is om in ’n sin voor te kom.
ŉ Behoefte aan kommunikasie gee dus aanleiding daartoe dat taal so effektief moontlik gebruik word ten spyte van die beperkings van die medium. Die beperkings van die medium gee aanleiding tot kreatiewe aanpassings in die taal (betekenis word steeds oorgedra in die konteks van die kommunikasie) om kommunikasie moontlik te maak. Vanuit die bogenoemde blyk dit dus dat kommunikasie slegs kan plaasvind as betekenis deur middel van die taal oorgedra wordIn hierdie afdeling oor sintaksis is ondersoek ingestel na sinstipes, sinseinde, die gebruik van lidwoorde, die gebruik van woordsoorte en die invloed van Engels op die sinstruktuur.

7.3.1 Sinstipes

In die analise van sinstipes is die verskillende soort sinne ondersoek wat die deelnemers in hulle twiets gebruik. Ten einde die sinstipes te kon ondersoek, is daar na Ponelis (1979:375) se tradisionele onderskeid na sinstipes verwys. Volgens Ponelis (1979:375) is daar vier verskillende sinstipes, naamlik:

- **ŉ Stelsin**: Ponelis (1979:376) definieer ŉ stelsin as ŉ sin waarin ŉ stelling informeel verstaan word (deur ontvangers van die boodskap) as ŉ beskrywing van die stand van sake. Hierdie soort sinne dien “as ŉ beskrywing van die stand van sake, dit wil sê om inligting oor die stand van sake mee te deel,” (Ponelis, 1979:376). Hierdie sinne kan ook mededelings genoem word.
- **ŉ Vraagsin**: Volgens Ponelis (1979:378) beliggaam ŉ vraagsin “ŉ appèl om inligting op die aangesprokene.” Daar word onderskei tussen algemene en spesifieke vraagsinne. Algemene vraagsinne kan volgens Ponelis (1979:378) eenvoudig deur ŉ *ja* of *nee*

[274] Dit is belangrik om in ag te neem dat die woordsoortetiketteerder wat in die studie gebruik is volgens CTexT nie alle woordsoorte korrek etiketteer nie. Daar is ’n 85-90% suksesratio wanneer ’n normale, Standaardafrikaanse korpus geanaliseer word. Wanneer hierdie woordsoortetiketteerder op ’n korpus wat vanuit sosialemediaplatforms saamgestel is, geëtiketteer word, kan die suksesratio tot minder as 60% daal. Daar is besluit om in die geval van my studie met die fout saam te werk ten einde ondersoek te kon instel na die woordsoorte. Die feit dat die beskikbare woordsoortetiketteerder nie geskik is vir die analise van tekste vanaf sosialemedia nie, is ’n leemte wat sonder twyfel verder navorsing en ontwikkeling van die gepaste programmatuur vereis.

[275] Sien hoofstuk 5 (Metodologie).

beantwoord word. Spesifieke vraagsinne vra ŉ spesifieke vraag en om die vraag te antwoord, moet ontvangers op die vraagwoord let.

Vir die doel van hierdie studie is daar nie tussen die verskillende soort vraagsinne onderskei nie, daar is slegs bepaal of vraagsinne deur die onderskeie outeurs gebruik word.

- **ŉ Bevelsin**: Volgens Ponelis (1979:383) is bevele en versoeke ŉ omvangryke sinstipe met ŉ groot verskeidenheid vorms en betekenisse. Die doel van ŉ bevelsin is om aangesprokenes tot aksie of optrede oor te haal (Ponelis, 1979:383).
- **ŉ Uitroepsin**[276]: Ponelis (1979:382) definieer ŉ uitroepsin as ŉ stelling met ŉ sterk affek (verbasing, ingenomendheid, afkeur of blydskap). Uitroepsinne kan vraagvorme met "nie" insluit, uitroepe met ŉ naamwoordstruktuur en ŉ "wat" as ŉ voorbepaling.

Ten einde hierdie analise te kon uitvoer, is *ATLAS.ti* gebruik. Elke sin is gemerk met ŉ kode (SinS, SinB, SinV en SinU). Hierdie kodes kan soos volg geïnterpreteer word:

- SinS – Hierdie kode verwys na stelsinne. Alle stelsinne is met SinS gemerk.
- SinB – Hierdie kode verwys na bevelsinne wat in twiets gebruik word. Alle bevelsinne is met SinB gemerk.
- SinV – Hierdie kode verwys na vraagsinne. Alle vraagsinne is met SinV gemerk.
- SinU – Hierdie kode verwys na uitroepsinne. Alle vraagsinne is met SinU gemerk.

Nadat al die sinne in *ATLAS.ti* gekodeer is, is die resultate as ŉ Microsoft Excel-dokument gestoor. Hierdie resultate word in Tabel 7.9 weergegee:

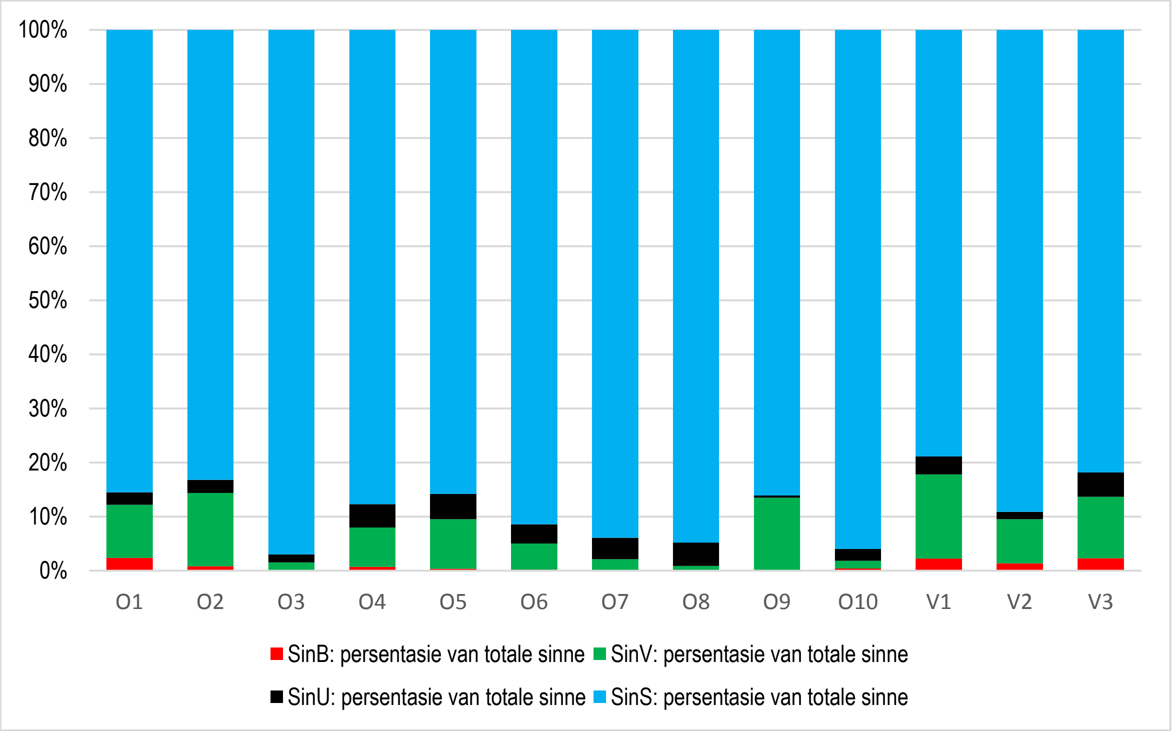
Tabel 7.9 Frekwensie van die voorkoms van die sinstipes

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	SinB	SinB (genormaliseer tot 1 000 sinne)	SinV	SinV (genormaliseer tot 1 000 sinne)	SinU	SinU (genormaliseer tot 1 000 sinne)	SinS	SinS (genormaliseer tot 1 000 sinne)
O1	669	944	8 942	22	23,31	93	98,52	22	23,31	807	854,87
O2	589	1 577	12 259	13	8,24	214	135,70	38	24,10	1 312	831,96
O3	206	266	3 379	0	0,00	4	15,04	4	15,04	258	969,92
O4	484	978	7 110	7	7,16	71	72,60	42	42,94	858	877,30
O5	333	556	4 579	2	3,60	51	91,73	26	46,76	477	857,91
O6	768	1 657	16 340	3	1,81	80	48,28	59	35,61	1 515	914,30
O7	457	790	5 655	1	1,27	16	20,25	31	39,24	742	939,24
O8	308	443	3 690	0	0,00	4	9,03	19	42,89	420	948,08
O9	275	467	4 918	0	0,00	63	134,90	2	4,28	402	860,81
O10	499	1 190	7 744	5	4,20	17	14,29	26	21,85	1 142	959,66
V1	42	90	844	2	22,22	14	155,56	3	33,33	71	788,89
V2	96	147	1409	2	13,61	12	81,63	2	13,61	131	891,16
V3	32	44	412	1	22,73	5	113,64	2	45,45	36	818,18

[276] Daar word ook na werkwoordlose uitroepe ondersoek ingestel, en daarom kan die uitroepsinne liewer as ortografiese eenhede beskou word.

Tabel 7.9 toon die aantal keer aan wat elk van die onderskeie outeurs die verskillende soort sinne gebruik en die frekwensie van die voorkoms van elk van die sinstipes, genormaliseer tot 1 000 sinne. By nadere ondersoek blyk dit dat die outeurs op Twitter meestal van stelsinne gebruik maak. Vraagsinne is ook gewild, gevolg deur bevelsinne en uitroepsinne. Die analise met betrekking tot die interne vergelyking van die onderskeie outeurs se korpora het aangedui dat die outeurs konsekwent van ongeveer dieselfde aantal stel-, uitroep-, vraag- en bevelsinne gebruik maak[277].

Daar is definitiewe verskille in die gebruik van die verskillende sinstipes tussen die verskillende outeurs. Figuur 7.1 dui die verhouding tussen die verskillende sinstipes aan:



Figuur 7.1 Die verhouding tussen die gebruik van die verskillend sinstipes deur die onderskeie outeurs

Figuur 7.1 dui die verhouding aan tussen die gebruik van die verskillende sinstipes deur die onderskeie outeurs in die korpora. Alhoewel dit duidelik is dat elk van die outeurs op ŉ unieke manier voorkeur gee aan sinstipes, is daar geen twyfel dat stelsinne die algemeenste sinstipe is wat op Twitter voorkom nie.

Stelsinne is die algemeenste sinstipe wat in die gespesialiseerde korpus gebruik word: O3 (96,99%[278]) en O10 (95,97%). Die laagste persentasie stelsinne wat in ŉ korpus voorkom, is in die korpus van O2 (83,20%). Selfs die laagste persentasie van die voorkoms van die gebruik

[277] Sien Bylaag F vir die interne vergelyking van die gespesialiseerde korpus.

[278] Ten einde die persentasie te kon bereken, is die hoeveelheid stelsinne wat gebruik is deur die totale hoeveelheid sinne wat in die korpus aangetref is, gedeel en met 100 vermenigvuldig. Hierdie formule is op elk van die sinstipes van elk van die outeurs uitgevoer.

van stelsinne is steeds meer as 50%. Daar kan vervolgens nie net aangevoer word dat outeurs van twiets na gelang van die aantal stelsinne wat in die korpus voorkom, geïdentifiseer kan word nie, maar ook dat Afrikaans soos dit op Twitter gebruik word ’n sterk aanwesigheid van stelsinne toon. Die ooreenkomste ten opsigte van die persentasie stelsinne wat in die verdagte tekste en die gespesialiseerde korpus gebruik word, is soos volg: V1 (78, 89%) stem ooreen met O2 (83,20%); V2 (89,12%) en O6 (91,43%) toon ooreenkomste terwyl V3 (73%) en O1 (85,49%) beduidend ooreenstem na gelang van die persentasie stelsinne wat in die korpus gebruik word.

Dit kom voor asof vraagsinne die tweede gewildste sinstipe op Twitter is alhoewel hierdie sinstipe nie meer as 13,57% (O2) van die sinstipes in die korpus uitmaak nie. Met die uitsondering van O10, blyk dit dat ’n hoër persentasie stelsinne per korpus aanleiding gee tot ’n laer persentasie vraagsinne. Hierdie tendens kan egter nie toegeskryf word aan die kategorieë waarin die onderskeie outeurs geplaas is nie of die rede waarom hulle Twitter gebruik nie, maar kan eerder as ’n persoonlike voorkeur beskou word. In die korpora van V1 is 15,56% van die totale sinne vraagsinne – ’n bevinding wat aanleiding daartoe gee dat O2 (13,5%) as die moontlike outeur van die eerste verdagte teks beskou kan word. 8,12% van die totale sinne in die korpus van V2 is vraagsinne, ’n persentasie wat vergelyk kan word met die voorkoms van vraagsinne in O4 (7,26%). In die derde verdagte teks (V3) is 11,36% van die totale sinne vraagsinne, ’n persentasie wat ooreenstem met die persentasie vraagsinne in O1 (9,9%).

Die hoogste persentasie uitroepsinne is 4,29% en word in die korpus van O8 aangetref (sien Tabel 7.10). Die laagste persentasie word in die korpus van O9 aangetref (0,43%). Wanneer die genormaliseerde waarde van die voorkoms van uitroepsinne in die verdagte tekste (sien Tabel 7.10) vergelyk word met die genormaliseerde waarde van die voorkoms van uitroepsinne in die gespesialiseerde korpus, blyk dit dat V1 (3,33%) en O6 (3,56%), V2 (1,36%) en O3 (1,56%) en V3 (4,55%) en O5 (4,68%) ooreenstem met betrekking tot die gebruik van uitroepsinne na gelang van die persentasies sinne vanuit die korpora wat uitroepsinne bevat.

Daar is bevind dat bevelsinne die ongewildste sinstipe is. Die grootste persentasie bevelsinne is in die korpus van O1 (2,33%) aangetref. Die persentasie sinne vanuit die korpora van die verdagte tekste wat bevelsinne is, is V1 – 2%, V2 – 1% en V3 – 2%. V1 en V2 stem ooreen met O1 (2,33%) terwyl daar nie ’n outeur in die gespesialiseerde korpus is wat met die persentasie bevelsinne wat in V3 aangetref word, ooreenstem nie.

Die ondersoek na die gebruik van die verskillende sinstipes in die gespesialiseerde korpus en die verdagte tekste toon aan dat daar wel ooreenkomste waargeneem kan word wanneer die

frekwensie gebruik word om die onderskeie sinne te ondersoek. Vervolgens is daar na stylkenmerke van die verskillende sinstipes ondersoek ingestel.

7.3.1.1 Stelsinne

Stelsinne verwys na sinne waarin die stand van sake uitgebeeld word. In hoofstuk 2[279] word daar verwys na Twitter se definisie wat in 2011 verander is na “What’s happening” en die beginsels dat Twitter beskou word as “the best way to discover what’s new in your world.” Hierdie twee aspekte van Twitter daag gebruikers dus uit om die vraag “what’s happening” met ŉ eenvoudige stelsin te beantwoord. In afdeling 7.3.1.1.1 word die gebruik van stelsinne deur die outeurs vanuit die gespesialiseerde korpus voorgehou en afdeling 7.3.1.1.2 fokus op die vergelyking met die verdagte tekste se gebruik van stelsinne.

7.3.1.1.1 Gespesialiseerde korpus

Die manier waarop die onderskeie outeurs stelsinne gebruik, pas aan by die oorheersende tema van elk van die gebruikers se twiets − die beantwoording van die vraag “what’s happening?”. Voorbeeld 132 tot 141 beeld die stelsinne uit wat deur die outeurs in die gespesialiseerde korpus gebruik word:

Voorbeeld 132: **O1** – My facial expression as iemand 'n geweer teen my kop hou: 🙄🔫

Voorbeeld 133: **O2** – Toe ek in 1982 verslag doen uit Pretoria vir Die #Kerkbode moes ek kopie per oornagpos stuur sodat dit vlg dag in Kaapstad geset kon word.

Voorbeeld 134: **O3** – @GEBRUIKER is by die perskonferensie waar ministers Nkoana-Mashabane, Chabane en Peters die media sal toespreek oor begrafnisreelings

Voorbeeld 135: **O4** – Eks bitter bly die Bokke kon die Sywurms wen.... As ons weer teen hulle moes verloor het ek my belt geëet sonder aromat

Voorbeeld 136: **O5** – Ek verstaan hoekom hulle dit "hartseer" noem. Want dit voel letterlik of my hart pyn.

Voorbeeld 137: **O6** – Voor jy 'n klip optel en gooi, onthou: Elke priester het 'n verlede en elke sondaar 'n toekoms...

Voorbeeld 138: **O7** – Twee lekker aande in @GEBRUIKER. Nog agt optredes oor.

Voorbeeld 139: **O8** – Baie dankie. Selfde daar

Voorbeeld 140: **O9** – Ek het alles uit my sak gegooi om my lighter te soek, en hierdie sleuteltjie gekry. Ek het hom nog nooit gesien nie.

Voorbeeld 141: **O10** – Die @GEBRUIKER se @GEBRUIKER het sopas sy @GEBRUIKER debuut gemaak. #LIOvSHA @GEBRUIKER

---

[279] Afdeling 2.6.2.2 (Twitter).

Voorbeeld 132 tot 141 toon ’n enkele voorbeeld aan van ’n stelsin vanuit die korpora van elk van die tien outeurs vanuit die gespesialiseerde korpus. Die outeurs gebruik stelsinne ten einde ’n bepaalde boodskap aan ontvangers (volgelinge) oor te dra.

Die manier waarop die onderskeie outeurs stelsinne gebruik, kan nie as kenmerkend beskou word nie aangesien daar nie ’n definitiewe patroonmatigheid teenwoordig is nie. Die persentasie wat die stelsinne van die korpus se totale aantal sinne uitmaak (sien ook Tabel 7.10 en Figuur 7.1) asook die verhouding tussen die verskillende sinstipes kan wel as ’n aanduiding van outeurskap beskou word.

7.3.1.1.2 Verdagte tekste

Aangesien daar in die voorbeelde vanuit die gespesialiseerde korpus (voorbeeld 132 tot 141) bepaal is dat daar nie ’n beduidende patroonmatigheid in die gebruik van die stelsinne te bespeur is nie − buiten dat elk van die outeurs soorte sinne gebruik om ’n definitiewe boodskap aan lesers oor te dra – word daar nie voorbeelde vanuit die verdagte tekste voorgehou nie.

7.3.1.2 Vraagsinne

Volgens Ponelis (1979:378) beliggaam ’n vraagsin “’n appèl om inligting op die aangesprokene.” In afdeling 7.3.1 is daar bepaal dat vraagsinne die tweede gewildste sinstipe op Twitter is. Wanneer die verdagte tekste ondersoek is, ondersteun die bevindinge die bogenoemde stelling. Alhoewel daar ’n aansienlike verskil[280] tussen die gebruik van stelsinne en vraagsinne voorkom, is die vraagsin dus ook in die verdagte tekste die tweede gewildste sinstipe.

7.3.1.2.1 Gespesialiseerde korpus

Die ondersoek na vraagsinne toon aan dat elk van die outeurs wat in hierdie studie gebruik is vraagsinne vir ’n ander doel gebruik, hetsy om kommentaar te lewer, retoriese vrae te stel, ’n opinie te lewer of ’n spesifieke gebruiker se opinie oor ’n bepaalde kwessie in te win. Voorbeeld 142 tot 161 beeld die vraagsinne uit soos dit in die gespesialiseerde korpus aangetref word:

Voorbeeld 142: **O1** – **Al ooit** 'n regte besluit gemaak en dan voel jy so accomplished met jou lewe? Nah, ek ook nie.

Voorbeeld 143: **O1** – **Gedink** jy's klaar gemerk? Sorry, dink weer. Hermerk!

Voorbeeld 144: **O2** – Einde v nuusankers aangekondig in briljante Time-art deur James Poniewozik. By ons afgesluit toe Rian Cruy uitgetree het. **Maar** mis ons dit?

[280] In die geval van V1 is 78% van die sinne stelsinne terwyl slegs 16% van die sinne vraagsinne was.

Voorbeeld 145: **O2** – "@GEBRUIKER: Wearable tech: the small British companies taking on Samsung and Google http://fw.to/wo03pof " Fantasties! **Einde van selfone**?

Voorbeeld 146: **O3** – Waar is Blade? Hier is Blade.

Voorbeeld 147: **O4** – Boys, #varkhart?

Voorbeeld 148: **O4 – David van Rooyen????? David van Rooyen.....?????** Wie de fok is dit??? Het hy nie dalk vroeer jare vir die Griffons gespeel nie?

Voorbeeld 149: **O5** – **Hoe** gebeur dit dat jy en iemand waarmee jy elke dag gepraat het en al jou geheime ken net op hou praat??

Voorbeeld 150: **O5** – **Wat** het ek die naweek geleer? Kamp is nie vir my bedoel nie.

Voorbeeld 151: **O6** – Ek doen 'n 1958 'shoot' vanaand. **Kry** mens nog Brylcreem op die mark?

Voorbeeld 152: **O6** – Langnaweek!? **Wat** gaan jy hierdie naweek doen?

Voorbeeld 153: **O7** – Ek het my hart verloor op haar. **Wonder** of sy single is?

Voorbeeld 154: **O8** – **Wat** van suiker bo op die sjokoladesous, bo op die roomys? #susterskind #godchild #icecream #kids…https://www.instagram

Voorbeeld 155: **O8** – **Hospitaal?** Sterkte daar

Voorbeeld 156: **O8** – Maak vuur op die tradisionele manier met 'n Tonteldoos. **Het** jy dit al gebruik? #fire zarsg https://www.instagram.com

Voorbeeld 157: **O9** – Môre almal, **but can you not?**

Voorbeeld 158: **O9** – Never in my fucking life could I ever have dreamt up the shit storm that is today. **What the actual fuck**, Oom Satan? Ek is jammer, foktog.

Voorbeeld 159: **O10** – Peyper het sy hande vanaand behoorlik vol. Goeie, fisieke rugby wat hier plaasvind. #SHAvCRU **Wat** sê @GEBRUIKER?

Voorbeeld 160: **O10** – Lekker man, **sien** ek julle by seniorsrugby? 😬 https://twitter.com/CaputKoshuis/status/697120487864987648 …

Voorbeeld 161: **O10** – Ek gaan verseker daar wees **&** jy? #ToekaMakiti

O1 stel kenmerkend retoriese en sarkastiese vrae, en verskaf dan ook ŉ antwoord in die twiet. Die vraagsin word ook telkens aan die begin van ŉ twiet gebruik, gevolg deur ŉ stelsin. O2 gebruik frases of woorde om die twiet aan te vul en die bedoelde boodskap te bevraagteken. Deur hierdie vraagsinne (hetsy ŉ frase of ŉ woord is) word lesers van die twiet gevra om na te dink oor stellings en om deel te neem aan die gesprek. Daar kan dus aangevoer word dat O2 die vraagsinne gebruik ten einde reaksie uit te lok. Daar moet ook gelet word op die feit dat hierdie woorde konteksgebonde is, en afhanklik is van wat op die oomblik van die twiet as nuuswaardig beskou word. O2 plaas telkens eers ŉ stelsin met ŉ vraagsin aan die einde van ŉ twiet.

O3 se gebruik van vraagsinne kan nie as kenmerkend beskou word nie. Die outeur gebruik slegs vier vraagsinne (1,5% van O3 se totale sinne is vraagsinne – Tabel 7.10) en drie van die vraagsinne is aanhalings. Die enkele vraagsin wat deur O3 self geformuleer is, bevat ook ’n retoriese vraag wat deur die outeur self beantwoord word. Wanneer die outeur van ’n verdagte teks geïdentifiseer moet word wat baie min vraagsinne bevat (minder as 2% van die totale aantal sinne), kan O3 gevolglik as outeur van die teks oorweeg word.

O4 maak gebruik van vraagsinne om sosiale kommentaar te lewer. Die gebruik van die woord *varkhart* (sien voorbeeld 147) is eie aan O4 en wanneer die woord in die verwysingskorpus gebruik word, is dit telkens met verwysing na O4. *Varkhart* word egter slegs twee keer in ’n vraagsin gebruik (een keer as ’n hutswoord), maar word gereeld in die res van O4 se korpus (veral in stelsinne) aangetref. Daarom kan *varkhart* as ’n aanduiding van die outeurskap van O4 beskou word. In die geval van voorbeeld 148 gebruik O4 ’n oormaat leestekens, maar die outeur se vraag is ook tydgebonde en lewer sosiale kommentaar op onlangse nuuswaardige gebeure. Anders as in die geval van O2, kan O4 se kommentaar amper as ’n parodie beskou word. O4 is verder ook die enigste outeur vanuit die gespesialiseerde korpus wat drie vraagsinne in ’n enkele twiet gebruik (voorbeeld 148).

O5 se gebruik van vraagsinne stem baie ooreen met dié van O1, maar O5 beantwoord nie elke vraagsin self nie, en wil reaksie by volgelinge uitlok (sien voorbeeld 149). Vraagsinne word ook oorwegend aan die begin van ’n twiet geplaas en indien dit nie die enigste sin in ’n twiet is nie volg die outeur dit met ’n stelsin op. Beide O6 en O7 gebruik vraagsinne om volgelinge by ’n gesprek te betrek en beide die outeurs plaas nie vraagsinne aan die begin van ’n twiet nie. O7 se vraagsinne word deur stelsinne voorafgegaan.

O8 se vraagsinne betrek lesers van ’n twiet by die aksie wat deur O8 uitgevoer is en O8 vra lesers of hulle al só iets probeer het of in die geval van voorbeeld 155 is dit ’n vraag na ’n persoon se welstand en ’n wens volg die vraag. Met die uitsondering van voorbeeld 154, is dit kenmerkend dat O8 deurlopend van hutswoorde na vraagsinne gebruik maak om sodoende te verseker dat daar geen onduidelikheid oor die vraag wat gestel is, heers nie (sien voorbeeld 154 en 156).

O9 se vraagsinne sluit oorwegend Engelse woorde en afkortings in (sien voorbeeld 157 en 158). O9 se gebruik van kodewisseling[281] in die vraagsinne kan as ’n kenmerkende eienskap van hierdie outeur se taal op Twitter beskou word, veral as die gebruik van kraswoorde[282] in die sinne ondersoek word. Verder plaas O9 nooit die vraagsin aan die begin van ’n twiet nie, en die vraagsin word deur ’n stelsin voorafgegaan.

[281] Sien afdeling 7.5.1 (Kodewisseling).
[282] Sien afdeling 7.7 (Kraswoorde).

O10 gebruik vraagsinne om die opinie van spesifieke gebruikers te vra (voorbeeld 158 en 159). Voorbeeld 160 toon aan hoe O10 die ampersandteken gebruik om 'n vraag te vra, maar hierdie gebruik van die ampersandteken kom slegs een keer in die korpus van O10 voor. Dit is belangrik om daarop te let dat die twiet nie die 140 karakterperk oorskry nie, en dus kan daar nie aangevoer word dat die ampersandteken ter wille van karakterbesparing gebruik word nie. Daar is nie 'n voorbeeld in die verwysingskorpus waar die ampersandteken soos in voorbeeld 160 van O10 gebruik word nie[283]. Daar kan dus aangevoer word dat wanneer O10 'n verdagte outeur is, en die ampersandteken word in 'n vraagsin gebruik, O10 as die moontlike outeur beskou kan word.

Die ondersoek na die gebruik van vraagsinne in die gespesialiseerde korpus toon aan dat vraagsinne gebruik word om kommentaar te lewer, retoriese vrae te stel, 'n opinie te lewer of 'n spesifieke gebruiker se opinie oor 'n bepaalde kwessie in te win. Die meeste vraagsinne word voorafgegaan of gevolg deur stelsinne. Daar is geen vraagsinne in die gespesialiseerde korpus wat saam met uitroep- of bevelsinne gebruik word nie. Vervolgens is die drie verdagte tekste geanaliseer om moontlike ooreenkomste voor te hou met die gebruik van vraagsinne in die gespesialiseerde korpus.

7.3.1.2.2 Verdagte tekste

Ten einde te kon bepaal of daar ooreenkomste tussen die vraagsinne in die verdagte tekste en die gespesialiseerde korpus voorgekom het, word daar ses voorbeelde (voorbeeld 162 tot 165) van vraagsinne vanuit die verdagte tekste voorgehou alvorens die vraagsinne bespreek word:

Voorbeeld 162: **V1** – Dalk is ek ouskool en sien joernaliste dit vandag anders. Tog, volle aanhaalkoppe is selde presies akkuraat. **So, is naastenby nou oukei?**

Voorbeeld 163: **V1** – Tekens van tye: Beeld se hoofopskrifte op pp. 1, 2, 3, 4, 6 vandag in aanhaaltekens. O, ook p.10 en 11. **Is dit luiheid of soeke na emosie?**

Voorbeeld 164: **V2** – **Wtf Apple?** iOS 10 is amper so shit soos Trump.

Voorbeeld 165: **V3** – **Hoekom** haar seer maak as sy nog nét goed was vir **jou?**

Voorbeeld 162 en 163 dui aan hoe V1 vraagsinne gebruik. Vanuit die ondersoek na V1 se vraagsinne kan die afleiding gemaak word dat V1 deurlopend 'n stelsin in 'n twiet gebruik wat onmiddellik met 'n vraagsin gevolg word – 'n kenmerk wat ook in die korpus van O2 en O7 waargeneem is. V1 maak van frases of woorde gebruik om 'n twiet aan te vul en die bedoelde boodskap te bevraagteken. Deur hierdie vraagsinne ('n frase of 'n woord) word die lesers van 'n twiet gevra om na te dink oor die stelling en deel te neem aan die gesprek. V1 gebruik net soos O2 vraagsinne om reaksie uit te lok.

[283] Ten tyde van die navorsing (Julie 2017) was daar nie 'n voorbeeld op Twitter waar 'n gebruiker die ampersandteken as vraagmerker gebruik nie.

V2 plaas die vraagsin eerste (voorbeeld 164), gevolg deur ŉ stelsin. Die twiet kan as ŉ retoriese en sarkastiese vraag beskou word, gevolg deur ŉ antwoord – ŉ kenmerk wat ook in die korpus van O1 aangetref word. Alhoewel V2 ŉ afkorting en kodewisseling in die sin gebruik – soortgelyk aan O9 – kan O9 nie as die outeur van die teks beskou word na gelang van die voorbeeld nie aangesien O9 die vraagsin nooit eerste in ŉ twiet plaas nie.

V3 se gebruik van die vraagsin stem ooreen met O5 aangesien O5 nie elke vraagsin self beantwoord nie en reaksie van die volgelinge uitlok (sien voorbeeld 149). Die vraagsin word ook oorwegend aan die begin van die twiet geplaas en indien dit nie die enigste sin in ŉ twiet is nie, volg die outeur dit met ŉ stelsin op.

Die manier waarop vrae dus deur die onderskeie verdagte outeurs gestel word, toon ooreenkomste met outeurs vanuit die gespesialiseerde korpus, en kan gevolglik as ŉ middel tot outeurskapidentifikasie beskou word.

7.3.1.3 Uitroepsinne

Die derde gewildste sinstipe wat in die ondersoek in die gespesialiseerde korpus geïdentifiseer is, is die uitroepsin. ŉ Uitroepsin word beskou as ŉ sin waarin daar sterk verbasing, ingenomendheid, afkeer of blydskap uitgedruk word. Weer eens is daar geen patroonmatigheid bespeur om die keuse van hierdie sinstipe te regverdig nie.

7.3.1.3.1 Gespesialiseerde korpus

Voorbeelde van uitroepsinne vanuit die gespesialiseerde korpus:

Voorbeeld 166: **O1** – **OHMYWORD!** Dis die kuiken-ding van vooraf! @GEBRUIKER 😂😂😂😂😂😂😂😂😂😂😂😂

Voorbeeld 167: **O2** – TV-advertensies het koerante baie seergemaak. Vandaar toestemming aan persmaatskappye om M-Net in 1985 te begin. **Winste moes pers help!**

Voorbeeld 168: **O3** – **Hier gaan ons!** #DAdebat #Insig

Voorbeeld 169: **O4** – **Leeus Leeus Leeus!!!** Mens kan sien as 'n span vir mekaar speel!! well done @LionsRugbyUnion well done coach Akkers

Voorbeeld 170: **O5** – Ek het geweet dit gaan 'n ongelooflik moeilike WEEK wees... **Oops** !

Voorbeeld 171: **O6** – My vriend in Frankryk se seuntjie, Anton is nou 2 en half jaar oud. **Sy ouers stel hom bloot aan als!** Hy praat... http://fb.me

Voorbeeld 172: **O7** – **So!** Hoeveel brandewyn gooi mens by rys voor dit gaar is

Voorbeeld 173: **O9** – **JA!** <3 @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER @GEBRUIKER

Voorbeeld 174: **O10** – Ditto! ✌🏼

Vanuit voorbeeld 166 tot 173 blyk dit dat elk van die gebruikers (behalwe O9) wel van uitroepsinne gebruik maak, maar geen van die voorbeelde kan as uniek tot ŉ bepaalde outeur beskou word nie. Die frekwensie van die gebruik van bevelsinne (of die persentasie wat die sinstipe uitmaak) wanneer al die sinstipes ondersoek is, kan as ŉ moontlike aanduiding van outeurskap beskou word.

Dit blyk verder dat uitroepsinne gereeld met woorde (tussenwerpsels) soos *oi, wow, ai, oops, hehe* of *yae* vergesel word. Dít kan ook as ŉ kenmerk van Afrikaans beskou word soos die taal op Twitter gebruik word.

7.3.1.3.2 Verdagte tekste

Voorbeeld 184 tot 186 van die gebruik van uitroepsinne in die verdagte tekste:

Voorbeeld 184: **V1** – SAUK moet asb kennis neem dat daar groot verskil is tussen persentasie en persentasiepunte. **Kry dit reg by verkiesingsuitslae!**

Voorbeeld 185: **V2** – DIS 'N QWERTY KEYPAD. **TIK JOU WOORDE VOLLEDIG UIT!**

Voorbeeld 186: **V3** – Jannie jammergat? **TOTSIENS!!**

Voorbeeld 184, 185 en 186 toon aan hoe uitroepsinne in die verdagte tekste gebruik word. Wanneer die voorbeelde vergelyk word met die voorbeelde in afdeling 7.3.1.3.1 (uitroepsinne), is dit duidelik dat outeurs op Twitter uitroepsinne gebruik om verskillende soort boodskappe oor te dra. V1 laat lidwoorde uit in uitroepsinne. In die gespesialiseerde korpus is dit 'n kenmerk wat in die korpus van O2 (sien voorbeeld 167 en 184) waargeneem word. V2 is geneig om hele uitroepsinne in hoofletters te tik − 'n kenmerk wat ook in die korpus van O1 (voorbeeld 166 en 185) gevind word. V3 gebruik 'n enkele woord in hoofletters om die kern van die uitroep te beklemtoon − 'n kenmerk wat ook in die uitroepsinne van O5 (voorbeeld 170 en 186) gevind word.

Die ondersoek na die gebruik van uitroepsinne kan dus 'n waardevolle bydra lewer tot die identifisering van die outeurs van die drie verdagte tekste. Na gelang van die ondersoek kan

daar aangevoer word dat O2 as die moontlike outeur van V1 beskou kan word, O1 kan die outeur van V2 wees en O5 kan moontlik die outeur van V3 wees.

7.3.1.4 Bevelsinne

Die hoofdoel van bevelsinne is om aangesprokenes tot aksie of optrede oor te haal. Wanneer die bevelsinne van die onderskeie outeurs ondersoek is, blyk dit dat die bevel eerder as die woorde ŉ kenmerk van die outeurs se taalgebruik is.

7.3.1.4.1 Gespesialiseerde korpus

O1 het 22 keer van bevelsinne gebruik gemaak, en gebruik hierdie soort sin om te motiveer, as ŉ bewys van liefde, as gelukwensing of om ontvangers van die boodskap te kalmeer. O2 se korpus beskik oor 13 bevelsinne en die outeur beveel telkens lesers om iets spesifiek te doen. O4 se bevelsinne dra ŉ wens oor en daar word verwys na eerlikheid, motivering en ŉ opdrag (jy mag nie ____ nie). O5 gebruik slegs bevelsinne om ontvangers van die boodskap of bedoelde ontvangers te verwens. O6 gebruik bevelsinne om ŉ opdrag oor te dra terwyl O7 hierdie sinstipe gebruik om individue te oorreed om ŉ opdrag uit te voer. O10 gebruik bevelsinne as aanmoediging. O3, O8 en O9 maak glad nie van bevelsinne gebruik nie (sien voorbeeld 169 tot 178):

Voorbeeld 169: **O1** – Chill dit uit.

Voorbeeld 170: **O2** – Gaan kyk!

Voorbeeld 171: **O4** – Wees eerlik!!

Voorbeeld 172: **O4** – Wag eers!!!

Voorbeeld 173: **O5** – Gaan vry ŉ flippen pofadder.

Voorbeeld 174: **O6** – Kontak hom!

Voorbeeld 175: **O7** − Skryf hom net uit die hart uit Bra. Elke man se hart klink anders as die een langs jou

Voorbeeld 176: **O7** − Hy beter mooi kyk

Voorbeeld 177: **O10** – Jy bly stil.

Voorbeeld 178: **O10** − Kom ons hou duimvas.

Voorbeeld 169 tot 178 toon aan hoe die outeurs (vanuit die gespesialiseerde korpus en waar daar wel bevelsinne gebruik is) van bevelsinne gebruik maak. Die outeurs gebruik bevelsinne om ŉ spesifieke emosie (bemoediging, motivering, verwensing, oorreding) oor te dra aan lesers, maar behalwe vir die gebruik van bevelsinne is daar nie ŉ definitiewe patroonmatigheid by die outeurs se gebruik van bevelsinne geïdentifiseer nie. Daar kan wel aangevoer word dat

bevelsinne oor die algemeen korter sinne bevat en dat twiets dan meestal net uit ŉ enkele bevel bestaan.

7.3.1.4.2 Verdagte tekste

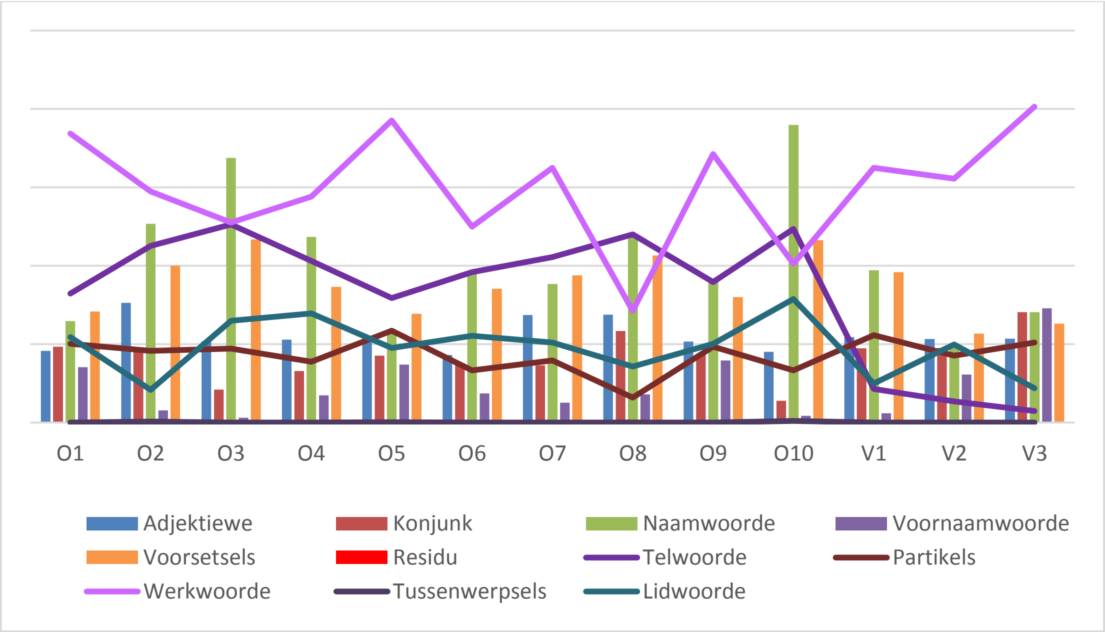
Soos reeds in afdeling 7.3.1.4.1 genoem, is bevelsinne nie net ongewild op Twitter nie, maar hierdie soort sinne het ook nie enige insiggewende inligting oor een van die outeurs openbaar nie. Indien daar ’n teks is waar die persentasie bevelsinne hoër is as byvoorbeeld die vraagsinne, kan dit ’n unieke kenmerk van ŉ spesifieke outeur wees, maar aangesien die bevindinge oor die gebruik van bevelsinne in die korpora wat ondersoek is, ooreenstem met die bevindinge vanuit die gespesialiseerde korpus, kan daar aangevoer word dat bevelsinne wat in Afrikaans op Twitter voorkom, gebruik word om ŉ spesifieke emosie (bemoediging, motivering, verwensing, oorreding) aan lesers oor te dra. Behalwe vir die gebruik van bevelsinne, kan daar nie ŉ definitiewe patroonmatigheid met betrekking to die gebruik van bevelsinne deur die outeurs geïdentifiseer word nie.

7.3.2 Woordsoorte

Die ondersoek na die woordsoortgebruik van die onderskeie outeurs is vlugtig uitgevoer om te bepaal watter woordsoorte deur die onderskeie outeurs op Twitter gebruik word[284]. Ten einde die analise te kon uitvoer, is die korpora in die rekenaarprogram NCHLT-Tagger geannoteer. Die verskillende woordsoorte is gemerk en die data is in ŉ Microsoft Excel-dokument gestoor en geanaliseer. Die verskillende woordsoortgroepe[285] (ten einde die analise te vergemaklik, is daar onderskeid getref tussen adjektiewe, konjunkte, naamwoorde, voornaamwoorde, residu, voorsetsels, telwoorde, partikels, werkwoorde, tussenwerpsels en punktuasie) is van mekaar geskei en die voorkoms van elk van die woordsoorte is genormaliseer tot 1 000 woorde.

Hierdie ondersoek is op die korpora van die outeurs van die gespesialiseerde korpus uitgevoer en ook op die drie verdagte tekste. Alvorens daar na die onderskeie woordsoorte in elk van die woordsoortgroepe ondersoek ingestel is, is die totaal van elk van die woordsoorte per outeur bereken (sien Figuur 7.2):

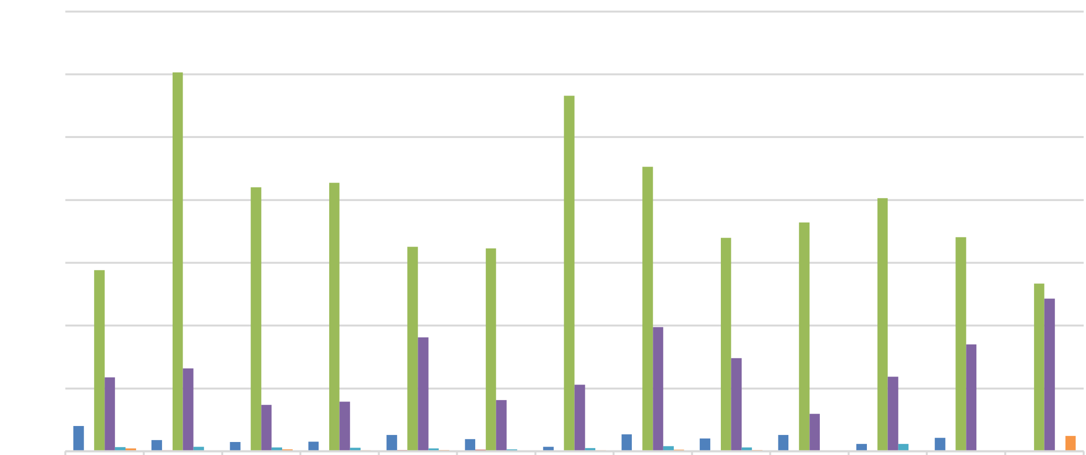
[284] Sien hoofstuk 8 (Verdere studie).
[285] Sien Bylaag F.



Figuur 7.2 Woordsoortgebruik

Figuur 7.2 toon die aantal woordsoorte aan wat elk van die outeurs in die gespesialiseerde korpus en die verdagte tekste gebruik. Alhoewel die gebruik van voorsetsels, residu, tussenwerpsels en telwoorde min of meer konstant is in die onderskeie korpora, is daar steeds ŉ verskil tussen die gebruik van die woorde in die onderskeie korpora waargeneem. Vervolgens is elk van die woordsoortgroepe analiseer om te kan bepaal of daar ŉ partoonmatigheid in die frekwensie van die woordsoorte te bespeur is. Ten einde hierdie analise uit te voer is daar van die rekenaarprogram NCHLT-Tagger gebruik gemaak. Die data is volgens die program geannoteer. Dit is belangrik om die leemte wat daar in die automatiese annoteerder voorkom te beklemtoon. In hierdie analise moes daar in ag geneem word dat die suksesratio van etikettering nie 100% is nie. Daarom het die moontlikheid bestaan dat van die woorde verkeerd geëtiketteer is. Die besluit is geneem om met die fout te werk ten einde te kon bepaal of díé soort analise wel ŉ bydra tot die identifisering van outeurs van mikroblogtekste kan lewer. Die etikettering van woordsoorte is ŉ duidelike leemte in die studie en die voorstel kan gemaak word dat ŉ etiketteerder ontwikkel word wat op sosialemediatekste fokus.

7.3.2.1 Adjektiewe



	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	V1	V2	V3
AOA	4.03	1.79	1.48	1.55	2.62	1.96	0.71	2.71	2.03	2.58	1.18	2.13	0.00
AOP	0.00	0.16	0.00	0.00	0.22	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASA	28.85	60.28	42.02	42.76	32.54	32.31	56.59	45.26	33.96	36.42	40.28	34.07	26.70
ASP	11.74	13.21	7.40	7.88	18.13	8.14	10.61	19.78	14.84	5.94	11.85	17.03	24.27
AVA	0.67	0.73	0.59	0.56	0.44	0.31	0.53	0.81	0.61	0.00	1.18	0.00	0.00
AVP	0.45	0.08	0.30	0.14	0.22	0.06	0.18	0.27	0.20	0.13	0.00	0.00	2.43

Figuur 7.3 Adjektiewe

Vanuit Figuur 7.3 blyk dit dat ASA[286] (adjektief: stellend/attributief) die algemeenste adjektief is wat in die gespesialiseerde korpus gebruik word, gevolg deur ASP (adjektief: stellend/predikatief – sien voorbeeld 179).

Die frekwensie (genormaliseer tot 1 000 woorde)[287] waarteen elk van die outeurs vanuit die gespesialiseerde korpus adjektiewe in hulle korpora gebruik (sien Figuur 7.3) dui aan dat O2 die meeste van ASA (60,28) gebruik maak terwyl O1 (28,85) die minste van ASA gebruik maak. O8 gebruik die meeste ASP (19,78 – sien voorbeeld 180) terwyl O10 slegs 5,94 (ASP) gebruik:

Voorbeeld 179: **ASA** – My honde tjank **elke** naweek as ek terug gaan universiteit toe en my hart breek **elke** keer. [O1]

Voorbeeld 180: **ASP** – Die gedagte van wat **moontlik** 'n straat weg van my plaasvind maak my siek. "Nearly 250 000 modern slaves in SA index" [O8]

Ten opsigte van die oorblywende adjektiewe, word daar aansienlik meer AOA's[288] deur O1 (4,03) gebruik terwyl O7 (0,71) weer in vergelyking met die ander outeurs, relatief min van AOA gebruik maak (die agt oorblywende outeurs gebruik onderskeidelik 1,55 en 2,62). Slegs O2, O5 en O6 gebruik AOP's in die gespesialiseerde korpus en die genormaliseerde waarde (tot 1 000

[286] Sien Bylaag F.
[287] Sein Bylaag F.
[288] Sien Bylaag F

woorde) wissel tussen 0,16, 0,22 en 0,24. Hierdie woordsoorte word dus relatief min deur outeurs op Twitter gebruik.

AVA's en AOP's word teen ŉ relatiewe konstante frekwensie deur die onderskeie outeurs in die gespesialiseerde korpus gebruik (sien Figuur 7.3). Daar kan dus aangevoer word dat tipe woordsoorte as ŉ redelike algemeen Afrikaanse gebruik van Twitter outeurs (soos waargeneem in die gespesialiseerde korpus) beskou kan word.

Figuur 7.3 toon aan dat daar aansienlik minder adjektiewe in die verdagte tekste voorkom as in die gespesialiseerde korpus. Daar word byvoorbeeld glad nie van AOA's gebruik gemaak nie terwyl daar slegs in V1 AVA's voorkom en V3 is die enigste outeur wat AVP's gebruik. In ooreenstemming met die gespesialiseerde korpus is ASA die algemeenste adjektief wat in die verdagte tekste aangetref word, gevolg deur ASP. Die frekwensie van gebruik, soos in die drie verdagte tekste gereflekteer, word in Tabel 7.11 voorgehou:

Tabel 7.10 Ooreenstemmende gebruik van adjektiewe

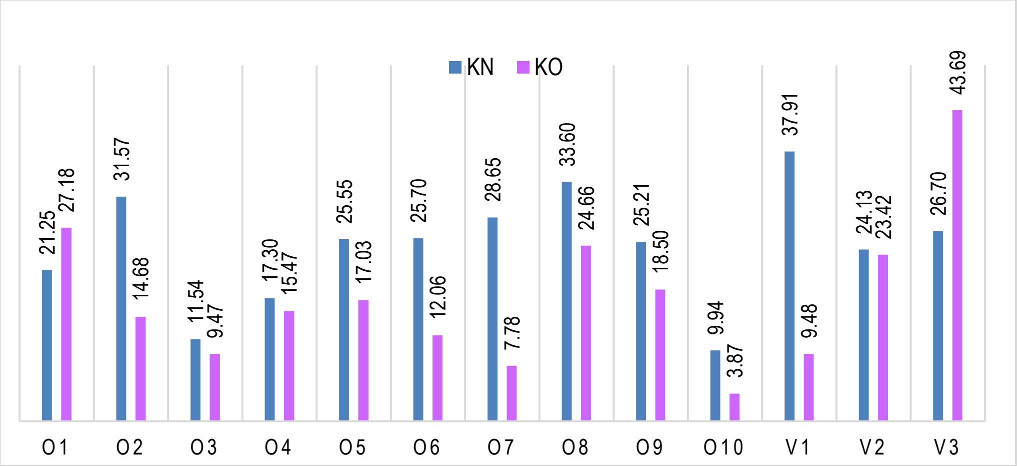
Korpus	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
AOA	1,18	O3 (1,48)	2,13	O6 (1,96)	0,00	O7 (0,71)
AOP	0,00	*	0,00	*	0,00	*
ASA	40,28	O3 (42,02)	34,07	O9 (33,96)	26,70	O1 (28,85)
ASP	11,85	O1 (11,74)	17,03	O5 (18,13)	24,27	O7 (19,78)
AVA	1,18	O7 (0,81)	0,00	O10 (0,00)	0,00	O10 (0,00)
AVP	0,00	O5 (0,06)	0,00	O5 (0,06)	2,43	O1 (0,45)

Tabel 7.10 dui die gebruik van adjektiewe in die verdagte tekste aan en die korpus vanuit die gespesialiseerde korpus wat ŉ ooreenkoms ten opsigte van die genormaliseerde waarde (tot 1 000 woorde) van die gebruik van die bepaalde woordsoortgroep aantoon. Alhoewel daar telkens een van die outeurs vanuit die gespesialiseerde korpus geïdentifiseer kan word wat ŉ ooreenkoms in die frekwensie (sien Tabel 7.10) van die bepaalde woordsoort in die verdagte tekste toon, is dit ook duidelik dat daar nie ŉ enkele outeur is wat meer as twee maal ooreenkomste met die verdagte tekste toon nie[289].

Daar kan wel aangevoer word dat die gebruik van adjektiewe kenmerkend van Afrikaans is soos dit op Twitter gebruik word en dat AOP's en AVA's nie as ŉ "gewilde" woordsoort op Twitter beskou kan word nie.

[289] In hierdie analise moes daar in ag geneem word dat die suksesratio van etikettering nie 100% reg is nie. Daarom het die moontlikheid bestaan dat van die woorde verkeerd geëtiketteer is. Die besluit is geneem om met die fout te werk ten einde te kon bepaal of díé soort analise wel ŉ bydra tot die identifisering van outeurs van mikroblogtekste kan lewer. Die etikettering van woordsoorte is egter ŉ duidelike leemte in die studie en die voorstel kan gemaak word dat ŉ etiketteerder ontwikkel word wat op sosialemedia tekste fokus.

7.3.2.2 Konjunkte



Figuur 7.4 Konjunkte

Figuur 7.4 toon die gebruik van konjunkte aan. Dit is duidelik dat daar in elk van die korpora van konjunkte (neweskikkend [KN] en onderskikkend [KO]) gebruik gemaak word. Die gebruik van konjunkte is dus kenmerkend van die Afrikaans wat op Twitter gebruik word (sien voorbeeld 181 tot 185):

Voorbeeld 181: **KN** – **En** dit was so [O7]

Voorbeeld 182: **KN** – **En** so sal ek **maar** die kaggel brand steek vir n voet vuur **want** die koue het kom nes skrop in elke deel van my binneste [O7]

Voorbeeld 183: **KN** – Haha. Nee jong, jy moet **maar** die punch vat 😆 [O8]

Voorbeeld 184: **KO** – "Liewe, Here. Seën hierdie studente, want hulle weet nie wat hulle doen nie. Gee hul wysheid, **sodat** hulle eendag... http://fb.me/4rpUsdU3Z [O6]

Voorbeeld 185: **KO** – Net **omdat** jy 'n karakter is, beteken dit nie jy het karakter nie. [O9]

O8 (33,60) gebruik die meeste KN (dit is ook opvallend dat O8 gereeld ŉ twiet met *en* begin – sien voorbeeld 181 en 182) terwyl O1 (27,18) se KO gebruik die hoogste is. O10 is die korpus waarin die minste konjunkte voorkom.

Soos sigbaar in Figuur 7.4 is V2 se gebruik van konjunkte relatief konstant – dit wil sê daar word ongeveer dieselfde aantal KN's en KO's deur die outeur gebruik. In beide V1 en V3 se tekste is daar ŉ aansienlike verskil in die frekwensie wat die gebruik van KO's en KN's betref. Dit is opvallend dat V1 voorkeur aan die gebruik van KN's gee. In teenstelling daarmee, gebruik

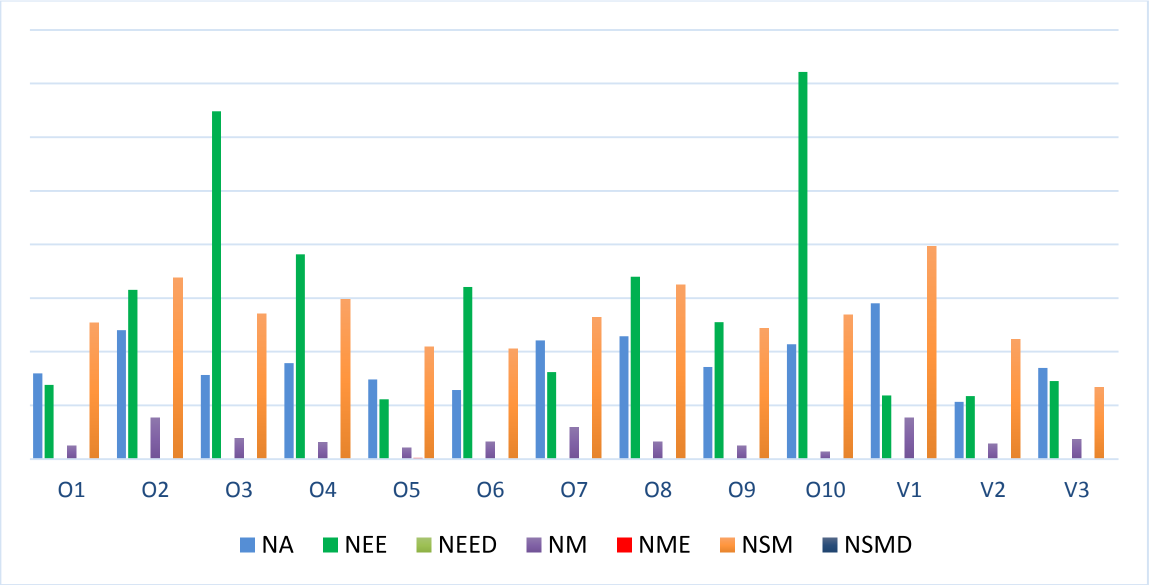
V3 meer KO's. Die ooreenkomste tussen die gebruik van konjunkte is geanaliseer en word in Tabel 7.11 voorgehou:

Tabel 7.11 Ooreenkomste met betrekking tot die gebruik van konjunkte

Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
KN	37,91	O8 (33,60)	24,13	O9 (25,21)	26,70	O6 (25,70)
KO	9,48	O3 (9,47)	23,42	O8 (24,66)	43,69	O1 (27,18)

Alhoewel daar ooreenkomste tussen die outeurs vanuit die onderskeie korpora waargeneem kan word, is dit telkens verskillende outeurs vanuit die gespesialiseerde korpus wat ŉ ooreenkoms met ŉ verdagte outeur toon en dus is die bevindinge nie konsekwent nie[290].

7.3.2.3 Naamwoorde



Figuur 7.5 Naamwoorde

Daar is tussen sewe verskillende naamwoorde[291] onderskei (sien Figuur 7.5). NM (massanaam) word die minste op Twitter gebruik terwyl NEE (naamwoord: eienaam/enkelvoud, basis) die meeste gebruik word.

O3 (129,62) en O10 (144,37) maak die meeste van NEE gebruik terwyl O5 (22,28) dit die minste gebruik. Wanneer die gebruik van NEE verder geanaliseer is, kan daar aangevoer word dat O5 se gebruik van die woordsoort NEE ooreenstem met die gemiddelde gebruik van NEE

[290] Die bevindinge kan weereens as ŉ eliminasie proses beskou word en daar kan aangevoer word dat outeurs wat nie ŉ ooreenkoms toon nie, nie meer as moontlike outeurs van die spesifieke verdagte teks beskou moet word nie.
[291] Sien Bylaag F.

mits die twee korpora (O3 en O10 wat van aansienlik meer NEE's in hulle korpora gebruik maak) nie by die berekening van die gemiddeld ingesluit word nie[292].

NSM (naamwoord: soortnaam/basis) is die tweede gewildste naamwoord op Twitter en kom die meeste in die korpus van O2 (67,62) voor. Dit is opvallend dat die woordsoort NSM relatief konstant teen dieselfde frekwensie deur die onderskeie outeurs in die gespesialiseerde korpus gebruik word[293]. Daarom kan die aanname gemaak word dat die woordsoort NSM as betreklik algemeen in Afrikaans soos dit op Twitter gebruik word, beskou kan word. Met betrekking tot die oorblywende naamwoorde, is dit slegs in NM wat O2 (15,42) en O7 (11,85) se gebruik van die woordsoort ŉ aansienlike verskil toon wanneer dit met die ander agt korpora vanuit die gespesialiseerde korpus vergelyk is[294].

Vanuit Figuur 7.5 blyk dit dat daar slegs vier van die sewe naamwoordgroepe in die verdagte tekste aanwesig is, hierdie bevinding is in teenstelling met die bevinding vanuit die gespesialiseerde korpus waar daar voorbeelde van elk van die sewe naamwoorde in die onderskeie korpora geïdentifiseer kan word. NM is as die naamwoord geïdentifiseer wat die minste deur die onderskeie outeurs van die gespesialiseerde korpora gebruik word. In die verdagte tekste kom NSMD, NME en NEED[295] glad nie in die tekste voor nie. NEE is die algemeenste naamwoord in die gespesialiseerde korpora, maar in die verdagte tekste word NSM die meeste deur die outeurs gebruik. Daar is dus ŉ verskil in die gebruik van naamwoorde tussen die gespesialiseerde korpora en die verdagte tekste wanneer die naamwoorde wat algemeen is die onderskeie korpora gebruik word, ondersoek is, maar (sien Tabel 7.12) daar is steeds ooreenkomste wanneer die individuele outeurs se gebruik van die onderskeie naamwoorde ondersoek word:

Tabel 7.12 Ooreenstemmende gebruik van naamwoorde

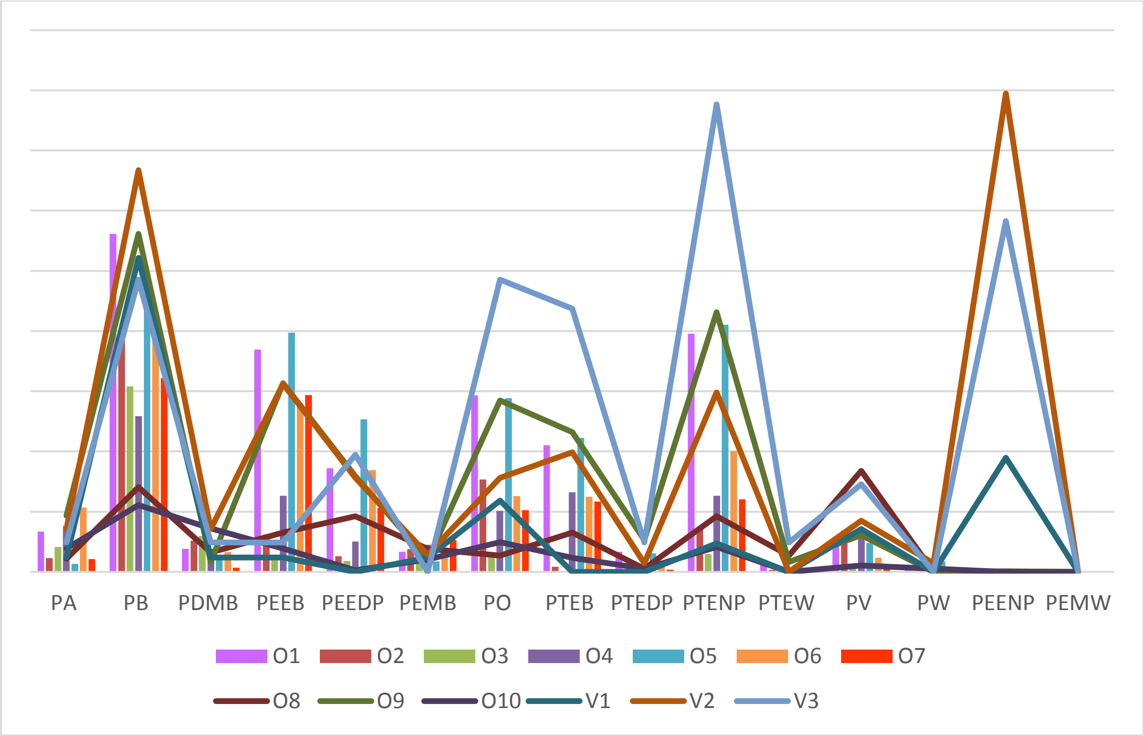
Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
NA	58,06	O2 (48,05)	21,29	O6 (25,70)	33,98	O4 (35,72)
NEE	23,70	O5 (22,28)	23,42	O5 (22,28)	29,13	O1 (27,62)
NEED	0,00	*	0,00	*	0,00	*
NM	15,40	O2 (15,42)	5,68	O1 (4,92)	7,28	O8 (6,50)
NME	0,00	*	0,00	*	0,00	*
NSM	79,38	O2 (67,62)	44,71	O9 (48,80)	26,70	O6 (41,06)
NSMD	0,00	*	0,00	*	0,00	*

[292] Die gemiddelde gebruik van NEE is 67 en 89 wanneer O3 en O10 by die berekening ingesluit is, en 50,62 wanneer die twee korpora uitgelaat word. O1, O2, O4, O5, O6, O7, O8 en O9 gebruik tussen 22,28 en 76,37 NEE per korpus. O3 en O10 kan vervolgens as uitskieters in terme van die gerbuik
van die NEE woordsoort beskou word.
[293] Die gebruik van NSM wissel tussen 41,06 en 67,62 en die gemiddeld is 53,56.
[294] O1, O3, O4, O5, O6, O8, O9 en O10 se gebruik van NM wissel tussen 4,15 en 7,69.
[295] Sien Bylaag F.

Vanuit Tabel 7.12 kan die afleiding gemaak word dat O2 ’n goeie kandidaat vir die outeurskap van V1 kan wees. O2 is in drie van die vier naamwoordgroepe wat deur V1 gebruik word die moontlike outeur na gelang van die frekwensie van die gebruik van die onderskeie naamwoorde. In die geval van V2 en V3 blyk dit dat – soortgelyk aan die analise van die gebruik van konjunkte – daar nie ’n enkele outeur vanuit die gespesialiseerde korpus is wat herhaaldelike ooreenkomste met V2 en V3 toon nie.

Die ondersoek na die gebruik van naamwoorde kan wel gebruik word wanneer die frekwensie van die gebruik van die verskillende naamwoordgroepe ondersoek word, en in die geval van V1 word O2 as ’n moontlike outeur van die teks beskou terwyl daar in die geval van V2 en V3 weer eens moontlike outeurs geëlimineer kan word (O10, O8, O7, O4, O3 en O1 kan byvoorbeeld nie as die outeur van V2 beskou word nie terwyl V3 nie deur O2, O3, O5, O6, O7, O9 of O10 geskryf is nie). Voornaamwoorde is die volgende woordsoortgroep wat ondersoek is ten einde te bepaal wat (indien enige) die ooreenkomste is tussen die verdagte tekste en die gespesialiseerde korpora.

7.3.2.4 Voornaamwoorde



Figuur 7.6 Voornaamwoorde

Daar is tussen 15 verskillende voornaamwoorde[296] onderskei[297], maar ten opsigte van die gebruik van voornaamwoorde op Twitter blyk dit dat PB (betreklike voornaamwoorde) die algemeenste voornaamwoord is wat op hierdie platform in Afrikaans gebruik word deur die outeurs vanuit die gespesialiseerde korpus – O1 (28,07) en O9 (28,06) gebruik die meeste PB.

Die woordsoorte PTENP (persoonlike voornaamwoord – tweede/enkelvoud/ongemarkeerd) en PEEB (besitlike voornaamwoorde – eerste/enkelvoud) blyk gereeld deur die onderskeie outeurs gebruik te word, en kom veral in die korpora van O1 (10,51 en 18,45), O5 (11,14 en 19,87) en O9 (11,59 en 15,66) voor. O2, O3, O4, O6, O7, O8 en O10 gebruik tussen 0,41 en 6,24 PTEB's (genormaliseer tot 1 000 woorde) en dus is die hoër frekwensie van gebruik wat in O1, O5 en O9 aangetref word, kenmerkend tot die outeurs se taalgebruik.

Die woordsoort PEEB word deur O2, O3, O4, O7 en O10 teen 'n relatiewe laer frekwensie gebruik ('n gemiddeld van 3,16 terwyl die gemiddelde gebruik van O1, O5, O6, O8 en O9 16.52 is). Daar kan vervolgens nie na gelang van die genormaliseerde waarde van die gebruik van PEEB aangevoer word dat 'n enkele outeur die outeur van een van die verdagte tekste is nie. Aangesien daar 'n ooreenkoms tussen die gebruik van PTENP en PEEB blyk te wees in die korpora van O1, O5 en O9 − wat nie in die oorblywende sewe korpora van die gespesialiseerde korpus gevind word nie − kan daar vervolgens aangevoer word dat indien PTENB sowel as PEEB se gebruik hoog is, O1, O5 of O9 moontlike outeurs van die teks kan wees.

