



Verb complementation patterns in Black South African English

M Partridge

 **orcid.org 0000-0003-1049-7957**

Thesis accepted in fulfilment of the requirements for the degree
Doctor of Philosophy in English
at the North-West University

Promoter: Prof AJ van Rooy

Co-Promoter:

Graduation: April 2019

Student number: 12837148

The job of the linguist, like that of the biologist or the botanist, is not to tell us how nature should behave, or what its creations should look like, but to describe those creations in all their messy glory.

—Arika Okrent (*In the land of invented languages*)

Acknowledgements

I thank our Heavenly Father who gave me the means, the support network, and the perseverance to complete this study. To Him be the glory.

I would also like to express my gratitude to the following people:

- My supervisor, Bertus van Rooy, not only for providing invaluable feedback on drafts of my work but also for taking over some of my teaching responsibilities allowing me time to work on this thesis.
- My parents, Cecil and Isabel Partridge, for their love, support, and understanding. Thank you for always believing in me (even when I stopped believing in myself), for reminding me of what is truly important in life, and for taking care of all the everyday-life things so that I did not need to.
- My sister, Charline Partridge, for always supporting me, patiently listening when I needed to vent, and for always finding something at which we can laugh. And thank you for taking care of so many dinners when I was working.

- My friends, especially Yolandi Nortjé, for enquiring about my progress and for checking in regularly.
- My colleagues at the School of Languages at the Vaal Triangle Campus of the North-West University.
- Dennis Mabena at the library of the Vaal Triangle Campus of the North-West University for always finding me the resources I need. I truly appreciate all your help.
- The North-West University for the financial support provided.

Abstract

In this thesis, some of the features pertaining to the verb-complementational patterns in the writing of experienced Black South African English (BSAE) speakers were investigated. These features are the ditransitive construction, anti-deletion in terms of the finite *that*-complementiser, and anti-deletion in terms of the non-finite infinitive *to*-marker. While the ditransitive construction has been investigated in other New Englishes, the construction has not yet been investigated in BSAE. Makalela (2013) and Mesthrie (2006) investigated anti-deletion in spoken corpora of BSAE. Both Makalela (2013) and Mesthrie (2006) found that the non-deletion of the *that*-complementiser is a feature of BSAE, while Mesthrie (2006) identified the undeletion of the infinite *to*-marker as another feature of BSAE.

These features were investigated by means of a parallel corpus containing original texts written by experienced BSAE writers and aligned with the same texts that were edited by professional South African language editors. The data from this corpus were compared to the data obtained from a corpus containing the writing of White South African English (WSAE) writers. Both corpora were collected within the Constrained Language project. In those cases where there was no significant difference in the use of the features between

the BSAE speakers and the WSAE speakers, the features were analysed in the Tswana Learner English (TLE) corpus as well to determine whether the feature that was expected but absent in the writing of the BSAE speakers may have been a learner feature. The results were analysed quantitatively from a usage-based perspective and interpreted against the background of the evolutionary theories of language change, theories that view language as an internal and external construct, and the models of New Englishes that describe the spread of New Englishes.

It was determined that the BSAE speakers use the prepositional dative allostruction more often than the WSAE speakers. In an attempt to gain insight into the phenomenon, a distinctive collexeme analysis was performed on the data of both the BSAE corpus and the WSAE corpus, but no significant differences were found. To determine whether the difference can be ascribed to variety, the ditransitive patterns were annotated for a number of predictors known to influence the choice between the double-object allostruction and the prepositional dative allostruction. The choices were then statistically modelled by means of a conditional inference tree and a random forest. The results revealed that the factors predicting dative alternation in WSAE also predict dative alternation in BSAE and that the more frequent use of the prepositional dative allostruction by the BSAE speakers may be the result of the uneven pressure exerted by the conditioning variables.

With regard to the non-deletion of the *that*-complementiser, it was found that the BSAE speakers omit the *that*-complementiser more often than the WSAE speakers in the academic and reportage registers. In the instructional and popular registers, though, the BSAE speakers retained the *that*-complementiser more often than the WSAE speakers. The difference in all the registers but the academic register was statistically significant. To determine whether the difference could be ascribed to variety, the *that*-complementisers were annotated for a number of predictors known to influence the choices made concerning the omission of the *that*-complementiser. The choices were statistically modelled by means of a conditional inference tree and a random forest, and the results revealed that many of the same factors that predict *that*-omission in WSAE, also predict *that*-omission in BSAE. A number of reasons were posited for the results, amongst which the BSAE speakers' experience in producing texts in specific registers.

It was also determined that there were not many differences in terms of the anti-deletion of the infinitive *to*-marker between the BSAE speakers and the WSAE speakers.