Die woordsoort PEMW (wederkeurige voornaamwoord – eerste/meervoud) kom slegs in die korpus van O1 (0,11 – sien voorbeeld 186) voor, en is die "ongewildste" voornaamwoord:

Voorbeeld 186: **O1** – Wanneer dit by liefde kom skiet ons almal maar **onsself** in die voet.

PEMW[298] (onsself) soos dit in voorbeeld 186 gebruik word, is die enigste voorbeeld van in die gespesialiseerde korpus. Dit kan toegeskryf word aan die manier waarop die outeurs Twitter gebruik – eerder as om oor hulle persoonlike ondervindinge of gebeure wat nuuswaardig is te twiet (wanneer nuuswaardige gebeure getwiet word, word daar veel eerder van eiename gebruik gemaak ten einde te verseker dat leser weet oor wie die twiet gaan).

Vanuit Figuur 7.6 kan daar aangevoer word dat V1 oor die algemeen van die minste voornaamwoorde gebruik maak. In ooreenstemming met die bevindinge oor die gespesialiseerde korpora is PB ook die gewildste voornaamwoord in die verdagte tekste terwyl PW (slegs in V2), PTEW (slegs in V3), PMEW (geen) en PEMB (slegs in V1 en V2) teen die laagste genormaliseerde (tot 1 000 woorde) waarde in die verdagte tekste gebruik word. Tabel 7.13 dui aan watter van die outeurs vanuit die gespesialiseerde korpora ooreenkomste toon met

[296] Sien ook afdeling 6.3.2 (Ratiobepalings: voornaamwoorde).
[297] Sien bylaag F.
[298] Sien Bylaag F.

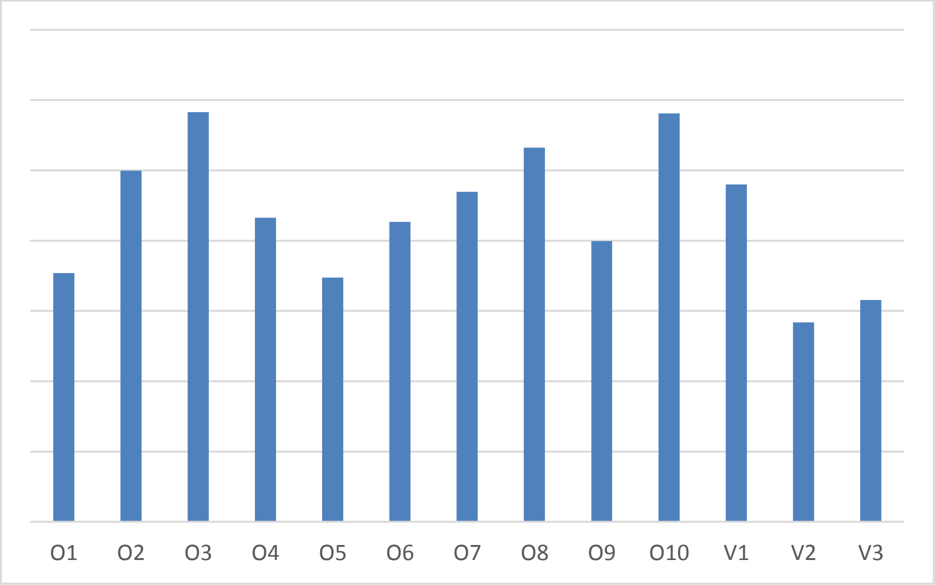
die gebruik van die onderskeie voornaamwoordgroepe soos deur die outeurs van die verdagte tekste:

Tabel 7.13 Ooreenstemmende gebruik van voornaamwoorde

Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
PA	1,18	O2 (1,14)	3,55	O1 (3,35)	2,43	O3 (2,07)
PB	26,07	O5 (24,46)	33,36	O1 (28,07)	24,27	O5 (24,46)
PDMB	1,18	O5 (1,31)	3,55	O3 (2,96)	2,43	O2 (2,61)
PEEB	1,18	O3 (1,18)	15,61	O9 (15,66)	2,43	O10 (1,94)
PEEDP	0,00	O10 (0,13)	7,81	O1 (8,61)	9,71	O1 (8,61)
PEMB	1,18	O3 (0,18)	1,42	O1 (1,68)	0,00	O5 (0,87)
PO	5,92	O6 (6,30)	7,81	O2 (7,67)	24,27	O1 (14,65)
PTEB	0,00	O3 (0,00)	9,94	O1 (10,51)	21,84	O9 (11,59)
PTEDP	0,00	O3 (0,00)	0,71	O5 (0,86)	2,43	O9 (2,64)
PTENP	2,37	O10 (2,07)	14,90	O6 (10,04)	38,83	O9 (21,55)
PTEW	0,00	O10 & O7	0,00	O10 & O7	2,43	O8 (1,36)
PV	3,55	O2 (3,43)	4,26	O2 (3,43)	7,28	O8 (8,40)
PW	0,00	O9, O8 & O7	0,71	O4 (0,84)	0,00	O9, O8 & O7
PEENP	9,48	*	39,74	*	29,13	*
PEMW	0,00	*	0,00	*	0,00	*

Ook in Tabel 7.13 kom dit duidelik na vore dat daar nie een van die outeurs vanuit die gespesialiseerde korpus is wat herhaaldelik ooreenkomste toon met die verdagte tekste nie (die gebruik van die verskillende voorsetselgroepe). O1 gebruik vyf keer voornaamwoorde (PA, PB, PEEDP, PEMP en PTEB) wat ooreenstem met die genormaliseerde waarde van V2. Daar kan vervolgens aangevoer word dat O1 definitief as ŉ moontlike outeur van V2 beskou kan word. In die geval van V1 en V3 kan daar weer van ŉ elimineringsproses gebruik gemaak word.

7.3.2.5 Voorsetsels



Figuur 7.7 Voorsetsels

Voorsetsels kom in al tien die gespesialiseerde korpora voor (sien Figuur 7.7). Soos in afdeling 6.3.3 bespreek, kan voorsetsels gebruik word om deur middel van die ratiobepalings die outeur van ŉ twiet te identifiseer. Verder blyk dit ook vanuit afdeling 6.3.5 dat die gebruik van voorsetsels as kenmerkend van die Afrikaans wat op Twitter voorkom, beskou kan word aangesien dit in elke korpus voorkom.

ŉ Afwyking vanaf die norm (ŉ korpus waarin daar nie voorsetsels voorkom nie) sal dus as ŉ kenmerkende patroon in die korpus beskou word. Vanuit hierdie ondersoek na die gebruik van voorsetsels is O10 (116,22) en O3 (116,60) die outeurs wat die meeste voorsetsels gebruik terwyl O5 slegs 69,45 voorsetsels gebruik (sien voorbeeld 187 tot 190):

Voorbeeld 187: **O10** – Jantjies kry 'n skopgeleentheid pale toe, maar is egter onsuksesvol. Telling steeds 19-0 **in** die guns van die Emirates Leeus. #LIOvBLU

Voorbeeld 188: **O1** – As ek ophou kan typos maak **in** my tweets, sal sit great wees.

Voorbeeld 189: **O2** – Noot vir noot nie my gunsteling nie. Maar STimes Lifestyle beledigend daaroor. Sal hulle so **oor** enige ander etniese groep skryf? #Afrikaans

Voorbeeld 190: **O5** – Ek is sommer gatvol vir die meisies wat altyd ander meisies judge **oor** hoe hulle lyk. Hulle weet dan hoe moeilik dit is om 'n meisie te wees!

Voorbeeld 187 tot 190 toon aan hoe van die outeurs vanuit die gespesialiseerde korpus voorsetsels gebruik. Hierdie analise sluit aan by die analise wat in afdeling 6.3.5 (Radiobepaling: voorsetsels) uitgevoer is en toon ook aan dat voorsetsels nie net deurlopend deur die verskillende outeurs op Twitter gebruik word nie, maar dat ŉ ondersoek na die gebruik van hierdie woordsoort wel aanleiding daartoe kan gee dat outeurs van tekste geïdentifiseer kan word.

Figuur 7.7 dui aan dat V1 die outeur is wat die meeste voorsetsels gebruik en V2 die minste. Aansien elk van die drie verdagte outeurs wel voorsetsels in hulle korpora gebruik, is daar nie ŉ afwyking van die norm nie. Ten einde te kon bepaal of daar ooreenkomste waargeneem kan word tussen die gebruik van voorsetsels in die gespesialiseerde korpus en die verdagte tekste, is die voorkoms van voorsetsels in die genoemde korpora met mekaar vergelyk (sien Tabel 7.14):

Tabel 7.14 Ooreenstemmende gebruik van voorsetsels

Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
SVS	95,97	O7 (93,90)	56,78	O5 (69,45)	56,78	O5 (69,45)

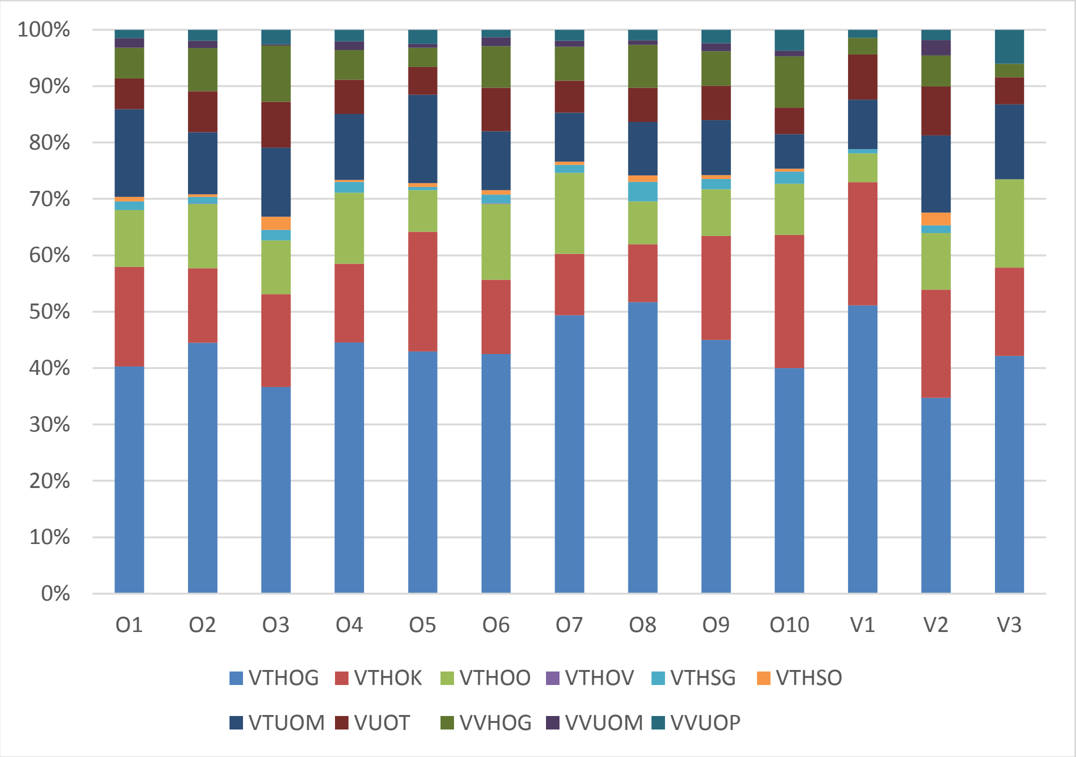
Vanuit Tabel 7.15 blyk dit dat O5 met beide V2 en V3 se gebruik van voornaamwoorde ooreenstem (sien voorbeeld 190, 194 en 195) terwyl daar ’n ooreenkoms tussen V1 en O7 waargeneem kan word:

Voorbeeld 169: **V2** – My header behoort julle baie te vertel **oor** myself.

Voorbeeld 170: **V3** – Jy worry altyd **oor** almal, vra of hulle okay is. Maar worry iemand **oor** jou? Vra hulle ooit of jy okay is??

Vanuit die voorbeelde kan daar aangevoer word dat V2 en V3 wel tematiese verskille toon, maar dat daar nie na gelang van die manier waarop die voorsetsels gebruik word, aangevoer kan word dat O5 eerder die outeur van V2 as V3 is nie. Soos reeds genoem, is al die outeurs (gespesialiseerde korpora) wat ooreenkomste met die verdagte tekste toon na gelang van die gebruik van voorsetsels by die lys van mootlike outeurs gevoeg. Werkwoorde is volgende ondersoek (Figuur 7.8).

7.3.2.6 Werkwoorde



Figuur 7.8 Werkwoorde: gespesialiseerde korpus

Vanuit Figuur 7.8 word daar tussen 11 verskillende werkwoorde onderskei. O5 (192,62) is die outeur wat van die meeste werkwoorde (genormaliseer tot 1 000) gebruik maak terwyl O8 slegs 71,27 werkwoorde gebruik.

Die werkwoord wat die meeste in hierdie gespesialiseerde korpus vanaf Twitter gebruik word, is VTHOG (werkwoord: teenwoordig/hoof/onskeidbaar/oorganklik) en dit is veral in O5 (82,77) wat hierdie woordsoort voorkom terwyl O9 slegs 36,86 VTHOG's gebruik. Die gebruik van werkwoorde kan dus ondersoek word ten einde te kan bepaal of die frekwensie waarteen die onderskeie outeurs in die gespesialiseerde korpus van die woordsoorte gebruik maak wel aanleiding kan gee dat outeurs van mikrobloginskrywings geïdentifiseer kan word[299].

Alhoewel daar tussen 11 verskillende werkwoorde onderskeid getref word, is daar in teenstelling met die bevindinge vanuit die gespesialiseerde korpus, slegs tien van die werkwoordgroepe wat in die drie verdagte tekste gevind kan word. VTHSG en VVUOM[300] kom slegs in die korpus van V2 voor terwyl VTHSG glad nie deur V3 gebruik word nie.

Soortgelyk aan die bevindinge vanuit die gebruik van woordsoorte in die gespesialiseerde korpus is VTHOG ook die gewildste werkwoord wat in die verdagte tekste gebruik word − 'n bevinding wat weer eens as 'n kenmerk van Afrikaans beskou kan word soos dit op Twitter gebruik word. Tabel 7.15 beeld die ooreenstemmende gebruik van die verskillende woordsoortgroepe aan tussen die verdagte tekste en die gespesialiseerde korpus:

Tabel 7.15 Ooreenstemmende gebruik van werkwoorde

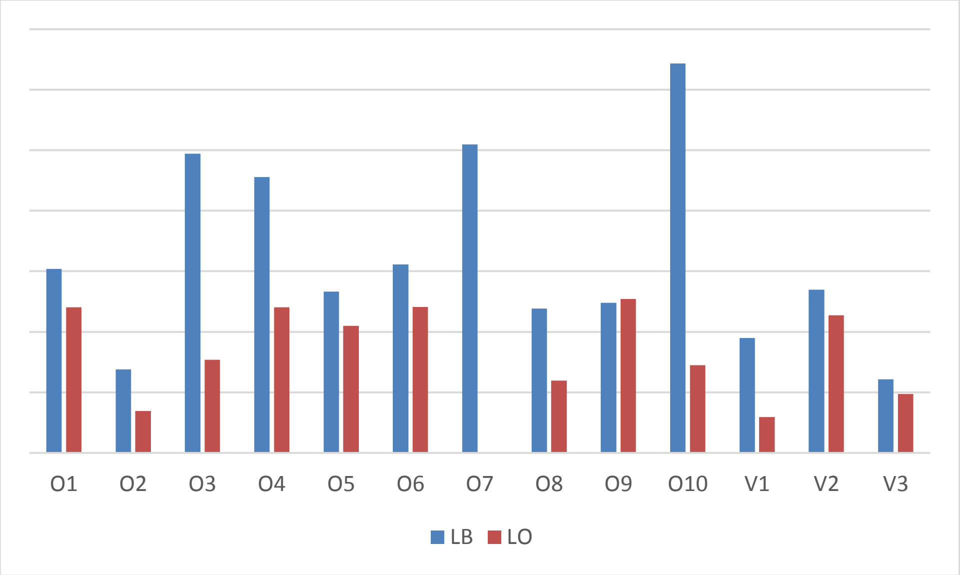
Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
VTHOG	82,94	O5 (82,77)	53,94	O6 (53,00)	84,95	O5 (82,77)
VTHOK	35,55	O1 (32,54)	29,81	O9 (31,52)	31,55	O9 (31,52)
VTHOO	8,29	O10 (9,17)	15,61	O2 (16,80)	31,55	O7 (23,34)
VTHOV	0,00	*	0,00	*	0,00	*
VTHSG	1,18	O2 (1,79)	2,13	O10 (2,20)	0,00	O5 (1,09)
VTHSO	0,00	O2 (0,65)	3,55	O3 (2,96)	0,00	O10 (0,52)
VTUOM	14,22	O7 (14,15)	21,29	O4 (16,88)	26,70	O1 (28,63)
VUOT	13,03	O2 (10,69)	13,48	O2 (10,69)	9,71	O5 (9,61)
VVHOG	4,74	O7 (5,42)	8,52	O10 (9,17)	4,85	O7 (5,42)
VVUOM	0,00	O3 (0,54)	4,26	O9 (2,44)	0,00	O3 (0,54)
VVUOP	2,37	O1 (2,80)	2,84	O2 (2,86)	12,14	O5 (4,80)

Tabel 7.15 dui weer eens aan dat nie een van die outeurs vanuit die gespesialiseerde korpus herhaalde ooreenkomste met die werkwoordgebruik van die verdagte tekste toon nie. Alhoewel O2 drie keer ooreenkomste met V1 toon[301], is O2 een van die outeurs wat oorweeg kan word vir outeurskap van V2 na gelang van die gebruik van werkwoorde. V3 gebruik vier keer werkwoorde en stem met O5 ooreen.

299 In afdeling 7.3.2.2 is die verdagte tekste se gebruik van woordsoorte ondersoek ten einde te bepaal of die ondersoek as suksesvol beskou kan word.
300 Sien Bylaag F.
301 Drie van die elf werkwoordgroepe van V1 stem met die werkwoordgebruik van O2 ooreen.

Die afleiding wat vanuit hierdie ondersoek gemaak kan word, is dat O1 as ’n moontlike outeur vir V1 en V2 beskou kan word terwyl O5 as die moontlike outeur van V3 geïdentifiseer kan word na gelang van die ondersoek van die genormaliseerde waarde (tot 1 000 woorde) van die onderskeie werkwoordgroepe. Die gebruik van bepaalde en onbepaalde lidwoorde (sien Figuur 7.9) is die laaste woordsoortgroep wat voorgehou word.

7.3.2.7 Lidwoorde



Figuur 7.9 Lidwoorde: gespesialiseerde korpus

Lidwoorde (bepaald en onbepaald) kom in elk van die gespesialiseerde korpora voor, maar daar kan duidelik onderskeid getref word tussen die aantal lidwoorde wat elk van die outeurs gebruik. O10 (64,31 genormaliseer tot 1 000 woorde) gebruik die meeste bepaalde lidwoorde terwyl O9 (25,42 genormaliseer tot 1 000 woorde) se onbepaalde lidwoordgebruik die hoogste is.

O2 maak deurlopend van minder lidwoorde (bepaald en onbepaald) gebruik of bring morfologiese veranderinge aan die lidwoord *die* aan[302] – sien voorbeeld 171 tot 173:

Voorbeeld 171: **O2** – So, nou is **d** een of ander bobbejaan se rassistiese foto op Facebook voorbladhoofstorie in **d** Sunday times.

Voorbeeld 172: **O2** – Een **v d** beste uitvindings ooit is **die** remote, **d** ding waarmee j kanale op TV verander sonder om op te staan. Probeer sonder een klaarkom.

Voorbeeld 173: **O2** – Klomp nuwe materiaal op #mediamense.co.za. Kyk uit vir lekker bydrae deur #GEBRUIKER, oudredakteur van Oggendblad by afdeling By Nabaat.

[302] Sien afdeling 7.4 (Morfologie).

Voorbeeld 171 tot 173 toon aan hoe O2 lidwoorde uitlaat of deur middel van weglating[303] die lidwoorde afkort. Hierdie kenmerk van lidwoordgebruik kan as kenmerkend tot die korpora van O2 beskou word. Dit word nie deur ander outeurs in die gespesialiseerde korpus gebruik nie.

Bepaalde lidwoorde is volgens die navorsing vir hierdie studie meer gewild. Daarom kan daar gestel word dat die bepaalde lidwoord meer algemeen in Afrikaans is soos dit op Twitter gebruik word − die gebruik van bepaalde lidwoorde is meer kenmerkend as die gebruik van onbepaalde lidwoorde in die Afrikaans wat op Twitter gebruik word. Hierdie kenmerk kan toegeskryf word aan die feit dat outeurs slegs ŉ bepaalde aantal karakters (140) het om ŉ boodskap oor te dra en daarom moet outeurs probeer om so duidelik as moontlik die boodskap oor te dra[304]. Ten einde te kon bepaal of daar ooreenskomste in die gebruik van lidwoorde waargeneem kon word, is die gespesialiseerde korpora met die verdagte tekste vergelyk (sien Tabel 7.16):

Tabel 7.16 Ooreenstemmende gebruik van lidwoorde

Woordsoort	V1	Soortgelyk aan V1	V2	Soortgelyk aan V2	V3	Soortgelyk aan V3
LB	18,96	O2 (13,79)	12,14	O2 (13,79)	26,97	O5 (26,64)
LO	5,92	O2 (6,93)	22,71	O1 (24,04)	9,71	O8 (11,92)

Tabel 7.16 toon aan dat O2 se gebruik van lidwoorde (bepaald en onbepaald) met die gebruik van lidwoorde deur V1 ooreenstem. Daarom kan O2 as die moontlike outeur van V1 beskou word na gelang van die gebruik van lidwoorde. In die geval van V2 stem die gebruik van bepaalde lidwoorde ooreen met O2 terwyl die onbepaalde lidwoord-gebruik met O1 ooreenstem. V3 kan na gelang van hierdie analise deur O5 of O8 geskryf wees. Wanneer daar na die gebruik van die bepaalde lidwoorde van V2 en O2 ondersoek ingestel is, blyk dit dat V2 nie bepaalde lidwoorde uitlaat nie en ook nie van weglating gebruik maak nie (sien voorbeeld 171 tot 173 en 174 en 175):

Voorbeeld 174: **V2** – Almal het hulle verslawing, of dit nou **'n** iets of **'n** iemand is.

Voorbeeld 175: **V2** – Hoe weet jy jy was lanklaas by **die** huis? Wanneer jy **die** verkeerde ligskakelaar vir **die** lig wat jy wil aansit, gebruik.

Wanneer voorbeeld 171 tot 173 met voorbeeld 174 en 175 vergelyk word, kan die afleiding gemaak word dat V2 nie deur O2 geskryf kan wees na gelang van die gebruik van lidwoorde nie. Die ondersoek na die gebruik van lidwoorde dui aan dat V1 deur O2, V2 deur O1 en V3 deur óf O5 óf O8 geskryf is.

303 Sien afdeling 7.4.3 (Weglating) en ook afdeling 7.5.2 (Die invloed van Engels).
304 Sien ook afdeling 7.5.2 (Die invloed van Engels).

Die sintaktiese ondersoek na die twiets van die gespesialiseerde korpora sowel as die drie verdagte tekste kan as suksesvol beskou word en kan sonder twyfel ŉ bydra lewer om die moontlike outeurs van mikrobloginskrywings te identifiseer. Soos met die vorige ondersoeke lewer hierdie analises nie individueel genoegsame bewyse van outeurskap nie, maar wanneer dit in kombinasie met die reeds bespreekte analises gebruik word, kan outeurskap wel bewys word.

7.3.3 Die invloed van Engels

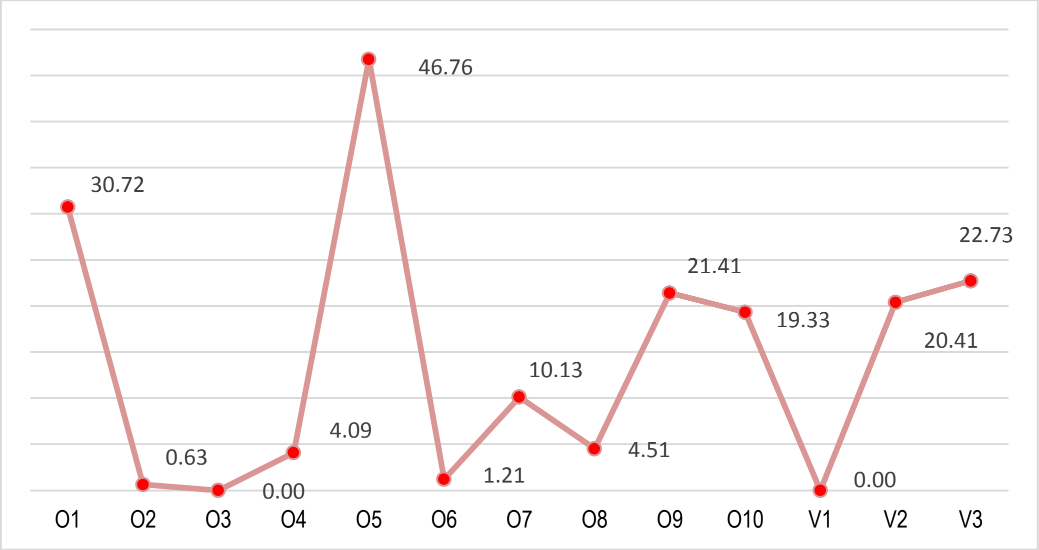
In die voorafgaande afdeling is die verskillende sinstipes en woordsoorte ondersoek. Die laaste aspek wat ondersoek word ten opsigte van sintaksis, is die invloed van Engels op Afrikaanse twiets[305]. Carstens (2011:314-320) stel dat die invloed van Engels veral op drie terreine in Afrikaans bespeur kan word, naamlik leenwoorde, direkte oornames en anglisismes.

Leenwoorde kan ŉ positiewe of negatiewe uitwerking op ŉ taal hê. Wanneer leenwoorde ŉ leemte vul wat nie reeds deur bestaande woorde in ŉ taal gevul kan word nie is die impak van leenwoorde positief (Carstens, 2011:318)[306]. ŉ Direkte oorname verwys na die gebruik van Engelse woorde in die plek van bestaande Afrikaanse woorde (Carstens, 2011:315-320)[307]. Die laaste aspek van die invloed van Engels waarna Carstens (2011:320) verwys, is anglisismes. Anglisismes verwys na woorde/frases/sinne wat nie net soos Afrikaans lyk nie, maar ook reg klink en ook die regte betekenis oordra (Carstens, 2011:320). In hierdie afdeling is daar ondersoek ingestel na die voorkoms van sinne wat as anglisisties beskou kan word in beide die gespesialiseerde korpus en die verdagte tekste (sien Figuur 7.10 – die invloed van Engels):

[305] Daar word in afdeling 7.6 (Kodewisseling) na IT en MT en dubbelmorfologie ondersoek ingestel.

[306] Carstens (2011:315-320) verwys na die gebruik van leenwoorde wat as Afrikaans beskou word. Voorbeelde sluit in *rugby* (direkte ontlening), *tjek* (indirekte ontlening – Engels: cheque) en *artifisieel* (kunsmatig word verkeerdelik as die korrekte Afrikaans – Dietse vorm − beskou alhoewel *artifisieel* die aanvaarbare Romaanse vorm is).

[307] Sien afdeling 7.5.1 (Kodewisseling).



Figuur 7.10 Invloed van Engels op sinne

7.3.3.1 Gespesialiseerde korpus: die invloed van Engels

Figuur 7.10 dui die genormaliseerde waarde per 1 000 sinne aan vir die voorkoms van sinne wat ŉ invloed vanaf Engels (anglisismes) toon. Dit blyk asof daar ŉ wesenlike verskil tussen die voorkoms van hierdie aspek is. O1 (30,72) en O4 (46,76) gebruik die meeste sinne waarin daar ŉ invloed van Engels sigbaar is. In die geval van kodewisseling is daar nie frases wat herhaaldelik gebruik word nie en wat nie as kenmerkend tot die outeurs se tekste beskou kan word nie. Voorbeelde van sinne vanuit die gespesialiseerde korpus waarin ŉ invloed vanaf Engels sigbaar is, sluit in:

Voorbeeld 169: **O7** – Mooi. **Nou stop om my te oordeel**. Want ek doen nie dit met jou nie

Voorbeeld 170: **O1** – As ek nog een bio lees met "papercut survivor" in, gaan ek death by papercuts **op myself gaan**.

Voorbeeld 171: **O10** – Die @GEBRUIKER gaan vandag teen UCT eerste kolf. W.Lubbe & J.Malan **vat die veld**.@GEBRUIKER #VarsityCricket

Voorbeeld 169 tot 171 toon die manier aan waarop die sinne met ŉ Engelse invloed gebruik word, maar hierdie ondersoek het aangetoon dat die frekwensie waarby die onderskeie outeurs sinne gebruik wat ŉ Engelse invloed toon as 'n groter aanduiding van outeurskap beskou kan word as die manier waarop hulle dit gebruik.

ŉ Verdere invloed van Engels is volgens Carstens (2011:65), die invloed van Engels op die gebruik van lidwoorde. Carstens (2011:65) stel dat lidwoorde óf toegevoeg word óf weggelaat word. In afdeling 7.3.2.7 is daar na die gebruik van lidwoorde ondersoek ingestel en daar is bevind dat O7 geen onbepaalde lidwoorde gebruik nie en O2 gebruik die minste bepaalde lidwoorde. In afdeling 7.4 is daar verder na O2 se geneigdheid ondersoek ingestel om van

weglating by die woorde *van* en *die* gebruik te maak. Aangesien die outeur ook lidwoorde uitlaat (*Byeenbring v nuus in orde, as hooggeregshofbevinding reg verstaan word*; *Fin24 het net in een geval Moneyweb se kopiereg geskend, bevind hof* of *Waar ons vandeesweek wyle Dene Smuts as politikus en joernalis eer, dink ook aan Afrikaanse joernaliste wat na Engelse pers moes uitwyk*) kan die faktor ook aan die invloed van Engels toegeskryf word.

Om te kon bepaal of die invloed van Engels wel in die ondersoek na die identifisering van outeurs van mikrobloginskrywings gebruik kan word, word die analise van die invloed van Engels in die verdagte tekste voorgehou.

7.3.3.2 Verdagte tekste: die invloed van Engels

Daar is bevind dat daar slegs in V2 (20,41) en V3 (22,73) voorbeelde voorkom van sinne waarin daar ʼn invloed van Engels waarneembaar is (sien Figuur 7.10). Vanuit die gespesialiseerde korpus stem O10 (19,33) en O9 (21,41) met V2 en V3 ooreen wanneer die genormaliseerde waarde van die voorkoms van sinne wat ʼn invloed vanaf Engels aantoon, ondersoek is.

Voorbeeld 172: **V2** – Kan ons **net gou praat oor** hoe mooi Potch se bome is.

Voorbeeld 173: **V3** – Baie dankie dat jy **tyd uit jou patetiese lewe vat** om oor my te praat.

Voorbeeld 172 en 173 is voorbeelde hoe die invloed van Engels in V2 en V3 daaruitsien. Die aantal keer wat die outeurs Afrikaanse sinne in Engels formuleer is van groter belang as die manier waarop hulle dit doen aangesien daar nie ʼn bepaalde patroonmatigheid voorkom in die korpora wat in hierdie studie gebruik is nie.

In die hieropvolgende afdeling is ondersoek na die morfologie ingestel ten einde te bepaal wat morfologies kenmerkend aan Afrikaans is soos dit op Twitter gebruik word en hóé, indien wel, die ondersoek na woorddele, vorm en betekenissamehang aanleiding tot die suksesvolle identifisering van mikroblog-outeurs kan gee.

7.4 Morfologie

Morfologie verwys na die woorddele, hulle vorm en betekenissamehang binne ʼn woord (Van Schoor, 1983:349). Van Huyssteen (2014:171) postuleer dat morfologie in leketerme as die beskrywing en bestudering van die struktuur en sistematiek van komplekse woorde soos *honde (hond* + *e)*, *hondmak (hond* + *mak)* en *hondetand (hond* + *e* + *tand)*. Van Huyssteen (2014:174) vul die bogenoemde definisie van morfologie aan en beskryf dit as “die studie van die vorm-betekenissisteem van komplekse”. Van Huyssteen (2014:198) verwys verder na die gebruik van

d vir *die*, *biqi* vir *bietjie*, *LMGA* vir *lag my gat af* en *lkr* vir *lekker* wat veral in die digitale era in SMS- en internettaal voorkom en plaas hierdie voorbeelde onder die reduksieprosesse wat in Afrikaans voorkom.

Dit is, ongeag hoe gebruikers ŉ woord verander ten einde ŉ boodskap oor te dra*,* belangrik dat die bedoeldebetekenis aan die leser (ontvanger) oorgedra word en semanties aanvaarbaar is. Gevolglik is daar ondersoek ingestel na die semantiese keuses wat mikrobloggebruikers in Afrikaans uitoefen. In die ondersoek na die morfologiese kenmerke van die korpora is daar ondersoek ingestel na die reduksieproses. Daar kan volgens Van Huyssteen (2014:198-200) in die reduksieproses[308] onderskei word tussen *afkorting, inkorting* en *truvorming*[309].

- **Afkorting**: Dit verwys na ŉ proses waardeur ŉ nuwe woord gevorm word "op basis van die beginletters van ŉ woord of woordgroep." Daar word onderskeid getref tussen "egte afkortings" en akronieme volgens Van Huyssteen (2014:198).

- **Inkorting**: "Ter wille van ekonomie in taalgebruik maak taalgebruikers dikwels in geskrewe taal gebruik van inkorting om iets kort en kragtig te stel" (Van Huyssteen, 2014:198). Daar word onderskeid getref tussen twee verskillende soorte inkorting, naamlik *knipsels sonder die inagneming van die morfologiese struktuur* en *knipsels met die inagneming van die morfologiese struktuur*[310].

Die bostaande prosesse maak deel uit van die reduksieproses in Afrikaans. Aangesien daar ŉ karakterbeperking op Twitter aangetref word, is die verwagting dat outeurs − waar moontlik − korter en eenvoudiger woorde sal gebruik. Daar word verder onderskeid getref tussen verkorting en weglating. Verkorting verwys volgens Olivier (2013:490) na die uitlaat van ŉ sillabe (soos aangetref in *bib* en *prof*[311]). Aangesien verkorting soortgelyk is aan knipsels sonder die inagneming van ŉ woord se morfologiese struktuur, is daar in die analise na die gebruik van weglating (soos deur Thiart, 2014:80 voorgehou) ondersoek ingestel. Daar is vervolgens ondersoek ingestel na die voorkoms van reduksieprosesse (afkortings, inkorting[312] en weglating) in die Afrikaans wat op Twitter gebruik word.

Deur die bogenoemde drie aspekte te ondersoek, het ek probeer om te bepaal of outeurs op só ŉ manier van hierdie prosesse gebruik maak dat dit as kenmerkend van hulle taalgebruik

---

308 Logogramme verwys na die verskynsel waar ŉ letter, syfer, simbool of teken ŉ woord of selfs ŉ frase verteenwoordig volgens Thiart (2014:78).
309 Vir die doel van hierdie studie is truvorming nie ondersoek nie.
310 Vir die doel van hierdie studie is daar nie onderskeid getref tussen knipsels met en knipsels sonder die inagneming van die morfologiese struktuur nie.
311 Hierdie aspek word reeds as deel van inkorting bespreek.
312 Sien afdeling 7.1 (Inleiding).

beskou kan word. Ten einde die voorkoms van die morfologiese veranderinge in die onderskeie korpora voor te hou, fokus ek eers individueel op afkortings, inkortings en verkorting.

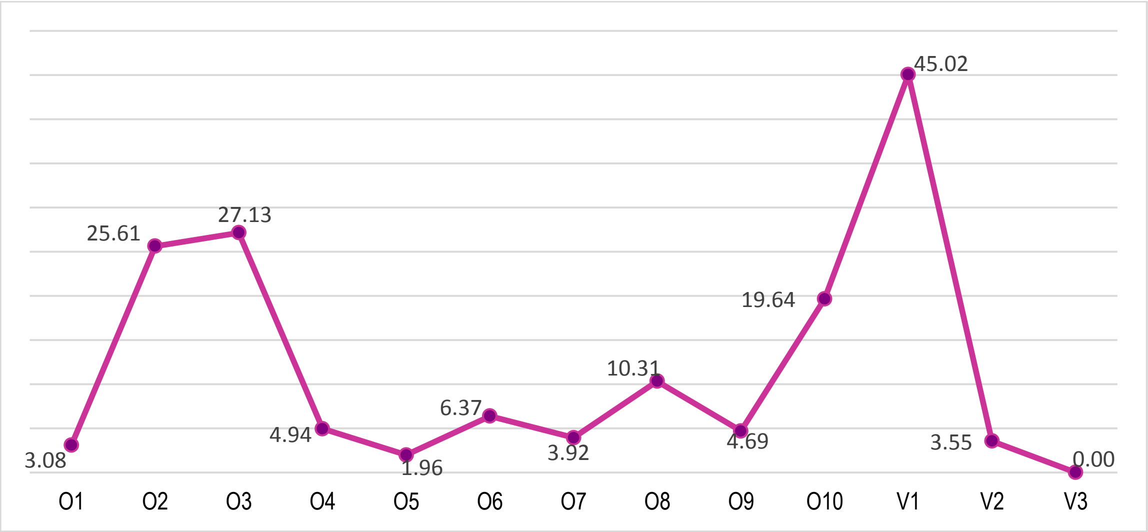
7.4.1 Afkortings

Soos bespreek in afdeling 7.4 (Morfologie) verwys afkortings volgens Van Huyssteen (2014:198) na ŉ proses waardeur ŉ nuwe woord gevorm word "op basis van die beginletters van ŉ woord of woordgroep." Daar word onderskeid getref tussen "egte afkortings" en akronieme (Van Huyssteen, 2014:198). Ten einde die stilistiese ondersoek uit te kon voer, is daar gekyk na die gebruik van afkortings, akronieme en afkortings soos daar na SMS-Afrikaans ondersoek ingestel is in die gespesialiseerde korpora en in die verdagte tekste (sien Tabel 7.17):

Tabel 7.17 Gebruik van afkortings in die gespesialiseerde korpora

Outeur	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	V1	V2	V3
Afkortings	28	310	89	35	9	103	22	38	23	149	38	5	0
Genormaliseerde waarde (1 000 woorde)	3,08	25,61	27,13	4,94	1,96	6,37	3,92	10,31	4,69	19,64	45,02	3,55	0,00

Vanuit Tabel 7.17 blyk dit dat elk van die outeurs (gespesialiseerde korpora en verdagte tekste) wel van afkortings gebruik maak en dat elk van die outeurs teen verskillende frekwensies van die afkortings gebruik maak (sien Figuur 7.11):



Figuur 7.11 Die gebruik van afkortings (gespesialiseerde korpora en verdagte tekste)

Figuur 7.11 toon aan hoe die genormaliseerde waarde (tot 1 000 woorde) vir elk van die tien outeurs in die gespesialiseerde korpora en die verdagte tekste gebruik is met betrekking tot afkortinggebruik. Hierdie grafiek toon duidelik aan dat elk van die outeurs oor ŉ unieke gebruik van afkortings beskik. Daar is verder individueel ondersoek ingestel na die gebruik van afkortings, akronieme en SMS-afkortings in die gespesialiseerde korpora en die verdagte tekste ten einde te bepaal of die morfologiese beginsels wel aanleiding kan gee tot die identifikasie van outeurs van mikrobloginskrywings. Vervolgens word die gebruik van afkortings (en voorbeelde van die afkortings) bespreek.

7.4.1.1 Gespesialiseerde korpus: foutlose afkortings

Ten einde die gebruik van afkortings in die onderskeie korpora te kon ondersoek, is daar na die gebruik van foutlose afkortings, afkortings waarin daar foute voorkom en die gebruik van afkortings eie aan sosiale media – SMS-afkortings – asook die gebruik van akronieme ondersoek ingestel. 29 afkortings (wat sonder foute voorkom) is in die onderskeie korpora geïdentifiseer[313] na gelang van die manier waarop die onderskeie outeurs van afkortings in die korpora gebruik maak (sien Figuur 7.12):

Figuur 7.12 Die persentasie waarby (foutlose) afkortings in die onderskeie korpora voorkom

Figuur 7.12 dui die persentasies van elk van die outeurs vanuit die gespesialiseerde korpora aan wat afkortings foutloos gebruik. Ten einde te kon bepaal watter persentasie afkortings deur

[313] Sien bylaag G – Foutlose gebruik van afkortings.

die onderskeie outeurs gebruik word (soos aangetoon in Figuur 7.12), is die onderstaande formule gebruik:

$$\text{Persentasie afkortings per korpus} = \frac{\textit{totale hoeveelheid afkortings per korpus}}{\text{totale aantal afkortings in al tien korpus}} \times 100$$

of

$$O2 = \frac{81}{212} \times 100$$

$$\therefore O2 = 38{,}21$$

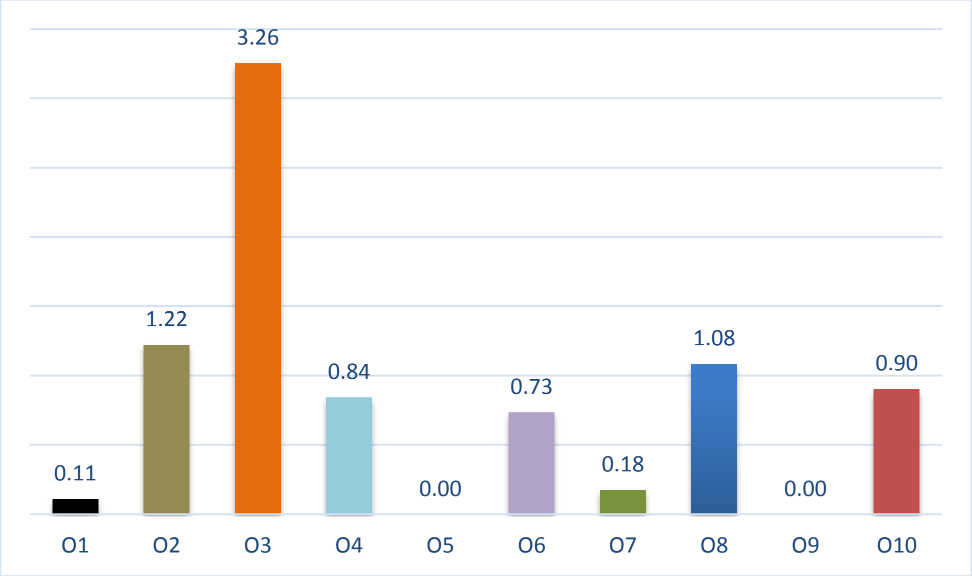
Wanneer die persentasies van elk van die outeurs se gebruik van afkortings bereken is, is dit duidelik dat die outeursafkortings (die korrekte gebruik van afkortings) só gebruik dat dit wel kenmerkend van die outeurs se skryfstyl beskou kan word. Vanuit Figuur 7.12 gebruik O2 verreweg die meeste afkortings terwyl O5 weer van geen afkortings gebruik maak nie. Hierdie aanname is verder getoets toe die verdagte tekste geanaliseer is. Die 29 (foutlose) afkortings wat deur die tien outeurs gebruik word[314], verskyn in Tabel 7.18:

Tabel 7.18 Afkortings (foutloos) wat in die gespesialiseerde korpora aangetref word

adv.	ens.	NS	RSA
a.g.v.	feb.	OB	SG
adj. SG	genl.	o.m.	TV
adj.pres.	gesk.	Okt.	vlg.
asb.	HK	p.p.	vs
BD	prof.	pres.	
CC	mnr.	Pta.	
Dr.	mun.	SA	

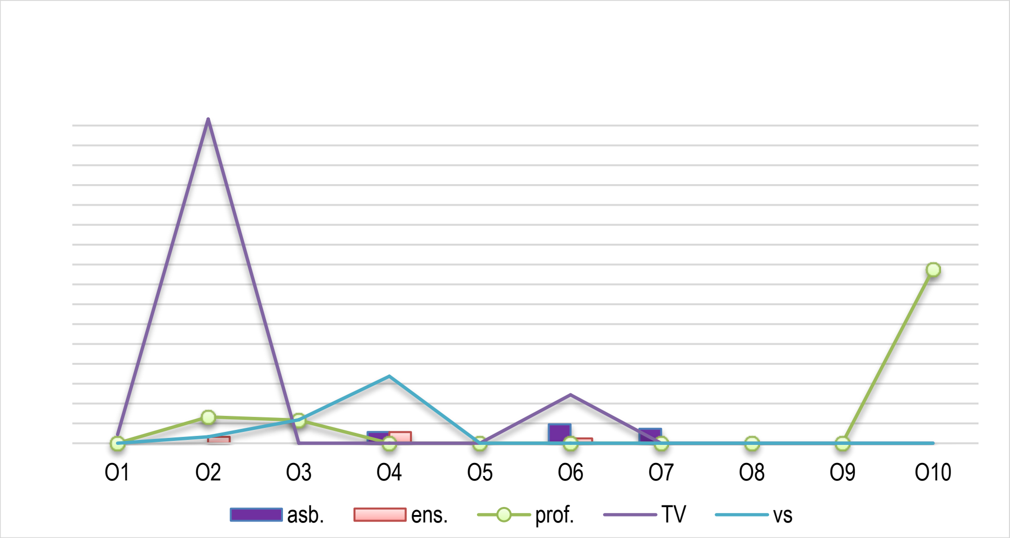
Vanuit die foutlose afkortings wat in Tabel 7.11 uitgebeeld word, word die afkorting *SA* deur nege van die tien outeurs gebruik en kan *SA* vervolgens as die gewildste afkorting beskou word, in die gespesialiseerde korpus (sien Figuur 7.13):

[314] Sien Bylaag G vir die onderskeie outeurs se gebruik van die afkortings.



Figuur 7.13 Die gebruik van die afkorting SA

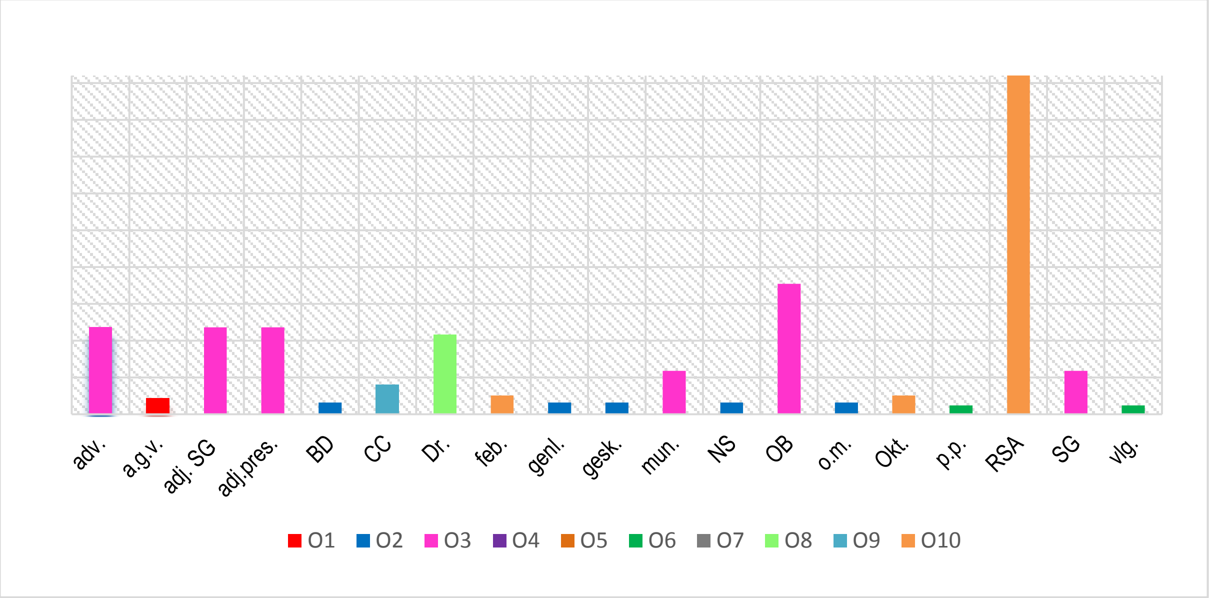
Soos gesien vanuit Figuur 7.13, blyk dit dat alhoewel die afkorting *SA* in nege van die tien korpora voorkom, O3 meer frekwent van hierdie afkorting gebruik maak. Vanuit die oorblywende 28 afkortings is *asb.* (O4, O6 en O7), *ens.* (O2, O4 en O6), *prof.* (O2, O3 en O10), *TV* (O1, O2 en O6), *vs* (O2, O3 en O4) – sien Figuur 7.14 – ook gewild en word hierdie afkortings deurlopend in drie van die korpora gevind.



Figuur 7.14 Genormaliseerde waarde van die afkortings wat deur meer as een outeur vanuit die gespesialiseerde korpora gebruik word

Figuur 7.14 toon die genormaliseerde waarde (tot 1 000 woorde) van die afkortings (*asb*, *ens.*, *prof. TV* en *vs*) aan. Vanuit Figuur 7.14 kon dit voor asof die gebruik van afkortings van outeur tot outeur verskil. *HK* (O1 en O10), *mnr.* (O2 en O8) en *pres.* (O2 en O3) word telkens deur

twee van die outeurs in die gespesialiseerde korpora gebruik. Die oorblywende afkortings (foutloos) word slegs deur een outeur in die ondersoek gebruik (sien Figuur 7.15):



Figuur 7.15 Die kenmerkende gebruik van (foutlose) afkortings in die gespesialiseerde korpora

Figuur 7.15 toon die afkortings en die genormaliseerde waarde aan waarteen die afkortings deur die onderskeie outeurs in die gespesialiseerde korpora gebruik word. Dit is duidelik dat die gebruik van afkortings verskil wanneer die tien outeurs se korpora met mekaar vergelyk word. Wanneer daar spesifiek ondersoek ingestel is na die afkortings wat slegs in ’n enkele korpora voorkom, kan daar aangevoer word dat O2 en O3 die meeste van afkortings gebruik maak wat nie deur een van die ander outeurs in die ondersoek gebruik word nie.

Verder is daar ook ondersoek ingestel na die mate waartoe die afkortings wat in die onderskeie korpora gevind word, uitgeskryf word. Die afkorting *asb.* word byvoorbeeld deur O4 (1), O6 (4) en O7 (1) gebruik terwyl die woord *asseblief* deur O1 (11), O4 (2), O5 (4), O6 (5) en O9 (4) gebruik word. O4 en O6 gebruik die woord *asseblief* dus meer keer as die twee outeurs wat van die afkorting gebruik maak (O4 – 1 afkorting en 2 keer die volledige woord; O6 – 4 afkortings en 5 keer die volledige woord – sien voorbeeld 174 tot 177):

Voorbeeld 174: **O4** – Boys, help ons met die Chat Bietjie Ball intro **asb**. Ons wil he julle moet julle bekke afnaam! Meer info hier: http://ow.ly/SFdjo

Voorbeeld 175: **O4** – Mens kannie genoeg Dankie sê Nie boys!! Ek gee nog R2000!! Help **asseblief** as julle kan!!!

Voorbeeld 176: **O6** – Waar kry Cash and carry my selfoonnommer? Ek het nog nooit daar gekoop! E-pos **asb**. vir my die dokument waar ek... http://fb.me/75kNjJmEc

Voorbeeld 177: **O6** – Liewe, vriendin. Moet my **asseblief** nooit weer saam nooi vir ete as jy jou kinders, ipad en 'chargers' gaan... http://fb.me /64sTj3Cp9

Voorbeeld 174 tot 177 toon aan hoe O4 en O6 van beide die afkorting en die volledige woord gebruik maak. In voorbeeld 174 is O4 se totale karaktertelling 128 terwyl die karaktertelling in voorbeeld 175, 89 is. Die karaktertelling van O6 in voorbeeld 176 is 136 en in voorbeeld 177 is dit 128. Dus kan daar (in dié geval) aangevoer word dat afkortings gebruik word om binne die karakterbeperking te bly:

Voorbeeld 178: **O2** – Bedanking v Jan-Jan Joubert as politieke redakteur van #Beeld moet 'n eerste in Naspers se *gesk.* wees. Waarop dui dit? Alles vloei (Thales)? [140]

Voorbeeld 179: **O2** – Dalk vertel volgende jr se nuwe *gesk* v #Naspers ons hoe internetgeld Afrikaanse joernalistiek gered het. Ek wag regtig om hul kant te hoor. [139]

Voorbeeld 180: **O2** – Gisteraand se opening *v* Spele was briljant. Dit help as mens so n' *geskiedenis* het. En as *d* kyker die musiek meestal ken. #Olimpics [131]

Voorbeeld 181: **O2** – Volksblad het op *Potch* onder *d* naam Het Westen begin, getrek *Bloem* toe, nuwe eienaar en naam gekry. Hierdie *geskiedenis* is ouer as 100 *jr*. [138]

Voorbeeld 178 tot 181 toon O2 se gebruik van die woord *geskiedenis* en die afkorting *gesk.* aan. Vanuit hierdie voorbeelde blyk dit dat ook dat hierdie outeur verkies om die afkorting te gebruik wanneer die karakterbeperking van 140 genader word. Die afleiding kan vervolgens gemaak word dat outeurs wat in Afrikaans twiet wel afkortings gebruik ten einde die bedoelde boodskap suksesvol oor te dra aan ŉ gekose teikengehoor ten spyte van die karakterbeperking van die medium. Voorbeeld 180 en 181 is veral interessant aangesien O2 van ander reduksieprosesse[315] gebruik maak ten einde steeds ŉ boodskap binne die 140 karakterbeperking te kan oordra.

Aangesien die outeurs (O1, O2, O3, O4, O6 en O10) in sommige gevalle slegs afkortings gebruik en in ander gevalle weer beide die afkorting en die woord gebruik, is daar nie noodwendig ŉ vaste patroon hier te bespeur nie – behalwe die manier waarop die outeurs ŉ koherente boodskap probeer oordra ten spyte van die karakterbeperking.

Vanuit hierdie ondersoek na die gebruik van (foutlose) afkortings, blyk dit dat ŉ ondersoek na die gebruik van afkortings wel insiggewend is wanneer outeurs van mikrobloginskrywings

[315] Sien afdelings 7.4.2 en 7.4.3.

geïdentifiseer moet word. Vervolgens is die gebruik van foutlose afkortings in die verdagte tekste ondersoek. ʼn Vergelyking is ten slotte getref tussen die gespesialiseerde korpora en die verdagte tekste se gebruik van foutlose afkortings om te bepaal of die ondersoek na die gebruik van foutlose afkortings wel aanleiding kan gee tot die identifikasie van mikroblogouteurs.

7.4.1.2 Verdagte tekste: foutlose afkortings

Hierdie afdeling (7.4.1.2) fokus spesifiek op die foutlose gebruik van afkortings in die verdagte tekste en die bevindinge word vergelyk met die bevindinge vanuit die gespesialiseerde korpus (afdeling 7.4.1.1).

In die verdagte tekste is daar vier foutlose afkortings wat deur drie van die outeurs gebruik word (sien voorbeeld 182, 183 en 184) – hierdie afkortings sluit in *pp.*, *p.*, *TV* en *ID* – en word slegs in die korpus van V1 en V3 aangetref:

Voorbeeld 182: **V1** – Tekens van tye: Beeld se hoofopskrifte op **pp.** 1, 2, 3, 4, 6 vandag in aanhaaltekens. O, ook **p.**10 en 11. Is dit luiheid of soeke na emosie?

Voorbeeld 183: **V1** – Maar daar was 'n tyd toe SAUK mense gehad het wat regtig **TV**-programme geproduseer het. Vandag? Nuus en aktualiteit. Tog is kantore vol.

Voorbeeld 184: **V3** – My **ID** sê 17 maar my lyf sê 70.

Die enkele gebruik van ʼn afkorting deur V3 verskaf nie genoegsame data om ʼn vergelyking ten opsigte van die gebruik van die afkorting te kan tref nie, maar O9 en O3 is die enigste outeurs in die gespesialiseerde korpora wat enigsins na “identiteitsdokumente” verwys – sien voorbeeld 185:

Voorbeeld 185: **O9** – Hoe is dit dat die persoon wie ek die meeste lief het en die persoon wie ek die meeste haat, 'n **identiteitsnommer** deel.

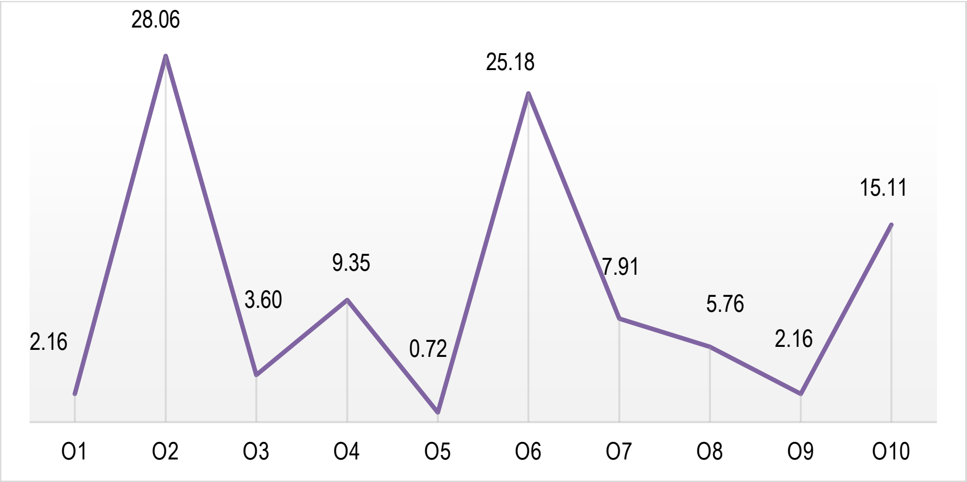
O3 – Moepya sê die departement van binnelandse sake sal oop wees oor die naweek vir kiesers om hul **ID'**s af te haal @GEBRUIKER

Alhoewel O3 ook *ID* gebruik, verskil die manier asook die konteks waarin die betrokke woord gebruik word van V3. Op grond van die tematiese ooreenkoms in die gebruik van die woord, kan O9 vervolgens as ʼn moontlike outeur van V3 beskou word. Wanneer die afkortings, soos voorgehou in voorbeeld 182 en 183, vergelyk word met die gebruik van foutlose afkortings in die gespesialiseerde korpus word die afkortings ook in die korpora van O2, O3, O4, O6, O7, O8 en O10 gevind. Die enigste outeur vanuit die gespesialiseerde korpus wat van al drie die afkortings wat in V1 gevind word, gebruik maak is O2. O2 kan vervolgens as die moontlike outeur van V1 na gelang van die ondersoek na die gebruik van foutlose afkortings beskou word.

7.4.1.3 Gespesialiseerde korpus: idiolektiese afkortings

In afdeling 7.4.1.1 en 7.4.1.2 is daar ondersoek ingestel na die onderskeie outeurs se gebruik van (foutlose) afkortings, in hierdie afdeling word daar gefokus op die gebruik van afkortings waarin daar wel foute voorkom, of te wel afwykings vanaf die norm wat dan as idiolektiese afkortings benoem kan word. Die ondersoek na idiolektiese afkortings sluit aan by Coulthard (2005:15) se verwysing na bevoegdheidsfoute − ŉ konsep wat verwys na die doelbewuste keuse van outeurs om ŉ reël te verbreek wanneer outeurs te kampe het met niestandaardvereistes wat deur ŉ kommunikasiemedium vereis word.

ŉ Totaal van 207 afkortings word in afdeling 7.4.1.1 bespreek. Na gelang van die ondersoek, soos uitgevoer in *ATLAS.ti* en WordSmith Tools, is daar ŉ bykomende 139 afkortings in die tien korpora geïdentifiseer. Hierdie 139 afkortings wyk almal af van die norm soos gestel in die Afrikaanse Woordelys en Spelreëls (2017). Figuur 7.16 toon die onderskeie outeurs se gebruik van idiolektiese afkortings aan:



Figuur 7.16 Persentasie (idiolektiese) afkortings wat in die gespesialiseerde korpora voorkom

Figuur 7.16 toon die persentasie idiolektiese afkortings aan wat in die tien korpora aangetref word. Die persentasies is uitgewerk soos in afdeling 7.4.1.1 (Gespesialiseerde korpora: idiolektiese afkortings:

$$(\text{Persentasie afkortings per korpora} = \frac{totale\ hoeveelheid\ afkortings\ per\ korpora}{\text{totale aantal afkortings in al tien korpora}} \times 100).$$

Soos duidelik vanuit Figuur 7.16, kom die meeste voorbeelde van foutiewe afkortinggebruik in die korpus van O2 voor (28,06), gevolg deur O6 (25,18) en O10 (15,11). Hierdie idiolektiese gebruik van afkortings stem ooreen met die bevindinge van afdeling 7.4.1.1 − O2 gebruik die

meeste afkortings foutloos, gevolg deur O10 en O6. Dit blyk dus dat die gebruik van afkortings, hetsy korrek of idiolekties (verkeerd), grotendeels ooreenstem in die onderskeie korpora. Dit is wel opvallend dat O5 slegs van idiolektiese afkortings (0,72%, oftewel 1 idiolektiese afkorting in die outeur se korpus) gebruik maak. Daar is 50 verskillende afkortings geïdentifiseer wat verkeerd gebruik word in die tien korpora wat ondersoek is (sien Tabel 7.19):

Tabel 7.19 Idiolektiese afkortings wat in die gespesialiseerde korpora voorkom

adv	Bv	Hk	Mev	pres
Afr	Cm	Juff	mil (miljoen)	prof
afr	Ea	Lg	milj (miljoen)	PTA
Antw	Ens	Lg	Mnr	pta
ARG	epos	JHB	Mr	Sept
a.k.a.	Feb	Jhb	Mrs	sg
asb	genl	jhb.	Mrt	tv
ASB	Gnr	Mej.	Nr	vlg
BD	gesk	Mej	NZ	vs
bib	Gr	mej[316]	P	v

In Tabel 7.19 kan die verskillende afkortings gesien word wat kenmerkend (dog verkeerd) gebruik word deur die onderskeie outeurs in die gespesialiseerde korpora. Vanuit Tabel 7.20 is dit duidelik dat die outeurs sommige afkortings op verskeie maniere verkeerd gebruik (sien byvoorbeeld *genl* en *gnr*, en *JHB*, *Jhb* en *jhb.*). Hierdie wisselvorme van afkortings en die gebruik van hierdie idiolektiese afkortings deur die onderskeie outeurs is ondersoek om te kon bepaal of daar ’n patroonmatigheid te bespeur is in die idiolektiese gebruik van afkortings.

Soos genoem, is daar 139 afkortings in die tien korpora geïdentifiseer wat foutief deur die outeurs gebruik word. Tabel 7.20 verskaf ’n opsomming van hoe elk van die onderskeie outeurs idiolektiese afkortings in hulle korpora gebruik:

Tabel 7.20 ’n Opsomming van die idiolektiese gebruik van afkortings deur die outeurs van die gespesialiseerde korpora

	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Idiolektiese afkortings in korpus	3	22	3	6	1	13	8	4	3	5
Totale idiolektiese afkortings in korpus	3	39	5	13	1	35	11	8	3	21
Totale idiolektiese afkortings in korpus (genormaliseer tot 1 000 woorde)	0,34	3,18	1,48	1,83	0,22	2,14	1,95	2,17	0,61	2,71
Idiolektiese afkortings uniek tot korpus	2	17	0	3	0	9	4	0	1	3
Idiolektiese afkortings algemeen tot Twitter	1	5	3	3	1	4	4	4	2	2
Totaal van unieke idiolektiese afkortings in die korpus	2	31	0	4	0	23	5	0	1	8

[316] Sien voorbeeld 201 tot 203 vir die wyse waarop O6 wisselvorme van die afkorting gebruik.

Totaal van unieke idiolektiese afkortings in die korpus (genormaliseer tot 1 000 woorde)	0,22	2,53	0,00	0,56	0,00	1,41	0,88	0,00	0,20	1,03
Totaal van idiolektiese afkortings algemeen tot Twitter	1	8	5	9	1	12	6	8	2	13
Totaal van idiolektieseafkortings algemeen tot Twitter (genormaliseer tot 1 000 woorde)	0,11	0,65	1,48	1,27	0,22	0,73	1,06	2,17	0,41	1,68

Soos gesien in Tabel 7.20 maak elk van die outeurs in hierdie studie van verskillende hoeveelhede idiolektiese afkortings gebruik. Dit blyk dat O2 die meeste idiolektiese afkortings gebruik wat slegs in hierdie outeur se korpus voorkom (31 idiolektiese afkortings − 2,53 afkortings genormaliseer tot 1 000 woorde). O2 gebruik meer idiolektiese afkortings wat uniek tot die outeur se korpus is as die kenmerkende idiolektiese afkortings wat op Twitter gebruik word. O5 (wat geen foutlose afkortings gebruik nie) maak wel van een (genormaliseer tot 1 000 woorde = 0,22) idiolektiese afkorting gebruik. O8 gebruik die meeste idiolektiese afkortings wat na gelang van hierdie studie as algemeen op Twitter beskou kan word (8 idiolektiese afkortings − 2,17 genormaliseer tot 1 000 woorde). Aangesien die studie gefokus het op kenmerkende eienskappe van ŉ bepaalde outeur om sodoende die outeur van ŉ bepaalde teks te kan identifiseer, word daar vervolgens op die idiolektiese afkortings gefokus wat uniek tot ŉ enkele outeur is.

O1 maak net van drie idiolektiese afkortings gebruik wat nie deur die ander outeurs in hierdie ondersoek gebruik word nie. Aangesien elk van hierdie afkortings ook slegs een maal of genormaliseer (tot 1 000 woorde) tot 0,11 in die korpora voorkom, is die afkortings nie so ŉ algemene voorkoms nie. Idiolektiese afkortings kan wel gebruik word indien O1 ŉ verdagte in die identifisering van ŉ outeur van ŉ mikrobloginskrywing is – sien voorbeeld 186 tot 188 vir O1 se gebruik van idiolektiese afkortings:

Voorbeeld 186: **O1** – As jy eerstejaar is en die **hk**[317] sê spring, vra jy hoe hoog en hoe lank moet jy in die lug bly. [92]

Voorbeeld 187: **O1** – PUK se **bib**[318] is my favourite plek op die hele kampus. [51]

Voorbeeld 188: **O1** – Universiteit **a.k.a.**[319] wat is slaap? [33]

Die bostaande drie afkortings is die enigste voorbeelde van idiolektiese afkortings wat in die korpus van O1 aangetref word. Alhoewel idiolektiese afkortings baie min in die korpus van O1 voorkom (slegs 0,11 wanneer die genormaliseerde waarde tot 1 000 woorde bereken is), kan

[317] HK (Afrikaanse Woordelys & Spelreëls, 2017:582).
[318] *bib.* of *bibl.* (Afrikaanse Woordelys & Spelreëls, 2017:570).
[319] Engels.

hierdie drie idiolektiese afkortings wel as ŉ kenmerk van die taalgebruik van O1 beskou word ten opsigte van Twitter. Dit is verder duidelik dat hierdie outeur nie van afkortings gebruik maak om karakters te bespaar nie aangesien nie een van die twiets waarin O1 idiolektiese afkortings gebruik oor meer as 100 karakters (van die moontlike 140) beskik nie. O2 maak van 17 idiolektiese afkortings (sien Tabel 7.21) gebruik wat as uniek tot die outeur se Twitter-korpus beskou kan word. Vanuit die 17 idiolektiese afkortings is daar 8 afkortings wat elk twee of meer keer in die korpus gebruik word (sien Figuur 7.17):

Figuur 7.17 Idiolektiese afkortings wat as uniek tot die korpus van O2 beskou kan word

Figuur 7.17 toon aan hoe O2 idiolektiese afkortings gebruik en watter idiolektiese afkortings die meeste deur O2 gebruik word. Vanuit Figuur 7.17 blyk dit dat O2 *Afr.* (Afrikaanse Woordelys & Spelreëls, 2017:566) op twee verskillende maniere spel (*Afr* en *afr*); *lg.* (Afrikaanse Woordelys & Spelreëls, 2017:590) word ook op twee maniere gespel (*lg* en *Lg*); Johannesburg (Afrikaanse Woordelys & Spelreëls, 2017:586*)* word ook op twee verskillende maniere gespel (*Jhb* en *jhb.*) – sien voorbeeld 189 tot 195:

Voorbeeld 189: **O2** – Kan subs by #Beeld **ea Afr** dagblaaie asb ophou verwys na Doer Onder. Australië lê min of meer oos v SA, nie suid nie. Dis koloniale denke. [137]

Voorbeeld 190: **O2** – So nou kan jy vir R99 vyftien meestal **afr** tydskrifte per mnd elektronies kry. Nou wie wil Sarie op skerm lees? [110]

Voorbeeld 191: **O2** – Heinrich Marnitz het eerste **Afrikaans** op TV gepraat. Uitsendings was al om die ander aand eers **Afrikaans**, dan Engels. [117]

Voorbeeld 192: **O2** – Opvallend hoe positief en vry van ras 'rassistiese' Radio Pretoria oor swart SA atlete berig. **#afrikaans** [104]

Voorbeeld 193: **O2** – http://PolitiFact.com en http://AfricaCheck.org gaan feite na van stories. Oud-Puk Anim van Wyk nou in **Jhb** aan **lg** verbonde. Gaan kyk gerus. [140]

Voorbeeld 194: **O2** – Kan media dit nie regkry nie? Persent en persentasiepunt NIE dieselfde nie. **Lg** d verskil tussen 2 persentasies. #Somme101 [121]

Voorbeeld 195: **O2** – Kunswerk oor Sharpville by SAUK in **jhb.** Sal hulle ook een van Marikane opsit? [77]

Voorbeeld 189 tot 195 toon aan hoe O2 die verskillende idiolektiese afkortings vir *Johannesburg* en *Afrikaans* gebruik. Dit blyk asof daar nie 'n definitiewe patroonmatigheid te bespeur is wanneer die wisselvorme van die gebruik van idiolektiese afkortings ondersoek is nie. Voorbeeld 194 en 197 toon aan hoe hierdie afkortings ook nie noodwendig gebruik word om karakters te bespaar nie aangesien die twiet in voorbeeld 190 oor 110 karakters beskik asook 'n idiolektiese afkorting terwyl die twiet in voorbeeld 191 oor 117 karakters beskik en geen afkortings gebruik word nie.

Aangesien die afkortings in Figuur 7.17 aangedui, slegs in die korpus van O2 voorkom, is hierdie idiolektiese afkortings veral in die verdagte tekste ondersoek aangesien die gebruik van hierdie afkortings op die outeurskap van O2 kan dui.

O4 maak van drie idiolektiese afkortings gebruik wat slegs in die korpus van hierdie outeur voorkom. Voorbeeld 196 tot 198 toon O4 se gebruik van hierdie idiolektiese afkortings aan:

Voorbeeld 196: **O4** – Al hierdie ou Springbok vs All Black games op Super Sport maak my so op my senuwees soos **Mr** Zuma voor 'n budget speech! #BoksInASICS [132]

Voorbeeld 197: **O4** – @GEBRUIKER **Meneer**. Jip. Dis belaglik. Speel op http://bokrapport.com . Dis al oplossing. [88]

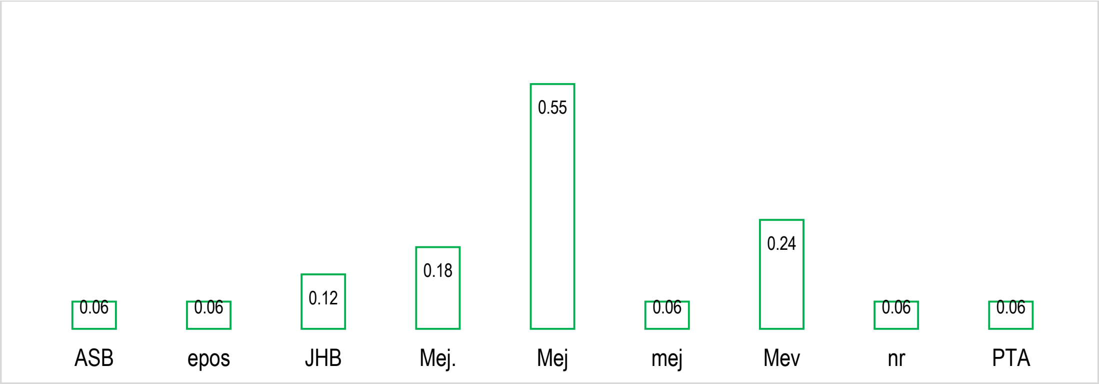
Voorbeeld 198: **O4** – Stem saam... Maar Lood se tackle count, ball carries, ruck count **ens** is net beter! En hy is solid in lynstane! [110]

Voorbeeld 196 tot 198 dui aan hoe O4 idiolektiese afkortings gebruik. Soos gesien in voorbeeld 196 gebruik O4 die Engelse afkorting *Mr* wanneer die outeur na President Zuma verwys. In voorbeeld 197 gebruik die outeur die woord *Meneer*. Dus kan daar nie aangevoer word dat die outeur slegs van afkortings (of Engelse afkortings) gebruik maak wanneer die woord *meneer* gebruik word nie.

Voorbeeld 198 dui die gebruik van die woord *ensovoorts* aan. Daar is geen voorkoms van die woord *ensovoorts* in enige van die oorblywende nege outeurs se korpora nie. O6 maak wel 'n enkele keer gebruik van die afkorting *ens.* (Afrikaanse Woordelys & Spelreëls, 2017:577), maar hierdie outeur gebruik die korrekte afkorting vir die spesifieke woord. Daarom kan die idiolektiese gebruik van die afkorting *ens* as 'n moontlike aanduiding van O4 se outeurskap in

hierdie studie beskou word. Soortgelyk aan die bevinding in die ondersoek na O1 se gebruik van afkortings (idiolekties) kan die karakterbeperking ook nie in die geval van O4 as die rede vir die gebruik van afkortings beskou word nie aangesien die langste twiet (voorbeeld 196) uit slegs 132 karakters bestaan.

Soos in Tabel 7.18 aangetoon, gebruik O6 35 verskillende idiolektiese afkortings. Slegs vier van hierdie afkortings (’n totaal van 12 afkortings of 0,73 − genormaliseer tot 1 000 woorde) kom ook in die ander nege korpora voor. Dit wil sê dat O6 nege keer (’n totaal van 23 of 1,41 – genormaliseer tot 1 000 woorde) idiolektiese afkortings gebruik wat in terme van hierdie studie as kenmerkend van O6 beskou kan word – sien Figuur 7.18. O6 maak ook van wisselvorme van idiolektiese afkortings gebruik en veral die afkorting mej. (Afrikaanse Woordelys & Spelreëls, 2017:593) word telkens deur die outeur gebruik (sien Figuur 7.18). Wanneer daar ondersoek ingestel is na die gebruik van *mejuffrou* blyk dit dat die woord glad nie in enige van die ander korpora gebruik word nie. Daarom kan die gebruik van hierdie afkorting asook die idiolektiese gebruik van die afkorting as kenmerkend tot die korpus van O6 beskou word.



Figuur 7.18 Die unieke gebruik van afkortings deur O6

Figuur 7.18 dui die unieke gebruik van afkortings (sommige idiolekties) aan soos dit deur O6 gebruik word. Die wisselvorme vir die afkorting van *mejuffrou* blyk die algemeenste afkorting in die outeur se korpus te wees. O6 maak van drie verskillende wisselvorme van die afkorting vir *mejuffrou* gebruik (sien voorbeeld 199 tot 201):

Voorbeeld 199: **O6** – Ek en Corlea is reg vir **mej** Veltieland in Upington. #backstage [62]

Voorbeeld 200: **O6** – Saam met die **Mej** Veltieland finaliste - (Hoërskool Duineveld) Dankie vir al die bederf, Upington! Was lekker om... http://fb.me/7vnKxb1rj [137]

Voorbeeld 201: **O6** – Kom skryf in vir Hartenbos se **Mej.** Bene! Of hoofrol in 49 Shades of Blue. 😝 [76]

Soortgelyk aan O2 kan karakterbeperkings nie as die rede vir die wisselvorme of selfs die gebruik van afkortings beskou word nie. Maar, soos in die geval van O6 kan die idiolektiese gebruik van afkortings outeurskap aandui en daarom is hierdie idiolektiese afkortings verder ondersoek wanneer die verdagte tekste geanaliseer is.

O7, O9 en O10 maak ook van idiolektiese afkortings in hulle onderskeie korpora gebruik, maar hierdie outeurs maak nie van wisselvorme van hierdie spesifieke afkortings gebruik nie. Verder gebruik O7 en O9 idiolektiese afkortings slegs een keer wat as kenmerkend tot hierdie twee outeurs se taalstyl beskou kan word. O9 en O10 maak slegs van die afkortings van die onderskeie woorde gebruik terwyl O7 die woorde ook volledig gebruik (sien voorbeeld 202 tot 20):

Voorbeeld 202: **O9** – Wanneer ek die woord 'hulle' of 'julle' gebruik sê ek in my kop "(h)uh vir haan" of "(j)i vir jas." Dankie Gr.2 **Juff**[320]**,** Adele. [124]

Voorbeeld 203: **O10** – Pollard se skopskoen vergroot die voorsprong vir RSA 10-0 teenoor **ARG**[321] #RWC2015 #RSAvARG [88]

Voorbeeld 204: **O10** − Tog so naby...tog sover!Geen verskonings. Credit to **NZ**[322]! A018B @GEBRUIKER @GEBRUIKER #RSAvsNZ #RWC201 [106]

Voorbeeld 205: **O7** – Ek lees **Gnr**[323] Christiaan de Wet se boek (ongelukkig die engelse een) "The three year War". Daai man was n reus. [134]

Voorbeeld 206: **O7** – My groot groot Oupa het die Joseph Roger in sy sak gedra. Volgens oorlewering het **Generaal** de Wet sy ingelsman geken [116]

Voorbeeld 207: **O7** – jay. kaap sien uit! @dieboerteater 22-25 **Mrt**[324] 021 979 1911 [57]

Voorbeeld 208: **O8** – @GEBRUIKER #Easter edition 26 & 27 **Maart** #spierwerfmarket Lekker!! [72]

Voorbeeld 209: **O7** – Dit stort neer hier in **pta**[325] [26]

Voorbeeld 202 tot 209 toon aan hoe O10, O9, O8 en O7 die idiolektiese afkortings (en in sommige gevalle die woord) gebruik. Dit blyk ook duidelik dat karakterbeperking ook nie in dié

[320] *juf.* (Afrikaanse Woordelys & Spelreëls, 2017:586).
[321] *Arg.* (Afrikaanse Woordelys & Spelreëls, 2017:567).
[322] *NS* (Afrikaanse Woordelys & Spelreëls, 2017:599).
[323] *genl.* (Afrikaanse Woordelys & Spelreëls, 2017:580).
[324] *Mrt.*(Afrikaanse Woordelys & Spelreëls, 2017:595).
[325] *Pta.* (Afrikaanse Woordelys & Spelreëls, 2017:606).

geval die gebruik van afkortings bevorder nie. Alhoewel daar in afdeling 7.4.1.1 bevind is dat O2 telkens van afkortings (foutloos) gebruik maak wanneer die 140 karakterbeperking genader word, blyk dit vanuit hierdie afdeling dat die outeurs nie noodwendig as gevolg van die karakterbeperking van afkortings gebruik maak nie.

Wat wel insiggewend blyk te wees vanuit hierdie afdeling is dat die idiolektiese gebruik van afkortings kenmerkend tot sommige outeurs se korpora beskou kan word. Nie al die outeurs gebruik konsekwent wisselvorme van afkortings of self net afkortings nie omdat daar in sommige korpora afkortings sowel as die woorde self voorkom. Daar kan sonder twyfel aangevoer word dat die idiolektiese gebruik van afkortings ʼn nuttige hulpmiddel is wanneer onbekende outeurs van mikrobloginskrywings geïdentifiseer moet word.

7.4.1.4 Verdagte tekste: idiolektiese gebruik van afkortings

In afdeling 7.4.1.3 word daar verwys na Coulthard (2005:15) se bevoegdheidsfoute. Vervolgens is daar in hierdie afdeling ondersoek ingestel na die gebruik van afkortings waarin daar foute voorkom in die verdagte tekste, of te wel afkortings waarin daar foute voorkom wat as idiolekties kenmerkend beskou kan word. Daar is vier afkortings geïdentifiseer in die drie verdagte tekste wat idiolekties gebruik word. Hierdie afkortings sluit in *antw*, *sg*, *vlg* en *milj* en word almal in die korpus van V1 aangetref (sien voorbeeld 210 [326], 211[327] en 212[328]):

Voorbeeld 210: **V1** – Die SAUK verloor geld weens wanbestuur, nie omdat hy radio- en TV-stasies het nie. Die **antw** is dus nie bloot privatisering nie, Dawie. [134]

Voorbeeld 211: **V1** – In d vorige bedeling was daar m ook hopeloos te veel **sg** SAUK-bestuurders. Eintlik net dophouers wat niks anders gedoen het as teedrink nie. [140]

Voorbeeld 212: **V1** – SAUK se salarisrekening beloop **vlg** STimes R3**milj** per jr. Watter deel daarvan kry mense wat inderdaad iets produseer wat uitsaaibaar is? [135]

In al drie die voorbeelde waar V1 idiolektiese afkortings gebruik, is die outeur baie naby aan die maksimum karakterdrumpel vir elk van die twiets, maar in die geval van voorbeeld 210 en 212, kon die outeur steeds die nodige punte by die afkortings geplaas het.

Vanuit die ondersoek na die gebruik van idiolektiese afkortings in die gespesialiseerde korpus is daar bevind dat daar etlike idiolektiese afkortings is wat slegs deur ʼn spesifieke outeur gebruik word. Die idiolektiese afkortings (al vier die foutiewe afkortings wat in die verdagte tekste gevind

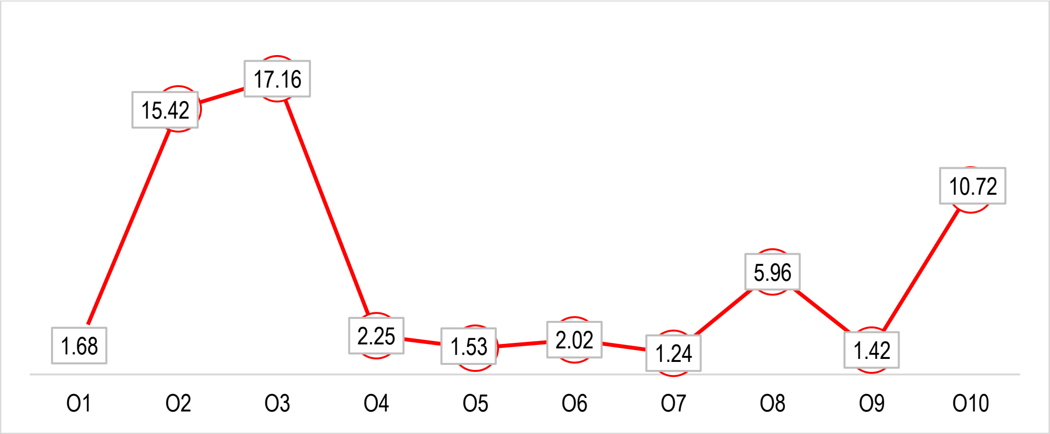
[326] **antw. (**Afrikaanse Woordelys & Spelreëls, 2017:567).
[327] **sg.** (Afrikaanse Woordelys & Spelreëls, 2017:613).
[328] **vlg.** (Afrikaanse Woordelys & Spelreëls, 2017:5621), **m.** (Afrikaanse Woordelys & Spelreëls, 2017:592).

word) wat in V1 gevind word, word slegs in die korpus van O2 gevind (sien Figuur 7.17). Alhoewel die uitlaat van die punte in die afkortings *antw.*, *sg.* en *vlg.* as 'n meer algemene bevoegdheidsfout beskou kan word, is dit veral die gebruik van die afkorting *milj* wat uniek tot die korpus van O2 blyk te wees en ook in die korpus van V1 voorkom wat as beduidend tot die outeurskap van die teks beskou kan word.

Die ondersoek na die idiolektiese gebruik van afkortings kan as suksesvol in die identifisering van outeurs van mikrobloginskrywings beskou word. Vervolgens is daar na die gebruik van akronieme in die gespesialiseerde korpora en die verdagte tekste ondersoek ingestel.

7.4.1.5 Gespesialiseerde korpus: akronieme

Die voorlaaste aspek wat in die afdeling oor afkortings ondersoek is, is die gebruik van akronieme. Soortgelyk aan die ondersoek oor die gebruik van afkortings (foutloos en idiolekties) is die akronieme in *ATLAS.ti* gekodeer en die gekodeerde dokumente is as Microsoft Excel-dokumente gestoor voordat die analises uitgevoer is. Figuur 7.19 toon die genormaliseerde waarde van die gebruik van akronieme aan:



Figuur 7.19 Genormaliseerde gebruik van akronieme deur O1 tot O10

Figuur 7.19 toon aan dat O3, O2 en O10 onderskeidelik van die meeste akronieme gebruik (O2 en O10 maak ook die meeste van afkortings – beide foutloos en idiolekties – gebruik). Daar is 134 akronieme geïdentifiseer in die tien korpora wat in die studie gebruik is en Tabel 7.21 verskaf 'n opsomming van die gebruik van akronieme:

Tabel 7.21 'n Opsomming van die gebruik van akronieme

	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Akronieme in korpus	15	189	58	16	7	33	7	22	7	83

Totale verskillende akronieme in korpus	9	28	16	14	6	16	6	11	7	20
Totale akronieme in korpus (genormaliseer tot 1 000 woorde)	1,68	15,62	17,16	2,25	1,53	2,02	1,24	5,96	1,42	10,72
Akronieme uniek tot korpus	6	12	3	6	1	7	2	5	3	11
Akronieme algemeen tot Twitter	3	16	13	8	5	9	4	6	5	9
Totaal van unieke akronieme in die korpus	9	19	7	7	2	10	2	5	3	42
Totaal van unieke akronieme in die korpus (genormaliseer tot 1 000 woorde)	1,01	1,55	2,07	0,98	0,44	0,61	0,35	1,36	0,61	5,42
Totaal van akronieme algemeen tot Twitter	6	170	51	9	5	23	5	17	5	41
Totaal van akronieme algemeen tot Twitter (genormaliseer tot 1 000 woorde)	0,67	13,87	879,31	1,27	1,09	1,41	0,88	4,61	1,02	5,29

Vanuit Tabel 7.21 kan die afleiding gemaak word dat elk van die outeurs in die gespesialiseerde korpora oor ŉ eiesoortige gebruik van akronieme beskik en ook dat daar wel akronieme gebruik word wat as algemeen tot Afrikaans soos dit op Twitter gebruik word, beskou kan word. Aangesien die doel van hierdie studie tweeledig was (om kenmerke van Afrikaans soos dit op Twitter gebruik word en kenmerke eie aan outeurs op die sosialemediaplatform te identifiseer), verwys ek eerstens na die algemene kenmerke wanneer akronieme op Twitter ondersoek is voordat ek na die kenmerkende gebruik van die onderskeie outeurs ondersoek ingestel het. Voorbeeld 213 tot 219 dui aan hoe hierdie akronieme in die gespesialiseerde korpus gebruik word:

Voorbeeld 213: **O9** – Tbh gee ek nie eintlik 'n fok om oor Inter/Nasionale Vrouedag nie. Waar is die "gelykheid" in dit? Feministe is erger as die **EFF**. [129]

Voorbeeld 214: **O4** – Ok... Daar gooi die Rebels gou die Crusaders rond soos **EFF** parlaments lede #SSRugby [83]

Voorbeeld 215: **O3** – .@GEBRUIKER Die **EFF**-leier, Julius Malema, staan die media te word [66]

Voorbeeld 216: **O2** – Dekking van gebeure in Parlement deeglik deur #media gedoen. En dit sluit #RSG en #SAfm beslis in. **DA**, **VF+** en **EFF** kry baie lugtyd. [130]

Voorbeeld 217: **O5** – As dit nie vir **Eskom** was nie, was daar seker nooit meer romantiese kerslig-etes nie. [83]

Voorbeeld 218: **O4** – Sien uit na die sharks game vanaand! **Eskom** moenie met hulle kak kom vanaand nie! As hulle iets wil shed, shed a tear!!@GEBRUIKER #SSRugby [138]

Voorbeeld 219: **O3** – Die **DA** se parlementêre-leier, Mmusi Maimane, is by **Eskom** se Megawatt-park om betoging te lei @GEBRUIKER [103]

Alhoewel daar agt akronieme geïdentifiseer is wat in drie of meer van die korpora aangetref word, toon voorbeeld 213 tot 219 die gebruik van akronieme aan wat eienskappe van die outeurs se taalgebruik openbaar. Wanneer daar na die gebruik van die akroniem *EFF* ondersoek ingestel is, blyk dit dat die kategorie[329] waarin die bepaalde outeur se Twitter-profiel geplaas is wel ’n invloed op die gebruik van die akroniem kan hê. Outeurs wat in die nuuskategorie geplaas is, gebruik die akroniem *EFF* in ’n politiese konteks terwyl die ander outeurs amper spottenderwys na *EFF* verwys ten einde ’n bepaalde boodskap oor te dra met ’n sosiaal gelaaide ondertoon (sien voorbeeld 213 tot 219).

Soortgelyk aan die gebruik van die akroniem *EFF* word *Eskom* meer formeel gebruik deur die outeurs wat in die nuuskategorie geplaas is terwyl outeurs wat in ander kategorieë val, meer spottenderwys na *Eskom* en die landwye “LOADSHEDDING-krises” verwys. Daar kan dus aangevoer word dat die gebruik van akronieme (wanneer ’n akroniem deur meer as een outeur gebruik word) meer lig werp op bepaalde outeurs se taalgebruik in die kategorie waarin die outeurs geplaas is.

Soos genoem, is daar 134 akronieme in die gespesialiseerde korpora geïdentifiseer waarvan 56 van hierdie akronieme deur slegs een van die outeurs gebruik word. Aangesien die akronieme wat slegs deur een van die outeurs gebruik word, belangrike inligting kan verskaf oor die outeurskap van die individu, word daar vervolgens na die kenmerkende gebruik van akronieme verwys. O9, O7, O5 en O3 se korpora beskik nie oor meer as drie akronieme nie – die gebruik van akronieme is dus nie kenmerkend tot hulle taalgebruik nie behalwe vir O3 se gebruik van *Sanral* (5 keer − 1,48 genormaliseer tot 1 000 woorde). Ten einde te kon bepaal of daar wel ooreenkomste tussen die gebruik van akronieme in die gespesialiseerde korpora en die verdagte tekste voorkom, is daar ondersoek ingestel na die gebruik van akronieme in die verdagte tekste.

7.4.1.6 Verdagte tekste: akronieme

Die ondersoek na die gebruik van akronieme in die verdagte tekste is uitgevoer soos uiteengesit in afdeling 7.4.1.5 (Gespesialiseerde korpora: akronieme). V1 maak 30 keer van akronieme gebruik (sien Tabel 7.22) terwyl V2 slegs een akroniem (*iOS*) gebruik. V3 gebruik geen akronieme in die teks wat vir die doel van die studie geanaliseer is nie. Die akroniem wat deur V2 gebruik word, kom glad nie in die gespesialiseerde korpus voor nie, maar hierdie

[329] Twitter-gebruikers is in vier groepe gedeel (sien afdeling 5.4 – Data-insameling en -verwerking).

akroniem kan nie as ŉ aanduiding van outeurskap beskou word nie aangesien dit wel in die verwysingskorpus en op Twitter voorkom.

Dit blyk dus dat die ondersoek na die gebruik van akronieme deur V2 en V3 nie enige aanduiding gee van wie die moontlike outeurs van die twee tekste kan wees nie. Daar word agt verskillende akronieme deur V1 gebruik en hierdie akronieme kan as insiggewend tot die identifikasie van outeurs van tekste beskou word.

Tabel 7.22 Die gebruik van akronieme in die verdagte tekste en die ooreenstemmende akronieme in die gespesialiseerde korpora

Verdagte teks	Akroniem	Frekwensie	Genormaliseerde waarde (1 000 woorde)	Gespesialiseerde korpus	Frekwensie	Genormaliseerde waarde (1 000 woorde)
V1	ANC	9	10,66	O2	19	1,57
V1	eNCA	1	1,18	*	*	*
V1	NP	1	1,18	O2	4	0,33
V1	RSA	1	1,18	*	*	*
V1	RSG	2	2,37	O2	30	2,48
V1	SAUK	19	22,51	O2	18	1,49
V1	sms	1	1,18	O7 en O8	1 en 1	0,27

V1 gebruik sewe verskillende akronieme waarvan vier ook in die korpus van O2 gevind kan word (sien Tabel 7.22). Daar kan dus aangevoer word dat 50% van die akroniemgebruik van V1 ook in die korpus van O2 voorkom. Een (*sms*) van die oorblywende vier akronieme word ook in die korpora van O7 en O8 aangetref, maar *eNCA* en *RSA* word slegs in die korpus van V1 aangetref (hierdie akronieme kom wel in die verwysingskorpus en op Twitter voor) – sien voorbeeld 2200 tot 223:

Voorbeeld 220: **V1** – Moegsien Williams is nou hoofred v Gupta-media; Anton Harber hoofred v **eNCA**. Destyds was albei redakteurs v alternatiewe blaaie vs regering.

Voorbeeld 221: **O2** – Eks-meisie "has again broken down", sê **eNCA**. OP se eks het bloot begin snik. Wanneer word snik "ineenstort". Of vertaal ek verkeerd? #Oscar

Voorbeeld 222: **V1** – In die 1990s is 'n string radiostasies vd SAUK geprivatiseer. Onthou Dawie Roodt dit nie? Buitendien is **RSG** geen las nie en maak geld.

Voorbeeld 223: **O2** – Wat probeer die SAUK bereik deur Kader Asmal se begrafnis op **RSG** uit te saai!

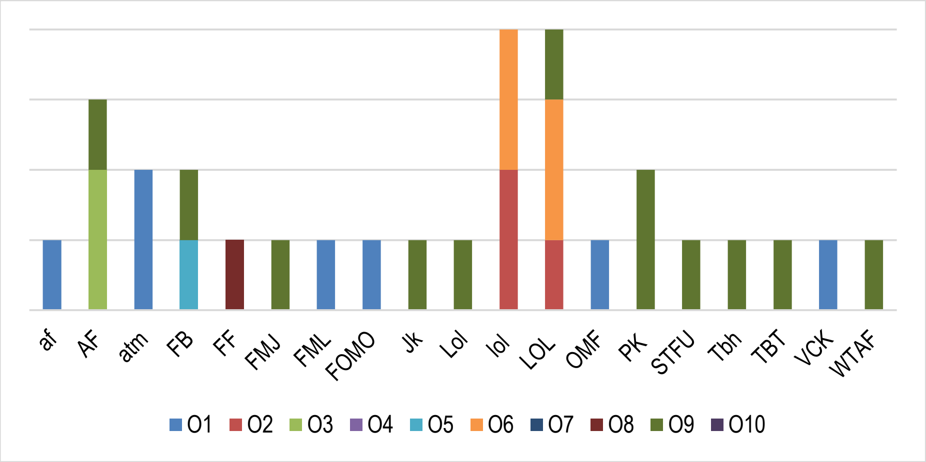
Voorbeeld 220 tot 223 toon aan hoe V1 en O2 akronieme in dieselfde konteks gebruik. Veral in voorbeeld 222 en 223 is dit interessant (en kenmerkend tot die outeur se taalstyl) dat *RSG* en *SAUK* saam in ŉ twiet gebruik word om sodoende die verband tussen die twee instansies te beklemtoon. Ook voorbeeld 220 en 221 stem ooreen aangesien beide die twiets oor ŉ

nuuswaardige aangeleentheid handel, maar in beide die twiets lewer die outeurs (V1 en O2) kommentaar op die nuuswaardige gebeure.

Na gelang van die analise van die gebruik van akronieme kan daar dus aangevoer word dat V1 deur O2 geskryf is. Ten einde die ondersoek na afkortings te kon voltooi, het ek die gebruik van SMS-afkortings geanaliseer wat deur die onderskeie outeurs van die gespesialiseerde korpora en in die verdagte tekste gebruik word.

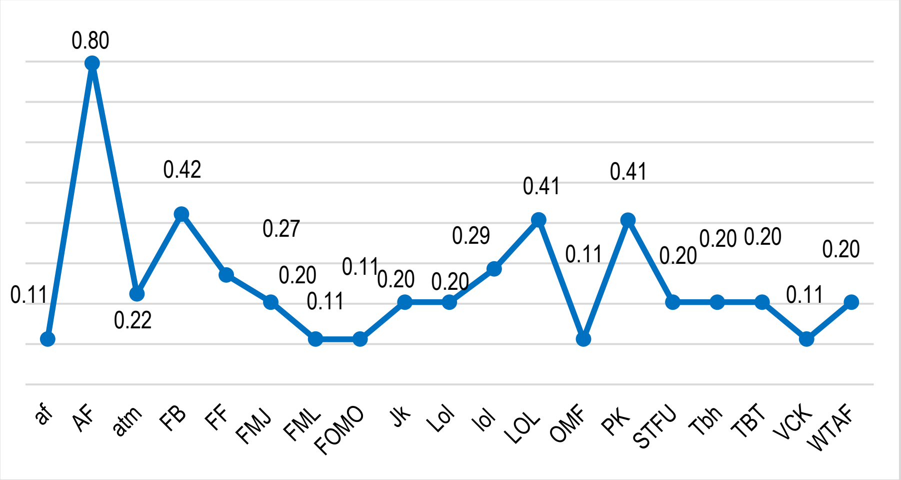
7.4.1.7 Gespesialiseerde korpus: SMS-afkortings

Soos in afdelings 7.4.1.1 tot 7.4.1.6 is die onderskeie korpora in *ATLAS.ti* geanaliseer en elke SMS-afkorting wat geïdentifiseer is, is gekodeer. Hierdie gekodeerde SMS-afkortings is as ’n Microsoft Excel-dokument gestoor. ’n Totaal van 18 verskillende SMS-afkortings is in die korpora van die tien outeurs geïdentifiseer (sien Figuur 7.20):



Figuur 7.20 Die gebruik van SMS-afkortings in die gespesialiseerde korpora

Figuur 7.19 dui die onderskeie outeurs se gebruik van die 18 verskillende SMS-afkortings aan. Vanuit Figuur 7.19 blyk dit dat O9 en O1 die meeste van SMS-afkortings in hulle korpora gebruik maak (genormaliseer tot 1 000 woorde). O9 gebruik 2,24 en O1 gebruik 0,78 SMS-afkortings terwyl O4, O7 en O10 geen SMS-afkortings gebruik nie. Alvorens daar na die kenmerkende gebruik van SMS-afkortings deur die sewe outeurs ondersoek ingestel is wat wel van hierdie soort afkortings gebruik maak, word die SMS-afkortings eers self voorhou (sien Figuur 7.21):



Figuur 7.21 Gewildste SMS-afkortings (genormaliseer tot 1 000 woorde) in die gespesialiseerde korpora

Figuur 7.21 toon aan dat *AF* (*as fuck*) die gewildste SMS-afkorting vanuit die gespesialiseerde korpora is wat in hierdie studie gebruik is (sien voorbeeld 224 en 225), gevolg deur die afkorting *FB* (vir Facebook – voorbeeld 226 tot 227) en *LOL* (*laugh out loud* – voorbeeld 228 tot 230):

Voorbeeld 224: **O9** – Me, stoned ***AF*** at Burger Bistro: "Excuse me Sir, what's the time frame on your Take Overs?" "Jammer, wat?" Pause. "Ek bedoel Take Aways."

Voorbeeld 225: **O1** – Victoria's Secret Fashion Show is bewys dat meisies die wêreld kan oorneem en sexy

af lyk terwyl hulle dit doen.

Voorbeeld 226: **O9** – As dit 2008 was en jy het vir my 'n geheime nommer geInbox op **FB** om op 'n status te sê wat ek rEgtIgiiz* van jou dink: 528262- Jy's 'n Poes.

Voorbeeld 227: **O5** – *gaan op **FB*** "ek iis soooo Liefiies ViiR jww my bokii beKi bAby** jYs mysE aLlEs!! MwAZzz* my eNgel liiEfling!!*" *skud kop* *hardloop weg*

Voorbeeld 228: **O9** – Fokken **Lol**.

Voorbeeld 229: **O6** – Dozi, Lydia en hul body guard. (wat amper so lank soos Dozi is) **lol** http://fb.me/4kf KovqUB

Voorbeeld 230: **O2** – Herschel Gibbs gekritiseer oor sy **LOL** aan einde v twiet oor Oscar se kolf. Neem darem in ag hy is net 'n eks-krieket- en rugbyspeler. **Lol**

Voorbeeld 224 tot 230 toon die verskillende maniere aan waarop die gewildste SMS-afkortings vanuit die gespesialiseerde korpora in hierdie studie gebruik is. Dit blyk asof die onderskeie outeurs hierdie gewilde SMS-afkortings op dieselfde manier gebruik − om meer emosie in ŉ bepaalde twiet oor te dra. O2 se gebruik van die SMS-afkorting *LOL* kan beskou word as sosiale kommentaar op die twiet van Herschel Gibbs waarin hy *LOL* gebruik het. O2 gebruik

geen ander SMS-afkortings nie en *LOL* word slegs in verwysing na die bogenoemde nuuswaardige voorval gebruik.

Ten einde te kon bepaal of enige van die outeurs SMS-afkortings in só ’n mate gebruik dat dit kenmerkend van ’n bepaalde outeur se taal- en skryfstyl is, het ek ondersoek ingestel na die SMS-afkortings wat slegs deur ’n enkele outeur gebruik word. Slegs drie van die tien outeurs wat in hierdie studie ondersoek is, gebruik SMS-afkortings wat in geen van die ander outeurs se korpora voorkom nie.

O1 se korpus bevat sewe SMS-afkortings waarvan slegs ses deur hierdie outeur gebruik word (sien voorbeeld 231, 232 tot 235):

Voorbeeld 231: **O1** – Emosionele toestand **atm**: 6 stukke koek opgeëet. *[at the moment]*

Voorbeeld 232: **O1** – Toe my wekker afgegaan het was my eerste woorde "**FML**". Dit behoort julle 'n idee te gee oor my mood vandag. [*fuck my life]*

Voorbeeld 233: **O1** – **FOMO**-lyer [*fear of missing out]*

Voorbeeld 234: **O1** – **OMF**😂😭😂😭😂😭😂😭😂 [*oh my fuck]*

Voorbeeld 235: **O1** – **VCK** MY NOU! [*vacation*]

Die SMS-afkortings wat deur O1 gebruik word, kom algemeen in die verwysingskorpus voor en dus is hierdie afkortings nie uniek tot die taalgebruik van O1 nie, maar aangesien O1 die enigste outeur in die gespesialiseerde korpus is wat hierdie afkortings gebruik, kan daar aangevoer word dat wanneer ’n verdagte teks hierdie afkortings bevat, kan O1 as ’n moontlike verdagte ondersoek word. O1 gebruik die afkortings om emosie in spesifieke twiets oor te dra. Daar kom slegs een SMS-afkorting in die korpus van O8 voor en hierdie afkorting word slegs deur O8 gebruik (sien voorbeeld 236):

Voorbeeld 236: **O8** – Ditto **#FF** [*follow Friday]*

O8 maak gebruik van die hutswoord *FF* (*follow Friday*). Die doel van die hutswoord is om ’n groep Twitter-gebruikers in die twiet te noem (deur van die @-gebruikerskonvensie gebruik te maak) en sodoende individue aan te moedig om mekaar te volg. Die manier waarop O8 van *#FF* gebruik maak, sluit nie by die norm van die hutswoord aan nie en kan dus as kenmerkend beskou word van O8 se taalgebruik.

O9 gebruik 12 SMS-afkortings waarvan 8 net in die outeur se korpus voorkom (sien voorbeeld 237 tot 242):

Voorbeeld 237: **O9** – **WTAF**. Hoe sterk kan iemand se lighter game wees? Is 'n lighter 'n valid rede om verlief te raak op iemand? **#FMJ** FOK. [*what the actual fuck* en *full metal jacket*]

Voorbeeld 238: **O9** – **Jk**. Wat was op met Spice Girls skoene, en waarom het ons ouers dit toegelaat? [*just kidding*]

Voorbeeld 239: **O9** – As nog een persoon "Gelukkige Valentynsdag" sê asof dit Kersfees in Februarie is gaan ek **PK**'s begin uitdeel. [*poes klap*]

Voorbeeld 240: **O9** – As jy weet jy kry vinnig koud moet jy dalk nie aantrek asof jy 'n prostituut deur Egipte se warmste somer is nie. Kry 'n trui en **STFU** [*shut the fuck up*]

Voorbeeld 241: **O9** – **Tbh** gee ek nie eintlik 'n fok om oor Inter/Nasionale Vrouedag nie. Waar is die "gelykheid" in dit? Feministe is erger as die EFF. [*truth be told*]

Voorbeeld 242: **O9** – I say I don't love you, but you know I'm a liar. `Cause when we kiss... Oooh...Fire. 🔥🎵👌 **#TBT** #Fok#Ja#VolumeOop [*throw back Thursday*]

O9 se gebruik van *FMJ* (voorbeeld 237) kan dalk ’n spelfout wees – die outeur wou dalk *FML* (*fuck my life*) gebruik het, want die J en L is langs mekaar op ’n QWERTY sleutelbord. Vanuit die konteks van die twiet wil dit tog voorkom asof *FML* meer gepas sou wees. Voorbeeld 237 tot 242 toon O9 se gebruik van SMS-afkorting aan. Dit is duidelik dat die outeur oorwegend van SMS-afkortings gebruik maak waarin daar kraswoorde voorkom. Hierdie tendens van O9 kan as ’n kenmerkende eienskap van die outeur se taalgebruik op Twitter beskou word waar daar ook SMS-afkortings met kraswoorde voorkom wat deur van die ander outeurs in die gespesialiseerde korpus (O1) gebruik word. Omdat hierdie soort SMS-afkortings algemeen in die verwysingskorpus gevind word, gee O9 voorkeur aan die gebruik van hierdie soort SMS-afkortings.

Vanuit die ondersoek na die gebruik van SMS-afkortings kan daar dus aangevoer word dat sekere afkortings meer gewild is as ander en dat die gebruik van hierdie SMS-afkortings as ’n kenmerk van Afrikaans soos dit op Twitter gebruik word, beskou kan word. Wanneer die onderskeie outeurs se “unieke” gebruik van SMS-afkortings ontleed is, blyk dit dat − soortgelyk aan die bevindinge in afdelings 7.4.1.1 tot 7.4.1.6 − die voorkoms van enkele van die afkortings in ŉ verdagte teks beskou word as ’n moontlike aanduiding van outeurskap en moet die teks verder ondersoek word om te kan bepaal of daar meer ooreenkomste is tussen die verdagte teks en die outeur wat van die SMS-afkorting gebruik maak alhoewel die afkortings ook in die verwysingskorpus en op Twitter self voorkom.

7.4.1.8 Verdagte tekste: SMS-afkortings

Die ondersoek na die gebruik van SMS-afkortings is uitgevoer soos in afdeling 7.4.1.7 voorgehou. Hierdie ondersoek is nie net belangrik om die ondersoek na die gebruik van afkortings te kon voltooi nie, maar is van belang aangesien die sosiale aard van die gekose platform aanleiding daartoe gee dat “nuwe” en “vindingryke” afkortings – soos die afkortings wat in SMS-taal gebruik word – ook op Twitter gebruik word.

ʼn Totaal van vier SMS-afkortings is in die drie verdagte tekste gevind en al vier hierdie afkortings kom in V2 voor. Daar is nie enige ooreenkomste in die gebruik van hierdie SMS-afkortings wat verband hou met ʼn spesifieke outeur vanuit die gespesialiseerde korpus nie, maar die gebruik van die SMS-afkortings sluit wel aan by die sosiale aard van die platform. Dit kom dus voor dat die lae frekwensie van die gebruik van SMS-afkortings in die verdagte tekste die gebruik van hierdie morfologiese aspek in die identifisering van outeurs van mikrobloginskrywings onmoontlik maak. Na gelang van die bevindinge in afdeling 7.4.1.7 is dit steeds ʼn waardevolle metode van ondersoek wat sonder twyfel ʼn bydra kan lewer in die identifisering van outeurs van mikrobloginskrywings.

7.4.2 Inkorting

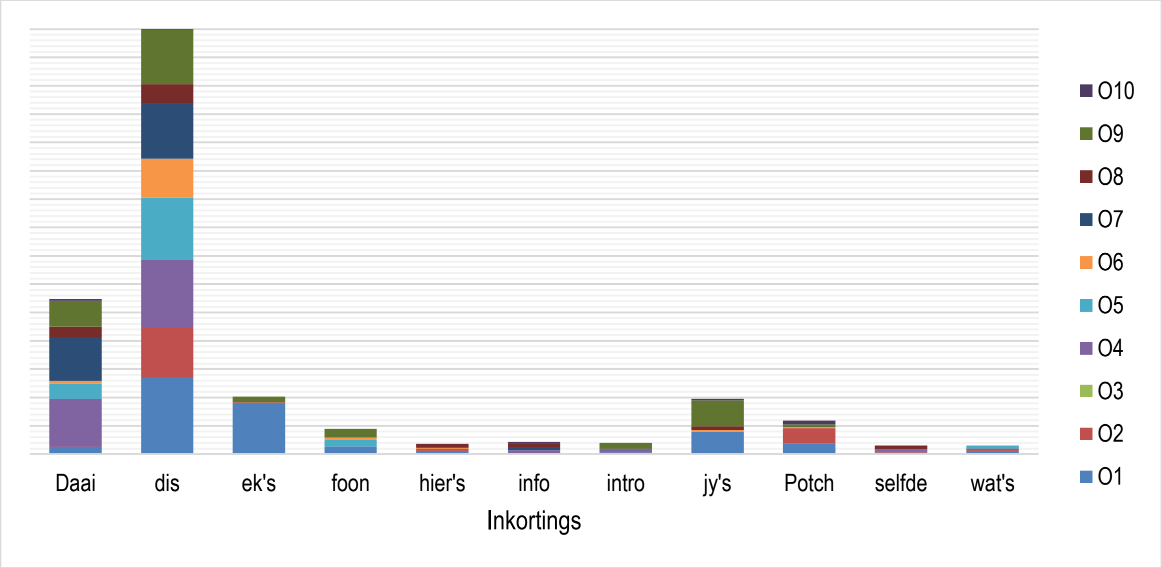
Van Huyssteen (2014:198) argumenteer dat taalgebruikers dikwels ter wille van ekonomie in taalgebruik van inkorting gebruik maak om iets kort en kragtig te stel. Met die argument van Van Huyssteen in gedagte, kan die aanname gemaak word dat Twitter-gebruikers veral van inkorting gebruik sal maak om hulle taalgebruik meer ekonomies te maak (veral wanneer die karakterbeperking van Twitter in gedagte gehou word).

In die korpora van die tien outeurs wat ondersoek is vir die doel van hierdie studie is daar 541 gevalle van inkorting geïdentifiseer[330]. Die inkortings is in *ATLAS.ti* gekodeer en daarna as ʼn Microsoft Excel-dokument gestoor. Die inkortings is ook as ʼn txt-dokument gestoor en as ʼn stoplys in WordSmith Tools ingevoer en elk van die outeurs se korpora is ondersoek ten einde te verseker dat al die inkortings getel is. Ten einde die analise te kon uitvoer, word daar eerstens na die gebruik van inkortings in die gespesialiseerde korpus verwys alvorens die verdagte tekste geanaliseer is.

[330] Sien Bylaag G.

7.4.2.1 Gespesialiseerde korpus: inkorting

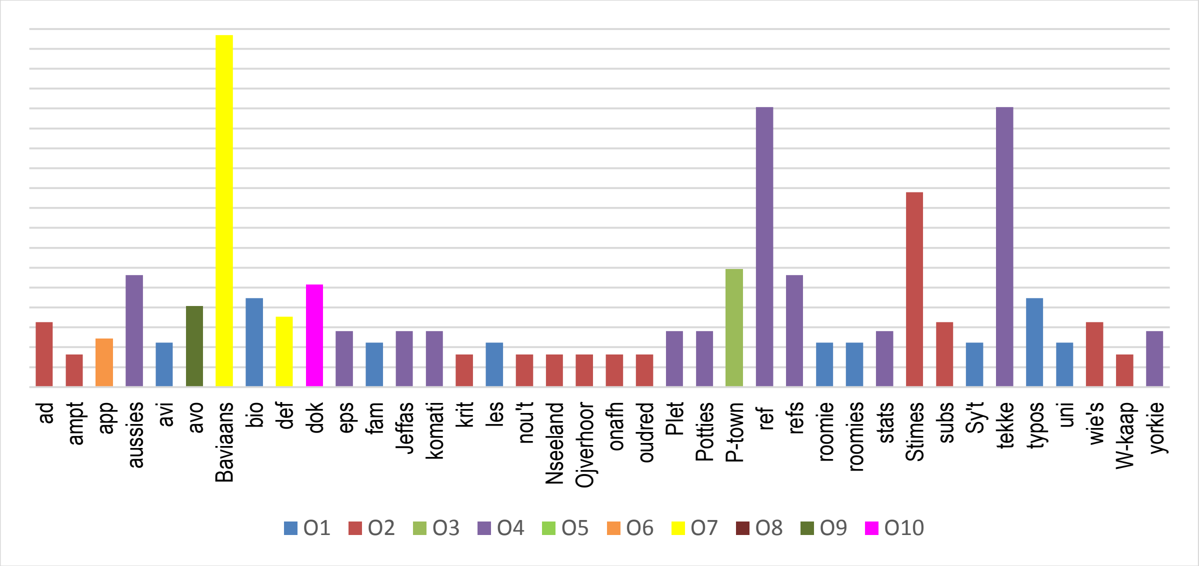
Die algemeenste voorbeelde van inkortings word in Figuur 7.22 uitgebeeld (inkortings wat deur drie of meer outeurs gebruik word):



Figuur 7.22 Gewildste inkortings wat deur die onderskeie outeurs in die gespesialiseerde korpus gebruik word

Soos sigbaar in Figuur 7.22 is daar 11 inkortings wat deur drie of meer outeurs gebruik word. Hierdie inkortings sluit in *daai*, *dis*, *ek's*, *foon*, *hier's*, *info*, *intro*, *jy's*, *Potch*, *selfde* en *wat's*. Die gewildste inkortings blyk *dis* (ʼn totaal van 30,23 genormaliseer tot 1 000 woorde) en *daai* (ʼn totaal van 10,96 genormaliseer tot 1 000 woorde) te wees. Hierdie twee inkortings word deur al die outeurs behalwe O3 gebruik, maar weer eens is daar ʼn duidelike onderskeid tussen die frekwensie waarteen die inkortings in die onderskeie korpora voorkom (sien ook Figuur 7.22).

Hierdie woorde soos uitgebeeld in Figuur 7.22 kan vervolgens as inkortings beskou word wat kenmerkend aan Afrikaans is soos dit op Twitter gebruik word. Aangesien daar in hierdie studie ook ondersoek ingestel is na die identifisering van die outeurs van mikrobloginskrywings, het ek vervolgens ondersoek ingestel na inkortings wat slegs deur een outeur in die studie gebruik word. 38 van die 61 inkortings word slegs in die korpus van ʼn enkele outeur aangetref (sien Figuur 7.23):



Figuur 7.23 Inkortings wat slegs aangetref word in die korpus van ’n enkele outeur vanuit die gespesialiseerde korpus

Die inkortings wat slegs in die korpus van ’n enkele outeur voorkom, word in Figuur 7.23 aangedui (genormaliseer tot 1 000 woorde). Vanuit hierdie analise blyk dit dat van die inkortings as uniek tot die korpora van ’n bepaalde outeur beskou kan word. Dit is veral O4 wat gereeld van inkortings soos *aussies*, *eps*, *Jeffas*, *komati*, *Plet*, *Potties*, *ref*, *refs*, *stats*, *tekke* en *yorkie* gebruik maak. Hierdie inkortings word in die verwysingskorpus ook aangetref, maar die frekwensie waarteen die inkortings in die onderskeie korpora voorkom kan as ’n middel gebruik word om outeurskap te bepaal.

7.4.2.2 Verdagte tekste: inkorting

In afdeling 7.4.2.1 is daar ondersoek ingestel na die gebruik van inkortings in die gespesialiseerde korpus. In hierdie afdeling is daar ondersoek ingestel na die gebruik van inkortings soos dit in die verdagte tekste voorkom. Ten einde te kon bepaal hoe die onderskeie outeurs van die verdagte tekste van inkortings gebruik maak, is die verdagte tekste geanaliseer soos in afdeling 7.4.2.1 voorgehou.

Tabel 7.23 Voorkoms van inkortings in die verdagte tekste

Inkorting	**Outeur**	**Frekwensie**	**Genormaliseerde waarde (1 000 woorde)**
Sol-Afriforum	V1	2	2,37
Stimes	V1	1	1,18
100ps	V1	1	1,18
hoofred	V1	2	2,37
Potch	V2	1	0,71
bio	V3	1	2,43
tho	V3	1	2,43

In Tabel 7.23 word die inkortings voorgehou soos dit in die verdagte tekste voorkom. Dit blyk dat V1 die meeste van inkortings gebruik maak – daar is vier voorbeelde van inkorting in die korpus van V1 terwyl V2 slegs een keer van inkortings gebruik maak en V3 twee keer. V1 gebruik ŉ totaal van ses inkortings (genormaliseer tot 7,11), V2 gebruik een inkorting (genormaliseer tot 0,71) en V3 gebruik twee inkortings (genormaliseer tot 4,85). Deur die genormaliseerde waarde van die voorkoms van inkortings in die drie verdagte tekste met dié van die gespesialiseerde korps te vergelyk, is daar bevind dat V1 (7,11) ooreenstem met O2 (7,10); V2 (0,71) en O3 (0,30) stem tot ŉ mate ooreen terwyl V3 (4,85) en O6 (4,22) as soortgelyk beskou kan word na gelang van die genormaliseerde waarde van die gebruik van inkortings.

Vanuit die vier inkortings wat deur V1 gebruik word (sien Tabel 7.23), kom slegs een in die gespesialiseerde korpus voor (*STimes* – O2). Hierdie inkorting word slegs deur O2 en V1 gebruik en dus kan O2 as die outeur van V1 beskou word. V2 gebruik die inkorting *Potch*, ŉ inkorting wat deur vyf van die outeurs in die gespesialiseerde korpus gebruik word en dus kan daar aangevoer word dat V2 se gebruik van inkortings nie as beduidend tot die identifisering van die outeur van ŉ teks beskou kan word nie. Van die twee inkortings wat deur V3 (*bio* en *tho*) gebruik word, kom *bio* in die korpus van O1 voor en dus kan O1 by die lys van moontlike outeurs van V3 gevoeg word. Die ondersoek na die gebruik van inkortings dra steeds by tot die ondersoek na die identifisering van outeurs van mikrobloginskrywings en kan vervolgens in die ondersoek na die outeurskapidentifikasie van mikroblogtekste ingesluit word.

7.4.3 Weglating

Weglating, soos in afdeling 7.4 voorgehou, word deur Thiart (2014:80) beskou as ŉ proses wat met sinkopee[331] in formele taalkunde vergelyk kan word en voer verder aan dat dit na die weglating van ŉ letter of ŉ klank in ŉ woord verwys. Ten einde die stilistiese analise oor die morfologie te kon voltooi, het ek ondersoek ingestel na weglating (soos deur Thiart voorgehou[332]).

Soos in afdelings 7.4.1 en 7.4.2 is die korpora van die tien outeurs wat die gespesialiseerde korpus vorm sowel as die verdagte tekste geanaliseer ten einde die gevalle waar weglating in die korpora voorkom, te identifiseer. Ten einde hierdie ondersoek te kon uitvoer, is daar eerstens na die analise van die gebruik van weglating in die gespesialiseerde korpus ondersoek ingestel alvorens die verdagte tekste geanaliseer is.

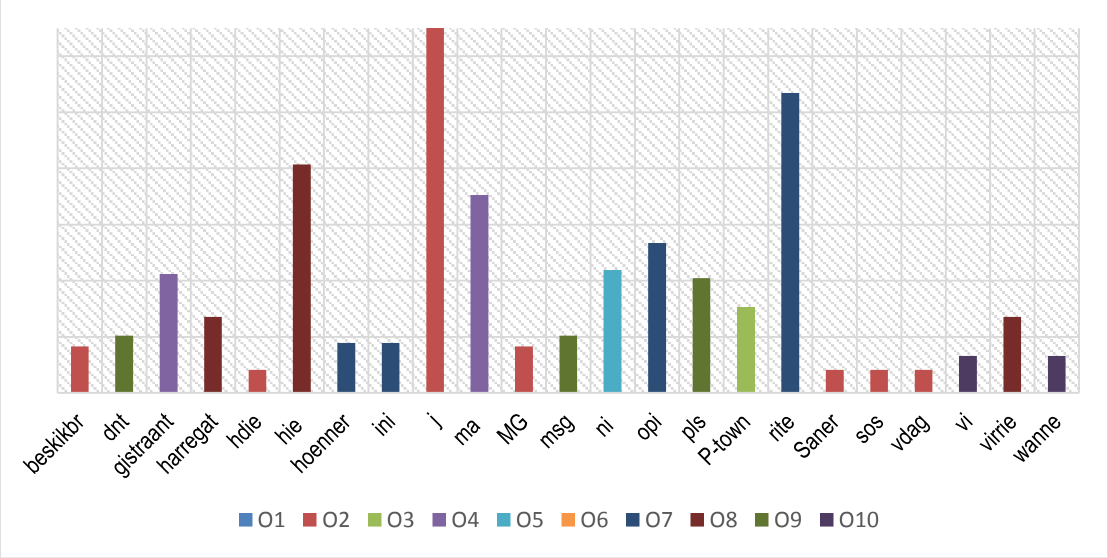
---

[331] Sinkopee verwys volgens Crystal (2008:469) na die weglating van 'n klinker in 'n woord.

[332] Voorbeelde van weglating sluit in *wanr* –wanneer, *hkm* – hoekom (Thiart, 2014:80). Crystal (2008:43) verwys ook na afkortings soos *brb – be right back*, *gtg – got to go*.

7.4.3.1 Gespesialiseerde korpus: weglating

In die gespesialiseerde korpus is daar 598 gevalle geïdentifiseer waar die outeurs van weglating gebruik maak[333]. Die weglatings is in *ATLAS.ti* gekodeer waarna dit as ŉ Microsoft Excel-dokument gestoor is. Ten einde seker te maak dat alle voorkomste van weglatings wel geïdentifiseer is, is ŉ lys van al die weglatings saamgestel wat as ŉ stoplys in WordSmith Tools ingevoer is. Figuur 7.24 dui die weglatings aan wat in die gespesialiseerde korpus geïdentifiseer is:



Figuur 7.24 Kenmerkende weglatings vanuit die gespesialiseerde korpus

Vanuit Figuur 7.24 blyk dit dat O2 die meeste van weglatings gebruik maak wat nie deur die ander outeurs gebruik word nie (sien voorbeeld 243 tot 246):

Voorbeeld 243: **O2** – Dane Beisheim v Ofm herinner my daaraan dat TV-uitsendings vandag 40 **j** gelede amptelik in SA begin het. [103]

Voorbeeld 244: **O2** – #Beeld se verlies aan Jan-Jan Joubert is wel Sunday Times se wins. Lees **j** graag daar, J-J! #SundayTimes [103]

Voorbeeld 245: **O2** – Wat ek bedoel is: ek is nie tipiese **SAner** nie en oorleef o.m. danksy wye aanbod v DStv. #Afrikaans #SkyNews #EngelsePrimierligaSokker [133]

Voorbeeld 246: **O2** – Lees Tim du Plessis se trefseker-artikel in **vdag** se #Beeld oor #OscarPistorius, rugby wit mans en die twak wat BBC en ander so graag spin. [138]

Die gebruik van weglatings in voorbeeld 243 tot 246 toon aan hoe O2 − ten spyte daarvan dat die twiets nie die karakterbeperking se afsnypunt nader nie − kies om van weglating gebruik te

[333] Sien Bylaag G (Inkorting).

maak. In die geval van voorbeeld 26 is die karaktertelling wel 138, maar selfs al het O2 van *vdag* (vandag) gebruik gemaak, sou die karaktertelling 140 gewees het. O2 maak dus die keuse om doelbewus van weglating gebruik te maak. Die gebruik van weglating kan as ʼn kenmerk van die sosiale aard van die platform beskou word asook ʼn kenmerk van Afrikaans soos dit op Twitter gebruik word. Bykomende voorbeelde van weglating soos deur ʼn enkele outeur in die gespesialiseerde korpus gebruik, sluit in:

Voorbeeld 247: **O9** – Wanneer jy met 'n babelas wakker gebel- en herinner word dat jy Stilfontein toe gaan. **Pls** tiny baby Jesus, no. [110]

Voorbeeld 248: **O8** – 'n Lekker soete **virrie** koue! 🍷 Nou's @GEBRUIKER reg vir die rugby [71]

Voorbeeld 249: **O4** – Woensdag! Afdraende van hier af... **Ma** dra maar nogsteeds 'n paar knee-en-elbow pads, 'n man se voet kan vas haak #sementburn [124]

Voorbeeld 250: **O1** – Redes hoekom ek nie dronk te word ni ; ek hou van weet wat ek doen en waar ek is. [81]

Voorbeeld 247 tot 250 dui aan hoe weglatings deur O9, O8, O4 en O1 gebruik word. Soortgelyk aan O2 kan daar nie aangevoer word dat die karakterbeperking die rede is waarom die outeurs van weglating gebruik maak nie. Alhoewel die weglatings wel in die verwysingskorpus aangetref word, is die weglatings wat slegs in die korpus van ʼn enkele outeur in die studie aangetref word insiggewend aangesien dit as ʼn manier beskou kan word om outeurskap te bewys. Daar is verder ondersoek ingestel na weglatings wat kenmerkend van ʼn enkele outeur se taalgebruik beskou kan word met die analisering van die verdagte tekste. Die verdagte tekste is vervolgens geanaliseer om te kon bepaal of weglating wel in die studie die outeurs van die verdagte tekste kon help met outeurskapidentifikasie.

7.4.3.2 Verdagte tekste: weglating

Die ondersoek na die voorkoms van weglatings in die verdagte tekste is uitgevoer soos voorgehou in afdeling 7.4.3. Tabel 7.24 beeld die resultate na die gebruik van weglatings in die verdagte tekste uit:

Tabel 7.24 Weglatings soos geïdentifiseer in die verdagte tekste

Weglating	Outeur	Frekwensie	Genormaliseerde waarde (1 000 woorde)
vd	V1	1	1,18
d	V1	3	3,55
v	V1	10	11,85
jr	V1	1	1,18

Vanuit Tabel 7.24 blyk dit dat daar slegs in die eerste verdagte teks (V1) voorbeelde van weglatings voorkom. Daar is vier verskillende weglatings geïdentifiseer. V1 maak gebruik van ʼn

totaal van 15 weglatings – genormaliseer tot 1 000 woorde − 17,77. Die enigste outeur vanuit die gespesialiseerde korpus wat min of meer dieselfde aantal weglatings gebruik, is O2 (13,96).

Die woorde waarin weglatings voorkom (sien Tabel 7.25), word gevind in die korpora van O2 (*d*, *vd* en *jr*), O6 (*jr* en *v*), O7 (*vd*) en O8 (*vd*). Alhoewel daar dus vier outeurs is wat ook van die weglatings gebruik wat in V1 voorkom, is dit O2 waarin daar die meeste ooreenkomste gevind kon word (sien voorbeeld 251 en 252):

Voorbeeld 251: **V1** – SAUK se salarisrekening beloop vlg STimes R3milj per **jr**. Watter deel daarvan kry mense wat inderdaad iets produseer wat uitsaaibaar is? [135]

Voorbeeld 252: **V1** – Ek vermoed minder as een derde **v** SAUK se mense dra regtig by tot uitsaai-inhou. Wat doen **d** res? Geboue vol mense wat salarisse kry vir wat? [139]

Vanuit die voorbeelde van V1 se gebruik van weglatings kan daar afgelei word dat die outeur (in die geval van voorbeeld 254) soms van weglatings gebruik maak wanneer die maksimum drempel vir die karakterbeperking genader word. Wanneer voorbeeld 251 en voorbeeld 243 met mekaar vergelyk word, kan die aanname gemaak word dat O2 en V1 weglating (*jr*) gebruik alhoewel die maksimum drempel vir die karakters nie oorskry word as *jaar* uitgeskryf sou word nie. Verder is daar ook ŉ tematiese ooreenkoms in die twiets aanwesig aangesien beide outeurs inligting oor die Suid-Afrikaanse media aan lesers oordra.

Voorbeeld 253 stem ook ooreen met die skryfstyl van O2 (vergelyk voorbeeld 253 met 246):

Voorbeeld 253: **O2** – Kern **v** saak lyk vir my **d** verskil tussen onwettige kommentaar en opmerkings wat bloot onwenslik is. Moet lg ook geskrap word? [124]

Wanneer voorbeeld 254 met 246 vergelyk word, kan daar weer eens ŉ ooreenkoms in die twee twiets waargeneem word. Nie net word daar vrae oor sosiale kwessies in die twiets gevra nie, maar die gebruik van die weglating blyk soortgelyk te wees. Na gelang van die ondersoek na die gebruik van weglatings kan O2 as die moontlike outeur van V1 beskou word.

Die morfologiese ondersoek toon dus aan dat outeurs van mikrobloginskrywings wel deur die analise van afkortings, inkortings en weglatings geïdentifiseer kan word. Alhoewel nie alle outeurs van reduksieprosesse − soos in afdeling 7.4 voorgehou − gebruik maak nie, is die ondersoek na die verskillende reduksieprosesse wel insiggewend en kan dus in ondersoeke as ŉ belangrike hulpmiddel gebruik word om die identifikasie van mikroblogouteurs te bepaal. In die hieropvolgende afdeling is daar ondersoek ingestel na die semantiese afwykings in die

gespesialiseerde korpus en die verdagte tekste om te kon bepaal of hierdie aspekte ook as ŉ suksesvolle manier beskou kan word om outeurs van mikrobloginskrywings te identifiseer.

7.5 Semantiek

Semantiek verwys na die betekenis in taal en kan ook beskou word as die analise van sinne en leksikale items (Crystal, 2008:429). Crystal (2011:8) verwys na die studie wat Charles Morris in 1946 gedoen het waarin semantiek beskryf word as die interpretasie van tekens − tekens verwys in ’n abstrakte manier na alles wat inligting oordra. De Stadler (1989:3, 9) definieer semantiek as die studie van betekenis waartydens betekenis gelykgestel word aan die verwysing van items en waar verwysings beskou word as die verhouding tussen woorde en die buitetalige werklikheid. Met betrekking tot die verhouding tussen woorde en die buitetalige betekenis, stel Akmajian *et al.* (2010:227) dat daar verskille in die betekenis van woorde kan heers soos dit deur verskillende groepe (dialekte) en individue (idiolekte) gebruik word om ŉ gewenste boodskap oor te dra.

Crystal (2011:75, 76) voer aan dat die internet sonder twyfel tot die ontstaan van nuwe linguistiese vorms van taal aanleiding gegee het en dat elke platform (tegnologiese ontwikkeling) oor ’n eiesoortige taalvorm/taalgebruik beskik – Crystal verwys na die taalvorms as variëteite – en dat hierdie taalvorms wissel in terme van fonologie, woordeskat en diskoers. Die semantiese waarde word dus deur verskeie faktore beïnvloed en dit is moontlik dat individue ’n eiesoortige betekeniswaarde aan bepaalde woorde (morfeme) kan heg, maar selfs indien individue se betekeniswaarde verskil van dít wat as die norm aanvaar word, is dit moontlik dat die boodskap (kommunikasie) steeds suksesvol in ŉ bepaalde konteks oorgedra kan word.

Die gebruik van emotikons word vervolgens ondersoek, ten einde die semantiese waarde en uniekheid van hierdie tekens in twiets te bepaal.

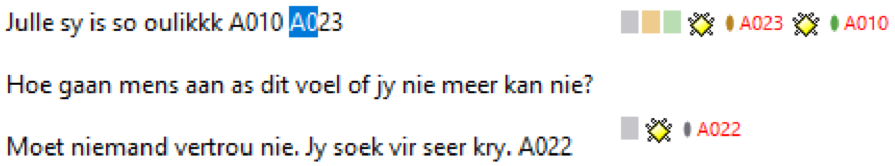
7.5.1 Emotikons

Emotikons is deur Go *et al.* (2009) as een van die vyf unieke kenmerke van Twitter geïdentifiseer[334]. Volgens Go *et al.* (2009) maak gebruikers deurentyd van emotikons gebruik om emosies uit te druk. Go *et al*. (2009) het emotikons vanuit die datastelle verwyder ten einde hulle navorsing uit te voer. Crystal (2011:46) stel dat emotikons wel as ŉ bykomende kenmerk beskou kan word in die identifisering van outeurs se taalgebruik en dat die gebruik van emotikons ŉ positiewe of negatiewe gevoel oordra. Ook Zappavigna (2015:71) voer aan dat die

[334] Sien hoofstuk 2 (afdeling 2.6.2.2 – Twitter).

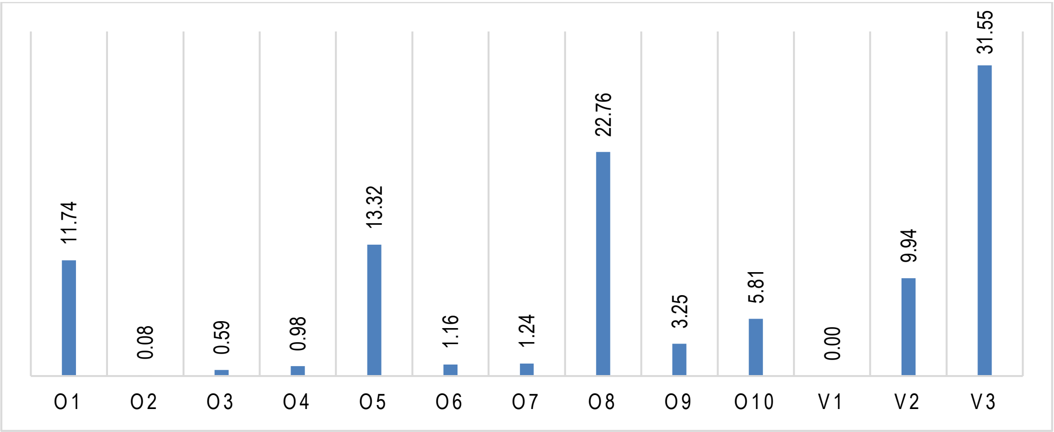
gebruik van emotikons 'n algemene kenmerk van mikroblogs is wat stilistiese en tekstuele voorstellings van gesigsuitdrukkings is en dat die gebruik van emotikons met sprekers se emosionele ingesteldheid, die boodskap en die gevoel verband hou wat hulle probeer om oor te dra. Die ondersoek na die gebruik van emotikons het nuttige inligting oor die boodskap wat outeurs wil oordra en die identiteit van 'n outeur aan die lig gebring.