Consequently, data from the TLE were also analysed. The results showed that while there were no significant differences between the BSAE speakers and the WSAE speakers and the learner BSAE speakers and the experienced BSAE speakers, the differences between the learner BSAE speakers and the WSAE speakers were significant. These results suggest that the BSAE speakers are gradually moving towards accepting the WSAE norm. To determine whether variety played a role in the differences, the data pertaining to the *help* + infinitive marker constructeme were annotated for a number of predictors known to influence the choices speakers make in terms of omitting the infinitive marker. The choices were statistically modelled by means of a conditional inference tree and a random forest, and the results revealed that the same factors that predict infinitive *to*-omission in the *help* + infinitive marker constructeme in WSAE predict infinitive *to*-omission in the *help* + infinitive marker constructeme in BSAE.

On the whole, the results indicate that as the BSAE speakers become more experienced and the features become more entrenched in their internal grammar, they seem to accommodate the WSAE speakers by selecting those linguistic features contributed by the WSAE speakers (and possibly the more experienced speakers of other L2 varieties) to the linguistic feature pool. These findings are attributed to the fact that English does not form part of the BSAE speakers' identity and only serves the purpose of social mobility. These findings are similar to recent findings from phonetic data, e.g. Mesthrie (2010) and Mesthrie, Chevalier and Dunne (2015). Furthermore, considering the minimal changes made by the editors to the writing of the BSAE speakers with regard to these features, it would seem as if the BSAE users' endonormative use of the features are accepted by the linguistic gatekeepers of the publishing industry.

Given these findings, it seems as if the norms (as they pertain to the linguistic features investigated in this thesis) are aligned across the STL and IDG components of the broader SAE community, and that BSAE learner usage is not in the process of establishing a radically different norm from the WSAE norm. This means that BSAE is still situated in Phase 3 of Schneider's Dynamic Model of the Evolution of Postcolonial Englishes.

Keywords: Black South African English (BSAE), White South African English (WSAE), learner speakers, experienced speakers, ditransitive construction, dative alternation, *that*-complementiser, *to*-infinitive marker, Construction Grammar, corpus linguistics, World Englishes, New Englishes, language change

Opsomming

In hierdie proefskrif is ondersoek ingestel na verskynsels wat verband hou met die komplementasiepatrone van werkwoorde in die skryfwerk van ervare Swart Suid-Afrikaanse Engelse taalgebruikers. Hierdie verskynsels sluit in die dubbeloorganklike konstruksie, anti-skrappingsverskynsels in terme van die *that*-onderskikker (die ekwivalent van die *dat*-onderskikker in Afrikaans), en anti-skrappingsverskynsels in terme van die *to*-infinitiesmerker (die ekwivalent van die infinitiefmerker *om te* in Afrikaans). Terwyl hierdie dubbeloorganklike konstruksie al in ander nuwe variëteite van Engels ondersoek is, is die konstruksie nog nie in Swart Suid-Afrikaanse Engels ondersoek nie. Makalela (2013) en Mesthrie (2006) het ondersoek ingestel na die anti-skrappingsverskynsel van die *that*-onderskikker in gesproke korpora van Swart Suid-Afrikaanse Engels en bevind dat dié taalverskynsel kenmerkend is van Swart Suid-Afrikaanse Engels. Mesthrie (2006) het verder bevind dat die anti-skrappingsverskynsel van die *to*-infinitiesmerker ook kenmerkend van Swart Suid-Afrikaanse Engels is.

Hierdie verskynsels is ondersoek met behulp van 'n parallelle korpus wat bestaan uit oorspronklike tekste wat deur ervare Swart Suid-Afrikaanse Engelse taalgebruikers

geskryf is, en wat belyn is met dieselfde teks wat geredigeer is deur 'n professionele Suid-Afrikaanse taalversorger. Die data wat van die korpus verkry is, is vergelyk met die data van 'n korpus wat bestaan uit oorspronklike tekste wat deur Wit Suid-Afrikaanse Engelse taalgebruikers geskryf is. Beide korpora is versamel in die *Constrained Language* projek. In gevalle waar daar nie 'n beduidende verskil in die gebruik van verskynsels deur die Swart Suid-Afrikaanse Engelse taalgebruikers en die Wit Suid-Afrikaanse Engelse taalgebruikers waargeneem kon word nie, is die verskynsels ook in die *Tswana Learner English* (TLE) korpus ondersoek ten einde vas te stel of die verskynsels nie moontlik 'n aanleedersverskynsel is nie. Die resultate is kwantitatief geanaliseer en vanuit 'n gebruiksgesbaseerde perspektief geïnterpreteer teen die agtergrond van evolusionêre teorieë van taalverandering, teorieë wat verband hou met taal as 'n interne en eksterne konstruk, en modelle wat die verspreiding van nuwe variëteite van Engels beskryf.