Ten einde die gebruik van emotikons te kon analiseer, is daar 'n kode aan elk van die emotikons toegeken[335]. Die emotikons is vervolgens met 'n toepaslike kode vervang – sien Bylaag C – ten einde die teks leesbaar vir Wordsmith Tools te maak. Die onderskeie tekste is (as txt formaat) in *ATLAS.ti* ingevoer en die emotikons is gemerk met 'n *ATLAS.ti* kode (sien Skermskoot 7.2) om te kon bepaal hoeveel keer die onderskeie emotikons gebruik word. Die konkordansie van die korpora is vervolgens ook geanaliseer om ondersoek in te stel na die konteks waarin die outeurs die emotikons gebruik.



Skermskoot 7.2 'n Voorbeeld van die manier waarop die emotikons in *ATLAS.ti* gemerk is

Nadat die emotikons in *ATLAS.ti* gemerk is, is die data as 'n Microsoft Excel-lêer gestoor. Soos in die vorige afdelings is daar eerstens 'n interne vergelyking uitgevoer om te kon bepaal of daar ooreenkomste tussen die onderskeie outeurs van die gespesialiseerde korpus se eie tekste is met betrekking tot die gebruik van emotikons. Daarna is die gebruik van emotikons tussen die verskillende outers van die gespesialiseerde korpus vergelyk en laastens, is die gebruik van emotikons in die verdagte tekste geanaliseer. Sien Figuur 7.25:



Figuur 7.25 Gebruik van emotikons in die gespesialiseerde korpus sowel as die verdagte tekste

[335] Sien afdeling 5.12.2 (Dataverwerking).

In Figuur 7.25 word die genormaliseerde waarde (tot 1 000 woorde) van die gebruik van emotikons in beide die gespesialiseerde korpus en die verdagte tekste aangedui. Vanuit die grafiek is dit duidelik dat elk van die outeurs emotikons teen ŉ unieke frekwensie gebruik. In afdeling 7.5.1.1 en 7.5.1.2 word ŉ opsomming van die bogenoemde ondersoek se bevindinge voorgehou.

7.5.1.1 Gespesialiseerde korpus: emotikons

ŉ Vergelyking tussen die gebruik van emotikons in die interne korpora van die gespesialiseerde korpus is uitgevoer om te kon bepaal of daar wesenlike ooreenkomste in die korpus van ŉ enkele outeur voorkom. Die aantal emotikons wat in elk van die korpora voorkom, is geïdentifiseer en die genormaliseerde waarde waarby elk van die emotikons voorkom, is ook bereken.

Eerstens, is daar bevind dat elk van die outeurs emotikons (sien Figuur 7.25) teen ŉ unieke frekwensie gebruik. Tweedens, het die interne vergelyking van die gespesialiseerde korpus aangetoon dat die outeurs se gebruik van emotikons wel as ooreenstemmend beskou kan word. Dus kan die frekwensie waarteen die onderskeie outeurs emotikons gebruik wel as ’n moontlike aanduiding van outeurskap beskou word. Die totale gebruik van emotikons is ondersoek om te bepaal of daar onderskeid getref kan word tussen die verskillende outeurs in die gespesialiseerde korpus.

Soos sigbaar in Figuur 7.25 maak O8 (22,76) die meeste van emotikons gebruik. O1 (11,64) en O5 (13,32) gebruik ook emotikons aansienlik meer frekwent as die oorblywende sewe korpora in die gespesialiseerde korpus. O2 (0,08), O3 (0,59) en O4 (0,98) toon die laagste waardes vir die gebruik van emotikons. Hierdie ondersoek na die totale gebruik van emotikons dui aan dat ŉ analise daarvan wel gebruik kan word vir die identifisering van outeurs van mikrobloginskrywings[336].

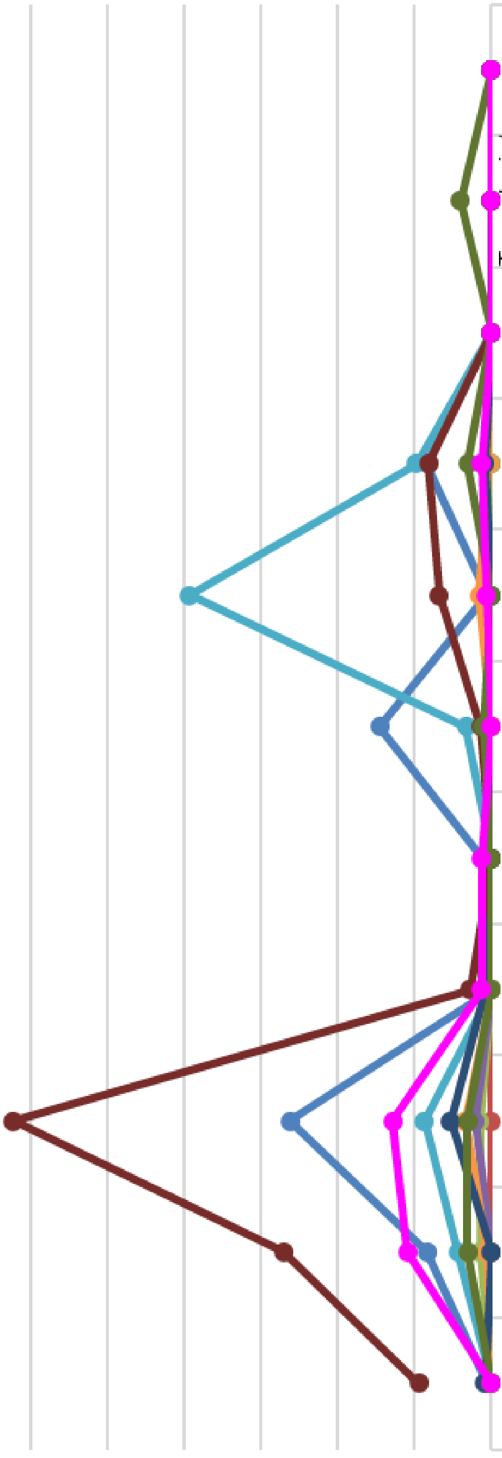
Aangesien daar verskillende soorte emotikons op Twitter gebruik word, het ek die emotikons in 11 verskillende groepe verdeel om te kon bepaal of ’n sekere outeur voorkeur aan ’n bepaalde groep emotikons gee. Hierdie groepe is geklassifiseer op grond van die emotikons se voorkoms, dit wil sê hoe die emotikons lyk, en nie die emosie wat voorgestel word nie. Die elf groepe is soos volg:

- **Diere**: Verwys na emotikons wat deur diere uitgebeeld word.
- **Gebare**: Verwys na emotikons wat gebare bevat.
- **Gesigte**: Verwys na emotikons wat soos gesiggies lyk.

[336] Die gebruik van emotikons deur die outeurs van die verdagte tekste word in afdeling 7.8.2 (Verdagte tekste: emotikons) voorgehou.

- **Kos**: Verwys na emotikons wat kos, koeldrank of drank uitbeeld.
- **Mense:** Verwys na emotikons wat mense uitbeeld.
- **Piktogramme**: Verwys na emotikons wat nog deur simbole uitgebeeld word.
- **Natuur**: Verwys na emotikons wat die natuur uitbeeld.
- **Lukraak**: Verwys na emotikons wat nie in enige van die ander groepe pas nie.
- **Sport**: Verwys na emotikons wat met sport verband hou.
- **Tegnologie/vermaak**: Verwys na emotikons wat met tegnologie of vermaak verband hou.
- **Vervoer**: Verwys na emotikons wat verskillende maniere van vervoer uitbeeld.

Daar is vir elk van die outeurs ŉ genormaliseerde waarde per 1 000 woorde vir die totale voorkoms van emotikons in die korpus bereken en ook vir elk van die bogenoemde elf groepe wat soorte emotikons bevat. Sien Figuur 7.26:



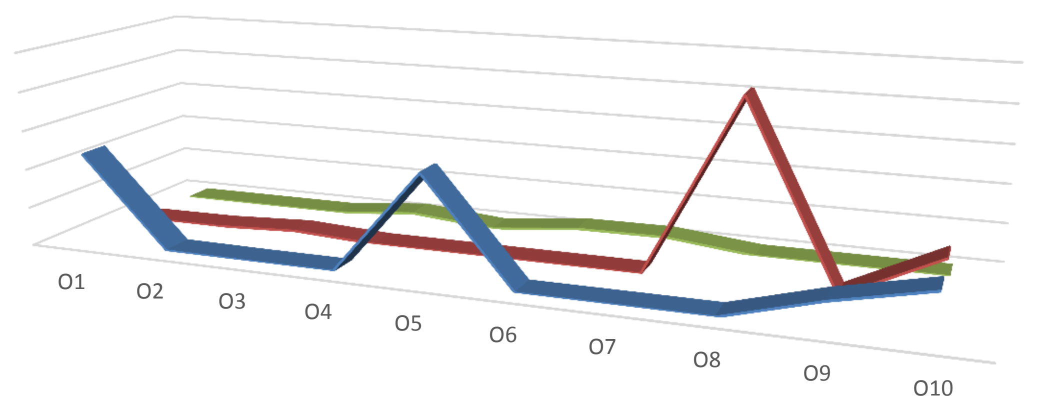
	Diere	Gebare	Gesig	Kos	Mense	Natuur	Piktogram	Lukraak	Sport	Tegnologie/ vermaak	Vervoer
O1	0,11	1,68	5,26	0,00	0,22	2,91	0,11	1,68	0,00	0,00	0,00
O2	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,00	0,00	0,00	0,00
O3	0,00	0,30	0,30	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
O4	0,14	0,00	0,42	0,00	0,00	0,00	0,28	0,14	0,00	0,00	0,00
O5	0,00	0,87	1,75	0,00	0,00	0,66	7,86	1,97	0,00	0,00	0,00
O6	0,00	0,12	0,67	0,00	0,00	0,00	0,31	0,06	0,00	0,00	0,00
O7	0,18	0,00	1,06	0,18	0,00	0,18	0,00	0,18	0,00	0,00	0,00
O8	1,90	5,42	12,47	0,54	0,00	0,27	1,36	1,63	0,00	0,00	0,00
O9	0,00	0,61	0,61	0,00	0,00	0,20	0,00	0,61	0,00	0,81	0,00
O10	0,00	2,20	2,58	0,26	0,26	0,00	0,13	0,26	0,00	0,00	0,00

Figuur 7.26 Genormaliseerde waarde per 1 000 woorde vir die verskillende soorte emotikons wat in die gespesialiseerde korpus gebruik word

Vanuit Figuur 7.26 blyk dit dat daar ŉ waarneembare verskil is tussen die genormaliseerde waarde vir die gebruik van emotikons vanuit die verskillende groepe deur die onderskeie outeurs. O8 maak byvoorbeeld van aansienlik meer emotikons gebruik wat in die "gesig"-

kategorie van emotikons geplaas is. O5 maak weer die meeste van piktogramme gebruik terwyl O1 veral van baie natuur-emotikons gebruik maak. Die gebruik van die verskillende soorte emotikons kan dus as ŉ bykomende manier beskou word om te onderskei tussen verskillende outeurs. Ten einde die ondersoek na die gebruik van die verskillende soorte emotikons te kon afhandel, is daar laastens ondersoek ingestel na die soort emotikons wat meer frekwent in die gespesialiseerde korpus gebruik word (sien Figuur 7.26).

Wanneer die gebruik van emotikons in terme van die kategorieë[337] ondersoek word, kan daar vanuit hierdie ondersoek aangevoer word dat van die kategorieë meer geneig is om van emotikons gebruik te maak. Die outeurs met die laagste genormaliseerde voorkoms van emotikons is in die vermaaklikheidspersoonlikhede- en blogger/nuuskategorie geplaas. Figuur 7.27 dui die gebruik van emotikons in terme van die kategorieë aan:

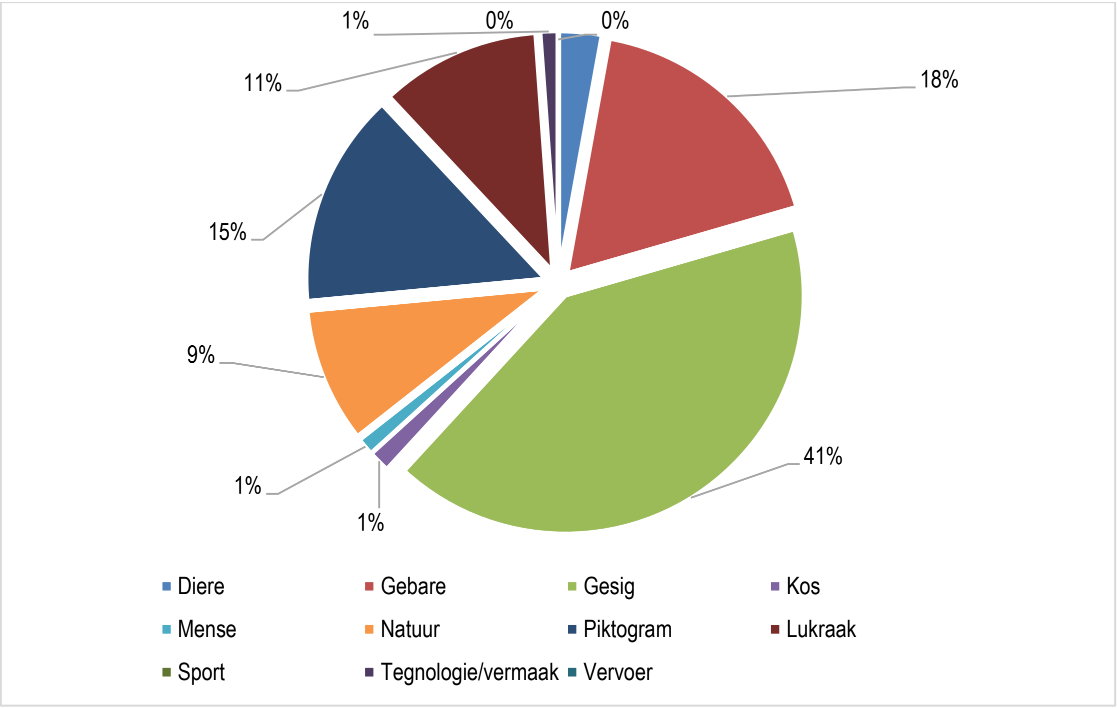


	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Bloggers	11.74	0.08	0	0	13.32	0	0	0	3.25	5.81
Nuus	0	0.08	0.59	0	0	0	0	22.76	0	5.81
Vermaak	0	0	0	0.98	0	1.16	1.24	0	0	0

Figuur 7.27 Emotikons soos gebruik per kategorie in die gespesialiseerde korpus

Figuur 7.27 toon die gebruik van emotikons per kategorie aan. O8 (wat as ŉ nuuskategorie geklassifiseer is) maak van die meeste emotikons gebruik, gevolg deur O1 en O5 (bloggerkategorie). Ná verwagting sou die bloggerskategorie ŉ meer informele taalstyl gevolg het as outeurs in die nuuskategorie, maar dit blyk dat O8 in dié geval nie soos die ander outeurs (O2, O3 en O10) in hierdie kategorie van minder emotikons gebruik maak nie.

[337] Sien hoofstuk 5 (afdeling 5.4 – Data-insameling en -verwerking) vir ŉ uiteensetting van die kategorieë.



Figuur 7.28 Verspreiding van emotikongroepe

Vanuit Figuur 7.28 blyk dit dat die meeste emotikons wat gebruik word, deel uitmaak van die gesigte-kategorie (41%), gevolg deur die gebare- (18%), piktogramme (15%) en lukraak-kategorie (11%) – hierdie kan dus as ŉ algemene tendens van Afrikaans beskou word soos dit op Twitter gebruik word. Ten einde te kon bepaal hoe die outeurs die emotikons gebruik, is daar na die groepering van die emotikons ondersoek ingestel (sien voorbeeld 254 tot 262):

Voorbeeld 254: **O1** – My facial expression as iemand 'n geweer teen my kop hou: 🙄🔫

Voorbeeld 255: **O3** – 👇😂

Voorbeeld 256: **O4** – Tjom...... Moenie vir my so sê nie 🙈

Voorbeeld 257: **O5** – God is besig om my dorp te seën 😃💧☁️

Voorbeeld 258: **O6** – Kontak hom! Dit kan 3 maande se huispaaiemente vir jou betaal!😉

Voorbeeld 259: **O7** – Ek het jou eers raak gesien toe ek langs jou was.🤗 jyt darm maar n grênd bakkie

Voorbeeld 260: **O8** – Geniet jou naweek :)

Voorbeeld 261: **O9** – Oooh... Fire. 🔥🎵👌 #TBT #Fok #Ja #VolumeOop

Voorbeeld 262: **O10** – Ek sien uit @GEBRUIKER 👏👏

Voorbeeld 254 tot 262 toon die verskillende maniere aan waarop die outeurs emotikons in hulle tekste gebruik. Hierdie voorbeelde is slegs verskaf as ŉ oorsig van die outeurs se emotikongebruik en toon aan dat alhoewel daar ŉ duidelike onderskeid tussen die verskillende outeurs getref kan word met betrekking tot die frekwensie waarteen die gebruikers emotikons gebruik, is daar nie ŉ bepaalde patroonmatigheid aanwesig wanneer die manier waarop die emotikons gebruik word, ondersoek is nie.

Zappavigna (2015:71) en Crystal (2011:46) verwys na emotikons as kenmerke van sosialemediaplatforms wat bydra tot die algehele gevoel wat met boodskappe oorgedra word – hetsy positief of negatief.

Vanuit die voorbeelde blyk dit dat Zappavigna (2012:71) en Crystal (2011:46) se argumente oor die gebruik van emotikons ook op Afrikaans soos dit op mikroblogs gebruik word van toepassing is. Die emotikons verskaf diepte aan boodskappe wat oorgedra word deur óf dit wat in die tekste gesê word te beklemtoon óf die emotikons dien as ŉ aanduiding van die emosie wat outeurs ervaar. Tydens die ondersoek was daar is nie ŉ enkele outeur wat emotikons slegs "negatief" of slegs "positief" gebruik het nie.

In die verwysingskorpus word emotikons op ŉ soortgelyke manier gebruik as in die korpora en dus kan die ondersoek na die gebruik van emotikons as ŉ aanduiding tot outeurskap beskou word, maar dit kan op sigself nie veel meer aandui as die "gevoel" van boodskappe wat oorgedra word nie. Die gebruik van emotikons word in die hieropvolgende afdeling bespreek om te kon bepaal of die gebruik van emotikons wel suksesvol gebruik kan word in die identifikasie van mikroblogouteurs.

7.5.1.2 Verdagte tekste: emotikons

In afdeling 7.5.1.1 word daar tussen verskillende soorte emotikons onderskei en die aantal keer wat die onderskeie outeurs van emotikons gebruik maak is geanaliseer − emotikongebruik kan wel as ŉ kenmerkende eienskap van outeurs beskou word.

In die ondersoek is daar ook bepaal dat die kategorie waarin die bepaalde Twitterouteurs geplaas is, kan ŉ verdere invloed op die gebruik van emotikons hê. Vervolgens is die gebruik van emotikons geanaliseer volgens die onderskeie emotikongroepe wat in die verdagte tekste voorkom (sien Tabel 7.25):

Tabel 7.25 Emotikongebruik in die verdagte tekste

	V1	Genormaliseer tot 1 000 woorde	V2	Genormaliseer tot 1 000 woorde	V3	Genormaliseer tot 1 000 woorde
Totaal	0	0,00	14	9,94	14	33,98
Diere	0	0,00	0	0,00	0	0,00
Gebare	0	0,00	2	1,42	8	19,42
Gesigte	0	0,00	12	8,52	3	7,28
Kos	0	0,00	0	0,00	0	0,00
Mense	0	0,00	0	0,00	1	2,43
Piktogramme	0	0,00	0	0,00	0	0,00
Natuur	0	0,00	0	0,00	0	0,00
Lukraak	0	0,00	0	0,00	2	4,85
Sport	0	0,00	0	0,00	0	0,00
Tegnologie/vermaak	0	0,00	0	0,00	0	0,00
Vervoer	0	0,00	0	0,00	0	0,00

Tabel 7.25 toon aan hoe die drie onbekende outeurs (V1, V2 en V3) van emotikons in hulle korpora gebruik maak. V1 gebruik geen emotikons nie terwyl V2 en V3 onderskeidelik van 9,94 en 33,98 emotikons genormaliseer tot 1 000 woorde gebruik maak. Wanneer die genormaliseerde waarde van die gebruik van emotikons vergelyk word met die gespesialiseerde korpora, blyk dit dat V1 (0,00) en O2 (0,08) en V2 en O1 (11,74) ooreenstem. O8 (22,76) toon die grootste ooreenkoms met V3, maar ten einde te kon bepaal of die genormaliseerde waarde waarby emotikons in die verdagte tekste gebruik word wel as ŉ aanduiding van outeurskap beskou kan word, is daar vervolgens na die soort emotikons en die manier waarop en wanneer die emotikons gebruik word, ondersoek ingestel.

Soos in Tabel 7.25 aangetoon, gebruik V2 emotikons wat in die gebare- (1,42) en gesigte-kategorie (8,52) geplaas is. Die emotikons wat in hierdie kategorie in die korpora van O1 gebruik word, stem weer eens beduidend ooreen met die genormaliseerde waarde van V2. Behalwe vir O3 (gebare – 0,3; gesigte – 0,3 en dus ŉ baie laer frekwensie as wat in V1 voorkom), is daar geen korpora wat slegs van emotikons gebruik maak wat in die gebare- en gesigte-kategorie geplaas is nie.

V3 (sien Tabel 7.25) gebruik emotikons wat in die gebare- (19,42), gesigte- (7,28), mense- (2,43) en lukraak-kategorie (4,85) geplaas is. In die gespesialiseerde korpora is daar nie ŉ korpus waarin die emotikongebruik met dié van V3 ooreenstem wanneer die onderskeie kategorieë ondersoek is nie. In Figuur 7.28 is dit duidelik dat die emotikons wat in die gesigte-kategorie geplaas is die gewildste blyk te wees op Twitter en ook in die geval van die verdagte

korpora. Hierdie kenmerk kan dus as ’n algemene kenmerk van Afrikaans beskou word soos dit op Twitter gebruik word. Voorbeeld 263 tot 266 beeld die manier uit waarop emotikons in die verdagte tekste se twiets gebruik word:

Voorbeeld 263: **V2** – Julle.😂😂😂😂😂😂

Voorbeeld 264: **V2** – Ek love hoe mense op Twitter so sensitief is deesdae. Sê net een ding, dan bitch en moan almal daaroor.😂👏🏼😂👏🏼Rustig, life isn't that bad.

Voorbeeld 265: **V3** – Okay so @GEBRUIKER het my klaar oortuig Dat goeie mense nog bestaan That bio tho😱👌👏

Voorbeeld 266: **V3** – Ek het omgegee.. nou is ek net van fok jou ✌🏼

Die voorbeelde (263 tot 266) wat in hierdie afdeling voorgehou word, sluit aan en bevestig die stelling in afdeling 7.5.1.1 dat emotikons gebruik word om die gevoel wat in twiets oorgedra word, te versterk en ondersteun. Tog blyk dit dat emotikons wel as ’n aanduiding van outeurskap kan dien.

Voorbeeld 267: **O1** – OHMYWORD! Dis die kuiken-ding van vooraf! @GEBRUIKER😂 😂😂😂😂😂😂😂😂😂😂

Voorbeeld 268: **O1** – Ek👏🏼is👏🏼so👏🏼gatvol👏🏼vir👏🏼hierdie👏🏼shit.

Voorbeeld 269: **O5** – Julle sy is so oulikkk😊😋

Voorbeeld 270: **O5** – Maak haar seer en ek belowe ek maak jou dood 👌

Wanneer voorbeeld 263 en 264 met voorbeeld 267 en 268 vergelyk word, is dit duidelik dat daar ’n ooreenkoms is in die manier waarop emotikons deur V2 en O1 gebruik word om die boodskap te versterk. Dieselfde emotikon word ook herhaaldelik gebruik. In voorbeeld 265, 266, 269 en 270 kan die ooreenkoms tussen die gebruik van emotikons in die korpora van V3 en O5 waargeneem word. Die emotikons word gebruik ten einde die boodskap te versterk en ’n bepaalde emosie oor te dra.

Vanuit hierdie analise kan daar vervolgens aangevoer word dat V2 deur O1 geskryf is en V3 deur O5. Aangesien V1 nie van emotikons gebruik maak nie, kan daar aangevoer word dat O2 die outeur van die teks is omdat O2 ook minimaal van emotikons gebruik maak. Ten opsigte van die Twitter-kategorieë kan daar aangevoer word dat V2 en V3 in die bloggerskategorie

tuishoort terwyl V1 vermoedelik in die nuuskategorie sal val na gelang van die gebruik van emotikons. Die gebruik van emotikons kan na gelang van die bevindinge van die studie wel as ’n manier beskou word waarop outeurs van mikrobloginskrywings geïdentifiseer kan word.

7.6 Kodewisseling

In aansluiting by die semantiese ondersoek (wat na die gebruik van tekens verwys het), moet daar ook ondersoek na die gebruik van verskillende kodes in ’n enkele twiet ondersoek ingestel word. Kracht en Klein (2014:315) argumenteer individue se taalkundige vermoë word nie deur die gebruik van ’n enkele taal se grammatika weerspieël nie, maar eerder individue se vermoë om tussen ’n versameling van verskillende tale/kodes (hierby kan kodewisseling ingesluit word) se grammatika te kan kies en steeds ’n verstaanbare boodskap aan ’n gehoor oor te dra. Sprekers moet dus oor genoegsame kennis beskik om die verskillende kodes te kan gebruik en steeds betekenis oor te kan dra.

Ten einde die kwalitatiewe analise te kon uitvoer, is daar ondersoek ingestel na kodewisseling. Hierdie ondersoek is eerstens op die gespesialiseerde korpus uitgevoer en daarna is die verdagte tekste geanaliseer.

Kodewisseling verwys volgens Du Plessis (1995:24-25) na die wisseling wat sprekers in ’n gesprek toepas tussen die variëteite van dieselfde taal of tussen tale. Ponelis (2009) onderskei tussen twee variëteite van Afrikaans − die klassieke en die demotiese variëteit − en postuleer dat die demotiese variëteit die grootste invloed van Engels toon. Hy (Ponelis, 2009) voer aan dat demotiese Afrikaans ’n beduidende Engelse invloed toon, maar dat dit nie met Afrikaanse omgangstaal verwar moet word nie. Volgens Myers-Scotton (2001:18) beskik elke spraakgemeenskap oor meer as een manier van kommunikeer en hierdie verskillende maniere waarop kodes gebruik word om te kommunikeer, is telkens die gevolg van sosiale verskille en/of invloede. Myers-Scotton (1998:18-20) argumenteer verder dat sprekers wat kennis dra van die verskillende maniere waarop hierdie kodes toegepas kan word om duidelike en verstaanbare kommunikasie in ’n gegewe situasie te laat geskied, hierdie gebruik van verskillende kodes tot hulle voordeel kan gebruik. Lawrence (1998:20-29) stel dat kodewisseling op verskeie maniere kan plaasvind. Vir die doel van my studie is daar na die onderstaande gevalle van kodewisseling ondersoek ingestel:

- Die invoeging van ’n ingebedde taal-lekseem (IT) of die invoeging van ’n meertalige-eiland (MT):

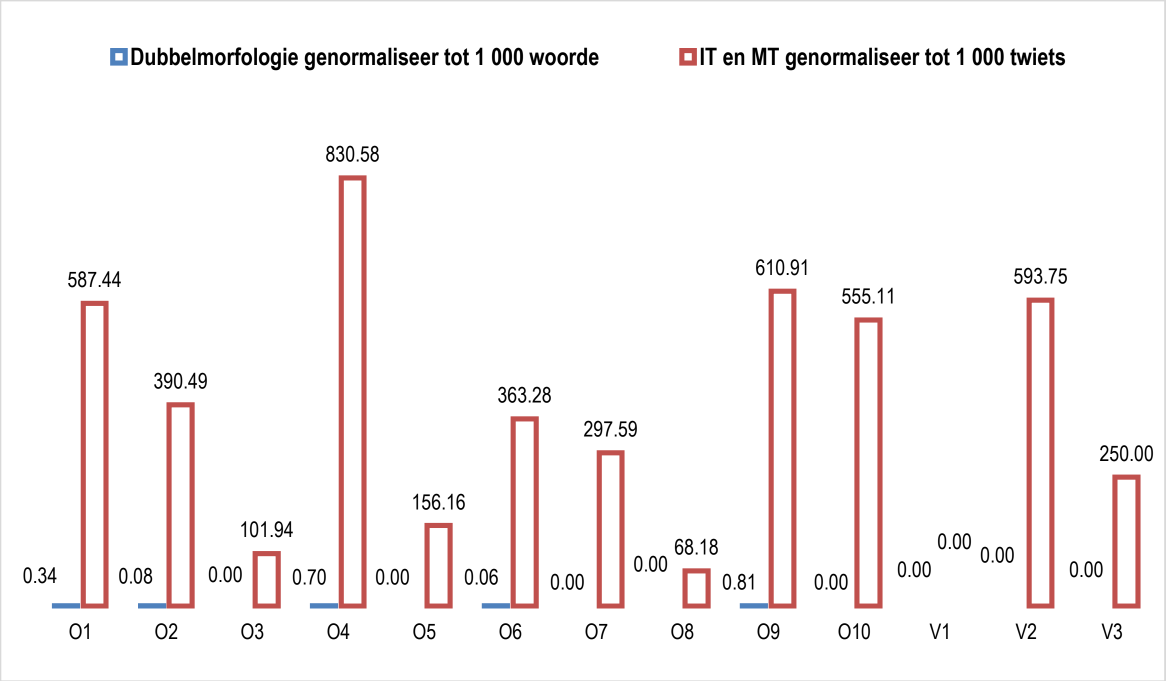
Voorbeeld 271: **O1** – (ingebedde taal-lekseem) Die **loser** het al haar hare afgesny en geskenk. Nou laat sy my nog toe om haar gesig te verf vir 'n taak. Beste mens.

Voorbeeld 272: **O1** – (meertalige-eiland) **It's all fun and games** om alleen te bly in 'n flat, totdat jy 'n bottel moet oopmaak wat vasklou soos Velcro.

- Wisseling binne een woord (dubbelmorfologie):

Voorbeeld 273: **O1** – Foto-tip 101: staan ver weg van die mense in 'n groepfoto, sodat wanneer jy on point lyk jy jouself ordentelik kan **uitcrop**.

Alhoewel die gebruik van kodewisseling en die onderskeid tussen die verskillende maniere waarop kodewisseling in 'n bepaalde situasie gebruik word, verwys ek vir die doel van my studie slegs oorkoepelend na die bogenoemde voorbeelde[338] – sien Figuur 7.29 vir 'n opsomming van die voorkoms van kodewisseling in die gespesialiseerde korpus en die verdagte tekste.



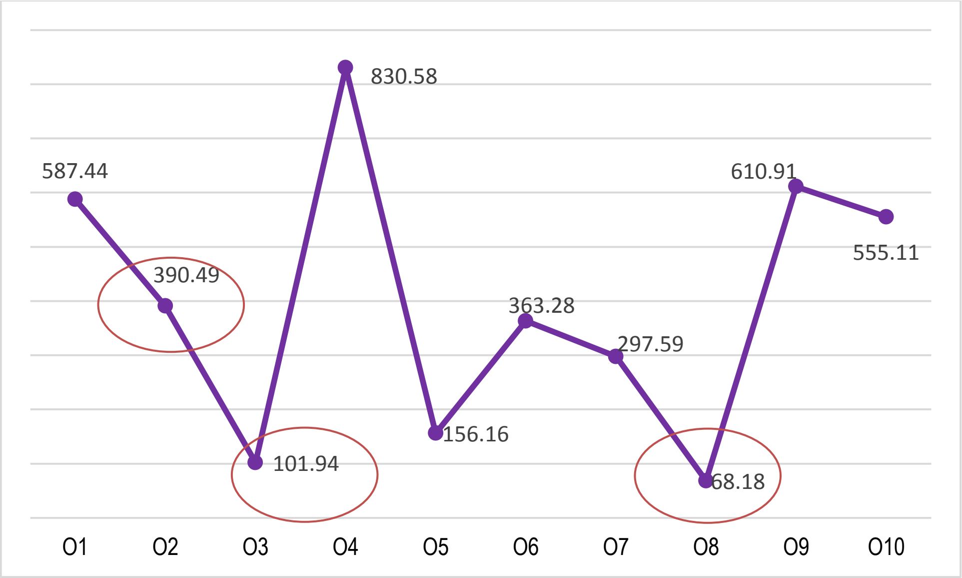
Figuur 7.29 Kodewisseling in die gespesialiseerde korpus

Twitter-gebruikers maak wel van kodewisseling gebruik en hierdie gebruikers wissel willekeurig tussen die gebruik van enkele woorde of Engelse sinne soos sigbaar in Figuur 7.29. Ten einde die voorkoms van kodewisseling te kan bespreek, fokus ek eers op die gebruik van IT en MT en daarna op dubbelmorfologie.

[338] Die gebruik van kodewisseling regverdig 'n omvattende studie (sien hoofstuk 7), maar vir die doel van outeurskapidentifikasie het ek slegs na die voorkoms van die genoemde voorbeelde ondersoek ingestel.

7.6.1 Gespesialiseerde korpus: IT en MT

IT en MT verwys na die invoeging van ŉ ingebedde taal-lekseem (IT) of die invoeging van ŉ meertalige-eiland (MT) – sien voorbeeld 256 en 257. In Figuur 7.30 word die gebruik van IT en MT aangedui:



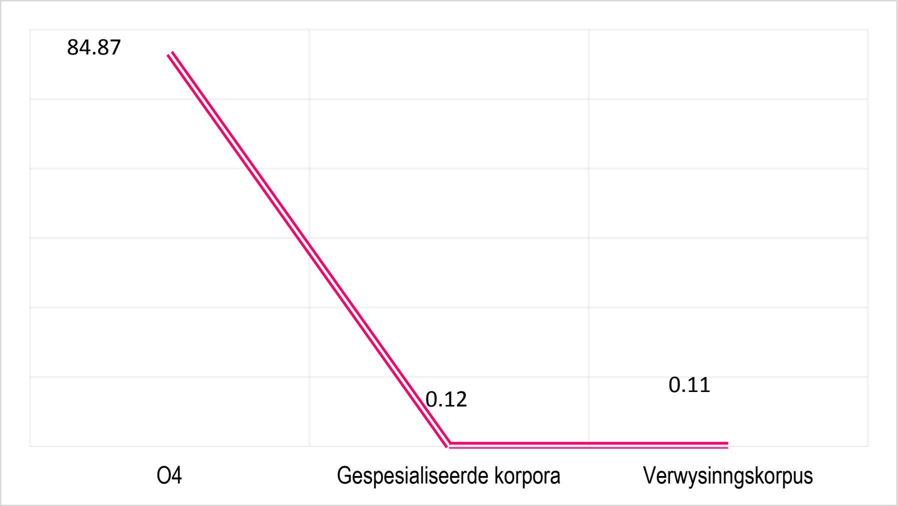
Figuur 7.30 Die voorkoms van IT en MT in die gespesialiseerde korpus

Vanuit Figuur 7.30 kan O1, O4 en O9 as die drie outeurs geïdentifiseer word wat die meeste van IT en MT gebruik maak in hulle twiets. Hierdie outeurs is in die blogger- (O1 en O9) en vermaaklikheidspersoonlikheidskategorie (O4) geplaas. Aangesien die verhouding tussen die gebruik van IT en MT nie ooreenstem met al die outeurs wat in ŉ spesifieke kategorie geplaas is nie[339], kan die gebruik van kodewisseling nie aan ŉ spesifieke Twitter-kategorie gekoppel word nie. Dit wil voorkom of die outeurs wat in die nuuskategorieë geplaas is, geneig is om minder van die invoeging van IT en/of MT gebruik te maak as die outeurs in die bloggerkategorie.

Na aanleiding van die ondersoek na die gebruik van IT en MT in die gespesialiseerde korpus, kan daar aangevoer word dat nie een van die outeurs wat deel uitmaak van die gespesialiseer korpus Afrikaans puristies benader nie. Hierdie benadering kan vervolgens as ŉ kenmerk van Afrikaans soos dit op Twitter voorkom, beskou word en kan toegeskryf word aan die sosiale aard van die betrokke platform.

[339] Verwys na afdeling 5.4 (Data-insameling en -verwerking).

Die konkordansie-ondersoek na die IT en MT in die onderskeie korpora het aangetoon dat daar slegs in O4 ŉ IT geïdentifiseer (sien Figuur 7.31) kan word wat herhaaldelik deur die outeur gebruik word:



Figuur 7.31 Genormaliseerde waarde per 1 000 woorde vir *boys* soos gebruik deur die gespesialiseerde korpus

Figuur 7.31 toon die genormaliseerde waarde per 1 000 woorde aan soos die woord *boys* in die onderskeie korpora gebruik word. ’n Soektog is ook in die verwysingskorpus uitgevoer ten einde te bepaal of O4 se gebruik van die woord (sien voorbeeld 276, 277 en 278) werklik as kenmerkend beskou kan word. *Boys kom* 28 keer in die gespesialiseerde korpus voor teen ’n genormaliseerde waarde per 1 000 woorde – 0,11 – teenoor O4 se 83 keer – 84,87. Dit blyk dus dat die manier waarop die woord gebruik word ook verskil. Daar is ook na voorbeelde van die gebruik van die woord *boys* in die res van die gespesialiseerde korpus ondersoek ingestel – sien voorbeeld 277 en 278:

Voorbeeld 274: **O4** – Baie geluk aan @GEBRUIKER @GEBRUIKER met julle Nuwe wereld rekord! 3100 flips in net oor die 5 ure!! Flippit **boys**!!! Well done!

Voorbeeld 275: **O4** – **Boys** wie dink julle gaan impress die naweek, Behalwe die Leeus?

Voorbeeld 276: **O4** – 70% van die kaartjes is reeds verkoop vir Life201 by @GEBRUIKER Parkview in Pretoria die 13, 20, 27 ste Februarie.. maak plan **boys**!

Voorbeeld 277: **O10** – Ek hou saam met jou duim vas. Let's go **boys**. 👍 @GEBRUIKER @GEBRUIKER #BackYourBoytjies

Voorbeeld 278: **O1** – Shame **boys**, die arme ou.

O4 gebruik *boys* as ’n aanspreekvorm (dit is asof *boys* ’n versamelnaam en ’n eienaam word). In die gespesialiseerde korpus word die woord deurgaans gebruik om na jong seuns te verwys en word verder as ’n hutswoord gebruik. Die manier waarop die woord *boys* gebruik word, verskil gevolglik wanneer O4 se gebruik van *boys* met die ander voorkoms van die woord vergelyk word; daarom kan O4 se gebruik daarvan as ’n kenmerk van outeurskap beskou word.

ʼn Verdere ondersoek na die voorkoms van IT en MT in die gespesialiseerde korpus het geen ander bevindinge (soos die gebruik van *boys*) opgelewer nie, maar voorbeelde van IT en MT is in elk van die gespesialiseerde korpora gevind (sien voorbeeld 279 tot 289):

Voorbeeld 279: **O1** – Ek het die afgelope paar maande net **junk food** geëet. Toe ek gesond begin eet toe word ek siek. **Moral of the story**: ek gaan nou 'n burger kry [MT]

Voorbeeld 280: **O1** – Ek dink studente luister meer vir die **DJ** se **instructions** as wat hulle luister na hulle eie dosente. [IT]

Voorbeeld 281: **O2** – "@GEBRUIKER: **Essex bus collides with lorry carrying steam engine** http://trib.al/ ". Gewonder wat is nuus? Hier's 'n voorbeeld. [MT]

Voorbeeld 282: **O2** – Koos Bekker **eat your heart out**. In 7e Laan gee hulle al lankal (en steeds!) dagblaaie uit met redaksies van 2 of 3. [MT]

Voorbeeld 283: **O3** – Prof. Christof Heyns sê die gebruik van afstandbeheerde robotvliegtuie (**drones**) hou gevaar in vir internasionale stabiliteit @GEBRUIKER [IT]

Voorbeeld 284: **O5** – Party meisies wys soveel **cleavage** in 'n foto, mens moet hulle gesig soek. [IT]

Voorbeeld 285: **O6** – Ek is SO moeg vir hierdie facebook goedjies wat sê: "I**n a relationship**." "**Together one year**!" bla bla bla. My... http://fb.me/8kntmFtm4 [MT]

Voorbeeld 286: **O7** – 14 man!!! Jy verloor teen 14 man. Ek gaan nou drink en **show**. Drink want daai **game** gaan homself nie vergeet nie [IT]

Voorbeeld 287: **O8** – Die gedagte van wat moontlik 'n straat weg van my plaasvind maak my siek. "**Nearly 250 000 modern slaves in SA index**" [MT]

Voorbeeld 288: **O9** – Elke dag wat ek nie hierdie stuk kak mens se skedel vergruis met my **stapler** nie. **#ThatIsStrengthThatIsGrace** [IT en MT]

Voorbeeld 289: **O10** – **#ClubChamps16** Die @GEBRUIKER wen met 8 paaltjies. Hulle is vanjaar se klubkampioene. Baie geluk aan @GEBRUIKER en die manne. #MyNWU [MT]

Voorbeeld 279 tot 289 kan as ŉ opsomming dien van die manier waarop die onderskeie outeurs in die gespesialiseerde korpus van IT en MT gebruik maak. Vanuit die voorbeelde is daar sekere tendense wat waargeneem kan word. Alhoewel daar – behalwe in die geval van O4 – nie IT's of MT's herhaal word en vervolgens as kenmerkend tot ŉ enkele outeur se korpus beskou kan word nie, kan die argument wel gemaak word dat outeurs wat in die nuuskategorie geplaas is (O2, O3 en O8[340]) geneig is om direkte aanhalings met betrekking tot nuuswaardige gebeure deur middel van MT (sien voorbeeld 281 en 287) aan ŉ teikengehoor oor te dra. IT word ook deur die outeurs (nuuskategorie – sien voorbeeld 268) gebruik.

Hierdie voorbeelde is wel met die verdagte tekste vergelyk om sodoende te kon bepaal of daar kenmerke is wat ooreenstem en gevolglik kan help om outeurs van verdagte tekste te kan identifiseer. Vervolgens is die voorkoms van IT en MT in die verdagte tekste geanaliseer.

7.6.2 Verdagte tekste: IT en MT

Soos in afdeling 7.6 voorgehou, verwys kodewisseling na die wisseling wat sprekers toepas in gesprekke tussen die variëteite van dieselfde taal of tussen tale (Du Plessis, 1995:24-25). In hierdie afdeling is die voorkoms van IT en MT in die verdagte tekste ondersoek.

Tabel 7.26 IT en MT soos gevind in die verdagte tekste

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	IT en MT	IT en MT genormaliseer tot 1 000 twiets
V1	42	90	844	0	0,00
V2	96	147	1409	57	593,75
V3	32	44	412	8	250,00

In Tabel 7.26 word die gebruik van IT en MT uitgebeeld soos dit in die verdagte tekste gevind word. Ten opsigte van die voorkoms van IT en MT in die verdagte tekste blyk dit dat die genormaliseerde waarde (tot 1 000 twiets) van V1 nie met enige van die ander outeurs in die gespesialiseerde korpus se gebruik van MT of IT ooreenstem nie. Die genormaliseerde waarde (tot 1 000 twiets) van V2 (593,75) en V3 (250,00) se gebruik van IT en MT stem onderskeidelik met O1 (587,44) en O7 (297,59) ooreen. Vervolgens verskaf ek voorbeelde vanuit die verdagte tekste om te kon bepaal of daar enige stylooreenkomste is en of enige woorde herhaaldelik deur die outeurs gebruik word:

Voorbeeld 290: **V2** – Ek's verlief op @GEBRUIKER se stem. Dankie @GEBRUIKER dat jy my **introduce** het aan sy **voice notes**. [IT]

[340] O2 is in die bloggers- en nuuskategorie geplaas na aanleiding van die gebruik van MTs soos in voorbeeld 308 – tog is dit belangrik om te let op die manier waarop die outeur steeds ŉ verband tussen die gewilde sepie *7de Laan* en joernalistiek maak.

Voorbeeld 291: **V2** − Universiteit tip: **Cite this for me** - doen alle verwysings met 'n volledige bronnelys en inteksverwysings. Plesier. [MT]

Voorbeeld 292: **V3** − Kry 'n man wat na **aftershave** ruik, nie 'n seuntjie ruik of hy in 'n blikkie Solo of Axe gebad het nie. [IT]

Voorbeeld 293: **V3** − Staan op. Gaan aan. Dis seer, ek weet, maar jys sterker as dit. **You got this**. [MT]

Voorbeeld 290 tot 293 toon dat daar in twee van die verdagte tekste wel voorbeelde van beide IT en MT voorkom. In afdeling 7.6.1 is daar bepaal dat dit slegs in die korpus van O4 teenwoordig is (sien voorbeeld 274 tot 276) waar ’n bepaalde IT voorkom wat kenmerkend tot die outeur se taalgebruik beskou kan word. Hierdie IT kom nie in die twee verdagte tekste voor nie, en O4 kan na aanleiding van hierdie ondersoek nie as die outeur van een van die tekste beskou word nie. Wanneer die manier waarop IT en MT in die verdagte tekste gebruik word met die gebruik daarvan in die gespesialiseerde korpus vergelyk word, is daar sekere ooreenkomste wat waargeneem kan word.

Daar is in geen van die verdagte tekste (IT- en MT-gebruik) voorbeelde van verwysings na nuuswaardige gebeure nie; daarom kan daar aanvaar word dat die outeurs van hierdie twee verdagte tekste nie in die nuuskategorie val nie. O2, O3 en O8 kan vervolgens nie as outeurs van V1 en V3 beskou word nie. Die manier waarop IT en MT gebruik word in die korpora van O2, O3 en O8 (sien voorbeeld 281 tot 283 en 289) stem nie ooreen met V2 en V3 nie. O6 kan ook nie as die outeur van V2 en V3 geïdentifiseer word nie op grond van die gebruik van die hiperskakel en die feit dat O6 aanhalingstekens gebruik wanneer ’n MT in ’n twiet gevoeg word (sien voorbeeld 287) – ’n kenmerk wat glad nie in V2 of V3 aanwesig is nie.

O7 kan na gelang van die genormaliseerde waarde (tot 1 000 twiets) waarby IT en MT in V2 se korpus voorkom as moontlike outeur van die teks beskou word. Alhoewel daar geen ooreenkoms tussen die IT is (die woordkeuse) wat deur O7 en V3 gebruik word nie, is daar op grond van die ondersoek na die gebruik van IT en MT nie genoegsame bewyse dat O7 ’n moontlike outeur van V3 is nie. Na gelang van die bogenoemde kan O7 nie as moontlike outeur van V2 uitgeskakel word nie.

O1 se korpus toon die grootste ooreenkoms ten opsigte van die genormaliseerde waarde (tot 1 000 twiets) van die voorkoms van IT en MT met V2. Met die vergelyking van die manier waarop IT en MT deur hierdie twee outeurs (sien voorbeeld 279 en 281, 292 en 293) gebruik word, is daar sterk ooreenkomste in die twiets van hierdie twee outeurs sigbaar. Nie net is daar tematiese ooreenkomste teenwoordig nie (studentelewe, stelling aangaande die individu se

lewe), maar die emotiewe bydra wat die IT en MT tot die twiet lewer kan as soortgelyk beskou word.

V3 en O7 stem ooreen na gelang van die genormaliseerde waarde (tot 1 000 twiets) van die voorkoms van IT en MT. O5 kan na gelang van die manier waarop IT en MT in die twiets van V3 gebruik word ook as ŉ moontlike outeur van die teks beskou word. Wanneer voorbeeld 288 met voorbeeld 292 en 293 vergelyk word, is daar geen rede om O7 nie as outeur van die teks te beskou na gelang van die voorkoms van IT en MT in die twiets nie, maar O5 (voorbeeld 286) stem meer ooreen. In voorbeeld 284, 292 en 293 is daar ŉ sterk negatiewe ondertoon en alhoewel voorbeeld 290 (O7) ook as ietwat negatief beskou kan word, is die teleurstelling as gevolg van ŉ rugbywedstryd en nie ŉ deurlopende ondertoon in al die twiets nie.

Die ondersoek na die invoeging van IT of MT kan na aanleiding van afdelings 7.6.1 en 7.6.2 as suksesvol in die identifisering van outeurs van mikrobloginskrywings beskou word. Alhoewel die bevindinge op sigself nie gebruik kan word om ŉ outeurskapidentifikasie uit te voer nie, kan dit wel in kombinasie met ander analises gebruik word. Daar kan nou gestel word dat O1 na gelang van die ondersoek na die gebruik van IT en MT as ŉ sterk moontlike kandidaat vir die outeurskap van V2 beskou kan word terwyl O5 en O7 moontlik V3 se outeurs kan wees. In die hieropvolgende afdeling word die voorkoms van dubbelmorfologie voorgehou − die tweede aspek van kodewisseling wat ondersoek is.

7.6.3 Gespesialiseerde korpus: dubbelmorfologie

Dubbelmorfologie verwys na wisseling binne een woord − Taal A word in ŉ woord gevoeg wat in Taal B gebruik word. Wanneer daar spesifiek na dubbelmorfologie ondersoek ingestel is, is daar 13 voorbeelde in totaal (dit sluit alle korpora in) in die gespesialiseerde korpus waar dit wel gebruik word. Vanuit hierdie 13 voorbeelde, word 10 van die dubbelmorfologiese woorde deur die voorvoegsel *ge* gevorm (sien voorbeeld 294):

Voorbeeld 294: **ge***zap*, **ge***fire*, **ge***-pritt*, **ge-***tapede*, **ge** *check*, **ge***bet*, **ge** *feature*, **ge***-PVR*, **ge***Inbox* en **ge***lace*.

- **O1** – Ek het nou net iemand **gezap** op die highway voor my ma as enige iemand wou weet hoe my Gautenger rage gaan.
- **O2** – Rapport se eerste vroulike redakteur is glo **gefire**. So wie gaan in haar plek kom? #afrikaans
- **O4** − Dit was gisteraand makliker as vanoggend om 4 uur op te staan vir gym.. Voel of iemand my oë toe **ge-pritt** het!Tekke aan, lets go! @GEBRUIKER
- **O4** – Ek is so lus vir die rugbyseisoen om te begin ek kyk sommer 'n ou **ge-tapede** skool game

- **O4** – Ek hoop Father Christmas het die bokke se olie en water **ge check** en die filters vervang!! Dis amper tyd
- **O4** − Ek wonder how ryk is die Bliksem wat **gebet** het op Japan...
- **O4** – Wie onthou toe dié ysters nog **ge feature** het op die troues se playlists?
- **O6** – Iemand vra my gister, hoe kry jy iets wat jy **ge-PVR** het op 'n DVD? Enige kenners? Ek onthou hoe ons VHS kasette... http://fb.me/1KMYDnakp
- **O9** – As dit 2008 was en jy het vir my 'n geheime nommer **geInbox** op FB om op 'n status te sê wat ek rEgtIgiiz* van jou dink: 528262- Jy's 'n Poes.
- **O9** – Waarom is 80% van alle Sondaeaande nog ooit **gelace** met 'n tipe nostalgie wat te maklik verwar kan word as kliniese depressie.

Vanuit hierdie tien voorbeelde is daar veral vier voorbeelde wat uitstaan in terme van die manier waarop dit gebruik word: **ge**-*pritt*, **ge-***tape**de***, **ge** *check* en **ge** *feature*. Die voorvoegsel word óf los van die woord gebruik óf dit word met ʼn koppelteken gebruik. Al vier hierdie voorbeelde word in die korpus van O4 aangetref. O6 gebruik slegs een keer die woord **ge**-PVR, maar aangesien dit ʼn verbinding met ʼn afkorting is, verskil die morfologiese struktuur van die stamwoord waarmee die verbinding met *ge-* gevorm is.

Die voorkoms van **ge-** en **ge** (soos in O4 gesien) is ook in die verwysingskorpus nagegaan en dit wil voorkom asof die **ge-** verbinding in dubbelmorfologie wel ook daar gevind word alhoewel slegs teen ʼn genormaliseerde waarde per 1 000 woorde van 0,01 wat aansienlik laer is as die 5,11 van O4. Die gebruik van **ge-** en **ge** as voorvoegsel tot ʼn woord wat as dubbelmorfologie beskou kan word, kan dus as ʼn aanduiding van outeurskap beskou word. O4 kan as die moontlike outeur van een van die verdagte tekste ondersoek word.

Kodewisseling kan − soos duidelik vanuit hierdie afdeling – nie as ʼn merker van outeurskap beskou word nie behalwe in die geval van O4 (dubbelmorfologie: **ge-** en **ge** en die woord ***boys***). Tog kan die voorkoms van kodewisseling en veral dubbelmorfologie wat deur die voorvoegsel ***ge*** gevorm word as ʼn kenmerk van Afrikaans beskou word soos dit op Twitter gebruik word.

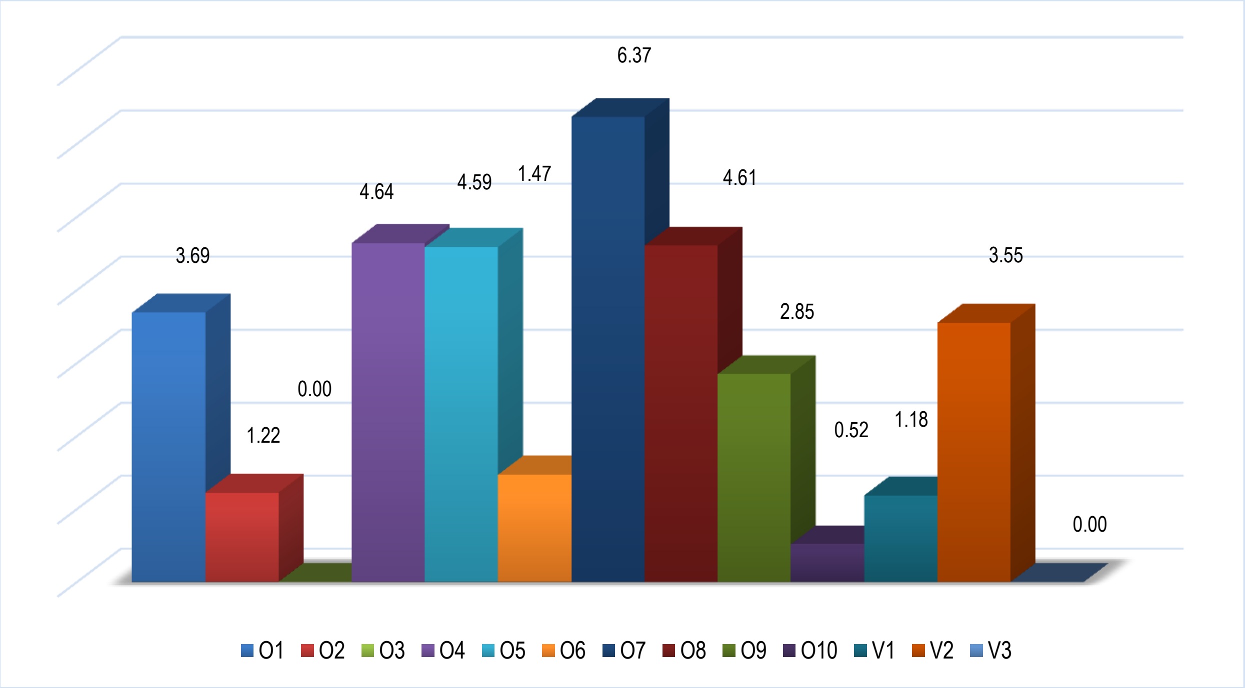
7.6.4 Verdagte tekste: dubbelmorfologie

Daar is in geen van die drie verdagte tekste ʼn voorbeeld waar daar van dubbelmorfologie gebruik gemaak is nie en dus kan daar aangevoer word dat O3, O5, O7, O8 of O10 – daar is ook nie voorbeelde van dubbelmorfologie in die outeurs se tekste nie – die outeurs van V1, V2 en V3 kan wees na gelang van die ondersoek na die gebruik van dubbelmorfologie.

Ten spyte daarvan dat daar nie in die verdagte tekste voorbeelde van dubbelmorfologie voorkom nie, kan die resultate in afdeling 7.6.3 as beduidend beskou word indien dubbelmorfologie deur 'n verdagte outeur gebruik word. Bevindinge vanuit die verdagte outeur se tekste sal vergelyk kan word. Die ondersoek, soos uitgebeeld in afdeling 7.6 (Kodewisseling), kan wel gebruik word om outeurs van mikrobloginskrywings te identifiseer. In afdeling 7.7 is die voorkoms van tussenwerpsels ondersoek.

7.7 Tussenwerpsels

Volgens die Afrikaanse Woordelys en Spelreëls (2017:734) kan tussenwerpsels (ook uitroepe genoem) na woorde of woordgroepe verwys wat los van 'n sin staan en dikwels 'n klanknabootsing, gevoelsuiting of waarskuwingsteken is. Tussenwerpsels verwys vir die doel van die studie na die gebruik van sinskonstruksies en leksikale items wat eie aan gesproke taal is en ooreenstem met randstukke wat deur Verhoef (2001:8-11) as een van die merkers van geskrewe Internetafrikaans geïdentifiseer is. Die ondersoek na tussenwerpsels is ook in *ATLAS.ti* uitgevoer en daar is soos met die ander eienskappe wat ondersoek is ook 'n interne vergelyking uitgevoer. Figuur 7.32 dui die genormaliseerde waarde (tot 1 000 woorde) vir die voorkoms van tussenwerpsels in die gespesialiseerde korpus en die verdagte tekste aan:



Figuur 7.32 Tussenwerpsels

7.7.1 Gespesialiseerde korpus: tussenwerpsels

Figuur 7.32 dui die genormaliseerde waarde (tot 1 000 woorde) aan vir die gebruik van tussenwerpsels in die gespesialiseerde korpus. O4, O5 en O9[341] se gebruik van tussenwerpsels stem ooreen (4,64, 4,59 en 4,61) en na gelang van die gebruik van tussenwerpsels kan die afleiding gemaak word dat die drie tekste moontlik deur dieselfde outeur geskryf is[342]. Die frekwensie waarby die oorblywende outeurs vanuit die gespesialiseerde korpus van tussenwerpsels gebruik maak, toon wel ŉ verskil aan en daarom is daar verder ondersoek ingestel na die manier waarop die outeurs die tussenwerpsels in hulle tekste gebruik. Vanuit die voorbeelde (sien voorbeeld 295 tot 300) blyk dit dat tussenwerpsels wel as kenmerkend van Afrikaans beskou kan word soos dit op Twitter gebruik word:

Voorbeeld 295: **O6** – Ek vertrou nie twitter. Ek volg mense wat ek nog nooit gevolg het nie. Ek blok mense, wat die volgende dag weer verskyn. **Mmmmm...**

Voorbeeld 296: **O8** – Ek dink dit klink heerlik! Met tamatiesous, **haha**

Voorbeeld 297: **O1** – **Ai my jinne**, jy's so pateties dis hartseer

Voorbeeld 298: **O5** – **Oeh** ! Jy skryf "jou" kar af.

Voorbeeld 299: **O1** – 7e Laan: Jong man flirt met jong kelnerin. **Wow**! Dit gaan die jong kykers trek. En omies opgewonde maak. Wat 'n resep.

Voorbeeld 300: **O9** – **YAAASSSSSSSSSSSSSSSSSSSSSSS** FOK FOK FOK @GEBRUIKER[343]

Die bogenoemde voorbeelde sluit enkele van die tussenwerpsels in wat in die onderskeie korpora voorkom. Meinard (2014:164) argumenteer dat tussenwerpsels eintlik ŉ hele sin se betekenis kan oordra of emosie wat ŉ sin se betekenis beïnvloed. Wanneer die voorbeelde vanuit die gespesialiseerde korpus ondersoek is, is dit duidelik dat tussenwerpsels wel gebruik word om emosie oor te dra en dat hierdie woorde telkens in die plek van ŉ langer sin geplaas word (voorbeeld 298 – *oeh* − die tussenwerpsel beeld verbasing en sarkasme uit). In die korpora wat ondersoek is, word verskillende vorme van *hahaha* algemeen gebruik – ŉ aspek wat weer eens nie net die klank van iemand wat lag naboots nie, maar word telkens gebruik om lesers in te lig dat die outeur iets snaaks of hartseer (veral sarkasme) vind. Die ondersoek na die gebruik van tussenwerpsels is dus insiggewend van hoe Afrikaans op Twitter gebruik word. Die voorkoms van tussenwerpsels in die verdagte tekste word nou voorgehou om sodoende te kon bepaal of die ondersoek na die gebruik van tussenwerpsels as ŉ suksesvolle middel in die identifisering van mikroblogouteurs beskou kan word.

[341] O4 is in die vermaaklikheidkategorie geplaas terwyl O5 en O9 in die bloggerskategorie val.

[342] Die interne vergelyking van die gespesialiseerde korpus het bewys dat sommige outeurs konsekwent met die gebruik van tussenwerpsels is of deurlopend van dieselfde frekwensie tussenwerpsels gebruik maak terwyl daar in ander interne vergelykings geen verband tussen die tekste van ŉ enkele outeur waargeneem kon word op grond van die frekwensie van die gebruik van tussenwerpsels nie.

[343] Sien ook afdeling 7.9 (Foute-analise).

7.7.2 Verdagte tekste: tussenwerpsels

Vanuit Figuur 7.32 is dit duidelik dat tussenwerpsels slegs in die korpora van V1 (1,18) en V2 (1,42) aangetref word. Die genormaliseerde waarde (tot 1 000 woorde) waarteen tussenwerpsels deur V1 en V2 gebruik word, stem onderskeidelik met O2 (1,22) en O1 (3,69) ooreen. Beide V3 en O6 maak nie van tussenwerpsels gebruik nie. Voorbeeld 301 tot 303 illustreer hoe die tussenwerpsels deur V1 en V2 gebruik word:

Voorbeeld 301: **V1** – Tekens van tye: Beeld se hoofopskrifte op pp. 1, 2, 3, 4, 6 vandag in aanhaaltekens. **O**, ook p.10 en 11. Is dit luiheid of soeke na emosie?

Voorbeeld 302: **V2** – Ek: "**Ai** julle, my lewe." My vriendin: "My lewer."

Voorbeeld 303: **V2** – Ek: "2016 gaan my jaar wees." Ook ek: "**Eh**, dalk volgende jaar."

Die vergelyking tussen die voorbeelde vanuit die verdagte tekste en die gespesialiseerde korpus dui duidelik aan dat tussenwerpsels gebruik word in Afrikaans soos dit op Twitter gebruik word, maar daar is nie ooreenkomste tussen die soort tussenwerpsels wat gebruik word nie. Die manier waarop V1 en V2 van tussenwerpsels gebruik maak, sluit dus by die algemene karakter van Afrikaans aan en na gelang van die frekwensie van die gebruik van tussenwerpsels kan O1 as moontlike die outeur van V2 beskou word terwyl O2 en V1 ooreenstem. Die ondersoek na die gebruik van kraswoorde word nou voorgehou. Hierdie analise het ook ŉ ondersoek na die semantiese aspekte ingesluit.

7.8 Kraswoorde

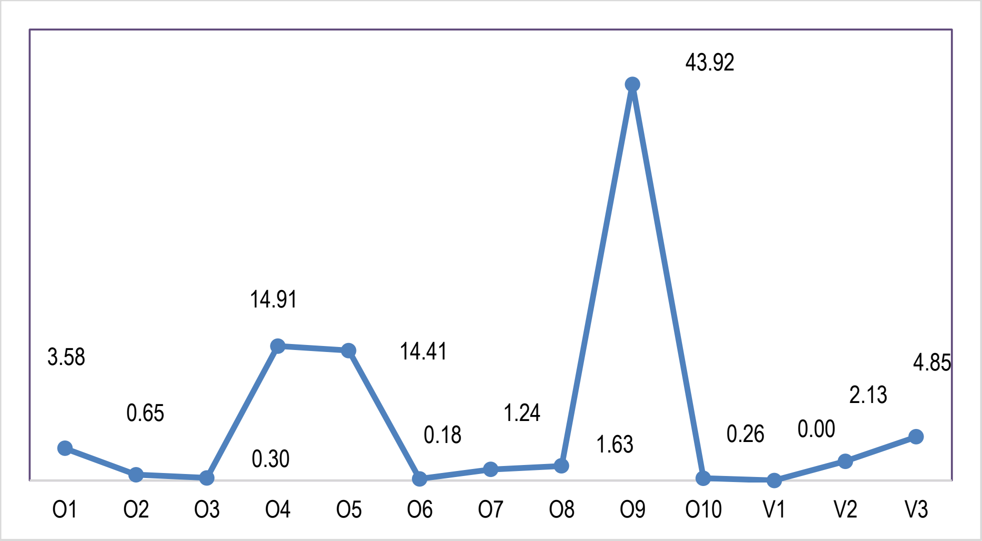
Kraswoorde kan gedefinieer word as ŉ versamelterm vir wat Jansen van Vuuren (2007:207) as kragwoorde beskryf. Kragwoorde verwys volgens Jansen van Vuuren (2007:207) na taboewoorde – “woorde wat op grond van godsdienstige oortuigings, moraliteit of smaak vermy word, en semi taboewoorde, veral met ŉ sterk negatiewe betekenis”. Wang *et al.* (2014:415) stel dat kraswoorde na die gebruik van woorde of frases verwys wat deur sommige individue as “rude, impolite, offensive, obscene, or insulting” beskou word.

Feinauer (1981:242) argumenteer dat die status, ouderdom, geslag van gesprekgenote, die rol van gespreksgenote in ŉ gemeenskap en die informaliteit van ŉ bepaalde situasie as “vloekgebruikbepalende” faktore in Afrikaans beskou kan word en dat vloekwoorde volgens haar navorsing hoofsaaklik gebruik word om ŉ negatiewe sprekersgesindheid of emosie oor te dra. Wang *et al.* (2014:417) het die voorkoms van vloekwoorde op Twitter ondersoek en het lyste saamgestel om die onderskeie kraswoorde te identifiseer. Die navorsers (Wang *et al., 2014*) het vreemde vorms van kraswoorde (*f_ck, f%ck, $hit* en *a55*) by die lys gevoeg ten einde te verseker dat alle kraswoorde geïdentifiseer word. Wang *et al.* (2014:218) het bevind dat

vloekwoorde sowat 1,15% van die woorde wat in korpora[344] ondersoek is, uitmaak of 7,73% (een uit dertien) van korpora bevat vloekwoorde.

In hierdie afdeling oor die gebruik van kraswoorde het ek vervolgens ondersoek ingestel na die voorkoms van kraswoorde in die gespesialiseerde korpus en die verdagte tekste om sodoende vas te stel of die gebruik van kraswoorde deur die onderskeie outeurs wat in hierdie studie gebruik is wel as kenmerkend beskou kan word. Daar is ook bepaal of daar kraswoorde is wat kenmerkend van Afrikaans is soos dit op Twitter gebruik word.

Die gebruik van kraswoorde is in *ATLAS.ti* geanaliseer en die voorkoms van kraswoorde (in beide die gespesialiseerde korpus en verdagte tekste) is as ʼn Microsoft Excel-dokument gestoor en die genormaliseerde waarde (tot 1 000 woorde) van die voorkoms van kraswoorde is bereken (sien Figuur 7.33):



Figuur 7.33 Kraswoorde in die gespesialiseerde korpus en verdagte tekste

Figuur 7.33 dui die genormaliseerde waarde per 1 000 woorde aan vir die gebruik van kraswoorde in die gespesialiseerde korpus en die verdagte tekste. Vervolgens is die gebruik van kraswoorde in die gespesialiseerde korpus ondersoek.

7.8.1 Gespesialiseerde korpus: kraswoorde

Vanuit Figuur 7.33 blyk dit dat elk van die outeurs wat deel van die gespesialiseerde korpus vorm, kraswoorde teen ʼn ander frekwensie gebruik. ʼn Interne vergelyking is ook op die gebruik van kraswoorde uitgevoer en vanuit hierdie analise kom dit voor asof die gebruik van

[344] Wang *et al.* (2014) se korpus het uit 51 miljoen twiets bestaan.

kraswoorde eweredig deur die korpora versprei is. Elk van die outeurs in die gespesialiseerde korpus gebruik kraswoorde op só ŉ manier in hulle twiets dat die frekwensie van die gebruik van hierdie woorde as kenmerkend tot elk van die outeurs beskou kan word.

O9 (43,92) is die outeur wat die meeste van kraswoorde gebruik maak (sien Figuur 7.33). Wanneer daar na die soort kraswoorde ondersoek ingestel is, blyk dit dat gebruikers nie huiwer om kraswoorde soos *fok, kak* en *doos* te gebruik om uiting aan hulle emosies te gee nie. Wanneer die konkordansie van die woorde ondersoek word, is dit duidelik dat die woorde in ŉ bepaalde konteks gebruik word om ekstra betekenis aan ŉ twiet te gee. Voorbeeld 304 tot 313 beeld die gebruik van kraswoorde uit:

Voorbeeld 304: **O9** – Want elke 2de **fok** is kwansuis ’n Akademikus
Voorbeeld 305: **O9** – **Fok** julle almal
Voorbeeld 306: **O5** – **fok** jou toe?
Voorbeeld 307: **O4** – Nie eers die Rand het so hard **gekak** soos ons vanaand nie
Voorbeeld 308: **O4** – herinner daar was op jou hart **gekak** nie
Voorbeeld 309: **O9** – Haha. Ek is die tipe **doos** wat aanneem maand-einde
Voorbeeld 310: **O9** – Ons almal is bereid om ’n **doos** vir iemand te wees
Voorbeeld 311: **O4** – Wat van jy hou op ’n dom **doos** wees?
Voorbeeld 312: **O9** – sal dit **bedoos** wees
Voorbeeld 313: **O9** – **Fucking** break kan gee sal dit bedoos wees

Die woord *doos* word teen ŉ genormaliseerde waarde per 1 000 woorde 23,55 keer in O9 aangetref, 1,02 keer in O4 en 0,04 keer in die verwysingskorpus. O9 is dus meer geneig om hierdie woord te gebruik. O9 maak verder ook van die woord *bedoos* (sien voorbeeld 312) gebruik. Hierdie verbuiging van die woord kom nie in die ander outeurs se korpora voor nie en is ook nie in die verwysingskorpus gevind nie.

O4 is in afdeling 7.6.4 (Gespesialiseerde korpus: dubbelmorfologie) geïdentifiseer as die outeur wat die meeste van die voorvoegsel ***ge*** gebruik maak (sien voorbeeld 296). In voorbeeld 307 en 308 gebruik die outeur weer eens die voorvoegsel ***ge***, maar word voor ŉ kraswoord geplaas. Aangesien geen van die ander outeurs in die gespesialiseerde korpus die voorvoegsel (***ge***) voor kraswoorde gebruik nie, kan dit as kenmerkend van die taalgebruik van O4 beskou word[345].

[345] Daar is geen voorbeelde gevind in die verwysingskorpus van die kraswoord ***gekak*** nie, maar ten tyde van die soektog (7 September 2017) is die woord wel op Twitter gevind, en word dit dus ook deur ander outeurs as O4 gebruik.

Daar word ook van kodewisseling gebruik gemaak (IT[346]) wanneer kraswoorde gebruik word. Hierdie woorde sluit *fucking*, *shit*, *bitch*, *fuck* en *fucked* in en word nie net deur O9 gebruik nie, maar ook meer frekwent. Daar is enkele kreatiewe verbuigings van kraswoorde gevind:

Voorbeeld 314: **O2** – Eenkeer het 'n onderwyser met 'n lisp vir ons klas gesê "**S(h)it** nou julle" en ek belowe ek het nog nooit so hard in my lewe gelag nie.

Voorbeeld 315: **O9** – As ons na haar verwys sê ons sommer **(e)SJ(y)fuckedup**.

Voorbeeld 316: **O1** – Mens kry dosente en dan kry mens **do(o)sente**

Die bogenoemde voorbeelde kan as hapax legomena[347] van die outeurs beskou word (elk van hierdie voorbeelde kom slegs een keer in elk van die korpora voor en is nie in die verwysingskorpus gevind nie. Daar is op Twitter ŉ basiese soektog uitgevoer ten einde te bepaal of hierdie woorde dalk deur ander Twittergebruikers ook gebruik word, maar daar kon nie ander voorbeelde gevind word nie behalwe in die twiets van die deelnemers[348].

Die hoë gebruiksfrekwensie van kraswoorde kan as kenmerkend van gesproke taal beskou word, en dui dus op die kombinasie van geskrewe en gesproke taal wat op RBK[349] gevind word en sluit dus aan by die bevindinge van Baron (1998:139-141) en Belmore en Collot (1996:25-27)[350]. In afdeling 7.8 is daar na Wang *et al.* (2014:218) se bevinding verwys dat vloekwoorde sowat 1,15% van die woorde uitmaak wat in korpora[351] ondersoek is. Wanneer die persentasie kraswoorde bereken is wat in die gespesialiseerde korpus voorkom, is daar bevind dat 0,60% van die woorde wat in die gespesialiseerde korpus voorkom kraswoorde is[352]. Die gebruik van kraswoorde in die Afrikaanse gespesialiseerde korpus wat vir hierdie studie gebruik word, blyk dus aansienlik laer te wees as die voorkoms van kraswoorde in die korpus van Wang *et al.* (2014:218), maar kraswoorde is steeds aanwesig in die gespesialiseerde korpus en kan vervolgens as ŉ kenmerk van Afrikaans beskou word soos dit op Twitter gebruik word. In die hieropvolgende afdeling is die gebruik van kraswoorde ondersoek.

[346] IT verwys na ŉ ingebedde-taal lekseem (sien afdeling 7.5.1 Kodewiselling).
[347] Hapax legomena verwys na woorde wat slegs een keer in ŉ teks voorkom (Coulthard & Johnson, 2010:526) – sien ook afdeling 3.4.1 (Benaderings tot outeurskapidentifikasie).
[348] Die soektog is op 24 Oktober 2016 uitgevoer. Dit is belangrik om in ag te neem dat daar elke sekonde van die dag nuwe inligting op Twitter geplaas word, en dus kan daar nou ander gebruikers wees wat ook die woorde soos genoem in voorbeel 304 tot 306 gebruik het.
[349] Rekenaarbemiddelde Kommunikasie
[350] Sien hoofstuk 3 (afdeling 3.3.3 – Forensiese linguistiek en informele tekste).
[351] Wang *et al.* (2014) se korpus het uit 51 miljoen twiets bestaan.
[352] Die berekening is gedoen deur die totale hoeveelheid kraswoorde deur die totale hoveelheid kraswoorde in die gespesialiseerde korpus te deel en met 100 te vermenigvuldig ten einde die persentasie te kry.

7.8.2 Verdagte tekste: kraswoorde

In afdeling 7.8.1 is daar ondersoek ingestel na die gebruik van kraswoorde soos in die gespesialiseerde korpus gebruik. Daar is bevind dat kraswoorde wel tot só 'n mate deur die onderskeie outeurs in die gespesialiseerde korpus gebruik word dat dit wel as 'n manier beskou kan word om outeurs van tekste te help identifiseer.

Wanneer die drie verdagte tekste ondersoek is ten opsigte van die gebruik van kraswoorde (sien Figuur 7.33), blyk dit dat daar slegs in die korpus van V2 (3 kraswoorde of 2,13 genormaliseer tot 1 000 woorde) en V3 (2 kraswoorde of 4,85 genormaliseer tot 1 000 woorde) is waar kraswoorde voorkom. In die vergelyking van die frekwensie van die gebruik van kraswoorde soos in die verdagte tekste met die gespesialiseerde korpus, is daar bevind dat V1 (0,00) en O6 (0,18), V2 (2,13) en O1 (3,58) en V3 (4,85) en O1 (3,58) ooreenstem. Vervolgens verskaf ek voorbeelde van die kraswoorde wat in die verdagte tekste voorkom om te kon bepaal of daar enige stylooreenkomste tussen die gebruik van kraswoorde in die verdagte tekste en die gespesialiseerde korpus is:

Voorbeeld 317: **V2** – Wtf Apple? iOS 10 is amper so **shit** soos Trump.

Voorbeeld 318: **V2** – A het seker **flippen** baie data, aangesien hy hulle altyd so baie kan message.

Voorbeeld 319: **V2** – Ek love hoe mense op Twitter so sensitief is deesdae. Sê net een ding, dan **bitch** en moan almal daaroor. 😂👏😂

Voorbeeld 320: **V3** - Ek het omgegee.. nou is ek net van **fok** jou ✌️

Voorbeeld 317 tot 320 toon die vier kraswoorde aan soos dit in die verdagte tekste gebruik word. Vanuit die voorbeelde blyk dit dat daar vier verskillende kraswoorde in die verdagte tekste voorkom (*shit*, *flippen*, *bitch* en *fok*). Wanneer die gebruik van dié kraswoorde verder ondersoek word, blyk dit dat wanneer voorbeeld 317 vergelyk word met O4 se gebruik van kraswoorde, die semantiese konteks waarin dié woorde gebruik word nie as ooreenstemmend beskou kan word nie. O4 gebruik kraswoorde om aan lesers vrae te vra of om 'n sterk emosie te beklemtoon. Soos in voorbeeld 320 aangetoon, gebruik V3 kraswoorde om 'n negatiewe gevoel te beklemtoon. In die korpus van O5 word *fok* gebruik om 'n negatiewe gevoel te beklemtoon (*… maar die wat dit kort .. Nee* ***fok*** *hulle sommer…*). Daar kan daarom aangevoer word dat O5 na gelang van die manier waarop die kraswoord gebruik word die moontlike outeur van die teks kan wees.

Die drie kraswoorde wat deur V2 gebruik word (*flipped*, *shit* en *bitch*), word ook deur O1 (*shit* en *bitch*), O5 (*flippen*, *shit* en *bitch*) en O9 (*bitch* en *shit*) gebruik. Daarom kan O1, O5 en O9 as

moontlike outeurs van V2 beskou word. Aangesien O1 die outeur is wat die grootste ooreenkoms met die kraswoordgebruik van V2 toon, is daar verder ondersoek ingestel na die manier waarop die kraswoorde deur die onderskeie outeurs gebruik word. V2 (sien voorbeeld 309) se gebruik van die woord *bitch* vergelyk met die kraswoordgebruik van O1 (*Jip, dis hoe siek ek is. Boonste lugweginfeksie is 'n* ***bitch****.*) en O9 (*My pa bel my om te* ***bitch*** *oor mense wat dit waag om Land Cruisers te badmouth*).

Die kraswoord *shit* word gebruik om aan te toon hoe sleg ŉ sekere aspek is. O1 (*Ek's te oud vir die* ***shit***) is die enigste outeur wat ook die kraswoord (en kodewisseling) gebruik om aan te dui hoe sleg ŉ bepaalde aspek is. Die laaste kraswoord (*flippen* – sien voorbeeld 318) word in die gespesialiseerde korpus slegs deur O5 gebruik. V2 gebruik die kraswoord as ŉ byvoeglike naamwoord van aantal en in die korpus van O5 word *flippen* nie op dieselfde manier gebruik nie.

Vanuit die ondersoek na die gebruik van kraswoorde − soos in die verdagte tekste gevind − kan die afleiding gemaak word dat V2 moontlik deur O1 geskryf is terwyl die outeur van V3 moontlik O5 kan wees. Daar kom nie kraswoorde in V1 voor nie en dus is die ondersoek nie op hierdie verdagteteks uitgevoer nie.

Na aanleiding van die analise van die semantiese aspekte met betrekking tot die voorkoms van kodewisseling (IT en MT asook dubbelmorfologie), die invloed van Engels, tussenwerpsels en ook die gebruik van kraswoorde kan hierdie aspekte as insiggewend beskou word en is noodsaaklik in die identifikasie van mikroblogouteurs.

Alhoewel daar nie vir elk van die verdagte tekste in die genoemde semantiese aspekte wat ondersoek is korpora vanuit die gespesialiseerde korpus geïdentifiseer kon word wat sonder enige twyfel as die outeurs van die verdagte tekste beskou kan word nie, kan die semantiese aspekte wel die aantal moontlike outeurs elimineer en telkens is daar ŉ maksimum van twee outeurs wat na gelang van hierdie ondersoek as moontlike outeurs van elk van die verdagte tekste beskou is. In die hieropvolgende afdeling is daar ondersoek ingestel na die gebruik van emotikons.

7.9 Foute-analise

Hubbard (1994:1) het ondersoek na ŉ foute-analise ingestel, maar bepaal dat daar nie genoegsame ooreenkomste tussen teks X en A voorkom om ŉ aanname te maak dat die patroonmatigheid van spelfoute as ŉ aanduiding van outeurskap gebruik kan word nie.

Lawrence en Solan (2004:457) het na die eklektiese benadering[353] verwys as 'n benadering waar spelling, punktuasie[354] en taalfoute ondersoek word en die foute vanuit verskillende tekste dan met mekaar vergelyk word ten einde ooreenstemmende foute en outeurs van spesifieke tekste te identifiseer. In hierdie afdeling is daar na spelfoute, die weglating van die tweede nie in die dubbelontkenning en letter-herhaling ondersoek ingestel.

Volgens Darics (2013:144) verwys letter-herhaling na die herhaling van letters om 'n geskrewe woord te skep wat gesproke eienskappe aan lesers oordra. Darics (2013:144) gaan voort deur te stel dat die woorde ortografies deur outeurs gemanipuleer word om klank en emosie aan lesers oor te dra, maar dat die soort woorde nie met onomatopee verwar moet word nie. Die woorde moet eerder as spelfoute beskou word, maar doelbewuste foute met 'n spesifieke funksie. Dit sluit dus aan by Coulthard (2005:15) se verwysing en ondersoek na bevoegdheidsfoute. 'n Bykomende aspek wat Coulthard (2005:16) uitlig in verband met die ondersoek van foute wanneer outeurskapidentifikasie op elektroniese tekste uitgevoer word, is die moontlikheid dat hierdie foute dalk slegs tikfoute kan wees. Ten einde die ondersoek te kon uitvoer, is die onderskeie korpora se woordelyste in WordSmith Tools geanaliseer en woorde wat verkeerd gespel is, is aangeteken. Die konkordansie opsie (WordSmith Tools) is verder gebruik om die uitlating van die ontkennende nie te identifiseer.

Die voorkoms van foute in die gespesialiseerde korpus toon aan dat die outeurs oor die algemeen probeer om die spelkonvensies van Afrikaans te volg. Daar is spelfoute in elk van die tien gespesialiseerde korpora geïdentifiseer, maar talle van die foute kan as tikfoute beskou word. Hierdie tikfoute sluit in *seminar* (O1), *ondehoud* (O4), *einlik* (O5)*, wonderike* (O6), *veseker* (O7), *afgwagting* (O8), *tiranie* (V1) en *inteksverwysing* (V2). Hierdie woorde (foute) word slegs een maal in elk van die korpora aangetref. Vervolgens word spelfoute in Tabel 7.27 voorgehou wat herhaaldelik voorkom:

Tabel 7.27 Spelfoute

Korpus	**Totale aantal twiets**	**Totale aantal sinne**	**Totale aantal woorde**	**Spelfout**	**Voorkoms in korpus**	**Genormaliseerde waarde tot 1 000 woorde van die spelfout**
O2	589	1577	12 259	wereld	2	0,16
O2	589	1577	12 259	vroeer	2	0,16
O2	589	1577	12 259	gese	2	0,16
O2	589	1577	12 259	se	10	0,82
O2	589	1577	12 259	he	3	0,24
O4	484	978	7 110	oppad	2	0,28
O4	484	978	7 110	gistraant	3	0,42
O4	484	978	7 110	Namibie	3	0,42

[353] Sien afdeling 3.4.1 (Benaderings tot outeurskapidentifikasie).
[354] In afdeling 7.2.2 en 7.2.3 is daar reeds na die gebruik van leestekens (afdeling 7.2.2.1 en 7.2.2.2) en hoofletters (afdeling 7.2.3.5 en 7.2.3.6) ingestel.

Korpus	Totale aantal twiets	Totale aantal sinne	Totale aantal woorde	Spelfout	Voorkoms in korpus	Genormaliseerde waarde tot 1 000 woorde van die spelfout
O6	768	1657	16 340	Kasette	2	0,12
O6	768	1657	16 340	geireteerd	2	0,12
O6	768	1657	16 340	erens	5	0,31
O7	457	790	5 655	reen	15	2,65
O7	457	790	5 655	droe	2	0,35
O7	457	790	5 655	oppad	11	1,95
O7	457	790	5 655	vroe	5	0,88
O7	457	790	5 655	komersieel	2	0,35
O7	457	790	5 655	musiekant	2	0,35
O7	457	790	5 655	wardeer	2	0,35

Soos in Tabel 7.27 voorgehou, is dit slegs O2, O4, O6 en O7 wat meer as een keer dieselfde woord verkeerd spel en wat dus nie as ’n tikfout beskou kan word nie. Indien die woorde wat in Tabel 7.27 gelys word in die verdagte tekste voorkom, kan die genoemde outeurs as moontlike outeurs van die tekste beskou word. ’n Verdere afleiding wat gemaak kan word, is dat die gebruik van kappies en deeltekens nie noodwendig plaasvind soos daargestel in die taalreëls nie.

Ten opsigte van die dubbelontkenning wat in Afrikaans gebruik word, blyk dit dat die outeurs wat in Afrikaans twiet, geneig is om die tweede nie uit te laat. Aangesien hierdie kenmerk in al die outeurs wat deel van die gespesialiseerde korpus uitmaak, gevind word, kan dit as kenmerkend beskou word van die Afrikaans wat op Twitter gebruik word.

Die korrekte gebruik van die onbepaalde lidwoord (’n) blyk ook ’n aspek te wees wat talle van die outeurs foutief gebruik, maar aangesien dit slegs in die korpora van O3, O8, O9 en O10 telkens korrek gebruik word, kan die outeurs as moontlike outeurs van een van die verdagte tekste oorweeg word indien daar nie foute ten opsigte van die onbepaalde lidwoordgebruik voorkom nie. Beide V1 en V2 gebruik telkens die onbepaalde lidwoord korrek en in die geval van V3 is daar slegs ’n enkele onbepaalde lidwoord wat as *n* gebruik word.

Die gebruik van letter-herhaling ten einde klank en emosie aan lesers oor te dra, is ’n verdere aspek wat telkens op Twitter gevind word (sien voorbeeld 321 tot 324):

Voorbeeld 321: **O1** – **Ekskuuuuus**

Voorbeeld 322: **O5** – Glimlag soos ’n **idioooot!!**

Voorbeeld 323: **O6** – Nooit te oud te leer, maar nóú moet ek **bittttttter** vinnig leer!

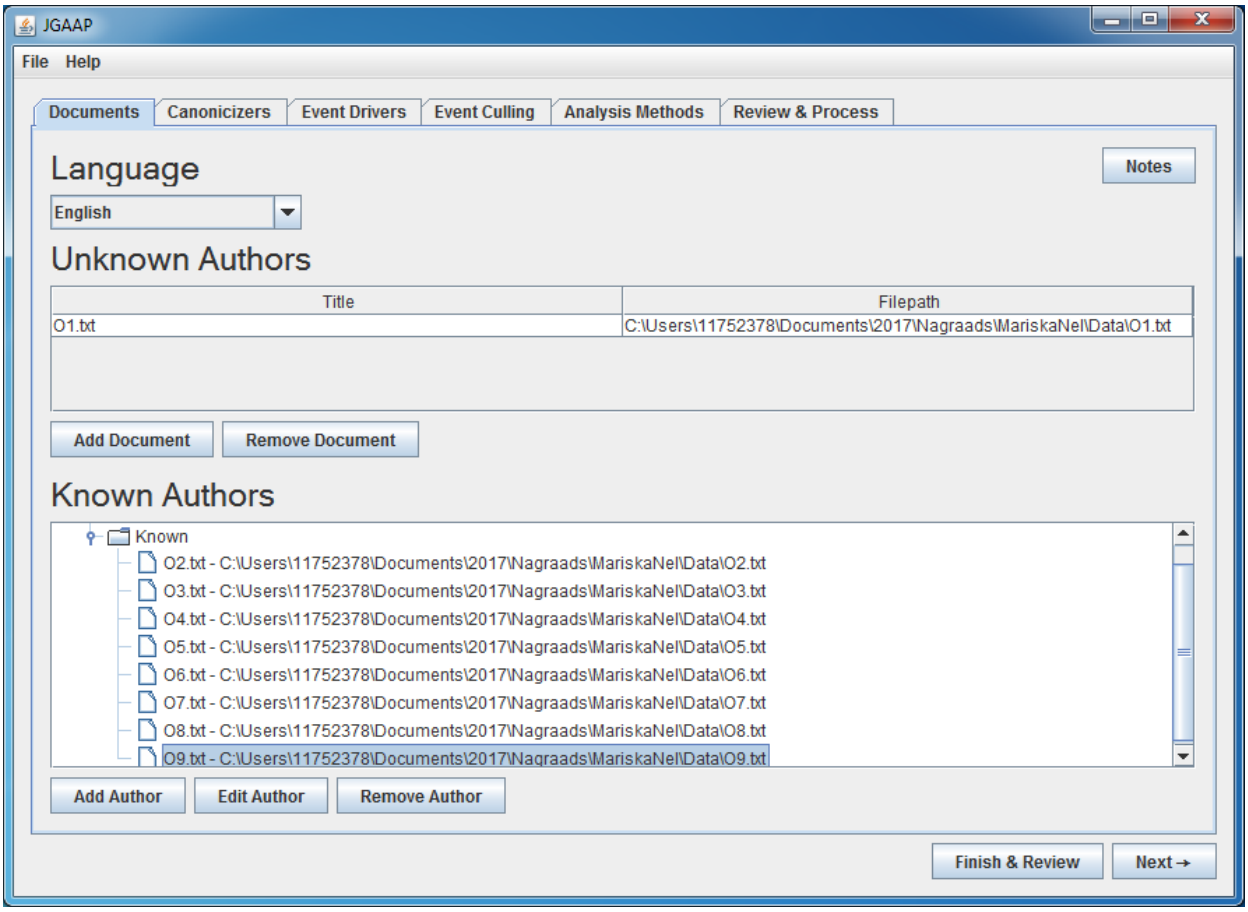
Voorbeeld 324: **O7** – **Reeeeeeeeeeeeeeen** asb Here laaaat dit net **reeeeen....**

Die gebruik van letter-herhaling blyk slegs ’n kenmerk van Afrikaans soos dit op Twitter gebruik word wat ooreenstem met internasionale navorsingsbevindinge. Net soos gebruikers vanuit

ander tale, maak Afrikaanse Twitter-gebruikers van letter-herhaling gebruik om ŉ informele gesprektaal te vestig en word klank en emosie aan leser hierdeur oorgedra.

7.10 ŉ N-gramanalise

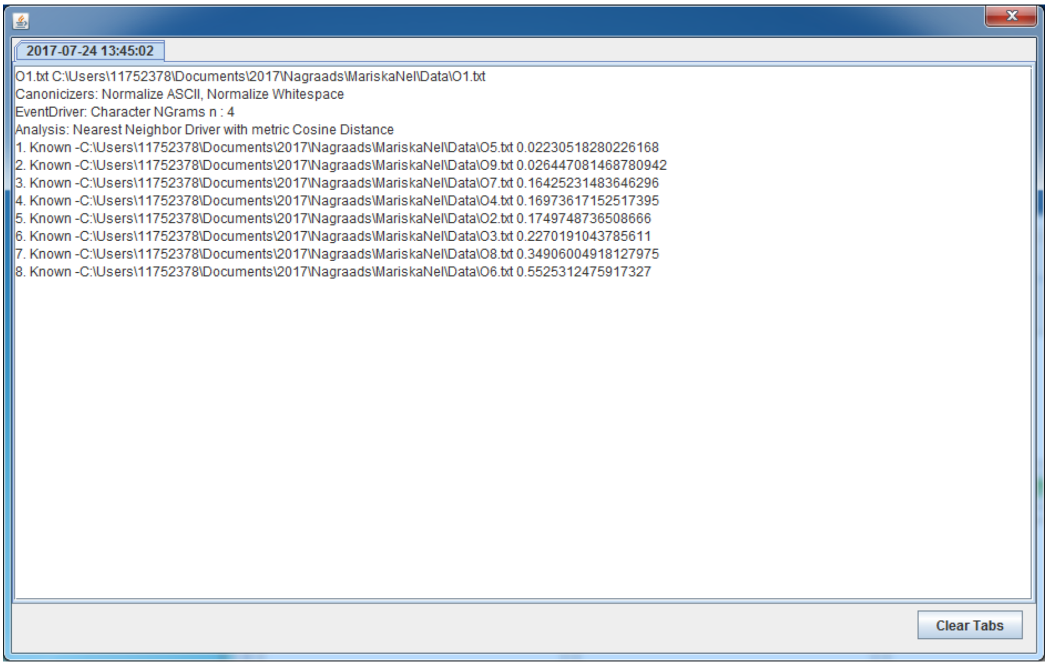
Volgens Thiart (2014:16) verwys ŉ n-gram in rekenaarlinguistiek na ŉ reeks items (*n*) wat in opeenvolging in gesproke of geskrewe taal voorkom[355]. ŉ N-gramanalise is met behulp van JGAAP uitgevoer. Die program *Java Graphical Authorship Attribution Program* is deur Evaluating Variation in Language (EVL) van Duquesne Universiteit ontwikkel. Skermskoot 7.3 is ŉ voorbeeld van hoe die program se koppelvlak lyk:



Skermskoot 7.3 JGAAP-koppelvlak

Ten einde die n-gramme te kon bereken, is die onderskeie korpora as txt-dokumente in die program ingevoer voor die analise kon plaasvind (sien Skermskoot 7.3):

[355] Sien afdeling 5.8.8.4 (n-gram).



Skermskoot 7.4 Txt-dokumente in JGAAP

O1.txt O1.txt
Canonicizers: Normalize ASCII, Normalize Whitespace
EventDriver: Character NGrams n : 4
Analysis: Nearest Neighbor Driver with metric Cosine Distance
1. Known - O5.txt 0.02230518280226168
2. Known - O9.txt 0.026447081468780942
3. Known - O7.txt 0.16425231483646296
4. Known - O4.txt 0.16973617152517395
5. Known - O2.txt 0.1749748736508666
6. Known - O3.txt 0.2270191043785611
7. Known - O8.txt 0.34906004918127975
8. Known - O6.txt 0.5525312475917327

O2.txt O2.txt
Canonicizers: Normalize ASCII, Normalize Whitespace
EventDriver: Character NGrams n : 4
Analysis: Nearest Neighbor Driver with metric Cosine Distance
1. Known - O3.txt 0.1672204211199676
2. Known - O1.txt 0.1749748736508666
3. Known - O9.txt 0.1829916711009305
4. Known - O5.txt 0.19326415492341142
5. Known - O7.txt 0.22984999742692191
6. Known - O4.txt 0.23807982411830242
7. Known - O8.txt 0.3224928939374625
8. Known - O6.txt 0.5429167688272186

Skermskoot 7.5 Afvoer vanaf JGAAP

Die bevindinge kon[356] soos voorgestel in Skermskoot 7.5 geanaliseer word. ŉ N-gramanalise is uitgevoer ten einde te kon bepaal of hierdie analise wel kan aandui of tekste vanaf Twitter (in Afrikaans) wel ooreenstem. In hierdie analise is die verdagte tekste met die gespesialiseerde korpus vergelyk – sien Tabel 7.28:

Tabel 7.28 ŉ N-gramanalise

N-gramanalise: V1							
V1: bi-gram		**V1: tri-gram**		**V1: vier-gram**		**V1: vyf-gram**	
O2	0,0478	O2	0,1929	O2	0,3110	O2	0,2835
O9	0,0820	O9	0,2151	O9	0,3719	O7	0,3467
O3	0,0883	O1	0,2453	O3	0,3890	O4	0,3852
N-gramanalise: V2							
V2: bi-gram		**V2: tri-gram**		**V2: vier-gram**		**V2: vyf-gram**	
O9	0,0477	O9	0,1502	O9	0,2456	O3	0,2557
O1	0,0539	O1	0,1685	O5	0,2909	O7	0,2635
O5	0,0548	O5	0,1774	O1	0,3117	O4	0,2819
N-gramanalise: V3							
V3: bi-gram		**V3: tri-gram**		**V3: vier-gram**		**V3: vyf-gram**	
O5	0,0547	O5	0,1880	O5	0,3185	O2	0,3681
O9	0,0665	O9	0,2119	O9	0,3455	O5	0,3735
O1	0,0675	O1	0,2131	O1	0,3889	O1	0,4505

Vanuit Tabel 7.28 blyk dit dat V1 telkens (bi-gram, tri-gram, vier-gram sowel as die vyf-gram) met O2 ooreenstem, en daarom kan O2 as die moontlike outeur van V1 beskou word na gelang van die n-gramanalise. V2 en O9 stem ooreen behalwe in die geval van die vyf-gram (O3 kan moontlik hier die outeur wees), maar aangesien O9 in die meeste gevalle (drie van die vier analises) as die outeur geïdentifiseer is wat die grootste ooreenkoms met V2 toon, kan O9 dus na gelang van die n-gramanalise as die moontlike outeur van V2 beskou word. V3 en O5 stem ooreen in terme van die bi-, tri- en vier-gram analises terwyl O2 en V3 ten opsigte van die vyf-gram analise ooreenstem. O5 en O2 (in die vyf-gram analise met V3) verskil slegs met 0,0054. O5 is dus − volgens die n-gramanalise − die outeur van V3.

Die n-gramanalise se resultate is relatief konsekwent (buiten in die geval van V2) om ŉ enkele outeur as die beste kandidaat vir outeurskap van ŉ bepaalde verdagte teks aan te dui en dus kan die gebruik van ŉ n-gramanalise as nog ŉ metode beskou word wat in kombinasie met verdere ondersoeke gebruik kan word ten einde outeurs van mikrobloginskrywings te identifiseer.

356 Sien bylaag I vir die volledige bevindinge van die analise van die drie verdagte tekste vanuit JGAAP.

7.11 Sintese

In hierdie afdeling word die bevindinge van die analises van die bogenoemde kategorieë en die kenmerke van Afrikaans bespreek soos waargeneem in die korpora wat in die studie gebruik is. Laastens, word die bevindinge vanuit die stilometriese analise[357] weer voorgehou om sodoende te kon bepaal wie die outeurs van die verdagte tekste is. In hierdie hoofstuk is ŉ stilistiese en teksanalise eerstens uitgevoer om te bepaal of die outeurs van mikrobloginskrywings wel geïdentifiseer kan word en tweedens wat as kenmerke van Afrikaans beskou kan word soos dit op hierdie sosialemediaplatform gebruik word. Ten einde ŉ stilistiese en teksanalise te kon uitvoer, is elk van die datastelle (gespesialiseerde korpus en die verdagte tekste) geanaliseer ten opsigte van sewe verskillende kategorieë[358], naamlik algemene kenmerke (afdeling 7.2), sintaksis (afdeling 7.3), morfologie (afdeling 7.4), semantiek (afdeling 7.5), kodewisseling (afdeling 7.6), tussenwerpsels (afdeling 7.7), kraswoorde (afdeling 7.8, foute-analise (afdeling 7.9), en n-gramanalise (afdeling 7.10).

7.11.1 Kenmerke van Afrikaans op mikroblogs

In hoofstuk 1 word die onderskeie navorsingsvrae en die doelwitte van die studie uiteengesit (sien afdeling 1.3 en 1.4). Een van die navorsingsvrae stel die vraag of daar enige kenmerke in Afrikaans sigbaar is in die taal soos dit op mikroblogs voorkom. Hierdie vraag is ondersoek ten einde ŉ norm daar te kan stel van hoe Afrikaans lyk wanneer dit op mikroblogs (in die geval van my studie – Twitter) gebruik word.

Alhoewel die navorsingsvraag ietwat verskil van die forensiese analise wat in die studie uitgevoer is, was die identifisering van kenmerke van Afrikaans tegnies die eerste stap in die outeurskapidentifikasieproses. Tot op hede bestaan daar nog nie ŉ korpus waar Afrikaans ingesluit is soos dit op ŉ sosialemediaplatvorm voorkom nie.

Ten einde kenmerke van Afrikaans te kon identifiseer soos dit op Twitter voorkom, is daar telkens in die empiriese analise van die stilistiese analise na die tekskenmerke verwys. In afdeling 7.1 (inleiding) van hierdie hoofstuk word daar na sewe kategorieë verwys waarvolgens die ondersoeke uitgevoer is. Vervolgens word die kenmerke van Afrikaans voorgehou soos dit in elk van die kategorieë geïdentifiseer is.

[357] Sien afdeling 6.8 (Sintese van die stilometriese analise).
[358] Sien afdeling 7.1 (Inleiding).

7.11.1.1 Algemene kenmerke

Algemene kenmerke verwys na Twitter-spesifieke kenmerke en die leestekengebruik soos dit op Twitter gebruik word. Alhoewel die ondersoek na die teksformaat volgens McMenamin (2002:216) na die uitleg van dokumente verwys – die manier waarop die datum geskryf word, leestekengebruik by aanspreekvorms en selfs die lynspasiëring – word die uitleg van twiets deur die platform self bepaal[359]. In die ondersoek na algemene kenmerke oor die teksformaat is die gebruik van die hutsteken, hutswoorde, verwysings na gebruikers en die gebruik van hiperskakels geanaliseer.

Vanuit die ondersoek na die bogenoemde aspekte is daar eienskappe geïdentifiseer wat as kenmerkend van Afrikaans beskou kan word soos dit op Twitter gebruik word. Die eienskappe sluit in:

- ŉ Gemiddeld van twee **hutstekens** per twiet is kenmerkend van Afrikaans soos dit op Twitter gebruik word. Daar word nie spesifiek voorkeur gegee aan woorde of frases ná ŉ hutsteken nie, beide kom gereeld voor (#KCOM112 #syleerditbymy). Verder vind kodewisseling gereeld plaas (*#ekhetditdesign*) en afkortings en kraswoorde word as hutswoorde gebruik. Hutswoorde kom aan die begin, in die middel of aan die einde van ŉ twiet voor en ŉ twiet kan ook net uit hutswoorde bestaan.

- **Hutswoorde** word gebruik om onderwerpe te beklemtoon (*#Beeld*, *#PhilMolefe*) en as ŉ manier om twiets in konteks te plaas met betrekking tot nuuswaardige gebeure (*#Stem2014*, *#OscarPistorius*) of om met ŉ internasionale gehoor te skakel.

  Verder het hutswoorde ook ten doel om ŉ spesifieke gevoel oor te dra of om ŉ opinie oor ŉ spesifieke gebeurtenis oor te dra. Outeurs stel dan ook duidelik watter emosie deur die gebeurtenis ervaar word (kwaad – *#dingewatmyafpiss).*

  Die persentasie twiets waarin daar hutswoorde voorkom (20,5%) blyk effens hoër te wees as wat Hong *et al.* (2011:519) bevind het in ander tale (byvoorbeeld Duits – 18%, Engels – 14%, en Nederlands – 13%).

- **Verwysings na gebruikers** word nie in elke twiet gebruik nie. Die rede waarom daar na gebruikers verwys word, blyk ŉ bepalende faktor in die plasing van die @-gebruikerskonvensie te wees. Indien gebruikers se twiets aangehaal word (nie hertwiet

[359] Sien hoofstuk 2 (afdeling 2.7.2.2 – Twitter).

nie), word verwysings na gebruikers aan die begin van twiets geplaas, meestal binne aanhalingstekens (sien voorbeeld 30).

Die @-gebruikerskonvensie word ook gebruik om in gesprek[360] te tree met ander Twitter-gebruikers, en in só ʼn geval word verwysings na gebruikers in die sin gebruik as ʼn aanspreekvorm. Hierdie gebruik beaam die stelling dat taal op sosiale media as ʼn hibriede vorm van ʼn kombinasie tussen geskrewe en gesproke taal beskou kan word. Die laaste kenmerk van verwysings na gebruikers is die manier waarop outeurs hierdie konvensie gebruik om ander gebruikers bewus te maak van spesifieke stellings (twiets) om sodoende ʼn diskoers te inisieer waaraan verskeie gebruikers kan deelneem.

- 22,4% van die twiets wat in die studie ondersoek is, bevat verwysings na gebruikers. ʼn Persentasie wat ietwat laer blyk te wees as Hong *et al.* (2011:519) se bevinding, maar steeds ʼn ooreenkoms toon met Duits (36%).

- Die funksie van verwysings na gebruikers in Afrikaans stem ooreen met Honeycutt en Herring (2009:2) se bevinding dat daar tussen die @-ligging (verwysing na ʼn spesifieke plek) en ʼn persoonlike aanspreekvorm onderskei moet word.

- Die analise van die gebruik van **hiperskakels** toon aan dat daar geen vaste plek is waar hiperskakels in twiets gebruik word nie, maar dat gebruikers wat in Afrikaanse twiet wel van hiperskakels gebruik maak. Die soort hiperskakel wat gebruik word, wissel tussen die verskillende gebruikers, maar skakels na ander sosialemediaplatforms en nuuswaardige gebeure is baie gewild.

- Daar kom in 18% van die twiets wat ondersoek is hiperskakels voor wat ooreenkomste toon met die gebruik van hiperskakels in Maleis, Koreaans en Nederlands (17%) soos deur Hong *et al*. (2011:519) bevind.

- Die ondersoek na die gebruik van leestekens soos voorgehou in afdeling 7.2.2, sluit aan by die navorsing van McMenamin (2002:216-231) waar die manier waarop leestekens in ʼn teks gebruik word of die afwyking van die norm gebruik kan word om outeurs van tekste te identifiseer. Verhoef (2001:8) stel in haar ondersoek na Internetafrikaans dat skryftekengebruik spaarsamig is, maar vanuit hierdie ondersoek kan daar aangevoer

[360] Hierdie bevinding sluit aan by Honeycutt en Herring (2009) se argument dat die @-gebruikerskonvensie as ʼn aanspreekvorm beskou kan word.

word dat die outeurs wat inskrywings op hierdie spesifieke sosialemediaplatform maak oor die algemeen van punte aan die einde van sinne gebruik maak en ook van meer as een leesteken (oortollige leestekens) ten einde ŉ spesifieke emosie aan lesers oor te dra. Wanneer outeurs emotikons, hutswoorde, hiperskakels en/of @-gebruikerskonvensies ná sinne of woorde gebruik (aan die einde van twiets), blyk dit ŉ algemene gebruik om nie leestekens te gebruik nie.

- Soortgelyk aan Rao *et al.* (2010:41) se bevinding dat meer as een leesteken (veral uitroep- en vraagtekens) gebruik word om sprekers se emosie (intonasie van sprekers se stem) in gesproke gesprekke na te boots, is hierdie kenmerk ook in twiets aanwesig wat in Afrikaans geskryf word.

- Ten spyte van die sosiale aard van die platform wat ondersoek is, gebruik die outeurs in hierdie studie oorwegend wel hoofletters aan die begin van sinne − ŉ kenmerk wat as algemeen beskou kan word van Afrikaans soos dit op Twitter gebruik word. Die gebruik van hoofletters op Twitter volg dus oorwegend die taalreëls van Afrikaans.

- Hoofletters word ook in Afrikaans − soos in die navorsing van Ritter *et al.* (2011:1527) − gebruik om belangrike inligting te beklemtoon of om bepaalde emosies oor te dra.

Dit blyk dus na gelag van die ondersoek na die algemene kenmerke dat die Afrikaans wat op Twitter gebruik word nie net ooreenkomste met bevindinge vanuit internasionale navorsing rakende taalveranderinge wat in die taal wat op Twitter gebruik toon nie, maar dat Afrikaans veral met ander Wes-Germaanse tale ooreenstem. Die sintaktiese kenmerke van Afrikaans word voorts voorgehou.

7.11.1.2 Sintaktiese kenmerke

Die sintaktiese ondersoek word in afdeling 7.3 voorgehou. In die ondersoek is daar na die verskillende sinstipes, die gebruik van woordsoorte, en die invloed van Engels ondersoek ingestel. Soos in die ondersoek na die algemene kenmerke van Afrikaans is daar wel eienskappe geïdentifiseer wat kenmerkend is van Afrikaans soos dit op Twitter gebruik word. Die kenmerke sluit in:

- **Stelsinne** is die gewildste **sinstipe** wat op Twitter gebruik word en selfs die laagste persentasie stelsinne per korpus is hoër as 50%. Die gebruik van stelsinne kan ook

toegeskryf word aan Twitter se definisie wat die vraag stel: "What's happening". Deur stellings, kan gebruikers vrae makliker beantwoord.

- **Vraagsinne** is die tweede gewildste **sinstipe** en nie net tradisionele vraagwoorde word gebruik om hierdie vrae te vorm nie. Daar is 32 vraagmerkers geïdentifiseer wat frekwent deur meer as een van die outeurs gebruik word. Vanuit die ondersoek kan daar ook aangevoer word dat vraagsinne gebruik word om kommentaar te lewer, retoriese vrae te stel, opinies te lewer of spesifieke gebruikers se opinies oor bepaalde kwessies in te win.

- **Uitroepsinne** is as die derde gewildste **sinstipe** op Twitter geïdentifiseer, gevolg deur **bevelsinne**. Uitroepsinne word meestal gebruik wanneer daar ŉ tussenwerpsel (sien voorbeeld 166 tot 174) in ŉ sin gebruik word – weer eens ŉ kenmerk wat betrekking het op die gesproke aard van Afrikaans. Bevelsinne se sinslengte is oor die algemeen baie korter en dit is kenmerkend dat daar oorwegend slegs van ŉ enkele sin in twiets gebruik gemaak word wanneer bevelsinne voorkom. Bevelsinne word gebruik te einde aangesprokenes te motiveer om bepaalde aksies uit te voer of om spesifieke emosies (bemoediging, motivering, verwensing of oorreding) oor te dra.

- In die analise van die woordsoortgebruik[361], is daar bepaal dat die volgende woordsoorte as die gewildste beskou kan word:
  - Adjektiewe – ASA
  - Konjunkte – KN
  - Naamwoorde – NEE
  - Voornaamwoorde – PB
  - Telwoorde – THAO
  - Partikels – UPO
  - Werkwoorde – VTHOG
  - Lidwoorde – LB

Die bostaande lys van woordsoorte is ŉ oorsig van die woordsoorte wat in elk van die woordsoortgroepe as die gewildste woordsoorte geïdentifiseer is[362].

[361] Sien hoofstuk 8 (Afdeling 8.6 – Aanbevelings vir verdere navorsing).
[362] Sien Bylaag F – woordsoorte en hoofstuk 8 (Afdeling 8.5 –Aanbevelings vir verdere navorsing).

Vanuit die ondersoek na die invloed van Engels op Afrikaanse twiets blyk dit dat lidwoorde uitgelaat word, en soos in voorbeeld 172 en 173, is die aantal keer wat die outeurs Afrikaanse sinne in Engels formuleer is van groter belang as die wyse waarop hulle dit doen aangesien daar nie ŉ bepaalde patroonmatigheid voorkom in die korpora wat in hierdie studie gebruik is nie.

7.11.1.3 Morfologiese kenmerke

Die ondersoek na morfologiese kenmerke, die gebruik van afkortings (foutloos of idiolekties, akronieme sowel as SMS-afkortings), inkorting en weglating word in afdeling 7.4 bespreek, en daar is etlike kenmerke van Afrikaans wat na gelang van dié ondersoek voorgehou kan word. Die aanname is gemaak dat die karakterbeperking wat op Twitter voorkom[363], aanleiding daartoe sal gee dat outeurs meer van reduksieprosesse in hulle tekste gebruik sal maak om sodoende "karakters te spaar" of soos Verhoef (2001:14) in haar studie bevind het dat die vervanging van woordgrense met morfeemgrense en fonetiese skryfwyses algemeen sal voorkom. Die bevindinge oor die algemene morfologiese kenmerke van Afrikaans sluit in:

- Daar word van afkortings gebruik gemaak – beide foutloos en idiolekties soos in afdelings 7.4.1.1 tot 7.4.1.4 voorgehou. Daar is bepaal dat *SA* ŉ gewilde afkorting blyk te wees (sien ook Tabel 7.18 vir ander foutlose afkortings wat in die gespesialiseerde korpus aangetref word). Talle wisselvorme (op ŉ foutiewe manier gebruik) van afkortings word gebruik en in dié geval is dit veral punte wat weggelaat word. Hierdie weglating van punte kan moontlik die gevolg van die karakterbeperking op die platform wees, maar na gelang van die gebruik van leestekens wat so na as moontlik aan die norm gebruik word soos daargestel deur die Afrikaanse taalreëls, is dit hoogs onwaarskynlik dat karakterbeperking die oorsaak is van die foutiewe gebruik van afkortings. Akronieme word ook op die platform gebruik en in teenstelling met die verwagte bevinding (dat afkortings kenmerkend is van gebruik op sosialemediaplatforms), is die gebruik van SMS-afkortings nie so volop nie. Dit blyk dat SMS-afkortings gebruik word wanneer die outeurs van kraswoorde gebruik maak.

- Met betrekking tot die voorkoms van inkorting en weglatings, toon die bevindinge van hierdie studie in teenstelling met Verhoef (2001) se bevinding dat die vervanging van woordgrense deur morfeemgrense en die aanwesigheid van ŉ fonetiese skryfwyse nie so sterk in die korpora teenwoordig is nie. Alhoewel daar wel voorbeelde van inkorting (sien afdeling 7.4.2) en weglatings (sien afdeling 7.4.3) in die onderskeie korpora

[363] Sien hoofstuk 2 (afdeling 2.6.2.2 – Twitter).

aangetref word, is dit veel eerder ’n kenmerk van die betrokke outeurs as kenmerkend van die taalgebruik van die platform.

7.11.1.4 Semantiese kenmerke

Soortgelyk aan die ondersoek na die morfologiese kenmerke is daar na die gebruik van emotikons ondersoek ingestel. 7.11.1.7 Gebruik van emotikons

Emotikons word deur Go *et al*. (2009) as een van die vyf unieke kenmerke van Twitter beskou. Crystal (2011) en Zappavigna (2012) ondersteun Go *et al*. (2009) se bevinding. In afdeling 7.8 (emotikons) het ek ondersoek ingestel na die gebruik van emotikons.

Daar is bevind dat elk van die tien outeurs vanuit die gespesialiseerde korpus van emotikons gebruik maak. In die verdagte tekste word emotikons in 66,76% van die korpora aangetref (sien Figuur 7.29). Daar kan dus aangevoer word dat ten spyte daarvan dat die aantal emotikons wat deur die onderskeie outeurs gebruik word verskil, emotikons wel as ’n kenmerk van Afrikaans beskou kan word soos dit op Twitter gebruik word.

Aangesien daar verskillende soorte emotikons is, het ek 11 emotikongroepe geïdentifiseer (sien afdeling 7.8.1) om te kon bepaal watter emotikons meer frekwent gebruik word. Figuur 7.31 toon die bevindinge van die ondersoek en vanuit die Figuur kan daar aangevoer word dat die gebruik van emotikons wat in die “gesig”-kategorie geplaas is die algemeenste gebruik word. Emotikons word in Afrikaans − ook in die navorsing oor Twitter en Engels deur Zappavigna (2012) en Crystal (2011) − gebruik om boodskappe uit te brei deur óf dit wat in ’n teks gesê word te beklemtoon óf om die emosies wat outeurs ervaar, uit te beeld. Daar is nie ’n enkele outeur wat emotikons slegs “negatief” of slegs “positief” gebruik nie.

In hierdie afdelings (7.9.1.1 tot 7.1.9.6) is die kenmerke van Afrikaans voorgehou soos dit op Twitter gebruik word in die eksploratiewe studie wat gedoen is. Alhoewel die korpus waarmee ek in die studie gewerk het betreklik klein is (aangesien die fokus van die studie was om te bepaal of outeurs van mikrobloginskrywings wel geïdentifiseer kan word), is daar steeds kenmerke van Afrikaans geïdentifiseer wat ooreenstem met die resultate van studies oor Twitter en die gebruik van taal.

Hierdie bevindinge plaas Afrikaans vervolgens parallel met ander tale ten opsigte van die invloed wat mikroblogs op taal het. Daar kan vanuit hierdie ondersoek ook sonder twyfel aangevoer word dat die hibriede aard van taal en die argument dat ’n taal wat op sosiale media gebruik word iewers tussen die geskrewe en gesproke taal geplaas kan word ook van Afrikaans

waar is. Ten slotte blyk dit ook dat die karakterbeperking wat op die platform (Twitter) aanwesig is nie afbreek doen aan die verstaanbaarheid van boodskappe nie. Kommunikasie vind steeds moeiteloos plaas en Afrikaans word deur gebruikers aangepas om aan die nuwe vereistes te voldoen wat aan die taal gestel word om kommunikasie moontlik te maak. Vervolgens word die bevindinge oor die identifisering van outeurs op Twitter voorgehou.

7.11.1.5 Kenmerke soos vanuit die ondersoek na kodewisseling

In die ondersoek na die gebruik van kodewisseling (afdeling 7.6) is daar na die invoeging van ŉ ingebedde taal-lekseem (IT) en ŉ meertalige-eiland (MT) gekyk. Vanuit die ondersoek na die bogenoemde aspekte is die onderstaande kenmerke van Afrikaans geïdentifiseer:

- Afrikaans word nie puristies benader op Twitter nie en IT en MT word deurlopend in al die korpora aangetref.
- Alhoewel daar wel van die outeurs is wat van anglisistiese uitdrukkings gebruik maak, is die eienskap nie kenmerkend nie.

Dit blyk dus dat Afrikaans wat op mikroblogs gebruik word wel as ŉ demotiese Afrikaans beskou word, wanneer demotiese Afrikaans volgens Ponelis (2009) se definisie beskou word – demotiese variëteit vertoon die grootste invloed van Engels en dit word volgens Ponelis (2009) as omgangstaal en spreektaal gebruik. Ook hierdie bevinding, rakende die voorkoms van kodewisseling en die invloed van Engels op Afrikaans soos dit op Twitter gebruik word sluit by een van Aisopos *et al.* (2012:187) se vier fundamentele maniere waarop mikroblogs fundamenteel taalreëls verbreek, naamlik **multitaligheid**, aan. Weer eens kan die verandertinge wat in Afrikaans plaasvind wanneer dit op ŉ mikrobloog gebruik word dus parallel met die ontwikkeling van ander tale beskou word.

7.11.1.6 Gebruik van tussenwerpsels

Verhoef (2001) en Jansen van Vuuren (2005) het bevind dat die gebruik van tussenwerpsels ŉ kenmerk van Internetafrikaans is. In my studie is daar in afdeling 7.7 ondersoek ingestel na die gebruik van tussenwerpsels in Afrikaans soos dit op Twitter gebruik word. Daar is bevind dat nege van die tien outeurs vanuit die gespesialiseerde korpus en twee vanuit die drie verdagte tekste se outeurs tussenwerpsels gebruik. Die gebruik van tussenwerpsels kan dus ŉ kenmerk van Afrikaans as ŉ gesproke taal beskou word en is ook kenmerkend van Afrikaans soos dit op Twitter gebruik word. Hierdie bevinding sluit aan by internasionale studies oor RBK wat stel dat taal op hierdie soort platforms ŉ hibriede kombinasie tussen gesproke en geskrewe taal sal wees.

7.11.1.7 Gebruik van kraswoorde

In afdeling 7.8 is daar ondersoek ingestel na die gebruik van kraswoorde. Wang *et al.* (2014:218) het bevind dat kraswoorde sowat 1,15% van die woorde wat in die korpora[364] ondersoek is, uitmaak − 7,73% (een uit dertien) van die korpora bevat vloekwoorde. In my studie is daar bevind dat 0,6% van die woorde in die gespesialiseerde korpus kraswoorde is. Daar is verder gevind dat elk van die tien outeurs in die gespesialiseerde korpus van kraswoorde gebruik maak. Daar kan vervolgens aangevoer word dat dit blyk asof Twitter-gebruikers wat in Afrikaans twiet meer geneig is om kraswoorde te gebruik alhoewel kraswoorde teen ’n laer frekwensie gebruik word as in Engels[365]. Die gebruik van kraswoorde kan vervolgens beskou word as kenmerkend van die Afrikaans soos dit op Twitter gebruik word.

7.11.1.8 Foute analise

In die foute-analise is daar ondersoek na spelfoute, die weglating van die tweede nie in die dubbelontkenning en letter-herhaling ondersoek ingestel. Soos in Tabel 7.27 voorgehou, is dit slegs O2, O4, O6 en O7 wat meer as een keer dieselfde woord verkeerd spel en wat dus nie as ’n tikfout beskou kan word nie. Indien die woorde wat in Tabel 7.27 gelys word in die verdagte tekste voorkom, kan die genoemde outeurs as moontlike outeurs van die tekste beskou word. ’n Verdere afleiding wat gemaak kan word, is dat die gebruik van kappies en deeltekens nie noodwendig plaasvind soos daargestel in die taalreëls nie. Ten opsigte van die dubbelontkenning wat in Afrikaans gebruik word, blyk dit dat die outeurs wat in Afrikaans twiet, geneig is om die tweede nie uit te laat. Aangesien hierdie kenmerk in al die outeurs wat deel van die gespesialiseerde korpus uitmaak, gevind word, kan dit as kenmerkend beskou word van die Afrikaans wat op Twitter gebruik word.

Die korrekte gebruik van die onbepaalde lidwoord (’n) blyk ook ’n aspek te wees wat talle van die outeurs foutief gebruik, maar aangesien dit slegs in die korpora van O3, O8, O9 en O10 telkens korrek gebruik word, kan die outeurs as moontlike outeurs van een van die verdagte tekste oorweeg word indien daar nie foute ten opsigte van die onbepaalde lidwoordgebruik voorkom nie. Beide V1 en V2 gebruik telkens die onbepaalde lidwoord korrek en in die geval van V3 is daar slegs ’n enkele onbepaalde lidwoord wat as *n* gebruik word.

Die gebruik van letter-herhaling ten einde klank en emosie aan lesers oor te dra, is ’n verdere aspek wat telkens op Twitter gevind word (sien voorbeeld 321 tot 324). Hierdie letterherhaling

[364] Wang *et al.* (2014) se korpus het uit 51 miljoen twiets bestaan.
[365] Sien afdeling 8.6 (Aanbevelings vir verdere navorsing).

dien om emosie, wat gewoonlik maklik in gesproke taal oorgedra word, in geskrewe tekste uit te beeld. Net soos gebruikers vanuit ander tale, maak Afrikaanse Twitter-gebruikers van letter-herhaling gebruik om ŉ informele gesprektaal te vestig en word klank en emosie aan leser hierdeur oorgedra.

7.11.1.9 N-gramanalise

Vanuit die ondersoek blyk dit datV1 telkens (bi-gram, tri-gram, vier-gram sowel as die vyf-gram) met O2 ooreenstem, en daarom kan O2 as die moontlike outeur van V1 beskou word na gelang van die n-gramanalise. V2 en O9 stem ooreen behalwe in die geval van die vyf-gram (O3 kan moontlik hier die outeur wees), maar aangesien O9 in die meeste gevalle (drie van die vier analises) as die outeur geïdentifiseer is wat die grootste ooreenkoms met V2 toon, kan O9 dus na gelang van die n-gramanalise as die moontlike outeur van V2 beskou word. V3 en O5 stem ooreen in terme van die bi-, tri- en vier-gram analises terwyl O2 en V3 ten opsigte van die vyf-gram analise ooreenstem. O5 en O2 (in die vyf-gram analise met V3) verskil slegs met 0,0054. O5 is dus – volgens die n-gramanalise – die outeur van V3.

Die n-gramanalise se resultate is relatief konsekwent (buiten in die geval van V2) om ŉ enkele outeur as die beste kandidaat vir outeurskap van ŉ bepaalde verdagte teks aan te dui en dus kan die gebruik van ŉ n-gramanalise as nog ŉ metode beskou word wat in kombinasie met verdere ondersoeke gebruik kan word ten einde outeurs van mikrobloginskrywings te identifiseer.

7.11.2 Stilistiese en teksanalise en outeurskapidentifikasie

Soos in afdeling 7.11 voorgehou, was die doel van die navorsing tweeledig. In afdeling 7.11.1 is die kenmerke van Afrikaans voorgehou soos dit op Twitter voorkom. In hierdie afdeling word daar verwys na die identifisering van mikroblogouteurs.

Die manier waarop elk van die aspekte wat in afdelings 7.2 tot 7.10 deur die onderskeie outeurs gebruik word, is met die voorkoms en gebruik soos in die verdagte tekste met mekaar vergelyk en ten slotte is daar bepaal of die ondersoek wel in die identifikasie van mikroblogouteurs gebruik kan word.

Die analises is uitgevoer deur interne vergelykings van die genormaliseerde waarde met mekaar te vergelyk om sodoende te kon bepaal of die onderskeie outeurs wel deurlopend met dieselfde frekwensie van die aspekte gebruik maak. Verder het die ondersoek na die frekwensie van die gebruik van die aspekte aangedui dat daar telkens onderskei kan word tussen die

verskillende outeurs op grond van die genormaliseerde waarde (in die afdeling is die waardes tot 1 000 woorde genormaliseer) van dié aspekte. Die genormaliseerde waarde van elk van die aspekte wat ondersoek is, is ook met die waardes van die verdagte tekste vergelyk en laastens, is die onderskeie stylkenmerke van die onderskeie outeurs in die studie met mekaar vergelyk ten einde te kon bepaal of daar enigsins kenmerke is wat ooreenstem. Die resultate vanuit die sewe verskillende kategorieë[366] ten opsigte van die identifisering van outeurs van mikrobloginskrywings in Afrikaans word vervolgens voorgehou.

7.11.2.1 Algemene kenmerke

Die ondersoek na die algemene kenmerke het ŉ analise van die teksformaat, leestekengebruik en die gebruik van hoofletters ingesluit. Die aspekte is ondersoek om te kon bepaal of die frekwensie van die gebruik sowel as die ooreenkomste in die onderskeie outeurs se individuele stylkenmerke as sodanig beskou kan word dat dit as ŉ manier beskou kan word waarop outeurs van mikrobloginskrywings geïdentifiseer kan word (sien Tabel 7.29):

Tabel 7.29 Twitter-spesifieke kenmerke: samevatting van die moontlike outeurs van die verdagte tekste

Kenmerk ondersoek		Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Twitter-spesifieke kenmerke	Hutswoorde	V1	-	-
		V2	O7	O1, O7 & O8
		V3	O1	-
	@-gebruikerskonvensie	V1	O5	O2, O3 & O8
		V2	O1	O1
		V3	O2	O1, O5 & O9
	Hiperskakels	V1	O1	O2 & O3
		V2	O9	O1
		V3	O5	O5
	Leestekengebruik aan die einde van sinne	V1	O1	O2
		V2	O2	-
		V3	O10	-
	Leestekens ontbreek aan die einde van sinne	V1	O1	O2
		V2	O2	-
		V3	O10	-
	Gebruik van leestekens	V1	O2	O2
		V2	O9	O10, O8, O6 & O2
		V3	O7	O2, O4, O5, O6, O8 & O10
	Hoofletters aanwesig aan die begin van sinne	V1	O3	O2
		V2	O1	O1
		V3	O10	-
	Hoofletters afwesig aan die begin van sinne	V1	O3	O2
		V2	O2	-
		V3	O10	-
	Hoofletters vir beklemtoning	V1	O3	-
		V2	O2 & O7	O1
		V3	O6	O5

[366] Sien afdeling 7.1 (Inleiding).

Tabel 7.29 toon die ooreenkomste soos geïdentifiseer is afdeling 7.2 aan. Vanuit die Tabel blyk dit dat ten spyte van die feit dat daar meer as een outeur vanuit die gespesialiseerde korpus geïdentifiseer kan word met ooreenkomste ten opsigte van die algemene kenmerke soos gemerk in die verdagte tekste, is daar vir elk van die verdagte tekste telkens 'n enkele outeur met betrekking tot die meerderheid aspekte as die outeur van een van die verdagte tekste beskou kan word.

Na gelang van die analise van die genormaliseerde (tot 1 000 woorde of sinne) waarde van die onderskeie aspekte van die algemene kenmerke en die ondersoek na die stylkenmerke kan daar aangevoer word dat O2 die moontlike outeur van V1 is en V2 waarskynlik deur O1 geskryf is. V3 is die verdagte teks wat die meeste ooreenkomste toon met die kenmerke van Afrikaans soos dit op Twitter (Twitter-spesifieke kenmerke) gebruik word. Tog kan O5 as die heel moontlikste outeur van V3 geïdentifiseer word. Die bevindinge vanuit die algemene kenmerke van Twitter, soos ondersoek in afdeling 7.2, kan nie as genoegsame bewys van outeurskap beskou word nie, maar hierdie analises kan sonder twyfel in 'n ondersoek na outeurskapidentifikasie gebruik word. Vervolgens word 'n opsomming van die sintaktiese ondersoek voorgehou.

7.11.2.2 Sintaktiese ondersoek

Die ondersoek na die sintaktiese kenmerke is in afdeling 7.3 uitgevoer. Daar is ondersoek ingestel na die sinstipes, die algemene woordsoortgebruik, sowel as die invloed van Engels in elk van die korpora. Die bevindinge met betrekking tot die verskillende sinstipes en woordsoortgebruik word in Tabel 7.30 voorgehou:

Tabel 7.30 Samevatting van die sintaktiese analise

Kenmerk ondersoek		Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Sintaktiese kenmerke	Stelsinne	V1	O8	-
		V2	O1	-
		V3	O4	-
	Vraagsinne	V1	O2	O2
		V2	O4	O1
		V3	O1	O5
	Uitroepsinne	V1	O6	O2
		V2	O3	O1
		V3	O5	O5
	Bevelsinne	V1	O1	-
		V2	-	-
		V3	O1	-
	Adjektiewe	V1	O3	-
		V2	O5	-
		V3	O7 & O1	-

Kenmerk ondersoek		Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
	Konjunkte	V1	O8 & O3	-
		V2	O8 & O9	-
		V3	O1 & O6	-
	Naamwoorde	V1	O2	-
		V2	O1, O5, O6 & O9	-
		V3	O1, O4, O6 & O8	-
	Voornaamwoorde	V1	O3	-
		V2	O1	-
		V3	O5	-
	Voorsetsels	V1	O7	-
		V2	O5	-
		V3	O5	-
	Werkwoorde	V1	O2	-
		V2	O2	-
		V3	O5	-
	Lidwoorde	V1	O2	-
		V2	O1 & O2	-
		V3	O5 & O8	-
	Die invloed van Engels	V1	O2	-
		V2	O9 & O10	-
		V3	O9 & O10	-

Vanuit Tabel 7.30 is dit duidelik watter korpora met elk van die verdagte tekste ooreenstem na gelang van die sintaktiese analise. Vanuit die ondersoek na die frekwensie van die voorkoms van die verskillende sinstipes blyk dit dat V1 deur O3, O4 of O8 geskryf kan wees. O1 is die moontlike outeur van V2 terwyl V3 moontlik deur O1, O4 of O5 geskryf is. Die stylkenmerke toon aan dat O2 die outeur van V1 is, O1 van V2 en O5 die outeur van V3. Dus kan die gevolgtrekking gemaak word dat V1 deur O2, V2 deur O1 en V3 deur O5 geskryf is.

Die oorsigtelike ondersoek na die gebruik van woordsoorte dui aan dat O2 of O3 die outeurs van V1 kan wees, O1 en O5 en V2 stem ooreen terwyl O1 of O5 die outeurs van V3 is. Die bevindinge vanuit die algemene kenmerke van Twitter (ondersoek in afdeling 7.3) kan nie as genoegsame bewys van outeurskap beskou word nie, maar hierdie analise kan sonder twyfel in die ondersoek gebruik word. Daar kan aangevoer word dat V1 en O2, V2 en O1, V3 en O5 genoegsaam ooreenstem dat hulle as die outeurs van die verdagte tekste beskou kan word. Vervolgens word ŉ opsomming van die morfologiese ondersoek voorgehou.

Met betrekking tot die invloed van Engels blyk dit dat V1 en O2 se stylkenmerke ooreenstemmend is, terwyl die invloed van Engels in beide V2 en V3 ppreenstem met dit wat in O9 en O10 waargeneem word.

7.11.2.3 Morfologiese ondersoek

In afdeling 7.4 is daar ondersoek na die morfologiese kenmerke ingestel. Hierdie analise het die gebruik van afkortings, inkortings en weglatings ingesluit. Tabel 7.31 bied 'n opsomming van die bevindinge:

Tabel 7.31 Samevatting van die morfologiese analise

Kenmerk ondersoek		Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Morfologiese kenmerke	Foutlose afkorting	V1	O2	O2
		V2	O5	-
		V3	O1	O9
	Idiolektiese afkortings	V1	O3	O2
		V2	O5	-
		V3	O5	-
	Akronieme	V1	O2	O2
		V2	O9	-
		V3	-	-
	SMS-afkortings	V1	-	-
		V2	O1	-
		V3	-	-
	Inkorting	V1	O2	-
		V2	O3	-
		V3	O6	O1
	Weglatings	V1	O2	O2
		V2	O1	-
		V3	O1	-

Die morfologiese ondersoek se bevindinge (soos voorgehou in Tabel 7.31) toon aan dat die kenmerke wat in hierdie ondersoek geanaliseer is nie noodwendig in die verdagte tekste voorkom nie. Die afwesigheid van sommige morfologiese kenmerke kan toegeskryf word aan die korpus grootte, maar ten spyte daarvan is daar steeds kenmerke wat ooreenstem na gelang van die frekwensie van sommige kenmerke aanwesig en die stylkenmerke geïdentifiseer.

In die ondersoek na die gebruik van afkortings is daar bevind dat V1 en O2, V2 en O5, V3 en O1 of O5 ooreenstem na gelang van die frekwensie van die voorkoms van die onderskeie aspekte ten opsigte van die gebruik van afkortings. Die ondersoek na ooreenkomste in die stylkenmerke toon aan dat V1 en O2 sonder twyfel deur dieselfde outeur geskryf is terwyl V2 en O9 oor ooreenstemmende stylkenmerke beskik.

Vanuit die ondersoek na die voorkoms van inkorting kan daar aangevoer word dat V1 en O2 (na gelang van die frekwensie van die voorkoms van inkorting in die onderskeie korpora en ooreenstemmende stylkenmerke) asook V2 en V3 deur O1 (in die twee verdagte tekste stem geen van die stylkenmerke ooreen nie) geskryf is. Die laaste morfologiese aspek wat

geanaliseer is, is die voorkoms van weglating in die onderskeie korpora. Soortgelyk aan die bevindinge oor die gebruik van inkorting kan O2 as die outeur van V1 beskou word terwyl beide V2 en V3 moontlik deur O1 geskryf is.

Na gelang van die ondersoek na morfologiese kenmerke kan O2 as die outeur van V1 beskou word terwyl die bevindinge oor die outeurskapidentifikasie vir V2 en V3 bewys dat O1 en O5 as moontlike outeurs van hierdie tekste beskou kan word. Alhoewel die morfologiese kenmerke wat ondersoek is nie in alle verdagte tekste genoegsame bewyse vir die identifisering van outeurs gelewer het nie, moet hierde aspekte steeds ondersoek word.

7.11.2.4 Semantiese kenmerke

In afdeling 7.5 is daar ondersoek ingestel na die semantiese kenmerke van Afrikaans soos dit op Twitter gebruik word. In hierdie ondersoek is daar spesifiek na die gebruik van emotikons ondersoek – ŉ aspek wat as sinoniem met sosialemediatekste beskou kan word - ingestel. Nie net het hierdie ondersoek kenmerke gelewer oor die gebruik van emotikons in Afrikaans nie, maar dit kan ook as ŉ waardevolle aspek in die identifisering van mikroblogouteurs beskou word (sien Tabel 7.32):

Tabel 7.32 Samevatting van die gebruik van emotikons

Kenmerk ondersoek	Emotikons	Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Emotikons	Totaal	V1	O2	O2
		V2	O1	O1
		V3	O8	O5
	Diere	V1	O2, O3, O5, O6, O9 & O10	-
		V2	O2, O3, O5, O6, O9 & O10	-
		V3	O2, O3, O5, O6, O9 & O10	-
	Gebare	V1	O2, O4 & O7	-
		V2	O1	-
		V3	O5	-
	Gesigte	V1	O2	-
		V2	O1	-
		V3	O1	-
	Kos	V1	O1, O2, O3, O4, O5, O6 & O9	-
		V2	O1, O2, O3, O4, O5, O6 & O9	-
		V3	O7 & O1	-
	Mense	V1	O2, O3, O4, O5, O6, O7, O8 & O9	-
		V2	O2, O3, O4, O5, O6, O7, O8 & O9	-
		V3	O10	-
	Piktogramme	V1	O3, O7 & O9	-
		V2	O3, O7 & O9	-
		V3	O3, O7 & O9	-
	Natuur	V1	O2, O3, O4, O7 & O9	-
		V2	O2, O3, O4, O7 & O9	-
		V3	O2, O3, O4, O7 & O9	-
		V1	O2 & O3	-

Kenmerk ondersoek	Emotikons	Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
	Lukraak	V2	O2 & O3	-
		V3	O5	-
	Sport	V1	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-
		V2	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-
		V3	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-
	Tegnologie	V1	O1, O2, O3, O4, O5, O6, O7, O8 & O10	-
		V2	O1, O2, O3, O4, O5, O6, O7, O8 & O10	-
		V3	O1, O2, O3, O4, O5, O6, O7, O8 & O10	-
	Vervoer	V1	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-
		V2	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-
		V3	O1, O2, O3, O4, O5, O6, O7, O8, O9 & O10	-

In afdeling 7.5 is die frekwensie van die voorkoms van die verskillende kategorieë waarin die emotikons verdeel is, geanaliseer. Die bevindinge word in Tabel 7.32 voorgehou. Dit blyk dus dat alhoewel al die kategorieë nie deur die onderskeie outeurs gebruik word nie (diere, natuur, sport, tegnologie en vervoer), kan die gebrek aan emotikons ook as ŉ aanduiding van outeurskap dien. Só kan daar aangevoer word dat O1 nie as die outeur van V1 beskou kan word na gelang van die gebruik van emotikons nie. O2 is die mees waarskynlike outeur van V1, O1 van V2 en O5 van V3 na gelang van die ondersoek na die gebruik van emotikons.

7.11.2.5 Kodewisseling

In afdeling 7.6 is daar ondersoek na kodewisseling in Afrikaanse twiets ingestel. Kenmerke wat in hierdie analise ondersoek is, het kodewisseling (IT en MT) sowel as dubbelmorfologie ingesluit (sien Tabel 7.33):

Tabel 7.33 Samevatting van die semantiese analise

Kenmerk ondersoek		Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Semantiese kenmerke	IT en MT	V1	-	O2
		V2	O1	O1
		V3	O7	O5
	Dubbelmorfologie	V1	-	O3, O5, O7, O8 & O10
		V2	-	O3, O5, O7, O8 & O10
		V3	-	O3, O5, O7, O8 & O10

Die semantiese kenmerke wat ondersoek is (aangedui in Tabel 7.33) dui aan dat daar in sommige van die aspekte wat ondersoek is (IT en MT) wel ooreenkomste waargeneem is met betrekking tot die frekwensie van die voorkoms van kenmerke tussen die verdagte tekste en die gespesialiseerde korpus terwyl daar in die oorblywende aspekte (dubbelmorfologie) stylkenmerke waargeneem is wat só ooreenstem dat sekere outeurs as moontlike outeurs van die verdagte tekste geëlimineer kon word.

Na gelang van die semantiese analise kan daar aangevoer word dat O2 waarskynlik die outeur van V1 is terwyl O1, O3, O5, O7, O8 en O10 as moontlike outeurs van V1 oorweeg kan word. In die geval van V3 kan O3, O5, O7, O8 en O10 nie as moontlike outeurs van die teks uitgeskakel word nie. Vanuit die ondersoek na die semantiese kenmerke blyk dit dus dat alhoewel daar moontlike outeurs geïdentifiseer is, die semantiese kenmerke veel eerder kenmerkend tot Afrikaans is soos dit op Twitter gebruik word as dat ŉ enkele outeur as moontlike outeur van ŉ verdagte teks aangetoon kon word. Aangesien daar wel in die geval van V1 ŉ enkele moontlike outeur (O2) geïdentifiseer is, kan die analise nie uitgelaat word wanneer outeurskapidentifikasie op mikroblogs uitgevoer word nie.

7.11.2.6 Tussenwerpsels

Die ondersoek in afdeling 7.7 het die gebruik van tussenwerpsels in die onderskeie korpora geanaliseer. Vanuit hierdie analise is daar bevind dat tussenwerpsels deur die meerderheid van die outeurs gebruik word en dat die gebruik van tussenwerpsels in Afrikaans soos dit op Twitter gebruik word, aansluit by Lockyer (2014:69) en Biber *et al.* (2003:450) se definisie dat woorde (wat hoofsaaklik in ŉ informele diskoers gebruik word) emosie aan die gehoor (in dié geval lesers van twiets) oordra. Tabel 7.34 som die bevindinge van die analise op:

Tabel 7.34 Samevatting van die ondersoek na die gebruik van tussenwerpsels

Kenmerk ondersoek	Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Tussenwerpsels	V1	O2	-
	V2	O1	-
	V3	O3	-

Die ondersoek na die gebruik van tussenwerpsels in Afrikaans soos dit op Twitter gebruik word, het aangetoon dat alhoewel die gebruik van tussenwerpsels wel as ŉ kenmerk van Afrikaans beskou kan word, is daar nie stylkenmerke vanuit die analise sigbaar wat aanduiding daartoe gee vir die identifisering van moontlike outeurs van die verdagte tekste nie. Die frekwensie waarby elk van die outeurs in hierdie studie van tussenwerpsels gebruik maak, toon aan dat V1 deur O2, V2 deur O1 en V3 deur O3 moontlik geskryf is.

7.11.2.7 Kraswoorde

In afdeling 7.8 is daar ondersoek ingestel na die gebruik van kraswoorde in die onderskeie korpora (sien Tabel 7.35):

Tabel 7.35 Samevatting van die ondersoek na die gebruik van kraswoorde

Kenmerk ondersoek	Verdagte teks	Ooreenstemmende korpus na gelang van genormaliseerde waarde	Ooreenstemmende korpus na gelang van stylkenmerke
Kraswoorde	V1	O6	-
	V2	O1	O1
	V3	O1	O5

Soos in Wang *et al.* (2014) se studie, is die gebruik van kraswoorde ook in Afrikaans ŉ kenmerkende aspek van Twitter-taalgebruik. Verder kon daar nie net na gelang van die frekwensie waarteen kraswoorde deur die onderskeie outeurs gebruik word, ooreenkomste tussen moontlike outeurs van die verdagte tekste waargeneem word nie, maar daar is stylkenmerke wat só ooreenstem dat die outeur van die teks geïdentifiseer kon word. Die bevindinge (voorgehou in Tabel 7.33) toon gevolglik aan dat O6 as die outeur van V1 beskou kan word. V2 is na gelang van die ondersoek sonder redelike twyfel deur O1 geskryf terwyl beide O1 en O5 as outeurs van V3 oorweeg kan word.

7.11.2.8 Foute-analise

Die laaste analise wat in die stilistiese en teksanalise uitgevoer is, is die foute-analise. Vanuit hierdie ondersoek blyk dit dat alhoewel daar wel spelfoute in die onderskeie korpora voorkom, die foute nie genoegsaam herhaal dat dit kenmerkend van die taalgebruik van ŉ bepaalde outeur beskou kan word nie. Die korrekte gebruik van lidwoorde (*ŉ* en nie *n* nie) word in O3, O8, O9 en O10 aangetref en dié outeurs moet ook as moontlike outeurs vir V1 en V2 beskou word terwyl O1, O2, O4, O5, O6 en O7 dan ook as moontlike outeurs van V3 oorweeg kan word (die laasgenoemde outeurs gebruik beide *ŉ* en *n*).

7.11.2.9 N-gramanalise

Die laaste stilistiese analise wat uitgevoer is, was die n-gramanalise (afdeling 7.10). In hierdie analise is daar bevind dat V2 nie soveel ooreenkomste met ŉ enkele outeur vanuit die gespesialiseerde korpus toon wanneer daar ondersoek na die bi-, tri-, vier- en vyf-gramme ingestel is nie. Daar kan nie sonder redelike twyfel gestel word dat daar ŉ spesifieke moontlike outeur geïdentifiseer kan word vir die teks na gelang van die analise nie.

V1 stem telkens (bi-gram, tri-gram, vier-gram sowel as die vyf-gram) met O2 ooreen en daarom kan O2 as die moontlike outeur van V1 beskou word. V3 en O5 stem ooreen in terme van die bi-, tri- en vier-gram analises terwyl O2 en V3 ten opsigte van die vyf-gram analise ooreenstem. Tog blyk dit dat O5 en O2 (in die vyf-gram analise met V3) slegs met 0,0054 verskil. O5 kan dus volgens die n-gramanalise geïdentifiseer word as die outeur van V3.

Net soos in die ondersoek na die taalverhouding in die twiets kan daar dus na gelang van die n-gramanalise in die geval van twee (V2 en V3) van die verdagte outeurs moontlike outeurs (O1 en O5) vanuit die gespesialiseerde korpus geïdentifiseer word. Alhoewel daar nie in die geval van V1 genoegsame ooreenstemmende resultate is om ’n outeur te kan identifiseer nie, is die analise van die n-gramme steeds as suksesvol in die identifiseringsproses van outeurs van mikroblogouteurs beskou.

In hierdie afdeling is die bevindinge oor die stilistiese en teksanalise voorgehou. Vanuit elk van die kategorieë wat ondersoek is, is daar moontlike outeurs vir die drie verdagte tekste geïdentifiseer. Die hieropvolgende afdeling bied ŉ samevatting van hoofstuk 7.

7.12 Samevatting

Die doel van hierdie hoofstuk is om die stilistiese en teksanalise voor te hou om die navorsingsvrae te kon beantwoord soos daargestel in afdeling 1.3.

In afdeling 7.11.1 is daar etlike kenmerke van Afrikaans geïdentifiseer soos dit op Twitter gebruik word. Ten einde die ondersoek na outeurskapidentifikasie te kon voltooi, is die stilistiese en teksanalise uitgevoer. Elk van die datastelle is ten opsigte van agt verskillende kategorieë geanaliseer. Die kategorieë is gekies na gelang van navorsing wat reeds oor outeurskapidentifikasie gedoen is (hierdie kategorieë word in hoofstuk 5[367] bespreek). In dié bespreking is daar na McMenamin (2002), Coulthard en Johnson (2007), Michell (2013), Thiart (2014) en Kotzé (2015) se metodes van outeurskapidentifikasie verwys. Vanuit die bespreking is die agt kategorieë saamgestel. Hierdie kategorieë het die analise van algemene, sintaktiese, morfologiese en semantiese kenmerke, die gebruik van tussenwerpsels, kraswoorde en emotikons en laastens, ŉ foute-analise ingesluit.

In afdeling 7.11.1 en 7.11.2 is die bevindinge oor die identifisering van outeurs van mikrobloginskrywings asook die kenmerke van Afrikaans soos wat dit op Twitter gebruik word, voorgehou. Vanuit hierdie afdeling (sien ook Tabel 7.29 tot 7.35) is dit duidelik dat daar telkens ŉ moontlike outeur geïdentifiseer na gelang van die kenmerke wat ondersoek is. Die stilometriese analise[368] se sintese (afdeling 6.8) toon die outeurs aan wat vanuit die gespesialiseerde korpus as die moontlike outeurs van die drie verdagte tekste beskou kan word.

[367] Sien afdeling 5.6.1 (Stilistiese analise).
[368] Sien afdelings 6.2 tot 6.7.

In afdeling 5.12.3.3 word daar na die datastelle en die rede vir die samestelling van “verdagte tekste” verwys. In hierdie afdeling (5.12.3.3) word daar aangevoer dat 'n scenario geskep moes word waarin outeurs van mikrobloginskrywings geïdentifiseer moet word en daarom is daar vir drie van die deelnemers vanuit die gespesialiseerde korpus elk 'n bykomende datastel geskep. Hierdie verdagte tekste is baie kleiner as die tekste waaruit die gespesialiseerde korpus bestaan. Die grootste “verdagte teks” het uit 91 twiets of 1 409 woorde bestaan terwyl die kleinste “verdagte teks” uit 32 twiets of 412 woorde bestaan het. Die grootte van hierdie korpora is dus aansienlik kleiner as die korpora wat tot dusver in outeurskapidentifikasie in Afrikaans gebruik is. Die bevindinge vanuit hoofstuk 6 en 7 ten opsigte van die identifisering van outeurs van mikrobloginskrywings word vervolgens voorgehou:

Tabel 7.36 Moontlike outeurs van V1

Moontlike outeurs van V1	
Sleutelwaarde-analise	O2
Ratio-analise	O2 & O1
Sinne, woorde en karakters per twiet	O2
Taalverhouding	O1, O4, O9 & O10
Leesbaarheidstoetse	O6
Algemene kenmerke	O2
Sintaktiese kenmerke	O2
Morfologiese kenmerke	O2
Semantiese kenmerke: Emotikons	O2
Kodewisseling	O2, O3, O5, O7, O8 & O10
Tussenwerpsels	O2
Kraswoorde	O6
Foute-analise	O3, O8, O9 & O10
N-gramanalise	O2

Tabel 7.36 bied 'n opsomming van die bevindinge met betrekking tot die outeurskapidentifikasie van V1 (in hoofstuk 6 en 7 uitgevoer). 'n Totaal van 14 verskillende is uitgevoer. Vanuit hierdie Tabel is dit duidelik dat O2 in 10 van die 14 analises as die mees waarskynlike outeur van V1 geïdentifiseer is. Die afleiding kan daarom gemaak word dat na gelang van die stilometriese analise in kombinasie met die stilistiese en teksanalise, O2 (met 71,4% sekerheid) as die outeur van V1 geïdentifiseer kan word.

Tabel 7.37 Moontlike outeurs van V2

Moontlike outeurs van V2	
Sleutelwaarde-analise	O1 & O5
Ratio-analise	O1

Moontlike outeurs van V2	
Sinne, woorde en karakters per twiet	O4
Taalverhouding	O2
Leesbaarheidstoetse	O8
Algemene kenmerke	O1
Sintaktiese kenmerke	O1
Morfologiese kenmerke	O1 & O5
Semantiese kenmerke: Emotikons	O1
Kodewisseling	O1, O3, O5, O7 & O8
Tussenwerpsels	O1
Kraswoorde	O1
Foute-analise	O3, O8, O9 & O10
N-gramanalise	O1 & O5

Tabel 7.37 verskaf ’n opsomming van die bevindinge vanuit hoofstuk 6 en 7 oor die identifisering van die outeur van V2. Vanuit die tabel is dit duidelik dat O1 in 10 van die 14 analises wat in hierdie studie uitgevoer is, as die moontlike outeur van V2 geïdentifiseer is. Soortgelyk aan die bevindinge ten opsigte van die outeur van V1, kan daar in die geval van die outeurskapidentifikasie vir V2 aangevoer word dat die gekombineerde stilometriese, stilistiese en teksanalise in 71,4% van die analises aantoon dat O1 die mees waarskynlike outeur van V2 is.

Tabel 7.38 Moontlike outeurs van V3

Moontlike outeurs van V3	
Sleutelwaarde-analise	O1 & O5
Ratio-analise	O4, O7 & O8
Sinne, woorde en karakters per twiet	O7
Taalverhouding	O5
Leesbaarheidstoetse	O4
Algemene kenmerke	O5
Sintaktiese kenmerke	O5
Morfologiese kenmerke	O1 & O5
Semantiese kenmerke: Emotikons	O5
Kodewisseling	O3, O5, O7, O8 & O10
Tussenwerpsels	O3
Kraswoorde	O1 & O5
Foute-analise	O3, O8, O9 & O10
N-gramanalise	O5

Die outeurskapidentifikasie van V3 − soos voorgehou in Tabel 7.38 − toon dat O5 in 64,3% van die analises as die outeur van V3 geïdentifiseer is. Dit wil sê dat O5 in 9 van die 14 analises as die oorwegende outeur van V3 geïdentifiseer is.

O1, O2 en O5 is wel onderskeidelik die drie outeurs vir wie daar elk ŉ bykomende datastel geskep is, en dus is die outeurs van hierdie drie verdagte tekste wel suksesvol geanaliseer. Daar kan ten slotte gestel word dat daar dus nie net kenmerke van Afrikaans soos dit op Twitter gebruik word, geïdentifiseer is nie, maar dat dit wel moontlik is om outeurs van mikrobloginskrywings te identifiseer wannneer ŉ korpus vanuit die twiets saamgestel word wat oor ŉ maksimum van 1 409 of ŉ minimum van 412 woorde beskik.

Hoofstuk 8

SAMEVATTING

You sort of start thinking anything's possible if you've got enough nerve.
J.K. Rowling *Harry Potter and the Half-Blood Prince* (2005:655)

8.1 Inleiding

Hierdie studie, *Outeurskapidentifikasie en mikroblogs: ŉ eksploratiewe forensiese analise in ŉ digitale era*, het ten doel gehad om die eienskappe van Afrikaans soos dit op 'n mikroblog (Twitter) gebruik word te bepaal om eksploratief ondersoek te kon instel na die suksesvolle identifisering van mikroblogouteurs. Tot op hede is die eienskappe van Afrikaans op 'n mikroblog, soos Twitter, nog nie beskryf nie en is dit 'n belangrike bydrae wat hierdie studie maak tot die dokumentering van Afrikaans op hierdie betrokke aanlyn omgewing. Verder is 'n raamwerk vir die identifisering van mikroblogouteurs saamgestel wat ook 'n bydrae lewer tot die forensiese linguistiek binne die internetomgewing. In hierdie hoofstuk word ŉ samevatting van die studie voorgehou deur te fokus op die navorsingsvrae, die literatuurstudie en die empiriese studie. 'n Sintese van die antwoorde op die navorsingsvrae word aangebied en aanbevelings vir verdere navorsing word verskaf.

8.2 Oorsig van die studie

Gedurende die afgelope paar jaar is daar talle nuwe tegnologieë ontwikkel wat die manier waarop kommunikasie plaasvind, beïnvloed en wesenlik verander. In afdeling 2.5 (Kommunikasie en die internet) is daar verwys na RBK en na gelang van die definisie van hierdie term kan daar aangevoer word dat hedendaagse kommunikasie nie sonder digitale tegnologie kan bestaan nie. Tog is dit onafwendbaar dat die medium waardeur kommunikasie plaasvind die manier waarop Afrikaans gebruik word, sal beïnvloed.

Vandag bevind ons ons in die tydperk van Web 3.0 − die semantiese web – en Web 4.0 (simbiotiese web). Individue is nou in staat om self inhoud te skep en te deel (sien Tabel 2.4 vir die verskille tussen Web 1.0, Web 2.0 en Web 3.0). Die mobiele aard van Web 3.0 en Web 4.0 – die manier waarop gebruikers deur middel van verskeie toestelle internettoegang het (gebruikers is nie meer afhanklik van slegs ŉ rekenaar om die internet te kan gebruik nie) – gee aanleiding daartoe dat sosiale interaksie en kommunikasie verander.

In afdeling 2.6. is daar na sosialemediaplatforms (mikroblogs en Twitter) verwys, en dit is duidelik dat mikroblogs as 'n platform beskou kan word waar nuwe inligting kortliks weergegee word. Twitter[369] word as só 'n mikroblogplatform beskou en beskik oor 'n 140 karakterbeperking – 'n boodskap kan nie langer as 140 karakters (spasies ingesluit) wees nie.

Soos genoem, was die doel van hierdie studie om te bepaal of outeurs van mikrobloginskrywings suksesvol geïdentifiseer kan word. Alhoewel daar reeds navorsing (in Afrikaans) uitgevoer is oor die identifisering van outeurs, is daar (volgens die navorser se kennis) nie navorsing gedoen oor die identifisering van outeurs van Afrikaanse tekste wat op 'n sosialemediaplatform geplaas word nie. Daar is aangevoer dat die karakterbeperking, die interaktiewe aard van sosiale netwerke en die kommunikasiedoel moontlik aanleiding kan gee tot 'n "nuwe" vorm van Afrikaans om in die taalbehoeftes van gebruikers te voldoen.

Verder gee die toename in die gebruik van sosiale media aanleiding daartoe dat dit makliker is om iemand na te boots, om anoniem te bly en/of ongewensde boodskappe te stuur en aangesien tekste wat op sosiale media voorkom oor die algemeen uit korter tekste bestaan (in sommige gevalle – soos Twitter – is karakterbeperkings aanwesig), is dit belangrik om te bepaal of die gewone maniere van outeurskapidentifikasie (sien afdeling 3.4) wel suksesvol op hierdie soort tekste toegepas kan word.

Daar is drie navorsingsvrae geïdentifiseer om die bogenoemde te kon ondersoek, naamlik:

- Wat is die taalstilistiese en stylkenmerke van mikroblogs in Afrikaans?
- Kan die outeur van 'n mikrobloginskrywing suksesvol geïdentifiseer word?
- Wat is die forensiese beginsels wat gebruik kan word om outeurskapidentifikasie op mikroblogs toe te pas?

Ten einde te bepaal of die navorsingsvrae wel deur die navorsing beantwoord is, verwys ek vervolgens na die doelwitte van die studie, die literatuurstudie en die empiriese ondersoek.

8.2.1 Doelwitte van die studie

Die doelwitte van hierdie studie is in hoofstuk 1 (afdeling 1.5) uiteengesit en die studie het dus ten doel gehad om eerstens te bepaal of die Afrikaans wat op Twitter gebruik word as "eiesoortig" beskou kan word alvorens outeurskapidentifikasie op tekste vanaf die mikroblogplatform uitgevoer is. Die doelwitte (soos voorgehou in afdeling 1.5) kan soos volg opgesom word:

[369] Daar is verskeie ander mikroblogplatforms, sien afdeling 2.6 – Mikroblogs.

- Om te bepaal wat die taalstilistiese en stylkenmerke van Afrikaans is soos dit op mikroblogs voorkom.
- Om te bepaal wat die forensiese beginsels is wat gebruik kan word om outeurs van mikrobloginskrywings te identifiseer.
- Om te bepaal of outeurs van mikrobloginskrywings suksesvol geïdentifiseer kan word.

Alhoewel die drie doelwitte van hierdie studie uiteenlopend kan voorkom, moet die feit dat daar tot op hede nie ŉ korpus vir Afrikaans soos dit op sosiale media gebruik word, bestaan[370] nie, in gedagte gehou word. Die gebrek aan ŉ verwysingskorpus waaruit kenmerke geïdentifiseer kan word en as ŉ afwyking van die norm beskou kan word, het my genoop om ook ondersoek na die algemene kenmerke van Afrikaans soos dit op Twitter gebruik word, ondersoek in te stel alvorens die outeurskapidentifikasie kon plaasvind.

8.2.2 Literatuurstudie

Ten einde die navorsingsvrae te kon beantwoord, is daar eerstens ŉ literatuurstudie onderneem. Die doel van die literatuurstudie was om ŉ agtergrond oor kernbegrippe in te win waarna die empiriese analise uitgevoer is.

Ten opsigte van die ontwikkeling van Afrikaans is dit duidelik (sien afdeling 2.3) dat Afrikaans ŉ taal is waarin die standaardisering (as taalverandering) beskou kan word as die produk van opkomende linguistiese gewoontes en normatiewe aktiwiteite van 'n kollektiewe spraakgemeenskap. Afrikaans is die kreatiewe konstruk van doelbewuste aktiwiteite van ŉ spesifieke groep individue en verder is Afrikaans ŉ taal wat aanpas by die eise wat nuwe tegnologie aan die taal stel. Crystal (2004:4) stel dit duidelik dat individue deur middel van hierdie elektroniese medium kommunikeer op ŉ manier wat fundamenteel verskil van ander semiotiese situasies.

Verhoef (2001) het 15 opvallende taalveranderinge in Afrikaans geïdentifiseer soos dit in Internetafrikaans gebruik word (sien Tabel 2.1) en argumenteer dat Internetafrikaans al hoe meer tekens van gesproke Afrikaans toon. Janse van Vuuren (2007) het bevind dat die Afrikaans wat in e-posse, kletskamers en e-tydskrifte gebruik word nie voldoen aan die konvensies van standaardgeskrewe Afrikaans nie (sien Tabel 2.2). Hierdie bevindinge van Janse van Vuuren stem ooreen met bevindinge van internasionale navorsers soos Baron (2001), Crystal (2004), Danet *(*1997), Herring (2002) en Werry (1996) en sodoende word die universele karakter van Afrikaans bevestig wanneer die informele digitale register as die norm toegepas word.

---

[370] Sien ook beperkings van die studie (afdeling 8.4).

Die ontwikkeling van tegnologie en die manier waarop RBK verander is in sosiale media (afdeling 2.6) is bespreek. Navorsing deur Bubaš (2002) toon aan dat RBK nie net ŉ positiewe ontwikkeling is nie, maar dat daar veral vier uitdagings van belang is:

- ŉ Beperkte sosiale aanwesigheid.
- Anonimiteit − deelnemers se identiteit is nie duidelik sigbaar nie.
- Vertraagde terugvoer.
- Onpersoonlike kommunikasie.

Hierdie uitdagings gee aanleiding tot die moontlikheid van sosialemedianetwerk misdrywe wanneer RBK gebruik word omdat anonimiteit dit vergemaklik. Dit word dus al hoe belangriker om outeurs van kort tekste te kan identifiseer aangesien misdrywe wel op dié platforms plaasvind. Ten einde ondersoek na die identifisering van outeurs van Afrikaanse mikrobloginskrywings te kon instel, is daar op ŉ omvattende manier na die volgende kernbegrippe ondersoek ingestel: forensiese linguistiek (afdelings 3.2 en 3.3), outeurskapidentifikasie (afdeling 3.4), outeurskapidentifikasie in ŉ digitale era (afdeling 3.5), korpuslinguistiek (afdeling 4.1 tot 4.6) en ŉ forensiese en korpuslinguistiese benadering (afdeling 4.7).

Vanuit hierdie afdelings het dit duidelik geblyk dat nie net outeurskapidentifikasie nie, maar ook die korpora wat gebruik word geweldig verander het in die afgelope dekade. Bhargava *et al. (*2013:37) voer in dié verband aan dat outeurskapidentifikasie nie op tekste korter as 1 000 woorde in die vroeë 19de eeu uitgevoer kon word nie. Hierdie minimum teksdrempel het oor die jare afgeneem, en in die vroeë 21ste eeu is 250 woorde as die minimum drempel beskou. Tog moet hierdie minimum drempel deurlopend verlaag word en so is daar verskeie studies oor outeurskapidentifikasie met betrekking tot klein korpora onderneem (sien afdeling 4.7 – ŉ Forensiese en korpuslinguistiese benadering).

Daar is verder ook verwys na Green en Sheppard (2013:64) wat stel dat daar ŉ minimum tekslengte bestaan wanneer identifiserende karaktereienskappe nie meer sigbaar is nie. Ledger en Merriam (1994:240) voer aan dat die minimum tekslengte 500 woorde is terwyl Forsyth en Holmes (1996) die getal na 250 woorde verminder het. Green en Sheppard (2013:64) stel dat hierdie tekslengte steeds aansienlik hoër is as die gemiddelde lengte van ŉ twiet − die gemiddelde lengte van ŉ twiet is minder as 25 woorde (gewoonlik selfs minder as 10 woorde per twiet).

Ten opsigte van navorsing in Afrikaans oor outeurskapidentifikasie en forensiese linguistiek (afdeling 3.3.2) en ook bestaande korpora in Afrikaans (afdeling 4.2.4), is dit duidelik dat daar ŉ leemte is veral wanneer die identifisering van outeurs van korter sosialemediatekste en die beskikbaarheid van ŉ omvattende korpus wat Afrikaans vanaf sosialemediaplatforms insluit in ag

geneem word. Die studie het dus na die identifisering van outeurs van Afrikaanse tekste vanaf ŉ sosialemediaplatform ondersoek ingestel.

Aangesien hierdie studie as *ŉ eksploratiewe forensiese analise in ŉ digitale era* beskou word, was dit veral in afdeling 3.5 (Outeurskapidentifikasie in ŉ digitale era) van die uiterse belang om die nodige beginsels voor te hou wat reeds vir die identifisering van mikroblogouteurs bepaal is. Alhoewel daar al taamlik navorsing oor hierdie aspek in ander tale (veral Engels) uitgevoer is, is daar nie ŉ bestaande metode ten einde die outeurs van Afrikaanse mikrobloginskrywings te identifiseer nie. Navorsing in Suid-Afrika met betrekking tot forensiese linguistiek is nog redelik beperk en (soos aangedui in afdeling 3.3) word die meerderheid van forensieslinguistiese analises op Engelse tekste uitgevoer. Tog is die studies wat reeds in ander tale uitgevoer is van belang omdat die aspekte wat die onderskeie navorsers geïdentifiseer en gebruik het in die identifisering van outeurs (sien afdeling 3.4, 3.5 en 4.7) as grondslag vir die outeurskapidentifikasie in hierdie studie gebruik (sien afdeling 5.6, 5.7, 6.8 en 7.10).

In hoofstuk 5 is daar ook na die verskillende metodologieë verwys wat toegepas kan word ten einde die studie te kon uitvoer. Ek het op ŉ pragmatiese benadering besluit aangesien dié benadering fokus op die identifisering van die beste en gepaste metodes en teorieë om ŉ probleem op te los. Ek het verder van ŉ geïntegreerde gemengde metode gebruik gemaak waar kwantitatiewe of kwalitatiewe data vir 'n groter ontwerp gebruik word (eksperimentele ondersoek) en die data het 'n ondersteunende rol in die ontwerp van die navorsing gespeel. Die geïntegreerde, gemengde navorsingsontwerp van my studie kan soos volg uiteengesit word:

- Metodologie: gemeng (beide kwalitatief en kwantitatief).
- Ontwerp: eksperimenteel (ŉ geïntegreerde, gemengde navorsingsontwerp).
- Instrumente: tekste en korpora.

Die literatuurhoofstukke (hoofstuk 2 tot 4) het ten doel gehad om die agtergrond te skep waarteen die studie begrond is. Die verskillende metodes van outeurskapidentifikasie is bespreek asook die korpusbenadering tot forensiese linguistiek. Die metode van my studie − die manier waarop deelnemers vir die navorsing gekies is, die manier waarop die data verwerk is, die navorsingsmetodes, die kwalitatiewe en kwantitatiewe analise, die instrumente vir die studie, etiekkwessies en die samestelling van die korpora − is verder bespreek om ŉ empiriese ondersoek (hoofstuk 6 en 7) te kon uitvoer. In die hieropvolgende afdeling word die empiriese ondersoek opsommend weergegee.

8.2.3 Empiriese ondersoek

Die empiriese ondersoek is in twee dele uitgevoer. Eerstens, die stilometriese analise en daarna is die stilistiese en teksanalise uitgevoer. Tydens die empiriese analise is die gespesialiseerde korpus en die verdagte tekste (sien afdeling 5.12) telkens geanaliseer en met mekaar vergelyk om ooreenstemmende kenmerke te identifiseer wat outeurskapidentifikasie moontlik kon maak.

ŉ Stilometriese analise verwys na die numeriese analise van linguistiese data. Na gelang van die navorsing in hoofstuk 2 tot 5, is daar ondersoek ingestel na die sleutelwoorde, ratiobepaling van die grammatikale woorde, die aantal sinne, woorde en karakters per twiet, die taalverhoudings in die twiets, ŉ n-gramanalise asook leesbaarheidstoetse in die stilometriese analise om sodoende ooreenkomste tussen die outeurs vanuit die gespesialiseerde korpus en die drie verdagte tekste te kon identifiseer.

Die stilistiese en teksanalise (hoofstuk 7) verwys na die beskrywende, kwalitatiewe analise wat uitgevoer is om die verdagte outeurs te identifiseer. In hierdie analise het ek nie net ondersoek ingestel na die stylkenmerke nie, maar ek het telkens ook na die frekwensie verwys waarby elk van die onderskeie outeurs die aspekte gebruik wat in die analises ondersoek is. Die analise van die frekwensies is intern vergelyk (dit wil sê elk van die outeurs in die gespesialiseerde korpus se korpus is in die helfte gedeel en die twee dele van die korpora is met mekaar vergelyk) ten einde te kon bepaal of die kenmerke wel genoegsame ooreenkomste in ŉ enkele outeur se korpus getoon het om gebruik te kon word in die outeurskapidentifikasieproses. Daarna is die bevindinge met die verdagte tekste vergelyk en ooreenstemmende frekwensies is aangeteken. Die stylkenmerke is geïdentifiseer deur van *ATLAS.ti* en WordSmith Tools gebruik te maak. Voorbeelde van elk van die aspekte wat ondersoek is, is ook telkens weergegee.

Nege kategorieë is geïdentifiseer wat in die stilistiese en teksanalise ondersoek is. Hierdie kenmerke (sien ook afdeling 7.1) sluit in: die algemene kenmerke (afdeling 7.2); en ŉ ondersoek na sintaktiese (afdeling 7.3), morfologiese (afdeling 7.4) en semantiese kenmerke (afdeling 7.5). Verder is kodewisseling (afdeling 7.6), die gebruik van tussenwerpsels (afdeling 7.7), kraswoorde (afdeling 7.8), ŉ foute-analise (afdeling 7.9), sowel as ŉ n-gramanalise (afdeling 7.10) uitgevoer is. Deur die bogenoemde nege aspekte te analiseer, is daar nie net kenmerke van Afrikaans soos dit op Twitter gebruik word, geïdentifiseer nie, maar die outeur van elk van die verdagte tekste is ook bepaal. In die volgende afdeling word daar spesifiek verwys na die beantwoording van die navorsingsvrae.

8.3 Beantwoording van die navorsingsvrae

Soos reeds genoem, is daar drie verskillende navorsingsvrae gestel:

- Wat is die taalstilistiese en stylkenmerke van mikroblogs in Afrikaans?
- Kan outeurs van mikrobloginskrywings suksesvol geïdentifiseer word?
- Wat is die forensiese beginsels wat gebruik kan word om outeurskapidentifikasie op mikroblogs toe te pas?

Die vraag oor die taalstilistiese en stylkenmerke van mikroblogs word eerste beantwoord alvorens outeurskapidentifikasie aangespreek word.

8.3.1 Taalstilistiese en stylkenmerke van Afrikaans op mikroblogs

In terme van kenmerke van Afrikaans soos dit op Twitter gebruik word, is daar vasgestel dat daar wel unieke eienskappe geïdentifiseer kan word. Hierdie kenmerke is in afdeling 7.11 bespreek. Daar is spesifiek verwys na ooreenkomste tussen Afrikaans en navorsing wat in ander tale gedoen is. Afrikaanse Twitter-gebruikers, soos in ander tale, pas die taal aan volgens die eise wat nuwe tegnologie aan die taal stel. Tog is dié aanpassings van so ŉ aard dat kommunikasie steeds kan plaasvind − die boodskap aan die bedoelde gehoor word oorgedra.

Die interaktiewe aard van die platform en die Twitter-spesifieke elemente (soos die gebruik van die @-gebruikerskonvensie, hiperskakels en die hutsteken) sluit ook aan by internasionale bevindinge oor die gebruik van hierdie medium. Daar is verder vasgestel dat in teenstelling met bevindinge van onder andere Janse van Vuuren, die gebruik van leestekens en hoofletters nie spaarsamig is nie. Die veranderinge wat dus in Afrikaans aangetref word wanneer daar ondersoek ingestel is na die taalveranderinge, kan dus vergelyk word met veranderinge op internasionale vlak (sien afdeling 7.10).

8.3.2 Outeurskapidentifikasie op mikroblogs

Die tweede en derde navorsingsvrae het gefokus op die identifisering van mikroblogouteurs. In hierdie afdeling is daar verwys na die bevindinge met betrekking tot hierdie vrae. Daar word eerstens opsommend na die bevindinge oor die moontlike outeurs van die drie verdagte tekste verwys alvorens die metode bespreek word wat gebruik is om outeurs van mikrobloginskrywings te identifiseer.

8.3.2.1 Outeurs van die verdagte tekste

Vanuit die literatuurstudie is daar 14 verskillende aspekte geïdentifiseer[371]wat gebruik is ten einde outeurs van mikrobloginskrywing te identifiseer. Hierdie aspekte sluit in: ŉ sleutelwaarde-analise; ŉ ratio-analise; die sinne, woorde en karakters per twiet; die taalverhouding; leesbaarheidstoetse; algemene kenmerke; sintaktiese kenmerke; morfologiese kenmerke; semantiese kenmerke; kodewisseling; die gebruik van tussenwerpsels; kraswoorde; ŉ foute-analise en ŉ n-gramanalise.

In die geval van die eerste verdagte teks (V1) is daar bevind dat O2 in 10 van die 14 analises as die mees waarskynlike outeur van V1 geïdentifiseer is. Aangesien O2 wel die outeur was van wie daar ŉ ekstra korpus saamgestel is om V1 te vorm, kan die ondersoek as suksesvol beskou word. O1 is in 10 van die 13 analises as die mees waarskynlike outeur van die tweede verdagte teks (V2) geïdentifiseer. In die geval van V2 is O1 wel die outeur gewees vanuit wie se korpus daar ŉ ekstra korpus saamgestel is om die verdagte teks (V2) te vorm. Daar kan dus ook in die geval van die identifisering van die tweede verdagte teks aangevoer word dat die outeurskapidentifikasie suksesvol was. Die derde en laaste verdagte teks (V3) stem in 9 van die 14 analises met die korpus van O5 ooreen en gevolglik kan daar aangevoer word dat O5 sonder redelike twyfel die outeur van V3 is. O5 is inderdaad dan ook die outeur wat gebruik is om die derde verdagte teks (V3) se korpus saam te stel.

In al drie die ondersoeke na die identifisering van die outeurs van die verdagte tekste is die werklike outeur van die tekste deur middel van die verskeie analises geïdentifiseer. Die gevolgtrekking kan dus gemaak word dat die metode wat gebruik is in die *eksploratiewe forensiese analise in ŉ digitale era* wel as suksesvol beskou kan word. Vervolgens word die metodes voorgehou ten einde outeurs van mikrobloginskrywings te identifiseer.

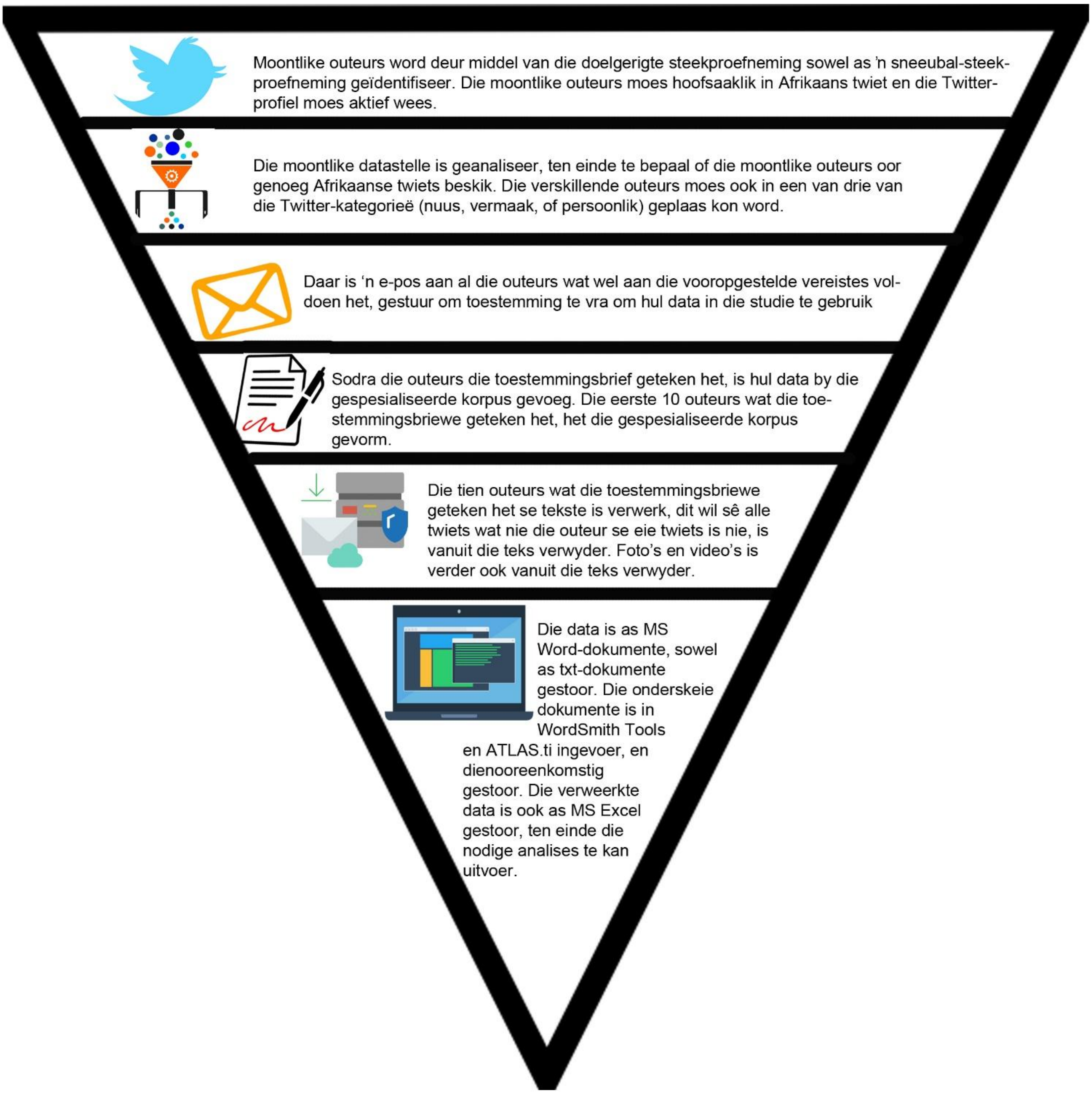
8.3.2.2 Metode vir die outeurskapidentifikasie van mikroblogouteurs

Soos reeds genoem, is daar van ŉ gemengde metode gebruik gemaak ten einde die outeurs van mikrobloginskrywings te kon identifiseer. Hierdie gemengde metode het uit ŉ stilometriese asook ŉ stilistiese en teksanalise bestaan. In hoofstukke 2, 3, 4 en 5 is daar na verskeie aspekte verwys wat ondersoek kan word ten einde die idiolek van ŉ individu te identifiseer. Aangesien daar tot op hede nie ŉ metode getoets is wat die outeur van ŉ kort teks in Afrikaans kan identifiseer nie, het ek nie net ŉ kombinasie van verskeie aspekte ondersoek in my eksploratiewe forensiese analise nie, maar ek ek het ook van ŉ kombinasie van benaderinge gebruik gemaak. Die proses van die identifisering van ŉ (Afrikaanse) outeur van ŉ mikrobloginskrywing kan in drie stappe uitgevoer

[371] Hierdie aspekte is geïdentifiseer soos aangedui in afdeling 8.3.2.

word. Die drie stappe sluit in; data-insameling en verwerking, die stilometriese analise en laastens die stilistiese en teksanalise.

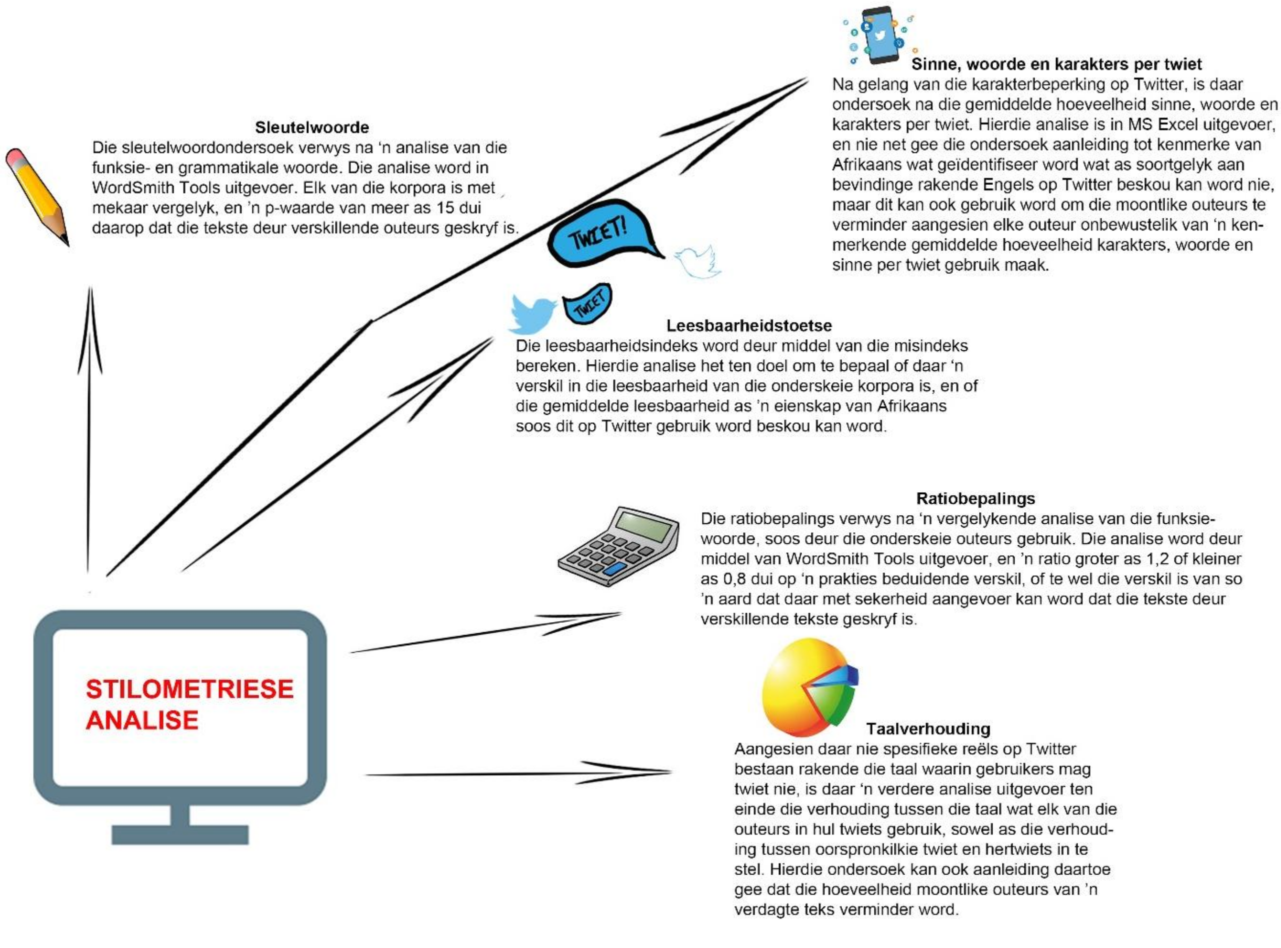
Figuur 8.1 bied ʼn opsomming van die metode wat gevolg is ten einde die data- insameling en verwerking in die proses van outeurskapidentifikasie te kon uitvoer:



Figuur 8.1 Proses van data-insameling en -verwerking

Figuur 8.1 is ʼn voorstelling van die proses van data-insameling en -verwerking. Nadat die data ingesamel en verwerk is, kon die analises gedoen word. Aangesien daar vir die doel van my studie

ŉ scenario geskep is waar drie outeurs as die verdagte tekste benoem is, dit wil sê daar was drie outeurs wat geïdentifiseer moes word wat ŉ moontlike strafregtelike misdryf begaan het. Indien die outeur van ŉ mikrobloginskrywing geïdentifiseer moet word vir strafregtelike redes sal die proses rakende die etiese klaring anders daaruitsien. Nadat die data ingesamel en verwerk is, kon die tweede stap aanvang neem. Die stilometriese analise word in Figuur 8.2 uitgebeeld en die stilistiese en teksanalise word in Figuur 8.3 voorgehou. Daarna volg ŉ opsomming van die metode wat gevolg is ten einde die outeurs van Afrikaanse mikrobloginskrywings te kon identifiseer (Figuur 8.4).

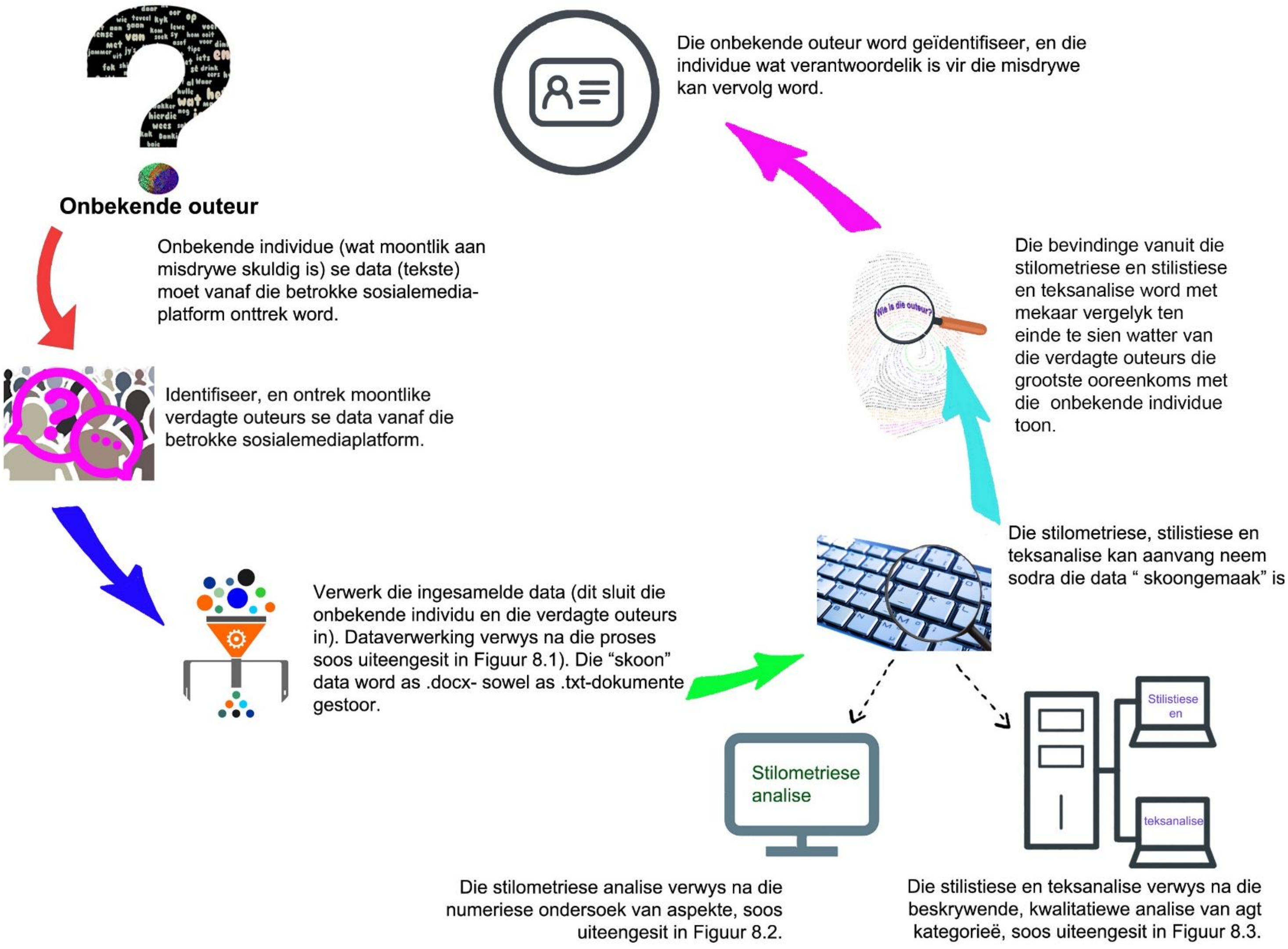


Figuur 8.2 Stilometriese analise

Die verwerkte data is telkens ondersoek aan die hand van die analises soos voorgehou in Figuur 8.2. Vanuit die stilometriese analise kon daar reeds afleidings gemaak word oor die outeurskap van die drie verdagte tekste. Die aspekte wat in die stilometriese analise ondersoek is, kan gevolglik as ŉ sukses beskou word ten opsigte van die identifisering van outeurs van mikrobloginskrywings.

Figuur 8.3 Stilistiese en teksanalise

Figuur 8.3 beeld die onderskeie aspekte uit wat in die stilistiese en teksanalise ondersoek is. Soos in hoofstuk 7 uiteengesit, kan elk van die aspekte wel ’n waardevolle bydra in die identifikasie van mikroblogouteurs lewer. Die aspekte wat dus in Figuur 8.2 en 8.3 weergegee word, kan wel in die suksesvolle identifisering van outeurs van Afrikaanse mikrobloginskrywings gebruik word. Die voorgestelde metode vir outeurskapidentifikasie word vervolgens voorgehou.



Figuur 8.4 Metode van outeurskapidentifikasie

Figuur 8.4 stel die metode voor wat gevolg kan word indien outeurs van Afrikaanse mikrobloginskrywings geïdentifiseer moet word. Figuur 8.1 tot 8.4 beeld dus die proses in geheel uit. Soos sigbaar in Figuur 8.4, is dit belangrik om tekste vanaf sosiale media in te samel en te verwerk van outeurs wat moontlike outeurs van anonieme tekste (tekste waarin misdrywe voorkom) kan wees. Die identifisering van outeurs kan nie uitgevoer word indien daar nie korpora is om die vergelykende analises mee uit te voer nie.

Die metode wat gebruik word, is ŉ gemengde metode waarin daar van beide stilometriese en stilistiese analises gebruik gemaak word. Die 13 hoofaspekte wat in die studie geïdentifiseer en geanaliseer is, blyk wel suksesvol te wees wanneer outeurs van Afrikaanse mikrobloginskrywings geïdentifiseer moet word.

8.4 Beperkings van die studie

Die studie, *Outeurskapidentifikasie en mikroblogs: ŉ eksploratiewe forensiese analise in ŉ digitale era*, se grootste beperking was die feit dat daar tot op hede nog nie ŉ korpus bestaan wat Afrikaans weerspieël soos dit op sosiale media gebruik word nie. ŉ Tweede beperking is die woordsoortetiketteerder wat in die studie gebruik is. Aangesien Afrikaans soos dit op sosialemediaplatforms gebruik word, aansienlik verskil van die Afrikaans wat tot dusver in Afrikaanse korpora beskikbaar is – die gebruik van Twitter-spesifieke tekselemente, kodewisseling en karakterbeperking op die platform het veral betrekking op hierdie verskil – is daar dus probleme met die groepering van die woordsoorte ondervind tydens die uitvoering van die studie. Selfs die informele Afrikaans wat in die WatKykJy-korpus op VivA beskikbaar is, is beperk.

Die verkryging van toestemming vanaf die outeurs wat as moontlike kandidate vir die studie geïdentifiseer is, kan as ŉ derde beperking beskou word. Aangesien die eerste tien outeurs wat die getekende toestemmingsbriewe teruggestuur het in die studie gebruik is, is daar nie genoegsame variasie in die variëteit van Afrikaans soos dit op Twitter gebruik word nie. Die aanwesigheid van meer taalvariëteite kan vanuit ŉ forensieslinguistiese oogpunt die identifisering van outeurs vergemaklik. Ten spyte van hierdie genoemde beperkings is daar wel in dié metode ruimte om die aspekte te kan identifiseer en analiseer.

8.5 Aanbevelings vir verdere navorsing

Ten spyte van die bogenoemde beperkinge tot hierdie studie is daar verskeie aspekte geïdentifiseer waaroor verdere navorsing onderneem kan word. Hierdie studie het die vorm van ŉ eksploratiewe studie oor die identifisering van outeurs (van kort tekste) in ŉ digitale era aangeneem. Aangesien die studie eksploratief van aard was en terselfdertyd ook kenmerkende eienskappe van Afrikaans soos dit op die gekose platform gebruik word, geïdentifiseer het, kan daar verder ondersoek ingestel word na die volgende:

- ŉ Diepgaande sintaktiese analise en ŉ manier om die bestaande programmatuur vir woordsoortetikettering te verbeter.
- Bevestiging van hierdie metode met betrekking tot unieke variëteite van Afrikaans.
- ŉ Vergelykende forensiese ondersoek tussen verskillende sosialemediaplatforms.
- ŉ Vergelykende forensiese ondersoek tussen Afrikaanse, Engelse en tweetalige mikrobloginskrywings.
- ŉ Ondersoek na die outomatisering van ŉ forensiese ondersoek deur middel van rekenaarlinguistiese algoritmes.

8.6 Samevatting en slot

In hierdie hoofstuk is daar ʼn oorsig van die studie verskaf alvorens die navorsingsvrae beantwoord is. Beperkinge van die studie is genoem en aanbevelings vir verdere navorsing is voorgestel. Met hierdie eksploratiewe, forensiese analise in ʼn digitale era is daar probeer om te bewys dat dit wel ook in Afrikaans moontlik is om outeurskapidentifikasie op (baie) korter tekste uit te voer en dat Afrikaans deurlopend deur gebruikers aangepas word ten einde aan hulle kommunikatiewe doelwitte te voldoen. Uit hierdie navorsing blyk dit dan ook duidelik dat dit wel moontlik is om die outeur van ʼn Afrikaanse mikrobloginskrywing te identifiseer mits die teks uit ʼn minimum van 408 woorde of 32 twiets bestaan. Aangesien die identifikasie van ʼn mikroblogouteur in Afrikaans wel suksesvol was, kan die aanname ook gemaak word dat hierdie metode ook op ander kort Afrikaanse tekste gebruik kan word. Dus berus die unieke bydrae wat hierdie studie lewer op beide die beskrywing van Afrikaans op Twitter sowel as die bevestiging van ʼn raamwerk van ʼn metode wat gevolg kan word om outeurskap van korpora wat van Twitter onttrek is te bepaal.

#taal #niekonstant #verander #behoefte #taalgebruiker #anonimiteit #identifikasie

Bronnelys

Abbasi, A. & Chen, H. 2005. Applying authorship analysis to extremist-group web forum messages. *IEEE Intelligent Systems*, 20(5):67-75.

Abbasi, A. & Chen, H. 2008. Writeprints: a stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems*, 26(2):1-29.

Aisopos, F., Papadakis, G., Tserpes, K. & Varvarigou, T. 2012. Content vs. context for sentiment snalysis: a comparative analysis over microblogs. Conference: 23rd ACM Conference on Hypertext and Social Media, June 25-28, 2012, Milwaukee, Wisconsin, USA. http://www.cs.uwm.edu/~munson/ht_12/ht/p187.pdf. Date of acces: 1 Dec. 2015.

Al-Salman, S.M. 2017. Has the power of language been compromised by the influence of social media? *Studies in English language teaching*, 5(2):172-186. http://dx.doi.org/10.22158/selt.v5n2p172 Date of access: 01 Oct. 2017.

An, J., Cha, M., Gummadi, K. & Crowcroft. J. 2011. Media landscape in twitter: a world of new conventions and political diversity. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.* https://www.aaai.org/ocs/index.php /ICWSM/ICWSM11/paper/viewFile/2825/3283 Date of access: 7 Feb. 2016.

Angouri, J. 2010. Quantitative, qualitative or both? Combining methods in linguistic research. (*In* Litosseliti, L., *ed.* Research methods in linguistics. New York: Continuum International Publishing Group. p. 29-45).

Anon. 2015. A brief history of the survey of English usage. http://www.ucl.ac..uk/english-usag Date of access: 1 Apr. 2016.

Anthony, L. 2009. Issues in the design and development of software tools for corpus studies: the case for collaboration. (*In* Baker, P., *ed.* Contemporary corpus linguistics. New York: Continuum. p. 87-104).

Anthony, L. 2013. Antconc: Readme. http://www.antlab.sci.waseda.ac.jp/software/antconc341/AntConc_readme.pdf Date of access: 20 Oct. 2015.

AntConc. 2014. AntConc Version 3.4.4. http://www.antlab.sci.waseda.ac.jp/software.html

Akmajian, A., Demers, R.A., Farmer, A.K. & Harnish, R.M. 2010. Linguistics: an introduction to language and communication. 6th ed. Cambridge: MIT Press.

Ariani, M.G., Sajedi, F. & Sajedi, M. 2014. Forensic linguistics: A brief overview of the key elements. *Procedia - Social and Behavioral Sciences,* 15:222-225.

Aston, G. 2011. Applied corpus linguistics and the learning experience. (*In* Vaina, V., Zyngier, S. & Barnbrook, G., *eds.* Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 1-17).

ATLAS.ti 2016. *ATLAS.ti* Version 7.5.10 (computer software). http://atlasti.com/ Date of access: 28 Jul. 2016.

Austin, J.L. 1962. How to do things with words. Oxford: Oxford University Press.

Baayen, H., Tweedie, F.J., Neijt, A.H., Halteren, H. & Van Krebbers, L. 2000. Back to the cave of shadows: stylistic fingerprints in authorship attribution. *The ALLC/ACH 2000* p. 156-158 University of Glasgow.

Baker, P. 2008. Pidgins versus creoles and pidgincreoles. (*In* Kouwenerg, S. & Singler, J.W., *eds.* The handbook of pidgin and creole studies. Oxford: Wiley Blackwell. p. 130-158).

Baker, P. 2010a. Sociolinguistics and corpus linguistics. Edinburg: Edinburgh University Press.

Baker, P. 2010b. Corpus methods in linguistics. (*In* Litosseliti, L., *ed.* Research methods in linguistics. New York: Continuum International Publishing Group. p. 93-116).

Baker, P. 2011. Social involvement in corpus studies. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds*. Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 18-27).

Baker, P., Hardie, A. & McEnery, T. 2006. A glossary of corpus linguistics. Edinburgh: Edinburgh University Press Ltd.

Barasa, S.N. 2010. Language, mobile phones and internet: a study of SMS texting, email, IM and SNS chats in computer mediated communication (CMC) in Kenya. Leiden: University of Leiden. (Dissertation − PhD).

Barlow, M. 2010. Individual usage: a corpus-based study of idiolects. 34th International LAUD Symposium, Landau, Germany. http://michaelbarlow.com/barlowLAUD.pdf Date of access: 12 Apr. 2016.

Baron, N.S. 1984. Computer-mediated communication as a force of change. *Visible language*, 18(2):120-145.

Baron, N.S. 1998. Letters by phone or speech by other means: the linguistics of email. *Language and communication*, (18):133-170.

Baron, N.S. 2001. Alphabet to email. How written English evolved and where it's heading. New York: Routledge.

Barry, K. & Luna, K. 2012. Stylometry for online forums. http:// cs.229.stanford.edu/proj2012/BarryLuna-StylometryforOnlineForums.pdf Date of access: 19 Aug. 2015.

Beckman, K. 2007. An immigrant's run-in with the law: a forensic linguistic analysis. New York: LFB Scholarly Publishing LLC.

Bednarek, M. 2009. Corpora and discourse: a three-pronged approach to analysing linguistic data. (*In* Haugh, M., Burridge, K., Mulder, J. & Peters, P.*, eds.* Selected proceedings of the 2008 HCSNet workshop on designing the Australian National Corpus. Somerville: Cascadilla Proceedings Project. p. 19-24.).

Beißwenger, M. & Storrer, A. 2008. Corpora of computer-mediated communication decisions (*In* Lüdeling, A. & Kytö, M., *eds.* Corpus linguistics: an international handbook. New York: Walter de Gruyter. p. 292-309).

Belmore, N. & Collot, M. 1996. Electronic language: a new variety of English. (*In Herring, S.C., ed.* Computer-mediated communication: linguistic, social and cross-cultural perspectives. Amsterdam: John Benjamins. p. 13-28).

Berber-Sardinha, T. 2001. Comparing corpora with WordSmith Tools: how large must the reference corpus be? www.aclweb.org. Date of access: 22 Sept. 2015.

Bhargava, M., Mehndiratta, P. & Asawa, K. 2013. Stylometric analysis for authorship attribution on twitter. (*In* Bhatnagar, V. & Srinivasa, S., *eds. Big data analytics: Second International Conference, BDA 2013 Mysore, India, December 2013 Proceedings*).

Biber, D. 2006. University language: a corpus-based study of spoken and written registers. Philadelphia: John Benjamins Publishing Co.

Biber, D., Conrad, S. & Leech, G. 2003. Longman student grammer of spoken and written English. Edinburg: Pearson Education limited.

Biber, D., Conrad, S. & Reppen, R. 2006. Corpus linguistics: investigating language structure and use. New York: Cambridge University Press.

Blackshaw, P. & Nazzaro, M. 2006. Consumer-generated media (CGM) 101: word-of-mouth in the age of the web-fortified consumer. New York: Nielsen Buzz Metrics.

Bloch, B. & Trager, G.L. 1942. Outline of linguistic analysis. Baltimore: Waverly Press.

Bloch, B. 1948. A set of postulates for phonemic analysis. *Language*, (24):3-46. http://www.jstor.org/stable/410284 Date of access: 5 Jan. 2016.

Bloomfield, L. 1950. Language. London: Allen & Unwin.

Bogers, T. & Björneborn, L. 2013. Micro-serendipity: meaningful coincidences in everyday life shared on twitter. iConference 2013 Proceedings. Fort Worth: Texas p. 196-208.

Bornoe, N. & Barkhuus, L. 2011. Privacy management in a connected world: students' perception of facebook privacy settings.(Paper presented at Workshop on Collaborative Privacy Practices in Social Media. : Part of the 2011 ACM conference on Computer supported cooperative work. Hangzhou:China). http://bornoe.org/papers/CSCW2011-Collaborative-Privacy-Workshop-bornoe.pdf Date of access: 1 Sept. 2015.

Borochovsky-Bar-Aba, E. & Kedmi, Y. 2010. The nature of SMS discourse: the case of Hebrew. *Folia Linguistica*, 44(1):1-30.

boyd, d.m. 2006. a bitty auto-biography / a smattering of facts. http://www.danah.org/aboutme.html Date of access: 22 Feb. 2015.

boyd, d.m. 2007. Why youth (heart) social networks: the role of networked public in teenage social life. (*In* MacAthur, D., *ed.* MacAthur Foundation Series on digital media and learning: youth, identity and digital media volume. Cambridge: MIT Press. p. 119-142).

boyd, d. 2014. It's complicated: the social lives of networked teens. London: Yale University Press.

boyd, d.m. & Ellison, N.B. 2007. Social networks sites: definition, history and scholarship. *Journal of computer-mediated communication*, 13(2):210-230.

boyd, d., Golder, S. & Lotan, G. 2010. Tweet, tweet, retweet: conversational aspects of retweeting on twitter. (Proceedings of HICSS-42, Persistent Conversation Track. Kauai, HI: IEEE Computer Society January 5-8, 2010). http://mikeg.typepad.com/perceptions/2011/10/abstract-tweet-tweet-retweet-conversational-aspects-of-retweeting-on-twitter.html Date of access: 12 Feb 2016.

Brennan, M., Afroz, S., & Greenstadt, R. 2011. Adversarial stylometry: circumventing authorship recognition to preserve privacy and anonymity*. ACM transactions on information and system security*, 15(3):1-22 http://doi.acm.org/10.1145/2382448.2382450 Date of access: 12 Feb. 2016.

Brocardo, M.L., Traore, I., Saad, S. & Woungang, I. 2013. Authorship verification for short messages using stylometry. (IEEE Intl. Conference on Computer, Information and Telecommunication Systems. Piraeus-Athens Greece). https://www.uvic.ca/engineering/ece/isot/assets/docs/Authorship_Verification_for_Short_Messages_using_Stylometry.pdf Date of access: 21 Feb. 2016.

Brunette, C. 2005. 'n Ondersoek na die invloed van rekenaar-bemiddelde kommunikasie op Afrikaans met verwysing na internasionale gevallestudies. Kaapstad: UCT. (Minitesis − MA).

Bruns, A. & Burgess, J. 2011. The use of Twitter hashtags in the formation of ad hoc publics. (Paper presented at 6th European Consortium for Political Research General Conference, August 25–27, University of Iceland, Reykjavik). http://snurb.info/files/2015/Twitter%20Hashtags%20from%20Ad%20Hoc%20to%20Calculated%20Publics.pdf Date of access: 12 May 2016.

Bruns, A. & Stieglitz, S. 2012. Quantitative approaches to comparing communication patterns on Twitter. *Journal of technology in human services*, 30(3-4):160-185. http://snurb.info/files/2012/Quantitative%20Approaches%20to%20Comparing%20Communication%20Patterns%20on%20Twitter.pdf Date of access: 20 Aug. 2014.

Bubaš, G. 2002. Computer mediated communiacation theories and phenomena: factors that influence collaboration over the Internet. CARNet Users Conference, Zaberg, September 2001. https://cuc.carnet.hr/cuc2001/papers/b1.pdf Date of access: 27 Jan. 2016.

Burrows, J.F. 1992. Word patterns and story shapes: the statistical analysis of narrative style. *Literary and linguistic computing*, 2:61-67

Butters, R.R. 2008. Trademarks and other proprietary terms. (*In* Gibbons, J. & Turell, T.M., *eds.* Dimensions of forensic linguistics. Philadelpia: John Benjamins Publishing Company. p. 231-247).

Butters, R.R. 2011. Forensic linguistics. *Journal of English linguistics*, 39(2):196-202.

Cabrio, E. 2010. Computational linguistics lab: corpus linguistics. Computational Linguistics Lab: WordNet. http://disi.unitn.it/~bernardi/Courses/CompLing/Slides_10_11/Lab4_WordNet.pdf Date of access: 17 Jan. 2015.

Carney, T.R. 2015. Twis om woorde: ’n Forensiese ondersoek na semantiese kwessies in hofsake. Bloemfontein: UFS (tesis – PhD).

Carrol, J. *ed.* 1991. Designing interaction: psychology at the human-computer interface. Cambridge: Cambridge University Press.

Carstens, W.A.M. 2011. Norme vir Afrikaans. 5e uitg. Pretoria: Van Schaik.

Carter, R. & McCarthy, M. 1995. Grammar and the spoken language. *Applied linguistics*, 16(2):141-58.

Castillo, E., Cervantes, O., Vilariño, D., Pinto, D. & León, S. 2014. Unsupervised methods for the authorship identification task. Notebook for PAN at CLEF 2014. p. 1035-1041.

Castillo, E., Cervantes, O., Vilariño, D. & Pinto, D. 2015. Author attribution using a graph based representation. *International Journal of Computer Applications.* 123(15):135-142. http://ceur-ws.org/Vol-1391/135-CR.pdf Date of access: 28 Jun. 2016.

Chafe, W. 1992. The importance of corpus linguistics to understanding the nature of language. (*In Svartvik, J. ed.,* Directions in corpus linguistics. Berlin: Mouton. p. 79-97).

Cavnar, W.B.& Trenkle, J.M. 1994. N-gram-based text categorization. http://odur.let.rug.nl/vannoord/TextCat/textcat.pdf Date of access: 04 Nov. 2017.

Chaski, C.E. 2001. Empirical evaluations of language based author identification techniques. *International journal of speech, language and the law*, 8(1):1-65.

Chaski, C.E. 2005. Who’s at the keyboard? Authorship attribution in digital evidence investigations. *Journal of digital evidence*, 4(1):1-13.

Chaski, C.E. 2007. The keyboard dilemma and forensic authorship attribution. *Advances in Digital Forensics III.*

Chaski, C.E. 2008. The computational-linguistic approach to forensic authorship attribution. (*In* Olsen, F., Lorz, A. & Stein, D., *eds*. Law and language: theory and practice. Düsseldorf: University Press. p. 1-14).

Chenault, B.G. 1997. Computer-mediated communication and emotion: developing personal relationships via CMC. http://www.neuage.org/php/vc/4-cmc.html Date of access: 12 Nov. 2015.

Chen, C., Zhang, J., Xie, Y., Xiang, Y. & Zhou, W. 2015. A performance evaluation of machine learning-based streaming spam tweets detection. *Computational Social Systems IEEE Transactions*, 2:65-76.

Cheng, W. 2012. Exploring corpus linguistics language in action. London: Routlege.

Chomsky, N. 1957. Syntactic structures. The Hague: Mouton.

Chomsky, N. 1958. Paper given at the Third Texas Conference on problems of linguistic analysis in English. Austin: University of Texas.

Choundhary, N. 2014. World Wide Web and its journey from Web 1.0 to Web 4.0. *International Journal of Computer Science and Information Technologies,* 5(6):8096-8100. http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf Date of access: 27 Oct. 2017.

Chun, D.M. 2013. Technology-supported materials for language teaching. (*In* Chapelle, C.A., *ed.* The encyclopedia of applied linguistics. Oxford: Blackwell Publishing Ltd. p. 6351-6380).

Ciardullo, M.A. & Romito, L. 2015. An interdisciplinary approach in forensic sciences: etiology, methodologies and importance of forensic linguistics. *Jacobs journal of forensic sciences*, 1(1):1-6.

Clark, H.H. 2009. Context and common ground. (*In* Mey, J. L., *ed.* Concise encyclopedia of pragmatics. Denmark: Elsevie. p. 116-119).

Coleridge, S.T. 1817. Biographia literaria; or biographical sketches of my literary life and opinions. London: Fenner.

Combrink, J. 1992. Hoe om ŉ verslag te skryf. Kaapstad: Tafelberg Uitgewers.

Conrad, S. 2011. Variation in corpora and its pedagogical implications. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds.* Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 44-62).

Conradie, J. & Groenewald, G. 2014. Die ontstaan en vestiging van Afrikaans. (*In* Carstens, W.A.M & Bosman, N., *reds.* Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik. p. 27-60).

Conover, M.D., Davis, C., Ferrara, E., McKelvey, K. & Menczer, F. 2013. The geospatial characteristics of a social movement communication network. *PLoS ONE*, 8(3):1-8.

Cornelius, E. 2012. ŉ Linguistiese ondersoek na die verstaanbaarheid van verbruikersdokumente vir die algemene Afrikaanssprekende publiek. Johannesburg: UJ (Proefskrif – PhD). https://ujcontent.uj.ac.za/vital/access/manager/Repository/uj:7366/CONTENT1 Datum van gebruik: 01 Apr. 2018.

Cornelius, E. 2013. Die benutting van leserskommentaar by die herskryf van ŉ verbruikersdokument in gewone Afrikaans. *LitNet Akademies*, 10(3). http://www.litnet.co.za/die-benutting-van-leserskommentaar-by-die-herskryf-van-n-verbruikersdokument-in-gewone-afr/ Datum van gebruik: 01 Apr. 2018.

Cornelius, E. 2015. Sinskonstruksies kenmerkend van Afrikaanse regstaal. *LitNet Akademies*, 12(3). http://www.litnet.co.za/sinskonstruksies-kenmerkend-van-afrikaanse-regstaal/ Datum van gebruik: 01 Apr. 2018.

Cotterill, J. 2004. Collocation, connotation, and courtroom semantics: lawyers' control of witness testimony through lexical negotiation. *Applied linguistics*, 25(4):513-537.

Cotterill, J. 2013. Corpus analysis in forensic linguistics. (*In* Chapelle, C.A., *ed.* The Encyclopedia of Applied Linguistics. Oxford: Blackwell Publishing Ltd. p. 1249-1256).

Coulthard, M. 1994. On the use of corpora in the analysis of forensic texts*. Forensic linguistics: the international journal of speech, language and the law*, 1(1):27-43.

Coulthard, D. 2004. Author identification, idiolect, and linguistic uniqueness. *Applied linguistics*, 25(4): 431-447.

Coulthard, M. 2005. Some forensic applications of descriptive linguistics. *Veredas*, 9(1):9-28.

Coulthard, M. & Johnston, A. 2007. An introduction to forensic linguistics: language in evidence. London: Routledge.

Coulthard, M. & Johnson, A. *reds*. 2010. The Routledge handbook of forensic linguistics. New York: Routledge.

Creswell, J.W. 2013. Research design: qualitative, quantitative and mixed methods approaches. 4th edition. London: SAGE.

Croker, R.A. 2009. An introduction to qualitative research. (*In* Heigham, J. & Croker, R.A., *eds.* Qualitative research in applied linguistics: a practical introduction. p. 3-24).

Cruse, A. 2000. Meaning in language: an introduction to semantics and pragmatics. Oxford: Oxford University Press.

Crystal, D. 1997. The Cambridge Encyclopedia of Language. 2nd edition. Cambridge: Cambridge University Press.

Crystal, D. 2004. Language and the internet. Cambridge: Cambridge University Press.

Crystal, D. 2006. Language and the internet. Cambridge: Cambridge University Press.

Crystal, D. 2008. A dictionary of linguistics and phonetics. 6th ed. Oxford: Blackwell Publishing.

Crystal, D. 2010. How is the internet changing language today? http://www.youtube.com/watch?v=P2XVdDSJHqY Date of access: 4 Mar. 2014.

Crystal, D. 2011. Internet linguistics: a student guide. London: Routledge.

Danet, B. 1997. Language, play and performance in computer-mediated communication: final report submitted to the Israel Science Foundation: February. http://atar.mscc.huji.ac.il./~msdanet/report95.html Date of access: 15 Nov. 2015.

Darics, E. 2013. Non-verbal signalling in digital discourse: the case of letter repetition. *Discourse, context & media*, 2(3):141-148. https://ac-els-cdn-com.nwulib.nwu.ac.za/S2211695813000330/1-s2.0-S2211695813000330-main.pdf?_tid=5dc9677e-ac1c-11e7-95dd-00000aab0f02&acdnat=1507462529_4e2feb496c672e78cae2fdfdb1ea7944 Date of access: 08 Oct. 2017

D'Arcy, A. & Young, T.M. 2012. Ethics and social media: implications for sociolinguistics in the networked public. *Journal of sociolinguistics*, 16(4):532-546.

Davenport, R.H. 1993. South Africa: a modern history. 4th ed. Hampshire: The Macmillan Press LTD.

Davenport, S.W., Bergman, S.M., Bergman, J.Z. & Fearrington, M.E. 2014. Twitter versus Facebook: exploring the role of narcissism in the motives and usage of different social media platforms. *Computers in human behavior*, 32:212-220.

Davies, A. 1999. An introduction to applied linguistics: from practice to theory. Edinburg: Edinburgh University Press.

Davis, A. 2005. A glossary of applied linguistics. Edinburg: Edinburg University Press.

Davies, M. 2011. Synchronic and diachronic uses of corpora. (*In* Viana, V., Zyngier, S. & Barnbrook,, G., *eds.* Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 62-80).

Davies, M. 2014. Corpora. http://corpus.byu.edu/corpora.asp Date of access: 10 Apr. 2016.

Davies, M. 2014. The Wikipedia Corpus. http://corpus.byu.edu/wiki/ Date of access: 09 Apr. 2016.

Davies, M. 2014. Corpus Of Global Web-Based English (GloWbE). http://corpus.byu.edu/glowbe/ Date of access: 09 Apr. 2016.

De Klerk, V. 2003. The language of truth and reconciliation: was it fair to all concerned? *Southern African linguistics and applied language studies*, 21(1):1-14.

Deller, R. & Hallam, S. 2011. Twittering on: audience research and participation using Twitter. *Participations: journal of audience & reception studies*, 8(1):216-245.

Derrida, J. 1974. Of Grammatology. John Hopkins University Press: Maryland

De Stadler, L.G. 1989. Afrikaanse semantiek. Johannesburg: Southern Boekuitgewers.

De Saussure, F. 1960. [1916] Course in general linguistics. Uit Frans vertaal deur Wade Baskin. New York: McGraw-Hill Book Company.

De Vel, O., Anderson, A., Corney, M. & Mohay, G. 2001. Multi-topic e-mail authorship attribution forensics. ACM Conference on Computer Security-Workshop on Data Mining for Security Applications. Philadelphia, PA.

De Villiers, M. 1983. Nederlands en Afrikaans. 4de uitg. Goodwood: Nasionale Opvoedkundige Uitgewer Beperk.

Decker, S. 2011. Semantic web. http://semanticweb.org/wiki/Main_Page.html Date of access: 28 Apr. 2016.

Den Besten, H. 1989. From khoekhoe foreignertalk via Hottentot Dutch to Afrikaans: the creation of a novel grammar. (*In* Pütz, M. & Dirven, R., *eds.* Wheels within wheels. Frankfurt: P. Lang. p. 207-249).

Denzin, N.K. 1978. The research act: a theoretical introduction to sociological methods. New York: Praeger.

Department Kuns en Kultuur & CTexT. 2013. Afrikaans NCHLT text corpora. Potchefstroom: CTexT, North-West University. ISLRN: 544-932-849-161-3. Available: http://rma.nwu.ac.za/index.php/resource-catalogue/afrikaans-nchlt-text-corpora.html

Deumert, A. 2004. Language standardization and language change: the dynamics of Cape Dutch. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Deumert, A. 2009. Semantics change: the internet and text messaging. (*In* Mey, J.L., *ed.* Cocise encyclopidia of pragmatics. Amsterdam: Elsevier. p. 909-912).

Dietrick, W., Miller, Z., Valyou, B., Dickinson, B., Munson, T., Hu, W., Garzia, F. ... Cusani, R. 2012. Gender identification on twitter using the modified balanced winnow. *Communications and network*, 4(3):189-195.

Dinh, T.L. 2014. Authorship identification for online text. Groningen: University of Groningen (Thesis – MA).

Donath, J. 1998. Identity and deception in the virtual community. (*In* Smith, M. & Kollock, P., *eds.* Communities in cyberspace. London: Routledge). http://smg.media.mit.edu /people/Judith/Identity/IdentityDeception.html Date of access: 3 Feb. 2016.

Donath, J.S. 1999. Identity and deception in the virtual community. (*In* Smith, M.A. & Kollock, P., *eds.* Communities in cyberspace. London: Routledge. p. 27-59).

Du Plessis, E. 2016. Etiese aspekte in navorsing. (*In* Joubert, I., Hartell, C. & Lombard. K., *reds.* Navorsing: 'n gids vir die beginnernavorser. Pretoria: Van Schiak p. 73-92).

Du Plessis, H. 1988. Variasietaalkunde. Pretoria: Serva Uitgewers.

Du Plessis, H. 1995. Variasietaalkunde. Pretoria: Kagiso.

Du Toit, P. & Smith-Müller, W. 2015. Stylboek: riglyne vir paslik skryf. Pretoria: Van Schaik.

Dunlap, J.C. & Lowenthal, P.R. 2009. Tweeting the night away: using Twitter to enhance social presence. *Journal of information systems education*, 20(2). http://patricklowenthal.com/publications/Using_Twitter_to_Enhance_Social_Presence.pdf Date of access: 19 Feb. 2016.

Eckert, P. 1995. (ay) Goes to the city: exploring the expressive use of variation. (*In* Guy, G., Feagin, C., Schriffen, D. & Baugh, J., *eds.* Towards a social science of language and society. Papers in honour of W. Labov. Vol 1. Variation and change in society. Amsterdam: John Benjamin. p. 47-68).

Eckert, P. 2008. Variation and the indexical field. *Journal of sociolinguistics*, 12/4:453-476.

Edwards, A. 2015. When is it ok to use informal language? https://www.grammarly.com/blog/when-is-it-ok-to-use-informal-language/ Date of access: 21 Feb. 2016.

Eggington, W. G. 2008. Deception and fruad. (*In* Gibbons, J. & Turell, T.M., *eds.* Dimensions of forensic linguistics. Philadelpia: John Benjamins Publishing Company. p. 249-264).

Ehrlich, K. & Shami, N.S. 2010. Microblogging inside and outside the workplace. *Association for the advancement of artificial intelligence*. http://www.cs.cornell.edu/~sadats/icwsm2010.pdf Date of access: 4 Feb. 2016.

Eisenstein, J. 2013. What to do about bad language on the internet. Proceedings of NAACL-HLT 2013:359-369. http://www.aclweb.org/anthology/N13-1037 Date of access: 31 Aug. 2016.

Elliot, T.S. 1943. Four Quartets. New York: Harcourt, Brace and Company.

Ellison, N.B. & boyd, d. 2013. Sociality through social network sites. (*In* Dutton, W.H., *ed.* The Oxford handbook of internet studies. Oxford: Oxford University Press. p. 151-172).

Feinauer, A. E. 1981. Die taalkundige gedrag van vloekwoorde in Afrikaans. Stellenbosch: SUN (Verhandeling – MA).

Fish, S. 1989. Doing what comes naturally: change, rhetoric, and the practice of theory in literary and legal studies. Durham: Duke University Press.

Fish, S. 1994. There's no such thing as free speech, and it's a good thing, too. Oxford: Oxford University Press.

Fischer-Starcke, B. 2010. Corpus linguistics in literary analysis: Jane Austin and her contemporaries. London: Continuum.

Fitzpatrick, E. & Bachenko, J. 2010. Building a forensic corpus to test language-based indicators of deception. (*In* Davies, M., Wulff, S. & Thomas, S., *eds.* Corpus-linguistic applications: current studies, new directions. Amsterdam: Rodopi. p. 183-96).

Fletcher, W.H. 2013. Corpus analysis of the world wide web. (*In* Chapelle, C.A., *ed.* The Encyclopedia of Applied Linguistics. Oxford:Blackwell Publishing Ltd. p. 1339-1347).

Flowerdew, L. 2004. The argument for using English specialized corpora to understand academic and professional settings. (*In* Connor, U. & Upton, T., *eds.* Discourse in the professions: perspectives from corpus linguistics. Amsterdam: John Benjamins. p. 11-33).

Forsyth, R. & Holmes, D. 1996. Feature finding for text classification. *Literary and linguistic computing*, 11(4):163-174.

Francis, W.N. 1992. Language corpora BC. (*In* Svartvik, J., *ed.* Directions in corpus linguistics. *Proceedings of Nobel symposium 82, Stockholm, 4-8 August, 1991*. Berlin: Mouton. p. 17-32).

Free large images. 2015. Twitter logo. http://www.freelargeimages.com/twitter-logo-hd-png-4557/ Date of access: 12 Aug. 2016.

Fries, S. 2014. *ATLAS.ti* 7 user manual. *ATLAS.ti* scientific software development GmbH http://atlasti.com/ Date of access: 28 Jul. 2016.

Frost, G. 2011. Twitter has evolved from a social network to the future of news. http://www.networkconference.netstudents.org Online Conference On Networks and Communities, 25 April 2011. Date of access: 3 Jul. 2014.

García-Gavilanes, R.O. 2015. User behavior in microblogs with a cultural emphasis.

Universitat Pompeu Fabra (Thesis − PhD). http://www.tdx.cat/handle/10803/287974 Date of access: 10 Feb. 2015.

Ghadessy, M., Henry, A. & Roseberry, R.L. 2001. Introduction. (*In* Ghadessy, M., Henry, A. & Roseberry, R.L., *eds.* Small corpus studies and ELT: theory and practice. Amsterdam: John Benjamins Publishing Company. p. xvii-xxii).

Gibbons, J. & Turell, T.M. 2008. Introduction. (*In* Gibbons, J. & Turell, T.M., *eds.* Dimensions of forensic linguistics. Philadelpia: John Benjamins Publishing Company. p. 1-4).

Giliomee, H. & Mbenga, B. 2007. Nuwe geskiedenis van Suid-Afrika. Kaapstad: Tafelberg.

Gilmore, P. 2011. We call it "Our Language": a children's Swahili pidgin transforms social and symbolic order on a remote hillside in up-country Kenya. *Anthropology & education quarterly*, 42(4):370-392.

Gleason, B. 2016. New literacies practices of teenage Twitter users. *Learning, media and technology*, 41(1):31-54. http://www-tandfonline-com.nwulib.nwu.ac.za/doi/pdf/10.1080/17439884.2015.1064955?needAccess=true Date of access: 01 Oct. 2017.

Go, A., Bhayani, R. & Huang, L. 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford. http://s3.eddieoz.com/docs/sentiment_analysis/Twitter_Sentiment_Classification_using_Distant_Supervision.pdf Date of access: 8 Jan. 2015.

Gong, W. & Ooi, V.B.Y. 2008. Innovations and motivations in online chat. (*In* Kelsey, S. & St. Amat, K., *eds.* Handbook of research on computer mediated communication. London: Information Science Reference. p. 917-933).

Grant, T. 2007. Quantifying evidence in forensic authorship analysis. *International journal of speech, language and the law*, 14(1):1-25.

Grant, T. 2008. Approaching questions in forensic authorship analysis. (*In* Gibbons, J. & Turell, T.M., *eds.* Dimensions of forensic linguistics. Philadelpia: John Benjamins Publishing Company. p. 215-230).

Grant, T. & Baker, K. 2001. Identifying reliable, valid markers of authorship: a response to Chaski. *International journal of speech, language and the law*, 8(1):66-79.

Granath, S. 2007. Size matters − or thus can meaningful structures be revealed in large corpora. (*In* Facchinetti, R., *ed.* Corpus linguistics 25 years on. New York: Rodopi. p. 169-185).

Grebe, H. 2010. Die Oosgrenshipotese as geografiese en teoretiese konstruk. *LitNet akademies*, 7(3):195-211.

Green, R.M. & Sheppard, J.W. 2013. Comparing frequency- and style-based features for twitter author identification. Twenty-Sixth International Florida Artificial Intelligence Research Society Conference. p. 64-69.

Gressel, G., Hrudya, P., Surendran, K., Thara, S. & Aravind, A. 2014. Ensemble learning approach for author profiling. Notebook for PAN at CLEF 2014.

Griffith, W. & Smalbergen, N. 2015. Can a tweet enjoy copyright protection? http://www.polity.org.za/article/can-a-tweet-enjoy-copyright-protection-2015-08-20 Date of access: 18 May 2016.

Gries, S. 2009. What is corpuslinguistics? *Language and linguistics compass*, 3:1-17. http://www.linguistics.ucsb.edu/faculty/stgries/research/2009_STG_CorpLing_LangLingCompass.pdf Date of access: 10 Jan. 2016.

Gries, S. 2011. Methodological and interdisciplinary stance in corpus linguistics. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds.* Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 81-98).

Grunig, J.E. & Hunt, T. 1984. Managing public relations. New York: Holt, Rinehart & Winston.

Grunig, L.A., Grunig, J.E. & Dozier, D.M. 2002. Excellent public relations and effective organizations: a study of communication management in three countries. Mahwah: Lawrence Erlbaum.

Halliday, M.A.K. 1973. Language in a social perspective: explorations in the functions of language. London: Edward Arnold.

Halliday, M.A.K. 1989. Spoken and written language. 2nd ed. Oxford: OUP.

Halliday, M.A.K., Teubert, W., Yallop, C. & Cermakova, A. 2004. Lexicology and corpus linguistics: an Introduction. London: Continuum.

Hardaker, C. 2013. "Uh.... not to be nitpicky,,,,,but...the past tense of drag is dragged, not drug." *Journal of Language Aggression and Conflict*, 1(1):58-86.

Hardaker, C. 2015. 'I refuse to respond to this obvious troll': An overview of responses to (perceived) trolling. *Corpora*, 10(2):201-229.

Hardaker, C. & McGlashan, M. 2015. "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics*, 91:80-93.

HAT. Handwoordeboek van die Afrikaanse Taal. 2015. 6de Uitgawe. Kaapstad: Pearson Education.

Henry, A. & Roseberry, R.L. 2001. Using a small corpus to obtain data for teaching a genre. (*In* Ghadessy, M., Henry, A. & Roseberry, R.L., *eds.* Small corpus studies and ELT: theory and practice. Amsterdam: John Benjamins Publishing Company. p. 93-135).

Herbig, F.J.W. & Joubert, S.J. 2006. Criminological semantics: conservation criminology − vision or vagary? *Acta criminologica*, 19(3):88-103.

Herring, S.C. 2001. Handbook of discourse analysis. http:l/www.scholar.google. com/url?sa=http;llwww.let.rug.nI/-redekerlherring.pdf. Date of access: 24 Jan. 2016.

Herring, S.C. 2002. Computer mediated communication on the internet. Annual review of information science and technology. *American society for information science*, 37:109-168. https://eric.ed.gov/?id=EJ645676 Date of access: 19 Feb. 2016.

Hesseling, D.C. 1923. [1899]. Het Afrikaansch. 2de uitg. Leiden: N.V. Boekhandel en Drukkerij.

Hayes, K.J. 2008. The road to Monticello: The life and mind of Thomas Jefferson. London: Oxford University Press.

Hoey, M. 2005. Lexical priming: a new theory of words and language. London: Routledge.

Holm, J. 2000. An introduction to pidgin and creoles. Cambridge: Cambridge University Press.

Holmes, D. 2005. Communication theory: media, technology, society. London: Sage Publications.

Holmes, D. 2009. Computer-mediated communication. (*In* Littlejohn, S.W. & Foss, K.A., *eds.* Encyclopedia of communication theory. Los Angles: SAGE Publications. p. 161-164).

Holmes, J. & Meyerhoff, M. *eds*. 2003. The handbook of language and gender. Oxford: Blackwell Publishers.

Honeycutt, C. & Herring, S. 2009. Beyond microblogging: conversation and collaboration in Twitter. (In Proceedings of the 42nd Hawaii International Conference on System Sciences). http://www.academia.edu/1067534/Beyond_microblogging_Conversation_and_collaboration_via_Twitter Date of access: 19 Feb. 2016.

Hong, L., Convertino, G. & Chi, E.D. 2011. Language Matters in Twitter: A Large Scale Study. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2856/3250/ Date of access: 29 Sept 2017.

Hoover, D.L., Culpeper, J. & O'Halloran, K. 2014. Digital literary studies: corpus approaches to poetry, prose, and drama. New York: Routledge.

Hubbard, E.H. 1994. Errors in court: a forensic application of error analysis. *South African journal of linguistics: supplement*, 12(20):3-16.

Hubbard, E.H. 1995. Linguistic fingerprinting? A case study in forensic stylometrics. *South African journal of linguistics: supplement*, 13(26):55-72.

Huberman, B.A., Romero, D. & Wu, F. 2008. Social networks that matters: Twitter under the microscope. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1313405 Date of access: 29 Sept. 2017.

Hughes, R. 1996. English in speech and writing: investigating language and literature. New York: Routledge.

Humphreys, L. 2010. Historicizing microblogging. International Communication Association 2012 Annual Conference, May 24-27, Phoenix, *AZ.* http://www.cs.utoronto.ca/~phillipa/papers/ICA2012.pdf. Date of access: 19 Feb. 2016.

Hundt, M., Nesselhauf, N. & Biewer, C. *eds*. 2007. Corpus linguistics and the web. New York: Rodopi.

Hunston, S. 2002. Corpora in applied linguistics. Cambridge: Cambridge University Press.

Hunston, S. 2008. Collection strategies and design decisions. (*In* Lüdeling, A. & Kytö, M., *eds.* Corpus linguistics: an international handbook. New York: Walter de Gruyter. p. 154-167).

Hyland, K., Huat, C.M. & Handford, M. 2012. Corpus applications in applied linguistics. New York: Continuum.

Hymes, D. 1971. Competence and performance in linguistic theory. (*In* Huxley, R. & Ingram, E., *eds.* Language acquisition: models and methods. New York: Academic Press. p. 3-28).

Hyon, S. 1996. Genre in three traditions: implications for ESL. *Tesol quarterly*, 30(4):693-722. http://www.jstor.org/stable/3587930 Date of access: 29 Aug. 2016.

Inches, G., & Crestani, F. 2011. Online conversation mining for author characterization and topic identification. (In *Proceedings of the 4th Workshop for Ph.D. Students in Information and Knowledge Management*, 28 Oct. p19-26. New York, NY: Association of Computing Machinery).

Internet Live Stats. 2015. http://internetlivestats.com Date of access: 24 Sept. 2015.

Internet World Stats. 2015. http://www.internetworldstats.com/stats.htm. Date of access: 31 Aug. 2015.

Ishihara, S. 2014. A likelihood ratio-based evaluation of strength of authorship attribution evidence in SMS messages using N-grams. *International journal of speech, language and the law*, 21(1):23-49.

Ivankova, N.V. & Creswell, J.W. 2009. Mixed methods. (*In* Heigham, J. & Croker, R.A., *eds.* Qualitative research in applied linguistics: a practical introduction. p. 135-164).

Ivankova, N.V., Creswell, J.W. & Clark, V.L.P. 2011. Foundations and approaches to mixed method research. (*In* Maree, K., *ed.* First steps in research. Pretoria: Van Schaik. p. 256-280).

Izarry, E. 1992. A computer-assisted investigation of gender-related idiolect in Octovio Oaz and Rosario Castellanos. *Computers and the humanities*, 26(2):103-117.

Jakobson, R. 1994. Language in literature. Cambridge: The Belknap Press.

Jansen van Vuuren, M. 2007. Die internet as platform vir 'n nuwe Afrikaanse spraakgemeenskap en 'n nuwe varieteit − 'n korpuslinguistiese ondersoek. Vanderbijlpark: Vaaldriehoekkampus: NWU (Proefskrif − PhD).

Java, A., Finin, T., Song, X. & Tseng, B. 2007. Why we Twitter: understanding microblogging usage and communities. Proceedings of the Joint 9th Webkdd and 1st Sna-Kdd Workshop on Web Mining and Social Network Analysis, ACM New York, p. 56-65. http://dx.doi.org/10.1145/1348549.1348556 Date of access: 15 Jan. 2015.

Jessen, M. 2010. The forensic phonetician: Forensic speaker identification by experts. (*In* Coulthard, M. & Johnson, A., *eds. The Routledge handbook of forensic linguistics. London: Routledge.* p: 378-395).

Johnson, A. & Wright, D. 2014. Identifying idiolect in forensic authorship attribution. *Language and law*, 1(1):37-69. http://ler.letras.up.pt/uploads/ficheiros/12684.pdf Date of access: 29 Apr. 2016.

Johansson, S. 2008. Some aspects of the development of corpus linguistics in the 1970's and 1980's. (*In* Lüdeling, A. & Kytö, M., *eds.* Corpus linguistics: an international handbook. New York: Walter de Gruyter. p. 33-52).

Johansson, S. 2011. A multilingual outlook of corpora studies. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds.* Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 115-129).

Johnstone, B. 2000. The individual voice in language. *Annual Review of Anthropology.* 2000, 29:405-24. http://www.annualreviews.org/doi/abs/10.1146/annurev.anthro.29.1.405 Date of access: 5 Jan. 2017.

Johnstone, B. 2011. Dialect enregisterment in performance. *Journal of sociolinguistics*, 15(5):657-679.

Jones, N.J. & Bennell. C. 2007. The development and validation of statistical predictionrules for discriminating between genuine and simulated suicide notes. *Archives of suicide research*, 11(2):1-15.

Joseph, H. 1996. Language history, language change and language relationship. Berlin: Mouton de Gruyter.

Juola, P. 2006. Authorship attribution. *Foundations and trends in information retrieval*, 1(3):233-334.

Juola, P. 2008. Authorship attribution. Hanover: Now Publishers Inc. http://books.google.co.za/books?id=_B2zDLdqe60C&pg=PA6&lpg=PA6&dq=authorship+attribution+definition&source=bl&ots=BdpZKOf7vp&sig=7ZZGiHiQEz1jS6FRH2N0MxLIgFY&hl=en#v=onepage&q&f=false Date of access: 20 Aug. 2014.

Kaplan, A.K. & Haenlein, M. 2010. Users of the world, unite! The challenges and opportunities of social media. *Business horizons*, (53):59-68. http://michaelhaenlein.eu/Publications/Kaplan,%20Andreas%20-%20Users%20of%20the%20world,%20unite.pdf Date of access: 28 Apr. 2016.

Kaschula,, R.H. & Ralarala, M.K. 2012. Language rights, intercultural communication and the law in South Africa. *South African Journal of African Languages*, 24(4):252-261.

Kelsey, T. 2010. Social networking spaces: from Facebook to Twitter and everything in between. New York: Springer-Verlag.

Kennedy, G. 1998. An introduction to corpus linguistics. London: Longman.

Kietzmann, J.H., Hermkens, K., McCarthy, I.P. & Silvestre, B.S. 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, (54)3:241-251. http://beedie.sfu.ca/files/PDF/research/McCarthy_Papers/2011_Social_Media_BH.pdf Date of access: 28 Apr. 2016.

Kilgarriff, A. & Grefenstette, G. 2003. Web as corpus: introduction. *Special issue. Computational linguistics*, 29(3):333-347.

Kirsten, J. 2013. Laaste spore van Nederlands in Afrikaanse werkwoorde. Vanderbijlpark: Vaaldriehoekkampus, NWU (Verhandeling − MA).

Kitade, K. 2001. Pragmatics of asynchronous computer-mediated communication. (*In* Carrol, J., *ed.* Designing interaction: psychology at the human-computer interface. Cambridge: Cambridge University Press. p. 4974-4979).

Klopper, R. 2009. The case of cyber forensic linguistics. *Alternation*, 16(1):261-294.

Koester, A. 2010. Building a small specialised corpora. (*In* O'Keeffe, A. & McCarthy, M., *eds.* The Routledge handbook of corpus linguistics. London: Routledge Taylor and Francis Group. p. 66-79).

Kopka, J. & Vogel, C. 2004. Testing the reliability of an authorship identification method. https://www.cs.tcd.ie/publications/tech-reports/reports.04/TCD-CS-2004-40.pdf Date of access: 14 May 2016.

Koppel, M., Argamon, S. & Shimoni, A.R. 2002. Automatically categorizing written texts by author gender. *Literary and linguistic computing*, 17(4):401-412.

Koppel, M. & Schler, J. 2003. Exploiting stylistic idiosyncrasies for authorship attribution. Referaat gelewer by die International Joint Conferences on Artificial Intelligence workshop on Computational Approaches to Style Analysis. Acapulco, Meksiko. http://citeseer.uark.ede:8080/citeseerx/viewdoc/summary;jsessionid=F8274EBE05D5F0E52BE08F1519BD458?doi=10.1.1.2.3019 Date of access: 2 Feb. 2015.

Koppel, M. & Schler, J. 2004. Authorship verification as a one-class classification problem. Proceedings of the 21st international conference on Machine learning, ICML, 2004, New York: ACM. http://www.machinelearning.org/proceedings/icml2004/papers/415.pdf Date of access: 5 Jan. 2016.

Koppel, M., Schler, J. & Argamon, S. 2011. Authorship attribution in the wild. *Language resources and evaluation*, 45(1):83-94.

Koppel, M., Schler, J. & Argamon, S. 2013. Authorship attribution: what's easy and what's hard? *Journal of law and policy*, (21)2:317-331. https://www.researchgate.net/publication/

251354683_Authorship_Attribution_What%27s_Easy_and_What%27s_Hard?enrichId=rgreq-3f9647a6d05e8fc3b578308a2e2e4dd1-XXX&enrichSource=Y292ZXJQYWdlOzI1MTM1NDY4MztBUzoxNjI1MzM1MzMwMzY1NDRAMTQxNTc2MjQyMjg0OQ%3D%3D&el=1_x_3 Date of access: 1 July 2016.

Kotzé, E.F. 2007. Die vangnet van die woord: forensies-linguistiese getuienis in ’n lastersaak. *Southern African linguistics and applied language studies (SALALS)*, 25(3):385-399.

Kotzé, E.F. 2010. Author identification from opposing perspectives in forensic linguistics. *Southern African linguistics and applied language studies (SALALS)*, 28(2):185-197.

Kotzé, E.F. 2012. Appropriate methodology for author identification − a comparative investigation into two divergent cases. Forensic linguistics / language and law: researching interdisciplinary dimensions and perspectives. Paper at 3rd European Conference of the International Association of Forensic Linguists, Porto, Portugal, 15-18 October 2012.

Kotzé, E.F. & Oosthuizen, J. 2015. Tel jou woorde voor jy skryf. *LitNet Akademies* http://www.litnet.co.za/tel-jou-woorde-voor-jy-skryf/ Datum van gebruik: 16 Jan 2016.

Kouwenberg, S. & Singler, J.W. *eds.* 2008. Introduction. (*In* Kouwenberg, S. & Singler, J.W., *eds*. The handbook of pidgin and creole studies. Oxford: Wiley Blackwell. p. 130-158.

Kracht, M. & Klein, U. 2014. The grammar of code switching. *Journal of logic, language and information*, 23:313-329. http://eds.b.ebscohost.com.nwulib.nwu.ac.za/eds/command/detail?vid=11&sid=bb69bef6-f09e-48b8-8154-df889a72b7d9%40sessionmgr103 Date of access: 07 Oct. 2017.

Kraljic, T., Brennan, E.S. & Samuel, A.G. 2007. Accommodating variation: dialects, idiolects, and speech processing. www.elsevier.com/locate/COGNIT Date of access: 18 Nov. 2015.

Krause, M. 2013. Stylometry-based fraud and plagiarism detection for learning at scale. https://www.researchgate.net/profile/Markus_Krause/publication/271836873_Stylometry-based_Fraud_and_Plagiarism_Detection_for_Learning_at_Scale/links/54d3727f0cf2b0c6146d8a94/Stylometry-based-Fraud-and-Plagiarism-Detection-for-Learning-at-Scale.pdf Date of access: 24 Dec. 2016.

Kredens, K. 2000. Forensic linguistics and the status of lingusitic evidence in the legal setting. University Lodz: Poland. (Dissertation − PhD).

Kriel, A.P. 2014. Alternatiewe funksies van Swartafrikaans. Potchefstroomkampus: NWU. (Verhandeling − MA).

Kübler, S. & Zinsmeister, H. 2015. Corpus linguistics and linguistically annotated corpora. London: Bloomsbury Academic.

Kuhl, J.W. 2003. The idiolect, chaos and language custom for from equilibrium: conversations in Morocco. The University of Georgia: Georgia (Thesis − PhD). http://www.joekuhl.com/page1.html. Date of access: 05 Jan. 2016.

Kwak, H., Lee, C., Park, H. & Moon, S. 2010. What is Twitter, a Social Network or a News Media? *Proceedings of the 19th international conference on World wide*. Raleigh, North Carolina, USA — April 26 – 30. webhttp://www.ambuehler.ethz.ch/CDstore/www2010/www/p591.pdf Date of access: 13 Feb. 2016.

Lake, T. 2010. Status report: Twitter nlp. Western Michigan University, Kalamazoo.

Lakshmi, Pateriya, P.K. 2012. A study on author identification through stylometry. *International journal of computer science & communication networks*, 2(6):653-657.

Lancashire, I. 1999. Probing Shakespeare’s idiolect in *Troilus and Cressida*. *University of Toronto quaterly*, 68(3):728.

Lapa-uitgewers & CTexT. 2015. NWU/Lapa-korpus 1.0. Potchefstroom: CTexT, Noordwes-Universiteit.

Larner, S. 2014a. A preliminary investigation into the use of fixed formulaic sequences as a marker of authorship. *International journal of speech, language and the law*, 21(1):1-22.

Larner, S. 2014b. Forensic authorship analysis and the word wide web. Basingstoke: Palgrave Pivot.

Larner, S. 2015. From intellectual challenges to established corpus techniques: introduction to the special issue on forensic linguistics. *Corpora*, 10(2)131-143. http://www.euppublishing.com/journal/cor Date of access: 20 Nov. 2015.

Lawrence, M.S. & Solan, P.M. 2004. Author identification in American courts. *Applied linguistics*, 25(4):448-465.

Layton, R., Watters, P. & Dazeley, R. 2010. Authorship attribution for twitter in 140 characters or less. 2010 Second Cybercrime and Trustworthy Computing Workshop. http://dl.acm.org/citation.cfm?id=1901988 Date of access: 12 Feb. 2016.

Ledger, G. & Merriam, T. 1994. Shakespeare, Fletcher, and the two noble kinsmen. *Literary and linguistic computing*, (9)3:235-248.

Lee, D.Y.W. 2010. What corpora are available? (*In* O'Keeffe, A. & McCarthy, M., *eds*. The Routledge handbook of corpus linguistics. London: Routledge Taylor and Francis Group. p. 107-121).

Leech, G. 2011. Principles and applications of corpus linguistics. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds*. Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 155-170).

Lemahieu, R., Van Canneyt, S., De Boom, C. & Dhoedt, B. 2015. Optimizing the popularity of twitter messages through user categories. IEEE International Conference on Data Mining Workshop (ICDMW). http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=7395833 Date of access: 31 Aug. 2016.

Leonard, A.R. 2005. Forensic linguistics: applying the scientific principles of language analysis to issues of the law. *International journal of the humanities*. (3)7:1447-9559.

Levon, E. 2010. Organizing and processing your data: the nuts and bolts of quantitative analyses. (*In* Litosseliti, L., *ed.* Research methods in linguistics. New York: Continuum International Publishing Group. p. 68-92).

Lexical Analysis Software & Oxford University Press. 2013. WordSmith Version 5.0. http://www.lexically.net/wordsmith/

Lilley, P. 2002. Hacked, attacked and abused: Digital Crime Exposed. London: Kogan Page Limited.

Lindqvist, M.E. 2011. Trickin', wanna and ain't: gender differences in the use of vernacular verb forms on twitter. http://www.diva-portal.org/smash/get/diva2:431162/FULLTEXT02.pdf Date of access: 14 Aug. 2014.

Lockyer, D. 2014. The emotive meanings and functions of English 'diminutive' interjections in Twitter posts. *SKASE Journal of theoretical linguistics*, 11(2):68-89. http://eds.a.ebscohost.com.nwulib.nwu.ac.za/eds/pdfviewer/pdfviewer?vid=4&sid=b3697a0b-8ea0-4b22-903c-e15c5d0e2ab1%40sessionmgr4006 Date of access: 25 Sept. 2017.

Lombard, H. 2016. Steekproefneming. (*In* Joubert, I., Hartell, C. & Lombard. K., *reds.* Navorsing: 'n gids vir die beginnernavorser. p. 93-108).

Louwerse, M.M. 2004. Semantic variation in idiolect and sociolect: corpus linguistic evidence from literary texts. *Computers and the humanities*, (38)2:207-221.

Love, H. 2002. Attributing authorship: an introduction. New York: Cambridge University Press.

Lüdeling, A. & Kytö, M. *eds*. 2008. Corpus linguistics: an international handbook. Volume 1. New York: Walter de Gruyter.

Lutz, W.D. 1992. Jargon. (*In* McArthur, T., *ed.* The Oxford companion to the English languages. Oxford: Oxford University Press. p. 543-544).

MacLeod, N. 2013. Forensic linguistics: overview. (*In* Chapelle, C.A., *ed.* The encyclopedia of applied linguistics. Oxford: Blackwell Publishing Ltd. p. 2160-2166).

MacLeod, N. & Grant, T. 2012. Whose tweet? Authorship analysis of micro-blogs and other short-form messages. Proceedings of the International Association of Forensic Linguists' tenth biennial conference. Aston University.

Malan, J.J. 1998. Oor die hofnotuleringstaal in die lig van die grondwet en na aanleiding van onlangse regspraak. *Tydskrif vir hedendaagse Romeins-Hollandse reg*, (61)4:696-705.

Malan, J.J. 2003. Die gebruik van Afrikaans as notulering van hofverrigtinge gemeet aan demokratiese standaarde. *Tydskrif vir regswetenskap*, 28(1):36-58.

Malathi, A. 2015. Linguistics: the study of language. *International journal of linguistics and computational applications (IJLCA)*, 2(1):30-32. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.5448&rep=rep1&type=pdf Date of access: 31 Aug. 2016.

Maree, K. & Pietersen, J. 2011. Sampling. (*In* Maree, K., *ed.* First steps in research. Pretoria: Van Schaik. p. 172-181).

Maroela Media en CTexT. 2015. NWU-Maroela Media-korpus 1.0. Potchefstroom: CTexT, Noordwes-Universiteit.

Martin, B. & Ringham, F. 2000. Dictionary of semiotics. New York: Cassell.

Matras, Y. 2009. Language contact. Cambridge: Cambridge University Press.

McEnery, T. & Wilson, A. 2001. Corpus linguistics. Edinburgh: Edinburgh University Press.

McEnery, T. & Hardie, A. 2012. Corpus linguistics: method, theory and practice. New York: Cambridge University Press.

McEnery, T., Xiao, R. & Tono, Y. 2006. Corpus-based language studies an advanced resource book. New York: Routledge Taylor & Francis Group.

McCarthy, M. & O'Keeffe, A. 2010. Historical perspective: what are corpora and how have they evolved. (*In* O'Keeffe, A. & McCarthy, M., *eds*. The Routledge handbook of corpus linguistics. London: Routledge Taylor and Francis Group. p. 3-14).

McKay, S. 2005. Sociolinguistics and second language learning. (*In* Hinkel, E., *ed.* Handbook of research in second language teaching and learning. London: Lawrence Erlbaum Associates Publishers. p. 288-291).

McMenamin, G.R. 1993. Forensic stylistics. London: Elsevier.

McMenamin, G. 2001. Style markers in authorship studies. *International journal of speech, language and the law*, 8(2):93-97.

McMenamin, G.R. 2002. Forensic linguistics: advances in forensic linguistics. California: CRC Press.

Meinard, M.E.M. 2014. Distinguishing onomatopoeias from interjections. *Journal of pragmatics*, (75):150-168. https://ac-els-cdn-com.nwulib.nwu.ac.za/S0378216614002422/1-s2.0-S0378216614002422-main.pdf?_tid=14e8335c-a21f-11e7-bacd-00000aacb35e&acdnat=1506364184_a29f2ae114c88d941b29b9b38182bfed Date of access: 25 Sept. 2017.

Mello, H. 2014. Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. (*In* Raso, T. & Mello, H., *eds*. Spoken Corpora and Linguistic Studies. Philidelphia: John Benjamins Publishing Company. p. 27 – 68).

Mey, J.L. 2009a. Adaptability in human-computer interaction. (*In* Mey, J.L., *ed.* Concise encyclopidia of pragmatics. Amsterdam: Elsevier. p. 8-12).

Mey, J.L. 2009b. Pragmatics: overview. (*In* Mey, J.L., *ed.* Concise encyclopidia of pragmatics. Amsterdam: Elsevier. p. 786-797).

Meyer, C. 2002. English corpus linguistics: an Introduction. Cambridge: Cambridge University Press.

Meyer, C.F. 2004. English corpus linguistics: an introduction. Cambridge: Cambridge University Press.

Meyer, C.F. 2008. Pre-electronic corpora. (*In* Lüdeling, A. & Kytö, M., *eds*. Corpus linguistics: an international handbook. Volume 1. New York: Walter de Gruyter. p. 1-14).

Michell, C.S. 2013. Investigating the use of forensic stylistics and stylometric techniques in the analyses of authorship on a publicly accessible social networking site (Facebook). Pretoria: UNISA. (Dissertation − MA).

Millican, P. 2003. Signature Stylometric System 1.0 http://www.signature.com Date of access: 10 Jun. 2013.

Moeketsi, R.H. 1997. Of African languages and forensic linguistics: the South African multicultural criminal courtroom. Pretoria: UNISA (Thesis – PhD).

Moeketsi, R.H. 1999. Discourse in a multilingual and multicultural courtroom: a court interpreter's guide. Pretoria: JL van Schaik.

Moeketsi, R.H. 2000. The do's and don'ts in court interpreting: a functional approach to a professional code. *Language matters*, 31(1):222-242.

Mollin, S. 2009. "I entirely understand" is a Blairism: the methology of identifying idiolect alcollocations. *International journal of corpus linguistics*, 14(3):367-392.

Morris, T. 2009. All a twitter: a personal and professional guide to social networking with twitter. Indianapolis: Pearson Education.

Mosteller, F. & Wallace, D.L. 1965. Interference in authorship problem: a comparative study of discriminating methods applied to the authorship of the disputed *Federalist* papers. *Journal of the American statistical association*, 58(203):275-309. https://www.stat.cmu.edu/Exams/mosteller.pdf Date of access: 24 Mrt. 2016.

Müller, D. 2007. Skryf Afrikaans van A tot Z. Kaapstad: Pharos

Muralidharan, S., Rasmussen, L., Patterson, D. & Shin, J.H. 2011. Hope for Haiti: an analysis of facebook and twitter usage during the earthquake relief efforts. *Public relations review*, 37(2):175-177.

Myers-Scotton, C. 2001. The matrix language frame model: developments and responses. (*In* Jacobson, R., ed. Codeswitching worldwide II. New York: Mouton de Gruyter. p. 23-58).

Namba, K. 2004. An overview of Myers-Scotton's Matrix Language Frame model. *Educational Research Bullitin Senri International School*. http://yayoi.senri.ed.jp/research/re09/namba.pdf Date of access: 25 Sept. 2017.

Naga Prasad, S., Narsimha, V.B., Reddy, P.V. & Babu, A.V. 2015. Influence of lexical, syntactical and structural features and their combination on authorship attribution for telugu text. *Procedia computer science*, 48:58-64. http://www.sciencedirect.com/science/article/pii/S1877050915006195 Date of access: 11 Jun. 2017.

Naone, E. 2008. Maintaining multiple personas online: a new site let users create profiles for the different sides of the personality. *Technology review*. Feb. http://www.technologyreview.com?Infotech/20183/page1/?a=f Date of access: 05 Sept. 2014.

Nel, M. & Jordaan, C. 2016. Die gemengde navorsingsmetode. (*In* Joubert, I., Hartell, C. & Lombard. K., *reds.* Navorsing: 'n gids vir die beginnernavorser. p. 377-394).

Nesselhauf, N. 2005. Corpus linguistics: a practical introduction. http.//www.as.uniheidel berg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf Date of access: 20 Apr. 2014.

Newman, P. 2010. The Routledge handbook of forensic linguistics. *The international journal of speech, language and the law*, 18(1):161-167.

Nguyen, D., Gravel, R., Trieschnigg, D. & Meder, T. 2013. "How old do you think I am?": a

study of language and age in twitter. Procceedings of the seventh International AAAI conference on weblogs and social media. https://pure.knaw.nl/portal/en/publications/how-old-do-you-think-i-am-a-study-of-language-and-age-in-twitter(88d77722-bde1-4d83-b7af-78b2b3136bdb).html Date of access: 31 Aug. 2016.

Nienaber, P.J., Senekal, J.H. & Bothma, T.C. 1963. Mylpale in die geskiedenis van die Afrikaanse letterkunde. Johannesburg: Afrikaanse Pers-Boekhandel.

Nini, A. 2014. Authorship profiling in a forensic context. Birmingham: Aston University (Thesis − PhD).

Nirkhi, S., Dharaskar, R.V. & Thakare, V.M. 2015. Authorship identification in digital forensics using machine learning approach. *International journal of latest trends in engineering and technology*, 5(1):369-372.

Oakes, M.P. 1998. Statistics for corpus linguistics. Edinburg: Edinburgh University Press.

Oakes, M.P. 2014. Literary detective work on the computer. Amsterdam: John Benjamins Publishing Company.

Odendal, F.F. & Gouws, R.H. 2015. Handwoordeboek van die Afrikaanse taal. 6de uitg. Kaapstad: Pearson.

Olivier, J. 2013. Die mate van konsekwentheid in SMS-Afrikaans. *LitNet akademies*, 10(2):479-505.

Olsson, J. 2005. What is forensic linguistics? www.thetext.co.uk/docs/what_is.doc Date of access: 28 Oct. 2016.

Olsson, J. 2008. Forensic linguistics. 2nd ed. Londen: Continuum International Publishing Group.

Olsson, J. 2009. Wordcrime: solving crime through forensic linguistics. London: Continuum.

Olsson, J. 2012. Some aspects of the use of corpora in forensiclinguistics. (*In* Hyland, K., Huat, C.M. & Handford, M., eds. Corpus applications in applied linguistics. New York: Continuum. p. 84-100).

O'Keeffe, A., McCarthy, M.J. & Carter, R.A. 2007. From corpus to classroom. Cambridge: Cambridge University Press.

O'Reilly, T. & Milstein, S. 2009. The twitter book. San Francisco: O'Reilly Media.

O'Reilly, T. & Battelle, J. 2009. Web Squared: Web 2.0 Five Years On. Web 2.0 summit. http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf Date of access: 15 Oct. 2014.

Oulasvirta, A., Lehtonen, E., Kurvunen, E. & Raento, M. 2009. Making the ordinary visible in microblogs. *Pers ubiquit comput.* https://pdfs.semanticscholar.org/b0e0/4806367ad13f819838ab041eff1d28c2cb12.pdf. Date of access: 25 Feb. 2016.

Palfrey, J. & Gasser, U. 2008. Born digital understanding the first generation of digital natives. New York: Basic Books.

Panicheva, P., Cardiff, J. & Rosso, P. 2010. Personal sense and idiolect: combining authorship attribution and opinion analysis. Proceedings 7th International Conference on Language Resource and Evaluation, LREC-2010, May 17-23, Malta. http://users.dsic.upv.es/~prosso/resource/PanichevaEtAl_LREC10.pdf Date of acess: 17 Nov. 2015.

Papworth, S.K., Nghiem, T.P.L., Chimalakonda, D., Posa, M.R.C., Wijedasa, L.S., Bickford, D. & Carrasco, L.R. 2015. Quantifying the role of online news in linking conservation research to facebook and twitter. http://onlinelibrary.wiley.com/doi/10.1111/cobi.12455/full Date of access: 31 Aug. 2015.

Passant, A., Hastrup, T., Bojars, U., & Breslin, J. 2008. Microblogging: a semantic and distributed approach. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.8430&rep=rep1&type=pdf Date of access: 20 Feb. 2015.

Pearson, S. & Watson, R. 2010. Digital triage forensics: processing the digital crime scene. Boston: Elsevier.

Peersman, C., Daelemans, W. & Van Vaerenberg, L. 2011. Predicting age and gender in online social networks. SMUC'11.

Perez, S. 2017. Twitter oficially expands iys character count to 280 starting today. https://techcrunch.com/2017/11/07/twitter-officially-expands-its-character-count-to-280-starting-today/ Date of access: 09 Nov. 2017.

Perry, F.L. 2005. Research in applied linguistics: becoming a discerning consumer. London: Lawrence Erlbaum Associates Publishers.

PharosAanlyn. https://www-pharosonline-co-za.nwulib.nwu.ac.za/home/ Datum van gebruik: 27 Okt. 2015.

Pilon, S. 2005. Outomatiese Afrikaanse woordsoortetikettering. Potchefstroom: NWU. (Verhandeling – MA).

Plester, B., Wood, C., & Bell, V. 2008. Txt msg n school literacy: Does texting and knowledge of text abbreviations adversely affect children's literacy attainment? *Literacy, 42*(3) 137–144.

Ponelis, F.A. 1979. Afrikaanse Sintaksis. Pretoria: J.L. van Schaik.

Ponelis, F.A. 1988. Afrikaans en taalversteuring. *Tydskrif vir geesteswetenskappe*, 28(2): 119-149.

Ponelis, F.A. 1989. Nederlands-Afrikaans: die Europese agtergrond van Afrikaans. (*In* Botha, T.J.R., *red.* Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. p. 38-71).

Ponelis, F. 2009. Die taal wat ons praat. http://www.volksblad.com/ByNuus/Die-taal-wat-ons-praat-20091218-2 Datum van gebruik: 29 Aug. 2016.

Prinsloo, S. 2002. Tekslinguistiek: van teorie tot praktyk. *Literator*, 23(2):105-127.

Prinsloo, D.J. & Prinsloo, D. 2011. WordSmit Tools. https://scholarspace.manoa.hawaii.edu/bitstream/handle/101254493/prinsloo.pdf?sequence=1 Datum van gebruik: 22 Mei 2016.

Protea Boekhuis & CTexT. 2015. PUK/Protea Boekhuis-korpus 2.0. Potchefstroom: CTexT, Noordwes-Universiteit.

Quirk, R. 1974. The linguist and the English language. London: Arnold

Radio Sonder Grense & CTexT. 2015. RSG Afrikaanse korpus 2.0. Potchefstroom: CTexT, Noordwes-Universiteit.

Ralarala, M.K. 2012. A compromise of rights, rights of language and rights to a language in Eugene Terre'Blanche's (ET) trail within a trail: evedince lost in translation. *Stellenbosch Papers in Linguistics,* 14(1):55-70.

Ralarala, M.K. 2014. Transpreters' translations of complainants' narrativess as evidence: whose version goes to court? *The Translator*, 20(3):377-395.

Rao, D., Yarowsky, D., Shreevats, A. & Gupta, M. 2010. Classifying latent user attributes in Twitter. *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*. New York: ACM. p. 37-44. https://www.cs.jhu.edu/~delip/smuc.pdf Date of access: 14 Sept. 2017.

Raidt, E.H. 1983. Einfuhrunq in geschichte und struktur des Afrikaans. Darmstadt: Wissenschaftliche Buchgesellschaft.

Raidt, E.H. 1989. Ontwikkeling van vroeë Afrikaans. (*In* Botha, T.J.R., *red.* Inleiding tot die Afrikaanse taalkunde. 2de uitg. Pretoria: Academica. p. 96-126).

Raidt, E.H. 1991. Afrikaans en sy Europese verlede. 3de uitg. Kaapstad: Nasionale Opvoedkundige Uitgewery Beperk.

Rasinger, S.M. 2010. Quantitative methods: concepts, frameworks and issues. (*In* Litosseliti, L., *ed*. Research methods in linguistics. New York: Continuum International Publishing Group. p. 49-68).

Renouf, A. 2007. Corpus development 25 years on: from super-corpus to cyber corpus. (*In* Facchinetti, R., *ed.* Corpus linguistics 25 years on. New York: Rodopi. p. 27-50).

Richardson, L-J. 2015. Micro-blogging and online community, internet archaeology 39. http://dx.doi.org/10.11141/ia.39.2 Date of acess: 9 Jan. 2016.

Rijhwani, S., Sequiera, R., Choundhury, M., Bali, K. & Maddila, C.S. 2017. Estimating code-switching on Twitter with a novel generalized word-level language detection technique. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics,* p. 1971-1982, Vancouver, Canada, July 30-August 4. http://www.aclweb.org/anthology/P17-1180 Date of access: 25 Sept. 2017.

Riley, J.A. 2011. Social media directory: the ultimate guide to Facebook®, Twitter®, and LinkedIn® resources. Indianapolis: Que Publishing.

Ritter, A., Clark, S., Etzioni, M. & Etzioni, M. 2011. Namedd entity recognition in tweets: an experimental study. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.* p. 1524-1534). http://delivery.acm.org/10.1145/2150000/2145595/p1524-ritter.pdf?ip=105.186.103.125&id=2145595&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=991456476&CFTOKEN=77797256&__acm__=1507017926_7fd7793bacbfde22b36c42dca628b01b Date of access: 03 Oct. 2017.

Roberge, P.T. 2003. Afrikaans. (*In* Deumert, A. & Vandenbussche, W., *eds*. Germanic standardizations. Amsterdam/Philadelphia: John Benjamins Publishing Company. p. 15-40).

Rocha, A., Schreirer, W.J., Forstall, C.W., Cavalcante, T., Theophilo, A., Shen, B., Ariadne, R.B. en Stamatatos. E. 2017. Authorship Attribution for social media forensics. *IEEE transactions on information forensics ans security,* 12(1):5-33.

Rose, P. 2002. Forensic speaker identification. New York: Taylor and Francis.

Rosen, A. 2017. Tweeting made easier. https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html Date of access: 09 Nov. 2017.

Ross, C., Terras, M., Warwick, C. & Welsh, A. 2011. Enabled backchannel: conference twitter use by digital humanists. *Journal of documentation*, 67(2):214-237.

Rowling, J. K. 2005. Harry Potter and the Half-Blood Prince. London: Bloomsbury.

Sampson, G. & McCarthy, D. *eds*. 2004. Corpus linguistics: readings in a widening discipline. London: Continuum.

Sanderson, P. 2007. Linguistic analysis of competing trademarks. *Language matters*, 38(1):132-149.

Sapir, E. 1921. Language an introduction to the study of speech. New York: Harcourt Brace.

Sardinha, T.B. 2011. Corpus linguistics in South America. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds*. Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 29-46).

Scott, M. 2011. WordSmith Tools: Version 5.0. Liverpool: Lexical Analysis Software Ltd.

Scott, M. 2012. WordSmith Tools: Version 6.0. Liverpool: Lexical Analysis Software Ltd.

Scholtz, J. du P. 1958. Die ontstaan van die Afrikaanse werkwoordelike vormstelsel. *Tydskrif vir wetenskap en kuns*, 18(1):61-99.

Schwarts, R., Tsur, O., Rappoport, A. & Koppel, M. 2013. Authorship identification of micro-messages. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, p. 1880-1891. Seattle, Washington.

Shezi, L. 2015. What do SA's 24.9 mil internet users spend most of their time doing online? http://www.htxt.co.za/2015/01/21/what-do-sas-24-9-mil-internet-users-spend-most-of-their-time-doing-online/ Date of access: 15 Feb. 2015.

SCSA (Supreme Court of South Africa). 2004. Judgement in the matter between The Stte and Luzuko Kerr Hoho. Case CC 132/02, p. 2814-2896 (Judge C.S. White).

Smith, S.S. & Spuy, R.W. 2002. PsycholinguisticsL Using Language Analysis for Identifying and Assessing Offenders. *71 FBI Law Enforcement Bullitin,* 16:16-21

Siegel, J. 2008. The emergence of pidgin and creole languages. New York: Oxford University Press.

Silva, R.S. & Laboreiro, G. 2011. “twazn me!!!” Automatic authorship analysis of micro-blogging messages. *Natural language processing and information systems*, 67(16):161-168.

Simpson, R.C. 2004. Stylistic features of academic speech: the role of formulaic expressions. (*In* Connor, U. & Upton, T., *eds*. Discourse in the professions: perspectives from corpus linguistics. Amsterdam: John Benjamins. p. 37-64).

Sinclair, J. 1991a. Corpus, concordance, collocation. Oxford: Oxford University Press.

Sinclair, J. 1991b. Shared knowledge. Georgetown University Round Table on Languages and Linguistics. 489-500. Washington D.C.: Georgetown University Press.

Sinclair, J. 1996. Preliminary recommendations on text typology. EAGLES Document EAG-TCWG-TTYP/P. http://www.ilc.pi.cnr.it/EAGLES96/texttyp/texttyp.html Date of access: 21 Jul. 2016.

Sinclair, J. 2004a. Trust the text: language, corpus and discourse. London: Routledge.

Sinclair, J. 2004b. Intuition and annotation: the discussion continues. (*In* Aijmer, K. & Altenberg, J., *eds*. Advances in corpus linguistics, p. 39-60. Amsterdam: Rodopi).

Sinclair, J. 2005. Corpus and text: basic principles. (*In* Wynne, M., *ed*. Developing linguistic corpora: a guide to good practice. Oxford: Oxbow Books. p. 1-16) http://ahds.ac.uk/linguistic-corpora/).

Sinclair, J. & Rockwell, G. 2016. Voyant-tools. http://voyant-tools.org/ Date of access: 25 May 2016.

Smuts, J.P. 1998. Die roman, ’n inleidende studie. Kaapstad: Human & Rosseau.

Solan, L.M. & Tiersma, P.M. 2004. Author identification in American courts. *Applied linguistics*, 25(4):448-465.

Sousa, R., Laboreiro, G., Sarmento, L., Grant, T., Oliveira, E., & Maia, B. 2011. ‘twazn me!!! ;(’ automatic authorship analysis of micro-blogging messages. (In *NLDB'11 Proceedings of the 16th international conference on Natural language processing and information systems*, Alicante: Spain, June 28 – 30. p.161-168). http://dl.acm.org/citation.cfm?id=2026029 Date of access: 29 Oct. 2016.

Statista. 2016. Internet Demographics. https://www.statista.com/statistics/265147/number-of-worldwide-internet-users-by-region/ Date of access: 04 Jan 2016.

Statistics. 2017. What is a F statistic. http://www.statisticshowto.com/f-statistic/ Date of access: 05 Nov. 2017.

Stamatatos, E. 2009. A survey of modern authorship attribution methods. *Journal of the American society for information science and technology*, 60(3):538-556.

Starbird, K., Palen, L., Hughes, A.L. & Vieweg, S. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. CSCW. February 6-10.

Statista. 2015. Statistics and facts about twitter. https://www.statista.com/topics/737/twitter/ Date of access: 31 Aug. 2015.

Strydom, H. 2012. Sampling in the quantitative paradigm. (*In* De Vos, A.S., Strydom, H., Fouché, C.B. & Delport, C.S.L., *eds*. Research at grass roots: for the social sciences and human service professions. 4th ed. Pretoria: Van Schaik. p.222-235).

Stubbs, M. 2007. Notes on the history of corpus linguistics and emirical semantics. (*In* Nenomen, M. & Niemi, S., *eds.* Collocations of idioms. Joensuu: Joensuu Yliopsito. p. 317-329).

Suid-Afrika. 1996. Grondwet van die Republiek van Suid-Afrika 1996.

Svartvik, J. 1967. The Evans statements: a case for forensic linguistics. GöteborgL Elanders Boktryckeri Aktibolag. http://www.thetext.co.uk/Evans%20Statements%20Part%201.pdf Date of access: 23 Mrt. 2016.

Spuy, R.W. 1993. Language Crimes: The use and Abuse of Language Evidence in the Courtroom. Oxford: Blackwell.

Spuy, R.W. 2007. Language in the American Courtroom. *Language and Linguistics Compass,* 1(1-2):100-114.

Spuy, R.W. 2010. The Language of Defamation Cases. Oxford: Oxford University Press.

Spuy, R.W. 2014. The Language of Murder Cases: Intentionality, Predisposition, and Voluntariness. Oxford: Oxford University Press.

Svartvik, J. 2007. Corpus linguistics 25+ years on. (*In* Facchinetti, R., *ed.* Corpus linguistics 25 years on. New York: Rodopi. p. 11-26).

Swales, J. 2011. A critical view on the use of corpora. (*In* Viana, V., Zyngier, S. & Barnbrook, G., *eds*. Perspectives on corpus linguistics. Philadelphia: John Benjamins Publishing Company. p. 220-229).

Taalkommissie (van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns). 2017. 11de uitg. Afrikaanse Woordelys en Spelreëls. Kaapstad: Pharos.

Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2011. Taalkommissiekorpus 1.1. Potchefstroom: CTexT, Noordwes-Universiteit.

Tannen, D. 2013. The Medium Is the Metamessage: Conversational Style in New Media Interaction. (*in* Tanner, D. Trester, A.M. *eds.* Discourse 2.0: Language and New Media. Washington, DC: Georgetown University Press p. 99 - 118).

Tannen, D. & Trester, A.M. *eds*. 2013. Language and new media. Washington: Georgetown University Press.

Tashakkori, A. & Creswell, J. 2007. The new era of mixed methods. *Journal of mixed methods research*, 1(1):3-8.

Tavakoli, H. 2012. A dictionary of research methodology and dtatistics in applied linguistics. Iran: Rahnama Press.

Teevan, J., Ramage, D. & Morris, M.R. 2011. #TwitterSearch: a comparison of microblog search and web search. WSDM’11, February 9-12, 2011, Hong Kong, China.

Teubert, W. 2004. Language and corpus linguistics. (*In* Halliday, M.A.K., Teubert, W., Yallop, C. & Čermáková, A., *eds*. Lexicology and corpus linguistics: an introduction. London: Continuum. p. 73-113).

Teubert, W. 2008. Some notes on the concept of cognitive linguistics. (*In* Gerbig, A. & Mason, O., *eds*. Language, people, numbers: corpus linguistics and society. New York: Rodopi. p 61-84).

Teubert, W. & Čermáková, A. 2004. Directions in corpus linguistics. (*In* Halliday, M.A.K., Teubert, W., Yallop, C. & Čermáková, A., *eds*. Lexicology and corpus linguistics: an introduction. London: Continuum. p. 113-165).

Thiart, L. 2014. Outeuridentifikasie: ʼn forensies-taalkundige ondersoek na Afrikaanse SMS-taal. Pretoria: UP. (Verhandeling – MA).

Thurlow, C. 2013. Fakebook: Synthetic Media, Pseudo-sociality, and the Rhetorics of Web 2.0 (*in* Tanner, D. Trester, A.M. *eds.* Discourse 2.0: Language and New Media. Washington, DC: Georgetown University Press p. 225 – 250).

Todd, L. 2005. Pidgins and creols. London: Routledge.

Tognini-Bonelli, E. 2001. Corpus linguistics at work. Philidelphia: John Benjamins.

Togneli-Bonelli, E. 2010. Theoretical overview of the evolution of corpus linguistics. (*In* O'Keeffe, A. & McCarthy, M., *eds*. The Routledge handbook of corpus linguistics. London: Routledge Taylor and Francis Group. p. 14-28).

Trudgill, P. 1992. Introducing language and society. London: Penguin Group.

Tsur, O. & Rappoport, A. 2012. What's in a Hashtag? Cotent-based prediction of the spread of ideas in microblogging communities. *WSDM*, '12:643-652. http://people.seas.harvard.edu/~orentsur/papers/wsdm12.pdf Date of access: 12 Oct. 2017.

Turell, M.T. 2008. Plagiarism. (*In* Gibbons, J. & Turell, T. M., *eds.* Dimensions of forensic linguistics. Philadelpia: John Benjamins Publishing Company. p. 265-299).

Twitter. 2016. https://twitter.com. Date of access: 1 Sept. 2015.

Van den Berg, M. 2000. Die konsep standaardafrikaans: ʼn kritiese oorsig en ʼn Hallidayaanse perspektief. Potchefstroom: PU vir CHO. (Ongepubliseerde doktorale proefskrif).

Van Der Merwe, H.J.J.M. 1970. Afrikaans sy aard en ontwikkeling. Pretoria: Van Schaik.

Van Heerden, E. 2000. Ontheemdes vind hul burg in kuberdorpies. *Beeld:* 4 Julie. http://www.152.111.1.88/argief/berigte/beeld/2000/07/04/4/11.html Datum van gebruik: 12 Nov. 2015.

Van Huyssteen, G.B. 2014. Morfologie (*In* Carstens, W.A.M & Bosman, N., *reds*. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik. p.171-208).

Van Niekerk, A. & Olivier, J. 2014. Pragmatiek (*In* Carstens, W.A.M & Bosman, N., *reds*. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik. p.257-310).

Van Rensburg, C. 2012. So kry ons Afrikaans. Pretoria: Lapa Uitgewers.

Van Rooy, B. 2014. Die Afrikaanse Taalkunde van toeka tot nou (*In* Carstens, W.A.M & Bosman, N., *reds*. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik. p.1-26).

Van Rooy, B. & Terblanche, L. 2006. A corpus-based analysis of involved aspects of student writing. *Language matters*, 37(2):160-182.

Van Schoor, J.L. 1983. Die Grammatika van Standaardafrikaans. Kaapstad: Lex Patria.

Van Wyk, G.J. red. 2003. Etimologiewoordeboek van Afrikaans. Stellenbosch: Buro van die WAT.

Váradi, T. 2001. The linguistic relevance of corpus linguistics. (*In* Rayson, P., Wilson, A., McEnery, T., Hardie, A. & Khoja, S., *eds.* Proceedings of the Corpus Linguistics 2001 Conference. Lancaster University: UCREL Technical Papers 13:587-593).

Varieng. 2016. Corpora in CoRD. http://www.helsinki.fi/varieng/CoRD/corpora/index.html Date of access: 12 Apr. 2016.

Venter, R.I. 2013. Die relevansie van forensies-linguistiese getuienis vir die doeleindes van persoonsidentifikasie by witboordjiemisdade. Potchefstroom: NWU. (Verhandeling – Mcom).

Verhoef, M.M. 1991. Taalbeplanning: die stand en die toekoms van taalbeplanning vir Afrikaans in ’n multitalige Suid-Afrika. Potchefstroom:NWU. (Proefskrif − PhD).

Verhoef. M. 2001. Die internet as dinamiese taalorngewing: taalveranderingsverskynsels in internetafrikaans. *Literator*, 22(3):1-19.

Viégas, F.B. 2005. Blogger’s expectations of privacy and accountability: an initial survey. *Journal of computer-mediated communication*, 10(3). http://onlinelibrary.wiley.com/doi/ 10.1111/j.1083-6101.2005.tb00260.x/full Date of access: 1 Sept. 2015.

Vikner, S. 1995. Verb movement and expletive subjects in the Germanic languages. New York: Oxford University Press.

Wang, W., Chen, L., Thirunarayan, K. & Sheth, A.P. 2014. Cursing in English on Twitter. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*:415-424. http://corescholar.libraries.wright.edu/knoesis/590 Date of access: 6 Sept. 2017.

WatKykJy. 2015. Watkykjy.co.za-korpus 1.0. Beskikbaar by: http://viva-afrikaans.org.

Webb, V.N. 1989. Die Afrikaanse variasietaalkunde. (*In* Botha, T.J.R., *red.* Inleiding tot die Afrikaanse taalkunde. Pretoria: Academica. p. 412-435).

Werry, C.C. 1996. Linguistic and interactional features of internet relay chat. (*In* Herring, S.C., *ed.* Computer-mediated communication. Lingistic, social and cross cultural perspectives. Phllidelphia: John Benjamins Publishing. p.130).

Williams, S.A., Terras, M. & Warwick, C. 2013. What people study when they study twitter: classifying twitter related academic papers. *Journal of documentation*, 69(3):1-74. http://discovery.ucl.ac.uk/13691572/2/WhatPeopleStudyWhenTheyStudyTwitter.pdf Date of access: 25 Aug. 2014.

Wilson, A., Archer, D. & Rayson, P. 2006. Corpus linguistics around the world. Amsterdam: Rodopi.

Winter, E. 1996. The statistics of analysing very short texts in a criminal context. (*In* Kniffka, H., Blackwell, S. & Coulthard, M., *eds*. Recent developments in forensic linguistics. Frankfurt am Main: Peter Lang. p. 141-180).

Wissing, D. 2014. Fonetiek. (*In* Carstens, W.A.M & Bosman, N., *reds*. Kontemporêre Afrikaanse taalkunde. Pretoria: Van Schaik. p.91-126).

Whiteman, N. 2012. Understanding ethics: rethinking practice in online research. New York: Springer.

Woods, A.F, Fletcher, P. & Hughes, A. 1989. Statistics in language studies. New York: Cambridge University Press.

Wood, A.F. & Smith, M.J. 2005. Online communication: linking technology, identity, and culture. New Jersey: Lawrence Erlbaum Associates Publishers.

World Wide Worx. 2016. Instagram, facebook, make waves in SA. http://www.worldwideworx.com/socialsa2016/ Date of access 31 Aug. 2016.

Wu, C.H., Lee, C. & Liang, H. 2009. Idiolect extraction and generation for personalized speaking style modeling. *Transactions on audio, speech, and language processing*, 17(1):127-137.

Wu, S., Hofman, J.M., Mason, W.A. & Watts, D.J. 2011. Who says what to whom on twitter. WWW 2011, March 28-April 1, 2011, Hyderabad, India.

Wunderlich, C. 2012. Empirical methods in discourse analysis and the internet. *Topics in linguistics: approaches to text and discourse analysis*, 9:42-45. http://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=13377590&AN=83800759&h=IPwveEuhpaoHFVV2IkVCKqIz%2bWsaLJ8BEL521XFd1a7sBHH2keH%2fH6fWJp1qQxmmde%2fuZGLNutJKZFzBZ%2b5PIw%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d13377590%26AN%3d83800759 Date of access: 30 Aug. 2015.

Wybenga, D.M. 1987. Diskoersanalise en Stilistiek: ŉ Inleiding. Pretoria: Serva-Uitgewers.

Yi, K., Choi, N. & Kim, Y.S. 2016. A content analysis of Twitter hyperlinks and their application in web resource indexing. *Journal of the Association for Information Science & Technology*, 67(8):1808-1821. http://onlinelibrary.wiley.com.nwulib.nwu.ac.za/doi/10.1002/asi.23508/epdf Date of access: 30 Sept. 2017.

Zappavigna, M. 2015. Searchable talk: the linguistic functions of hashtags. *Social Semiotics*, 25(3):274-291. http://eds.a.ebscohost.com.nwulib.nwu.ac.za/eds/pdfviewer/pdfviewer?vid=1&sid=bdab598a-e1bb-47cf-b647-e94af45bdc11%40sessionmgr4006 Date of access: 26 Sept. 2017.

Zeevat, H. 2009. Pragmatics: optimality theory. (*In* Mey, J.L., *ed.* Cocise encyclopidia of pragmatics. Amsterdam: Elsevier. p. 782-786).

Bylaag A: Etiekklaring

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Private Bag X6001, Potchefstroom,
South Africa, 2520

Tel: (018) 299-4900
Faks: (018) 299-4910
Web: http://www.nwu.ac.za

Institutional Research Ethics Regulatory Committee
Tel: +27 18 299 4849
Email : Ethics@nwu.ac.za

2016/03/31

ETHICS APPROVAL CERTIFICATE OF PROJECT

Based on approval by the **NWU Ethics Committee of the Language Matters (EC-LM)** on **06/04/2016**, the North-West University Institutional Research Ethics Regulatory Committee (NWU-IRERC) hereby **approves** your project as indicated below. This implies that the NWU-IRERC grants its permission that, provided the special conditions specified below are met and pending any other authorisation that may be necessary, the project may be initiated, using the ethics number below.

Project title: **Outeurskapidentifikasie en mikroblogs: ’n eksploratiewe forensiese analise in ’n digitale era**

Project Leader/Supervisor: **Prof. J.A.K. Olivier**
Student: **Ms M Nel**

Ethics number:	N	W	U	-	0	0	1	4	2	-	1	6	-	A	8
	Institution				Project Number						Year			Status	

Status: S = Submission; R = Re-Submission; P = Provisional Authorisation; A = Authorisation

Application Type: N/A
Commencement date: 2016-01-01 **Expiry date:** 2016-12-31 **Risk:** **N/A**

Special conditions of the approval (if applicable):

- Translation of the informed consent document to the languages applicable to the study participants should be submitted to the EC-LM (if applicable).
- Any research at governmental or private institutions, permission must still be obtained from relevant authorities and provided to the EC-LM. Ethics approval is required BEFORE approval can be obtained from these authorities.

General conditions:

While this ethics approval is subject to all declarations, undertakings and agreements incorporated and signed in the application form, please note the following:

- *The project leader (principle investigator) must report in the prescribed format to the NWU-IRERC via EC-LM:*
  - *annually (or as otherwise requested) on the progress of the project, and upon completion of the project*
  - *without any delay in case of any adverse event (or any matter that interrupts sound ethical principles) during the course of the project.*
  - *Annually a number of projects may be randomly selected for an external audit.*
- *The approval applies strictly to the protocol as stipulated in the application form. Would any changes to the protocol be deemed necessary during the course of the project, the project leader must apply for approval of these changes at the EC-LM. Would there be deviated from the project protocol without the necessary approval of such changes, the ethics approval is immediately and automatically forfeited.*
- *The date of approval indicates the first date that the project may be started. Would the project have to continue after the expiry date, a new application must be made to the NWU-IRERC via EC-LM and new approval received before or on the expiry date.*
- *In the interest of ethical responsibility the NWU-IRERC and EC-LM retains the right to:*
  - *request access to any information or data at any time during the course or after completion of the project;*
  - *to ask further questions, seek additional information, require further modification or monitor the conduct of your research or the informed consent process.*
  - *withdraw or postpone approval if:*
    - *any unethical principles or practices of the project are revealed or suspected,*
    - *it becomes apparent that any relevant information was withheld from the EC-LM or that information has been false or misrepresented,*
    - *the required annual report and reporting of adverse events was not done timely and accurately,*
    - *new institutional rules, national legislation or international conventions deem it necessary.*
- *EC-LM can be contacted for further information via Bertus.VanRooy@nwu.ac.za or 018 210 3482.*

The IRERC would like to remain at your service as scientist and researcher, and wishes you well with your project. Please do not hesitate to contact the IRERC or EC-LM for any further enquiries or requests for assistance.

Yours sincerely

Prof LA
Du Plessis

Digitally signed by Prof LA Du Plessis
DN: cn=Prof LA Du Plessis, o=North-West University, ou=Campus Rector, email=Linda.DuPlessis@nwu.ac.za, c=ZA
Date: 2016.04.18 09:59:03 +02'00'

Prof Linda du Plessis
Chair NWU Institutional Research Ethics Regulatory Committee (IRERC)

Bylaag B: Brief om toestemming vir gebruik van Twitter-data

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOMKAMPUS

Privaat sak X6001, Potchefstroom
Suid-Afrika 2520

Tel: 018 299-1111/2222

Navorsingseenheid vir Tale en Literatuur in die Suid-Afrikaanse Konteks

Promotor: Prof Jako Olivier
Tel: 018 285-2078
E-pos: Jako.Olivier@nwu.ac.za

Geagte

TOESTEMMING: NAVORSING RONDOM TWITTER-AFRIKAANS

Met hierdie brief word u versoek om deel te neem aan navorsing rondom Afrikaans op Twitter. U terugvoer sal van onskatbare waarde wees. Hierdie navorsing vorm deel van ’n projek, Outeurskapidentifikasie en mikroblogs: ’n eksploratiewe forensiese analise in ’n digitale era. In hierdie verband is die volgende etieknommer deur die Noordwes-Universiteit aan hierdie studie toegeken: NWU-00142-16-A8.

Wees verseker dat u response vertroulik hanteer sal word en dat alle vrae en refleksies slegs vir ontledingsdoeleindes is. Vir die doel van die studie sal twiets oor ’n bepaalde tydperk versamel en geanaliseer word. Die algehele resultate is per versoek, na afloop van die studie beskikbaar (e-pos: mariskanel28@gmail.com). Deelname aan die navorsing is ten volle vrywillig en u sal op geen manier benadeel word indien u deelneem of nie deelneem nie. U kan ook op enige tydstip aan die navorsing onttrek.

Indien u instem om aan hierdie navorsing deel te neem, voltooi asseblief die vorm hieronder.

Baie dankie vir u aandag in die verband.

Outeurskapidentifikasie en mikroblogs: ’n eksploratiewe forensiese analise in ’n digitale era

____

Vriendelike groete

Me M Nel
Navorser

Prof JAK Olivier
Promotor

Ek ____________________________________ (naam en van) gee hiermee toestemming dat my twiets vir navorsingsdoeleindes gebruik mag word.

______________________________ ________________

Handtekening Datum

Bylaag C: Sleutelwaardes

Die onderstaande Tabelle is 'n voorbeeld van hoe die sleutelwaardes bereken is.

O10_1 en O10_2						
Sleutelwoord	Frek.	%	Verw. Frek.	Verw. %	Sleutelwaarde	P
OM	11	0,29	31	0,82	8,81	0,00
HULLE	18	0,47	4	0,11	7,59	0,01
TEEN	11	0,29	1	0,03	6,68	0,01
TE	14	0,37	31	0,82	5,86	0,02
WORD	15	0,39	31	0,82	5,05	0,02
HIERDIE	2	0,05	11	0,29	5,00	0,03
VIR	100	2,63	71	1,88	4,45	0,03
DEUR	23	0,60	40	1,06	4,23	0,04
				Gemiddeld	5,96	

Die bostaande tabel dui die sleutelwaardes aan van O10 se interne vergelyking. Hierdie vergelyking is vir elk van die korpora uitgevoer ten einde te verseker dat die kopus van 'n enkele Twitter-outeur wel genoegsame ooreenkomste toon dat daar sonder redelike twyfel aangevoer kan word dat die teks deur dieselfde individu geskryf is.

Daarna is die tien outeurs vanuit die gespesialiseerde korpus se tekste met mekaar vergelyk en die bevindinge is in tabelvorm aangedui.

O1 en O3						
Sleutelwoord	Frek.	%	Verw. Frek.	Verw. %	Sleutelwaarde	P
EK	12	0,37	383	4,22	114,38	0,00
MY	7	0,21	243	2,68	72,54	0,00
JY	5	0,15	194	2,14	58,66	0,00
NA	25	0,76	11	0,12	31,91	0,00
IN	84	2,56	105	1,16	30,62	0,00
AS	13	0,40	134	1,48	22,99	0,00
VAN	55	1,68	67	0,74	20,77	0,00
KAN	4	0,12	80	0,88	19,47	0,00
SO	2	0,06	62	0,68	16,91	0,00
DEUR	12	0,37	5	0,06	14,75	0,00
				Gemiddeld	40,30	

In die tabelle wat die berekening van die sleutelwaardes aandui, is die kolom "sleutelwaardes" of rooi, of groen. Die rooi tabelle dui daarop dat die tekste deur verskillende outeurs geskryf is, terwyl die groen kolomme daarop dui dat die gebruik van die bepaalde word tot so 'n mate

ooreenstem dat daar sonder redelike twyfel aangevoer kan word dat die tekste wat vergelyk is deur dieselfde outeur geskryf is.

Laastens is die drie verdagte tekste ook met die tekste vanuit die gespesialiseerde korpus vergelyk.

V1_O3

Sleutelwoorde	Frekwensie	%	Verwysingskorpus frekwensie	Verwysingskorpus %	Sleutelwaarde	P
EK	383	4,22	12	0,37	164,31	0,000
MY	243	2,68	7	0,21	106,11	0,000
JY	194	2,14	5	0,15	87,20	0,000
JOU	108	1,19	0	0,00	66,97	0,000
NIE	355	3,91	42	1,28	64,07	0,000
SE	27	0,30	53	1,62	55,38	0,000
NET	96	1,06	5	0,15	32,88	0,000
DIS	48	0,53	0	0,00	29,68	0,000
AS	134	1,48	13	0,40	29,56	0,000
NA	11	0,12	25	0,76	28,89	0,000
				Totaal	66,51	

Die bevindinge is dan in ŉ enkele tabel geplaas.

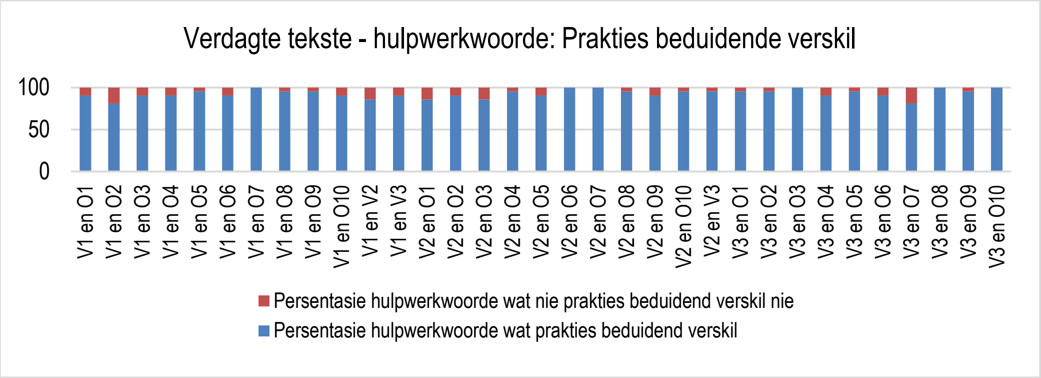
Vergelyking	Sleutelwaarde
V1_V2	19,71
V1_V3	20,28
V1_O1	18,23
V1_O2	0,00
V1_O3	66,51
V1_O4	78,36
V1_O5	13,92
V1_O6	30,99
V1_O7	28,16
V1_O8	44,85
V1_O9	17,98
V1_O10	140,67

Bylaag D: Ratio’s

Die onderstaande tabelle is voorbeelde van die aanvanklike ratio-berekeninge.Hierdie berekeninge is vir elk van die funksiewoord-groepe (lidwoorde, voegwoorde, hulpwerkwoorde, voornaamwoorde en voorsetsels) uitgevoer.

Woord	begin	behoort	bly	gaan	het	hoef	is
Ratio V3 en O1	0,00	0,00	0,00	0,59	1,24	4,34	1,60
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O2	0,00	0,00	0,00	0,94	1,34	0,00	2,41
	Verskillend	Verskillend	Verskillend	Dieselfde	Verskillend	Verskillend	Verskillend
Ratio V3 en O3	0,00	0,00	0,00	0,71	1,37	0,00	2,11
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O4	0,00	0,00	0,00	0,75	1,78	5,75	2,41
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O5	0,00	0,00	0,00	0,54	1,39	11,11	1,21
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O6	0,00	0,00	0,00	1,30	1,40	39,66	2,87
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O7	0,00	0,00	0,00	0,98	1,39	0,00	2,63
	Verskillend	Verskillend	Verskillend	Dieselfde	Verskillend	Verskillend	Verskillend
Ratio V3 en O8	0,00	0,00	0,00	5,97	3,48	0,00	6,72
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O9	0,00	0,00	0,00	1,49	1,25	5,97	1,69
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend
Ratio V3 en O10	0,00	0,00	0,00	1,21	2,92	0,00	1,80
	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend	Verskillend

In die voorbeeld word V3 se hulpwerkwoord gebruik met die gespesialiseerde korpus vergelyk. Vanuit die analise is die onderstaande afgelei.



Die berekeninge, soos in die vorige tabelle uitgebeeld is vir elk van die funksiewoord-groepe uitgevoer, en die bevindinge is soos volg gestoor.

Ratiobeplaings: Voornaamwoorde in die gespesialiseerde korpus

	O1 en O2	O1 en O3	O1 en O4	O1 en O5	O1 en O6	O1 en O7	O1 en O8	O1 en O9	O1 en O10
Verskillend	77	84	83	69	80	81	82	70	83
Dieselfde	9	2	3	17	6	5	4	16	3
Persentasie voornaamwoorde wat prakties beduidend verskil	89,5	97,7	96,5	80,2	93,0	94,2	95,3	81,4	96,5
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	10,5	2,3	3,5	19,8	7,0	5,8	4,7	18,6	3,5

	O2 en O3	O2 en O4	O2 en O5	O2 en O6	O2 en O7	O2 en O8	O2 en O9	O2 en O10	O3 en O4
Verskillend	79	79	74	72	80	80	76	83	81
Dieselfde	7	6	11	13	6	6	10	3	5
Persentasie voornaamwoorde wat prakties beduidend verskil	91,9	91,9	86,0	83,7	93,0	93,0	88,4	96,5	94,2
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	8,1	7,0	12,8	15,1	7,0	7,0	11,6	3,5	5,8

	O3 en O5	O3 en O6	O3 en O7	O3 en O8	O3 en O9	O3 en O10	O4 en O5	O4 en O6	O4 en O7
Verskillend	83	82	81	78	84	82	78	73	81
Dieselfde	3	4	5	8	2	4	8	13	5
Persentasie voornaamwoorde wat prakties beduidend verskil	96,5	95,3	94,2	90,7	97,7	95,3	90,7	84,9	94,2
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	3,5	4,7	5,8	9,3	2,3	4,7	9,3	15,1	5,8

	O4 en O8	O4 en O9	O4 en O10	O5 en O6	O5 en O7	O5 en O8	O5 en O9	O5 en O10	O6 en O7
Verskillend	82	80	81	79	79	81	72	85	78
Dieselfde	4	6	5	7	7	5	14	1	8
Persentasie voornaamwoorde wat prakties beduidend verskil	95,3	93,0	94,2	91,9	91,9	94,2	83,7	98,8	90,7
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	4,7	7,0	5,8	8,1	8,1	5,8	16,3	1,2	9,3

Ratiobeplaings: Voornaamwoorde in die verdagte teks

	V1 en O1	V1 en O2	V1 en O3	V1 en O4	V1 en O5	V1 en O6	V1 en O7	V1 en O8	V1 en O9	V1 en O10	V1 en V2
Verskillend	77	81	84	84	77	83	85	84	79	83	76
Dieselfde	9	5	2	2	9	3	1	2	7	3	10
Persentasie voornaamwoorde wat prakties beduidend verskil	89,5	94,2	97,7	97,7	89,5	96,5	98,8	97,7	91,9	96,5	88,4
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	10,5	5,8	2,3	2,3	10,5	3,5	1,2	2,3	8,1	3,5	11,6

	V1 en V3	V2 en O1	V2 en O2	V2 en O3	V2 en O4	V2 en O5	V2 en O6	V2 en O7	V2 en O8	V2 en O9	V2 en O10
Verskillend	76	69	78	82	75	83	83	81	81	83	85
Dieselfde	10	16	8	4	11	3	3	5	5	3	1
Persentasie voornaamwoorde wat prakties beduidend verskil	88,4	80,2	90,7	95,3	87,2	96,5	96,5	94,2	94,2	96,5	98,8
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	11,6	18,6	9,3	4,7	12,8	3,5	3,5	5,8	5,8	3,5	1,2

	V2 en V3	V3 en O1	V3 en O2	V3 en O3	V3 en O4	V3 en O5	V3 en O6	V3 en O7	V3 en O8	V3 en O9	V3 en O10
Verskillend	83	84	79	78	75	84	82	81	75	75	81
Dieselfde	3	2	7	8	11	2	4	5	11	11	5
Persentasie voornaamwoorde wat prakties beduidend verskil	96,5	97,7	91,9	90,7	87,2	97,7	95,3	94,2	87,2	87,2	94,2
Persentasie voornaamwoorde wat nie prakties beduidend verskil nie	3,5	2,3	8,1	9,3	12,8	2,3	4,7	5,8	12,8	12,8	5,8

Vanuit die tabelle is die grafieke wat in afdeling 6.3 gebruik is saamgestel.

Bylaag E: Algemene Kenmerke

In hierdie tabelle word die interne vergelykings van die gespesialiseerde korpus weergegee. Soos in afdeling 7.2.1 bespreek is daar 'n interne vergelyking tussen die tekste van elk van die outeurs in die gespesialiseerde korpus uitgevoer, ten einde te bepaal of die outeur se frekwensie van die gebruik van die algemene kenmerke (Twitter-spesifieke elemente) as sodanig beskou word dat dit as 'n kenmerk vir die identifisering van 'n mikroblog outeur beskou kan word.

Die gebruik van hutstekens in die gespesialiseerde korpus

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale hutswoorde in korpus	Maksimum hutswoorde per twiet	Hutswoorde genormaliseer tot 1 000 woorde	Gemiddelde hutswoorde per twiet
O1	669	944	8 942	24	4	2,68	0,04
O1_1	334	453	4 813	13	4	2,70	0,04
O1_2	335	491	4 129	11	1	2,66	0,03
O2	589	1 577	12 259	397	3	32,38	0,67
O2_1	302	743	6 211	136	2	21,90	0,45
O2_1	287	834	6 048	261	3	43,15	0,91
O3	206	266	3 379	37	2	10,95	0,18
O3_1	107	134	1 590	12	2	7,55	0,11
O3_2	99	132	1 789	25	2	13,97	0,25
O4	484	978	7 110	121	3	17,02	0,25
O4_1	243	467	3 388	34	3	10,04	0,14
O4_2	241	511	3 722	87	3	23,37	0,36
O5	333	556	4 579	14	2	3,06	0,04
O5_1	164	259	2 248	6	2	2,67	0,04
O5_2	169	297	2 331	8	2	3,43	0,05
O6	768	1 657	16 340	47	2	2,88	0,06
O6_1	387	858	8 167	42	2	5,14	0,11
O6_2	381	799	8 173	5	2	0,61	0,01
O7	457	790	5 655	4	1	0,71	0,01
O7_1	233	402	2 877	1	1	0,35	0,00
O7_2	224	388	2 778	3	1	1,08	0,01
O8	308	443	3 690	283	5	76,69	0,92
O8_1	158	224	1 846	125	4	67,71	0,79
O8_2	150	219	1 844	158	5	85,68	1,05
O9	275	467	4 918	16	4	3,25	0,06
O9_1	147	244	2 492	10	4	4,01	0,07
O9_2	128	223	2 426	6	1	2,47	0,05
O10	499	1 190	7 744	363	4	46,88	0,73
O10_1	235	579	3 777	229	4	60,63	0,97
O10_2	264	611	3 967	134	3	33,78	0,51

Die gebruik van verwysings na gebruikers in die gespesialiseerde korpus

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale verwysing na gebruikers in korpus	Maksimum verwysing na gebruikers per twiet	Verwysing na gebruikers per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal verwysings na gebruikers per twiet
O1	669	944	8 942	24	1	2,68	0,04
O1_1	334	453	4 813	14	1	2,91	0,04
O1_2	335	491	4 129	10	1	2,42	0,03
O2	589	1 577	12 259	28	1	2,28	0,05
O2_1	302	743	6 211	21	1	3,38	0,07
O2_1	287	834	6 048	7	1	1,16	0,02
O3	206	266	3 379	180	3	53,27	0,87
O3_1	107	134	1 590	92	2	57,86	0,86
O3_2	99	132	1 789	88	3	49,19	0,89
O4	484	978	7 110	233	6	32,77	0,48
O4_1	243	467	3 388	72	6	21,25	0,30
O4_2	241	511	3 722	161	5	43,26	0,67
O5	333	556	4 579	3	1	0,66	0,01
O5_1	164	259	2 248	2	1	0,89	0,01
O5_2	169	297	2 331	1	1	0,43	0,01
O6	768	1 657	16 340	30	2	1,84	0,04
O6_1	387	858	8 167	16	2	1,96	0,04
O6_2	381	799	8 173	14	2	1,71	0,04
O7	457	790	5 655	40	5	7,07	0,09
O7_1	233	402	2 877	21	5	7,30	0,09
O7_2	224	388	2 778	19	2	6,84	0,08
O8	308	443	3 690	299	7	81,03	0,97
O8_1	158	224	1 846	169	7	91,55	1,07
O8_2	150	219	1 844	130	6	70,50	0,87
O9	275	467	4 918	47	8	9,56	0,17
O9_1	147	244	2 492	29	8	11,64	0,20
O9_2	128	223	2 426	18	1	7,42	0,14
O10	499	1 190	7 744	904	5	116,74	1,81
O10_1	235	579	3 777	433	4	114,64	1,84
O10_2	264	611	3 967	471	5	118,73	1,78

Die gebruik van hiperskakels in die gespesialiseerde korpus

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van hiperskakels in die korpus	Hiperskakels per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hiperskakels per twiet
O1	669	944	8 942	13	1,45	0,02
O1_1	334	453	4 813	6	1,25	0,02
O1_2	335	491	4 129	7	1,70	0,02
O2	589	1 577	12 259	37	3,02	0,06
O2_1	302	743	6 211	23	3,70	0,08
O2_1	287	834	6 048	14	2,31	0,05
O3	206	266	3 379	7	2,07	0,03
O3_1	107	134	1 590	2	1,26	0,02
O3_2	99	132	1 789	5	2,79	0,05
O4	484	978	7 110	34	4,78	0,07
O4_1	243	467	3 388	19	5,61	0,08
O4_2	241	511	3 722	15	4,03	0,06
O5	333	556	4 579	0	0,00	0,00
O5_1	164	259	2 248	0	0,00	0,00
O5_2	169	297	2 331	0	0,00	0,00
O6	768	1 657	16 340	628	38,43	0,82
O6_1	387	858	8 167	328	40,16	0,85
O6_2	381	799	8 173	300	36,71	0,79
O7	457	790	5 655	16	2,83	0,04
O7_1	233	402	2 877	7	2,43	0,03
O7_2	224	388	2 778	9	3,24	0,04

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van hiperskakels in die korpus	Hiperskakels per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hiperskakels per twiet
O8	308	443	3 690	86	23,31	0,28
O8_1	158	224	1 846	37	20,04	0,23
O8_2	150	219	1 844	49	26,57	0,33
O9	275	467	4 918	3	0,61	0,01
O9_1	147	244	2 492	0	0,00	0,00
O9_2	128	223	2 426	3	1,24	0,02
O10	499	1 190	7 744	6	0,77	0,01
O10_1	235	579	3 777	4	1,06	0,02
O10_2	264	611	3 967	2	0,50	0,01

Die gebruik van leestekens aan die einde van sinne

Outeur	Aantal sinne	Leestekens gebruik aan die einde van ’n sin	Frekwensie van gebruik van leesteken aan die einde van 'n sin	Leestekens uitgelaat aan die einde van ’n sin	Frekwensie van die uitlating van leestekens aan die einde van 'n sin
O1	944	917	971,40	27	28,60
O1_1	453	437	964,68	16	35,32
O1_2	491	480	977,60	11	22,40
O2	1 577	1 499	950,54	78	49,46
O2_1	743	705	948,86	38	51,14
O2_2	834	794	952,04	40	47,96
O3	266	75	281,95	191	718,05
O3_1	134	41	305,97	93	694,03
O3_2	132	39	295,45	93	704,55
O4	978	759	776,07	219	223,93
O4_1	467	384	822,27	83	177,73
O4_2	511	405	792,56	106	207,44
O5	556	527	947,84	39	70,14
O5_1	259	238	918,92	21	81,08
O5_2	297	279	939,39	18	60,61
O6	1 657	1 217	734,46	440	265,54
O6_1	858	717	835,66	141	164,34
O6_2	799	539	674,59	260	325,41
O7	790	443	560,76	347	439,24
O7_1	402	228	567,16	174	432,84
O7_2	388	215	554,12	173	445,88
O8	443	198	446,95	245	553,05
O8_1	224	96	428,57	128	571,43
O8_2	219	102	465,75	117	534,25
O9	467	461	987,15	6	12,85
O9_1	244	231	946,72	3	12,30
O9_2	223	221	991,03	2	8,97
O10	1 190	1 021	857,98	169	142,02
O10_1	579	525	906,74	54	93,26
O10_2	611	534	873,98	77	126,02

Die gebruik van meer as een leesteken

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale gebruik van oortollige leestekens in die korpus	Oortollige leestekens per korpus genormaliseer tot 1 000 sinne	Gemiddelde aantal oortollige leestekens per twiet
O1	669	944	8 942	7	7,42	0,01
O1_1	334	453	4 813	3	6,62	0,01
O1_2	335	491	4 129	4	8,15	0,01
O2	589	1 577	12 259	1	0,63	0,00
O2_1	302	743	6 211	0	0,00	0,00
O2_1	287	834	6 048	1	1,20	0,00
O3	206	266	3 379	82	308,27	0,40
O3_1	107	134	1 590	13	97,01	0,12
O3_2	99	132	1 789	69	522,73	0,70
O4	484	978	7 110	326	333,33	0,67
O4_1	243	467	3 388	134	286,94	0,55
O4_2	241	511	3 722	192	375,73	0,80
O5	333	556	4 579	64	115,11	0,19
O5_1	164	259	2 248	16	61,78	0,10
O5_2	169	297	2 331	48	161,62	0,28
O6	768	1 657	16 340	77	46,47	0,10
O6_1	387	858	8 167	31	36,13	0,08
O6_2	381	799	8 173	46	57,57	0,12
O7	457	790	5 655	5	6,33	0,01
O7_1	233	402	2 877	1	2,49	0,00
O7_2	224	388	2 778	4	10,31	0,02
O8	308	443	3 690	7	15,80	0,02
O8_1	158	224	1 846	1	4,46	0,01
O8_2	150	219	1 844	6	27,40	0,04
O9	275	467	4 918	1	2,14	0,00
O9_1	147	244	2 492	0	0,00	0,00
O9_2	128	223	2 426	1	4,48	0,01
O10	499	1 190	7 744	35	29,41	0,07
O10_1	235	579	3 777	20	34,54	0,09
O10_2	264	611	3 967	15	24,55	0,06

Die gebruik van hoofletters in die onderskeie korpora

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Hoofletters aanwesig aan die begin van 'n sin	Hoofletters aan die begin van 'n sin per korpus genormaliseer tot 1 000 sinne	Hoofletters afwesig aan die begin van 'n sin	Hoofletters afwesig aan die begin van 'n sin per korpus genormaliseer tot 1 000 sinne
O1	669	944	8 942	924	978,81	20	21,19
O1_1	334	453	4 813	443	977,92	10	22,08
O1_2	335	491	4 129	481	979,63	10	20,37
O2	589	1 577	12 259	1521	964,49	56	35,51
O2_1	302	743	6 211	714	960,97	29	39,03
O2_1	287	834	6 048	768	920,86	27	32,37
O3	206	266	3 379	266	1 000,00	0	0,00
O3_1	107	134	1 590	134	1 000,00	0	0,00
O3_2	99	132	1 789	132	1000,00	0	0,00
O4	484	978	7 110	870	889,57	108	110,43
O4_1	243	467	3 388	413	884,37	54	115,63
O4_2	241	511	3 722	457	894,32	54	105,68
O5	333	556	4 579	509	915,47	47	84,53
O5_1	164	259	2 248	249	961,39	10	38,61
O5_2	169	297	2 331	260	875,42	37	124,58
O6	768	1 657	16 340	1477	891,37	180	108,63
O6_1	387	858	8 167	736	857,81	122	142,19
O6_2	381	799	8 173	741	927,41	58	72,59
O7	457	790	5 655	767	970,89	23	29,11

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Hoofletters aanwesig aan die begin van 'n sin	Hoofletters aan die begin van ’n sin per korpus genormaliseer tot 1 000 sinne	Hoofletters afwesig aan die begin van 'n sin	Hoofletters afwesig aan die begin van ’n sin per korpus genormaliseer tot 1 000 sinne
O7_1	233	402	2 877	397	987,56	5	12,44
O7_2	224	388	2 778	370	953,61	18	46,39
O8	308	443	3 690	437	986,46	6	13,54
O8_1	158	224	1 846	220	982,14	4	17,86
O8_2	150	219	1 844	217	990,87	2	9,13
O9	275	467	4 918	456	976,45	11	23,55
O9_1	147	244	2 492	237	971,31	7	28,69
O9_2	128	223	2 426	219	982,06	4	17,94
O10	499	1 190	7 744	1 085	911,76	105	88,24
O10_1	235	579	3 777	519	896,37	60	103,63
O10_2	264	611	3 967	566	926,35	45	73,65

Gebruik van hoofletters ten einde woorde te beklemtoon

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters vir beklemtoning in korpus	Totale aantal hoofletters vir beklemtoning in korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters vir beklemtoning per twiet
O1	669	944	8 942	39	4,36	0,06
O1_1	334	453	4 813	24	4,99	0,07
O1_2	335	491	4 129	15	3,63	0,04
O2	589	1 577	12 259	24	1,96	0,04
O2_1	302	743	6 211	7	1,13	0,02
O2_1	287	834	6 048	17	2,81	0,06
O3	206	266	3 379	0	0,00	0,00
O3_1	107	134	1 590	0	0,00	0,00
O3_2	99	132	1 789	0	0,00	0,00
O4	484	978	7 110	105	14,77	0,22
O4_1	243	467	3 388	69	20,37	0,28
O4_2	241	511	3 722	36	9,67	0,15
O5	333	556	4 579	23	5,02	0,07
O5_1	164	259	2 248	12	5,34	0,07
O5_2	169	297	2 331	11	4,72	0,07
O6	768	1 657	16 340	172	10,53	0,22
O6_1	387	858	8 167	49	6,00	0,13
O6_2	381	799	8 173	123	15,05	0,32
O7	457	790	5 655	13	2,30	0,03
O7_1	233	402	2 877	7	2,43	0,03
O7_2	224	388	2 778	6	2,16	0,03
O8	308	443	3 690	11	2,98	0,04
O8_1	158	224	1 846	3	1,63	0,02
O8_2	150	219	1 844	8	4,34	0,05
O9	275	467	4 918	23	4,68	0,08
O9_1	147	244	2 492	12	4,82	0,08
O9_2	128	223	2 426	11	4,53	0,09
O10	499	1 190	7 744	130	16,79	0,26
O10_1	235	579	3 777	59	15,62	0,25
O10_2	264	611	3 967	44	11,09	0,17

Hoofletters wat nie volgens die taalreëls gebruik is nie

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters nie volgens die taalreëls in korpus gebruik nie	Verwysing na hoofletters nie volgens die taalreëls gebruik nie per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters nie volgens die taalreëls gebruik nie per twiet
O1	669	944	8 942	4	0,45	0,01
O1_1	334	453	4 813	0	0,00	0,00
O1_2	335	491	4 129	4	0,97	0,01
O2	589	1577	12 259	1	0,08	0,00
O2_1	302	743	6 211	0	0,00	0,00
O2_1	287	834	6 048	1	0,17	0,00
O3	206	266	3 379	7	2,07	0,03
O3_1	107	134	1 590	4	2,52	0,04

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	Totale aantal hoofletters nie volgens die taalreëls in korpus gebruik nie	Verwysing na hoofletters nie volgens die taalreëls gebruik nie per korpus genormaliseer tot 1 000 woorde	Gemiddelde aantal hoofletters nie volgens die taalreëls gebruik nie per twiet
O3_2	99	132	1 789	3	1,68	0,03
O4	484	978	7 110	8	1,13	0,02
O4_1	243	467	3 388	4	1,18	0,02
O4_2	241	511	3 722	4	1,07	0,02
O5	333	556	4 579	0	0,00	0,00
O5_1	164	259	2 248	0	0,00	0,00
O5_2	169	297	2 331	0	0,00	0,00
O6	768	1 657	16 340	13	0,80	0,02
O6_1	387	858	8 167	4	0,49	0,01
O6_2	381	799	8 173	9	1,10	0,02
O7	457	790	5 655	81	14,32	0,18
O7_1	233	402	2 877	34	11,82	0,15
O7_2	224	388	2 778	47	16,92	0,21
O8	308	443	3 690	9	2,44	0,03
O8_1	158	224	1 846	9	4,88	0,06
O8_2	150	219	1 844	0	0,00	0,00
O9	275	467	4 918	0	0,00	0,00
O9_1	147	244	2 492	0	0,00	0,00
O9_2	128	223	2 426	0	0,00	0,00
O10	499	1 190	7 744	10	1,29	0,02
O10_1	235	579	3 777	5	1,32	0,02
O10_2	264	611	3 967	5	1,26	0,02

Bylaag F: Sintaksis

Die verskillende sinstipes in die gespesialiseerde korpus.

Korpus	Totale aantal Twiets	Totale aantal sinne	Totale aantal woorde	SinB	SinB (genormaliseer tot 1 000 sinne)	SinV	SinV (genormaliseer tot 1 000 sinne)	SinU	SinU (genormaliseer tot 1 000 sinne)	SinS	SinS (genormaliseer tot 1 000 sinne)
O1	669	944	8 942	22	23,31	93	98,52	22	23,31	807	854,87
O1_1	334	453	4 813	8	17,66	45	99,34	7	15,45	393	867,55
O1_2	335	491	4 129	14	28,51	48	97,76	15	30,55	414	843,18
O2	589	1577	12 259	13	8,24	214	135,70	38	24,10	1312	831,96
O2_1	302	743	6 211	7	9,42	99	133,24	20	26,92	617	830,42
O2_1	287	834	6 048	6	7,19	115	137,89	18	21,58	695	833,33
O3	206	266	3 379	0	0,00	4	15,04	4	15,04	258	969,92
O3_1	107	134	1 590	0	0,00	1	7,46	2	14,93	131	977,61
O3_2	99	132	1 789	0	0,00	3	22,73	2	15,15	127	962,12
O4	484	978	7 110	7	7,16	71	72,60	42	42,94	858	877,30
O4_1	243	467	3 388	3	6,42	38	81,37	19	40,69	407	871,52
O4_2	241	511	3 722	4	7,83	33	64,58	23	45,01	451	882,58
O5	333	556	4 579	2	3,60	51	91,73	26	46,76	477	857,91
O5_1	164	259	2 248	1	3,86	22	84,94	12	46,33	223	861,00
O5_2	169	297	2 331	1	3,37	28	94,28	14	47,14	454	1528,62
O6	768	1 657	16 340	3	1,81	80	48,28	59	35,61	1515	914,30
O6_1	387	858	8 167	2	2,33	45	52,45	33	38,46	778	906,76
O6_2	381	799	8 173	1	1,25	35	43,80	26	32,54	737	922,40
O7	457	790	5 655	1	1,27	16	20,25	31	39,24	742	939,24
O7_1	233	402	2 877	1	2,49	8	19,90	16	39,80	376	935,32
O7_2	224	388	2 778	0	0,00	8	20,62	15	38,66	366	943,30
O8	308	443	3 690	0	0,00	4	9,03	19	42,89	420	948,08
O8_1	158	224	1 846	0	0,00	2	8,93	12	53,57	210	937,50
O8_2	150	219	1 844	0	0,00	2	9,13	7	31,96	210	958,90
O9	275	467	4 918	0	0,00	63	134,90	2	4,28	402	860,81
O9_1	147	244	2 492	0	0,00	31	127,05	2	8,20	211	864,75
O9_2	128	223	2 426	0	0,00	32	143,50	0	0,00	191	856,50
O10	499	1 190	7 744	5	4,20	17	14,29	26	21,85	1142	959,66
O10_1	235	579	3 777	2	3,45	10	17,27	8	13,82	559	965,46
O10_2	264	611	3 967	3	4,91	7	11,46	18	29,46	583	954,17

Die woordsoort-etikette wat in die studie gebruik is, is vanuit Pilon (2005:130-133) geneem. Pilon verwys in haar studie na 133 woordsoort-etikette. Vir die doel van my studie is hierdie woordsoort-etikette gebruik om die geëtiketteerde tekste te analiseer.

Etiket	Kategorie
NSE	Naamwoord/soortnaam
NEE	Naamwoord/eienaam
NMS	Naamwoorde/soortnaam
NM	Naamwoord/massanaam
NSMD	Naamwoord/soortnaam
NSED	Naamwoord/soortnaam
NEM	Naamwoord/eienaam
NVE	Naamwoord/versamelnaam
NA	Naamwoord/abstrak
NVED	Naamwoorde/versamelnaam
NVM	Naamwoord/versamelnaam
NEED	Naamwoord/eienaam
NEMD	Naamwoord/eienaam
NME	Naamwoord/maatnaam
NMED	Naamwoord/maatnaam
NMM	Naamwoord/maatnaam
NMMD	Naamwoord/maatnaam
NVMD	Naamwoord/versamelnaam
VTHOG	Werkwoord/teenwoordig
VTHOO	Werkwoord/teenwoordig
VUOT	Werkwoord/hulp
VTHOK	Werkwoord/teenwoordig
VVHOO	Werkwoorde/verlede
VVHOG	Werkwoord/verlede
VTUOM	Werkwoord/teenwoordig - hulp
VVHOK	Werkwoord/verlede
VTUOP	Werkwoord/teenwoordig
VTUOA	Werkwoord/teenwoordig
VVUOM	Werkwoord/verlede - hulp
VVUOP	Werkwoord/verlede - hulp
VTHSG	Werkwoord/teenwoordig
VTHSO	Werkwoord/teenwoordig
VTHOV	Werkwoord/teenwoordig
VVUOA	Werkwoord/verlede - hulp
VVHOV	Werkwoord/verlede - hulp
ASA	Adjektief
ASP	Adjektief

Etiket	Kategorie
PTENP	Voornaamwoord/persoonlik
PDMNP	Voornaamwoord/persoonlik
PEEB	Voornaamwoord/besitlik
PTEB	Voornaamwoord/besitlik
PO	Voornaamwoord/onbepaald
PDEDP	Voornaamwoord/persoonlik
PDMB	Voornaamwoord/besitlik
PEEDP	Voornaamwoord/persoonlik
PEMB	Voornaamwoord/besitlik
PV	Voornaamwoord/vraend
PTEDP	Voornaamwoord/persoonlik
PDEW	Voornaamwoord/wederkerend
PTMNP	Voornaamwoord/persoonlik
PTMB	Voornaamwoord/besitlik
PEEW	Voornaamwoord/wederkerend
PW	Voornaamwoord/wederkerend
PEMW	Voornaamwoord/wederkerend
PDMW	Voornaamwoord/wederkerend
PTEW	Voornaamwoord/wederkerend
PTMW	Voornaamwoord/wederkerend
LB	Lidwoord/bepaald
LO	Lidwoord/onbepaald
BSG	Bywoord/graad
BST	Bywoord/tyd
BSD	Bywoord/modaliteit
BSW	Bywoord/wyse
BSO	Bywoord/ontkennend
BSP	Bywoord/plek
BSS	Bywoord/skakeling
BSF	Bywoord/frekwensie
BSM	Bywoord/maat
BVW	Bywoord/wyse
BVM	Bywoord/wyse
BSR	Bywoord/rigting
BOW	Bywoord/wyse
BOF	Bywoord/frekwensie
BOM	Bywoord/maat

Etiket	Kategorie
AOP	Adjektief
AVA	Adjektief
AVP	Adjektief
AOA	Adjektief
PB	Voornaamwoord/betreklik
PDENP	Voornaamwoord/persoonlik
PEENP	Voornaamwoord/persoonlik
PA	Voornaamwoord/aanwysend
PDEB	Voornaamwoord/besitlik
TRAO	Telwoord/rang - onbepaald
THBB	Telwoord/hoof - bepaald
TRBB	Telwoord/rang - bepaald
TRPB	Telwoord/rang - bepaald
THPO	Telwoord/hoof - onbepaald
THNB	Telwoord/hoof - bepaald
TRBO	Telwoord/rang - onbepaald
TRPO	Telwoord/rang - onbepaald
W	Tussenwerpsel
UPI	Uniek/partikel - infintief
UPO	Uniek/partikel - ontkenning
UPS	Uniek/partikel - genitief
UPV	Uniek/partikel - vergelyking
UPW	Uniek/partikel - werkwoord
UPD	Uniek/partikel - deel
UPB	Partikel - betreklik
UDS	Uniek - dis
UPE	Uniek/partikel - graad
UXD	Uniek/eksistensiële daar
UPG	Uniek/partikel - graad

Etiket	Kategorie
BVF	Bywoord/frekwensie
SVS	Voorsetsel
KN	Konjunk/neweskikkend
KO	Konjunk/onderskikkend
THAB	Telwoord/hoof - bepaald
THBO	Telwoord/hoof - onbepaald
THAO	Telwoord/hoof - onbepaald
TRAB	Telwoord/rang - bepaald
THPB	Telwoord/hoof - bepaald
UE	Uniek/enklities
RSE	Residu/simbool
RVE	Residu/vreemdetaal
RLE	Residu/letternaamwoord
RAE	Residu/afkorting
RWD	Residu/woorddeel
RKE	Residu/akroniem
RVM	Residu/vreemdetaalwoord
RLM	Residu/letternaamwoord
RSF	Residu/suffiks
RAM	Residu/afkorting
RFE	Residu/formule
RKM	Residu/akroniem
RO	Residu/ongeklassifiseerd
RPF	Residu/prefiks
RSM	Residu/simbool
ZE	Punktuasie - sinsmiddel
ZE	Punktuasie - sinseinde
ZPL	Puntuasie - links parentese
ZPR	Punktuasie - regs parentese

Bylaag G: Morfologie

Afkortings

Afkortings	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**	**O10**
adv		**1**								
Afr		**9**								
afr		**1**								
Antw		**1**								
ARG										**3**
a.k.a.	**1**									
asb		2				4	1			
ASB						1				
BD		**1**								
bib	1									1
bv		2								
Cm			1						1	
ea		2								
ens				1						
epos						1		1		
Feb							1			
genl		1					2			
gnr							1			
gesk		1								
gr					1				1	
hk	1									
Juff									1	
lg		2								
Lg		1								
JHB						2				
Jhb		2								
jhb.		1								
Mej.						3				
Mej						9				
mej						1				
Mev						4				
mil (miljoen)				1			1			
milj (miljoen)		2								
Mnr						3		4		
Mr				1		2				
Mrs							2			
Mrt							1			
nr						1				
NZ										3

Afkortings	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**	**O10**
p		1								
pres		1								
prof										
PTA						1				
pta							2			
Sept		1								
sg		2								
tv		2	3	2		1	2	2		
vlg		2				4				
vs		1	1	6				1		12
v										2
Totale afkortings	3	39	5	11	1	37	13	8	3	21

	O1	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**	**O10**
Foutiewe afkortings in korpus	3	22	3	6	1	13	8	4	3	5
Totale foutiewe afkortings in korpus	3	39	5	13	1	35	11	8	3	21
Totale foutiewe afkortings in korpus (genormaliseer tot 1 000)	0,34	3,18	1,48	1,83	0,22	2,14	1,95	2,17	0,61	2,71
Foutiewe afkortings uniek tot korpus	2	17	0	3	0	9	4	0	1	3
Foutiewe afkortings algemeen tot Twitter	1	5	3	3	1	4	4	4	2	2
Totaal van unieke foutiewe afkortings in die korpus	2	31	0	4	0	23	5	0	1	8
Totaal van unieke foutiewe afkortings in die korpus (genormaliseer tot 1 000)	0,22	2,53	0,00	0,56	0,00	1,41	0,88	0,00	0,20	1,03
Totaal van foutiewe afkortings algemeen tot Twitter	1	8	5	9	1	12	6	8	2	13
Totaal van foutiewe afkortings algemeen tot Twitter (genormaliseer tot 1 000)	0,11	0,65	1,48	1,27	0,22	0,73	1,06	2,17	0,41	1,68

Inkortings

Inkorting	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**	**O7**	**O8**	**O9**	**O10**
ad		2								
ampt		1								
app						2				
aussies				2						
avi	1									
avo									1	

Inkorting	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Baviaans							5			
bio	2									
bloem		1		1		2	3			
Daai	4	1		24	5	3	17	3	9	1
daar's	4					7			1	
def							1			
direksieverg		1								
dis	48	44		34	20	45	22	5	19	2
dok										2
ek's	32	1							2	
EP3										
eps				1						
fam	1									
fone	1					2				
foon	5				2	3			3	
hier's	2	2				1		1		
hy's	6								4	
info				2			1	1		1
intro	1			2					2	
jbay							1			
Jeffas				1						
jy's	14					2		1	9	1
jy't	4					1				
komati				1						
krit		1								
les	1									
nou't		1								
Nseeland		1								
Ojverhoor		1								
onafh		1								
oudred		1								
Plet				1						
Potch	7	13				1			1	2
Potties				1						
prag								1	1	
P-town			1							
ref				5						
refs				2						
roomie	1									
roomies	1									
selfde	1	1		1				1		
Somerset-W		1								
stats				1						

Inkorting	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Stimes		6								
subs		2								
sy's	1							1		
Sy't	1									
tekke				5						
typos	2									
uni	1									
varsity	1			2						
wat's	2	2			1					
wie's		2								
W-kaap		1								
yorkie				1						

Inkorting	Outeur	Frekwensie
Sol-Afriforum	V1	2
Stimes	V1	1
100ps	V1	1
hoofred	V1	2
Potch	V2	1
bio	V3	1
tho	V3	1

Weglating

Weglating	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
beskikbr		2								
d		130						1		
daars		1					1			
di					2	2		1		
dnt									1	
eks				5	9		4			
ekt				12			1			
gaat						1		2		
gistraant				3						
harregat								1		
hdie		1								
hie								3		
hoenner							1			
hyt				1			2			

Weglating	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
ini							1			
j		16								
jr		10				1				
jyt				3			1			
kannie				7						
koene				3		1				
lekke				15			1			
ma				5						
MG		2								
msg									1	
ni					2					
opi							3			
pls									2	
P-town			1							
rite							6			
Saner		1								
sos		1								
v						1				
vd							6	1		
vdag		1								
vi										1
virrie								1		
vlg		2				5				
wa							4		1	
wanne										1
willie				2						1

Weglating	Outeur	Frekwensie
vd	V1	1
d	V1	3
v	V1	10
jr	V1	1

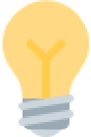
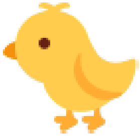
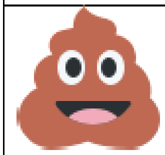
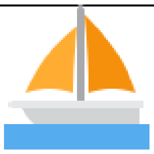
Bylaag H: Emotikons

A001	A007	A016	A024	A033	A042	A051
A002	A008	A017	A025	A034	A043	A052
A003	A009	A018A	A026	:| A035	;-) A044	A053A
A004	A010	A018B	A027	:) A036	A045	A053B
A005	A011	A019	♡ A028	□ A037	A046	A054
A006A	A012	A020	A029	♁ A038	A047	A055
A006B	A013	A021	A030	:D A039	A048	A056
A006C	A014	A022	A031	A040	A049	A057
A006D	A015	A023	A032	A041	A050	A058

A006E
A067
A074B
A083
A092
A101
<3
A110
A059
A068A
A075
A084
A093
A102
A111
A060
A068B
A076
A085
:-)
A094
A103
A112
A061
A069
A077
A086
A095
A104
A113
A062
A070
A078
A087
A096
A105
A114
A063
A071
A079
A088
A097
A106
A115
A064
A072
A080
A089
A098
A107
A116
A065
A073
A081
;)
A090
A099
☆
A108
A117
A066
A074A
A082
:(
A091
A100
A109
:')
A118

A119	A128	A137	A146	A155	A164A	A172
:-) A120	A129	A138	A147	A156	A164B	A173
:-\ A121	A130	A139	A148	A157	A165	A174
A122	A131	A140	A149	A158	A166	A175
A123	A132	A141	A150	A159	A167	A176
A124	A133	A142	A151	A160	A168	A177
A125	A134	A143	A152	A161	A169	A178
A126	A135	A144	A153	---> A162	A170	A179
A127	A136	A145	A154	A163	A171	A180

A181
A190
A199
A208
A217
A226
A335
A182
A191
A200
A209
A218
A227
=)
A336
A183
A192
(:
A201
COOL
A210
A219
A228
A337
A184
A193
B-)
A202
A211
A220
A229
A338
A185
A194
:O
A203
A212
A221
A230
A339
A186
-_-
A195
A204
A213
A222
A231
A340
A187
:v
A196
A205
A214
A223
A332
A341
A188
:3
A197
A206
A215
A224
A333
A342
:))
A189
:p
A198
A207
A216
A225
A334
A343

A344	A345	 A347	A346	A348	 A349	</3 A351
 A350	 A352					

Bylaag I: JGAAP

verdag1
n : 2
1. O2 - 0.04784074270420402
2. O9 - 0.08206599295080863
3. O3 - 0.0883059600503373
4. O4 - 0.0962157505605935
5. O5 - 0.0988332452811801
6. O1 - 0.09933313898322615
7. O6 - 0.10235315407453083
8. O7 - 0.10428014195109081
9. O8 - 0.18018833571655601

verdag1
n : 3
1. O2 - 0.19299503257676043
2. O9 - 0.2151439416540385
3. O1 - 0.24530213050035954
4. O5 - 0.2455283264530218
5. O4 - 0.29213993191253496
6. O3 - 0.29352967367593574
7. O7 - 0.3108284700592008
8. O6 - 0.3636756055664321
9. O8 - 0.5147283860075675

verdag1
n : 4
1. O2 - 0.31101289561159384
2. O9 - 0.3719563737396019
3. O3 - 0.3890014601186722
4. O5 - 0.4010594628741956
5. O1 - 0.43551668603199445
6. O4 - 0.4478294249902822
7. O7 - 0.47674440882727687
8. O6 - 0.6452912129191737
9. O8 - 0.6757057430681037

verdag1
n : 5
1. O2 - 0.28357873189164673
2. O7 - 0.34674440312959853
3. O4 - 0.3852054820982389
4. O3 - 0.3872833475239599
5. O5 - 0.40251645839138894
6. O9 - 0.42491134735977454
7. O1 - 0.4333878289137062
8. O8 - 0.45283874268664537
9. O6 - 0.752100708428903

verdag2
n : 2
1. O9 - 0.04477493176236802
2. O1 - 0.0539877058940047
3. O5 - 0.05482469828523351
4. O2 - 0.07498574772503397
5. O4 - 0.08254206948615483
6. O6 - 0.08259601973663167
7. O7 - 0.08405321288989764
8. O3 - 0.11256442452242255
9. O8 - 0.18726091763121966

verdag2
n : 3
1. O9 - 0.150237379470887
2. O1 - 0.168563754693984
3. O5 - 0.1774804251359665
4. O2 - 0.25811479133648707
5. O4 - 0.2777288646330929
6. O7 - 0.2976833458886651
7. O6 - 0.3299914504130912
8. O3 - 0.34403837561444295
9. O8 - 0.5333683475874283

verdag2
n : 4
1. O9 - 0.245606293077555
2. O5 - 0.29090229973695847
3. O1 - 0.311784983053806
4. O4 0.3594359993398525
5. O7 - 0.3990390018683284
6. O2 - 0.40071281815979665
7. O3 - 0.40475329922969727
8. O6 - 0.589199229749023
9. O8 - 0.6675919527914635

verdag2
n : 5
1. O3 - 0.25578303988412876
2. O7 - 0.2635207079278111
3. O4 - 0.28193909799517847
4. O5 - 0.33168528921404905
5. O2 0.3333393012953295
6. O9 - 0.341470083058116
7. O1 - 0.3618287093709076
8. O8 - 0.40621661837890877
9. O6 - 0.7658126631902982

verdag3
n : 2
1. O5 - 0.05474638301390322
2. O9 - 0.06655207433931076
3. O1 - 0.06759879985283634
4. O7 - 0.09976513150308342
5. O2 - 0.10010125163548844
6. O4 - 0.10629455859602654
7. O6 - 0.11372418857167532
8. O3 - 0.1462495589566608
9. O8 - 0.2173448002115631

verdag3
n : 3
1. O5 - 0.18800747604212653
2. O9 - 0.2119291709816814
3. O1 - 0.2131980649226024
4. O2 - 0.2763259826844078
5. O4 - 0.3404679427590517
6. O7 - 0.3562962521950781
7. O3 - 0.3837436176841057
8. O6 - 0.38779933461879423
9. O8 - 0.5790469854200642

verdag3
n : 4
1. O5 - 0.31857055856727556
2. O9 - 0.3455155782263106
3. O1 - 0.3889704385851397
4. O2 - 0.4388625561573457
5. O4 - 0.5375255210309589
6. O3 - 0.5544772259483711
7. O7 - 0.5557129096267253
8. O6 0.6950429116892723
9. O8 - 0.7684786773206874

verdag3
n : 5
1. O2 - 0.368136537269224
2. O5 - 0.3735923292922527
3. O1 - 0.450450543934357
4. O9 - 0.45098998145449265
5. O7 0.4908477731733878
6. O4 - 0.532692875061556
7. O3 - 0.5668199717568172
8. O8 - 0.5822911584099253
9. O6 - 0.8182999318310229

Bylaag J: Misindeks

Teks	Aantal woorde	Aantal sinne	Gemmiddelde sinslengte	3+ Lettergreep woorde / 2	Leesbaarheids indeks
O1	107	9	11,9	4	6,36
O1_1	101	7	14,4	2,5	6,77
O1_2	106	13	8,2	4	4,86
O2	106	11	9,6	6,5	6,45
O2_1	102	12	8,5	6,5	6,00
O2_2	102	13	7,8	6,5	5,74
O3	103	9	11,4	7	7,38
O3_1	105	6	17,5	7,5	10,00
O3_2	106	7	15,1	2,5	7,06
O4	102	13	7,8	3	4,34
O4_1	103	11	9,4	1	4,15
O4_2	106	14	7,6	2	3,83
O5	107	17	6,3	5	4,52
O5_1	103	10	10,3	1,5	4,72
O5_2	101	12	8,4	1	3,77
O6	106	14	7,6	6	5,43
O6_1	102	12	8,5	7,5	6,40
O6_2	102	13	7,8	8	6,34
O7	104	12	8,7	1,5	4,07
O7_1	104	14	7,4	2	3,77
O7_2	102	13	7,8	2,5	4,14
O8	104	11	9,5	4	5,38
O8_1	101	7	14,4	0,5	5,97
O8_2	101	14	7,2	3,5	4,29
O9	100	13	7,7	3,5	4,48
O9_1	101	14	7,2	1	3,29
O9_2	100	13	7,7	1,5	3,68
O10	101	16	6,3	3	3,73
O10_1	101	10	10,1	5	6,04
O10_2	108	13	8,3	10	7,32
V 1	102	13	7,8	6,5	5,74
V 2	109	14	7,8	5,5	5,31
V 3	100	11	9,1	1,5	4,24