Daar is vasgestel dat die Swart Suid-Afrikaanse sprekers die voorsetseldatief-konstruksie meer gereeld gebruik as die Wit Suid-Afrikaanse Engelse sprekers. In 'n poging om die verskynsel te verklaar, is 'n onderskeidende kolleksie-manalise uitgevoer op die data van onderskeidelik die Swart Suid-Afrikaanse Engelse korpus en die Wit Suid-Afrikaanse Engelse korpus, maar geen statisties beduidende verskil kon waargeneem word nie. Om vas te stel of die verskil moontlik toegeskryf kan word aan variëteit, is die dubbelloorganklike konstruksies gekodeer vir verskeie faktore wat bekend daarvoor is dat dit die keuse tussen die dubbelobjek-konstruksie en die voorsetseldatief-konstruksie voorspel. 'n Statistiese model wat die keuses uiteensit, is toe onttrek. Dit is gedoen met modelleringstegnieke wat bekend staan as besluitnemingsbome en ewekansige woude. Die resultate wys dat baie van die faktore wat die datiefalternasie in Wit Suid-Afrikaanse Engels voorspel, ook die datiefalternasie in Swart Suid-Afrikaanse Engels voorspel. Daarom word daar vermoed dat die hoër frekwensie van die gebruik van die voorsetseldatief-konstruksie onder Swart Suid-Afrikaanse Engelse taalgebruikers moontlik toegeskryf kan word aan die oneweredige druk wat deur die kondisionering-veranderlikes uitgeoefen word.

Met betrekking tot die anti-skrappingsverskynsels van die *dat*-onderskikker is daar bevind dat die Swart Suid-Afrikaanse taalgebruikers die *dat*-onderskikker meer gereeld weglaat as die Wit Suid-Afrikaanse sprekers in die akademiese en verslaggewing-registers. In die instruktiewe en populêre registers laat die Wit Suid-Afrikaanse Engelse taalgebruikers die *dat*-onderskikker meer gereeld weg as die Swart Suid-Afrikaanse

Engelse taalgebruikers. Die verskille tussen die twee groepe is statisties beduidend in al die registers behalwe die akademiese register. Om vas te stel of die verskille toegeskryf kan word aan variëteit, is die *dat*-onderskikers geannoteer vir 'n aantal faktore wat daarvoor bekend is dat hulle taalgebruikers se keuses ten opsigte van *dat*-weglating voorspel. 'n Besluitnemingsboom en 'n ewekansige woud is van die data onttrek met behulp van statistiese modelleringstegnieke. Die resultate toon dat die basiese faktore wat die weglating van die *dat*-onderskikker in Wit Suid-Afrikaanse Engels voorspel, ook die weglating van die *dat*-onderskikker in Swart Suid-Afrikaanse Engels voorspel. 'n Aantal redes is vir hierdie resultate aangevoer, waaronder die Swart Suid-Afrikaanse sprekers se ervaring met betrekking tot die produksie van tekste in sekere registers.

Daar is ook vasgestel dat daar nie baie verskille is in terme van die anti-skrapping van die infinitiefmerker tussen die Swart Suid-Afrikaanse Engelse sprekers en die Wit Suid-Afrikaanse Engelse sprekers nie. Gevolglik is data van die TLE korpus ook geanaliseer. Die resultate dui daarop dat terwyl daar nie enige statisties beduidende verskille tussen die Swart Suid-Afrikaanse Engelse taalgebruikers en die Wit Suid-Afrikaanse Engelse taalgebruikers is nie, en ook nie tussen die aanleeders van Swart Suid-Afrikaanse Engels en die Swart Suid-Afrikaanse Engelse taalgebruikers nie, daar wel 'n statisties beduidende verskil tussen die aanleeders van Swart Suid-Afrikaanse Engels en die Wit Suid-Afrikaanse Engelse taalgebruikers is. Hierdie resultate suggereer dat die Swart Suid-Afrikaanse taalgebruikers geleidelik besig is om die Wit Suid-Afrikaanse Engelse taalgebruikers se norme aan te neem. Om vas te stel of variëteit 'n rol speel in die verskille, is die data wat verband hou met die *help* + infinitief-konstruksie geannoteer vir 'n aantal faktore wat daarvoor bekend is dat dit die taalgebruikers se keuses ten opsigte van die weglating van die infinitiefmerker voorspel. 'n Besluitnemingsboom en 'n ewekansige woud is van die data onttrek met behulp van statistiese modelleringstegnieke. Die resultate toon dat dieselfde faktore wat die weglating van die infinitiefmerker in die Wit Suid-Afrikaanse Engelse taalgebruikers se taal voorspel, ook die weglating van die infinitiefmerker in die Swart Suid-Afrikaanse Engelse taalgebruikers se taal voorspel.

In die geheel dui die resultate daarop dat soos die Swart Suid-Afrikaanse Engelse taalgebruikers meer ervare raak en die taalverskynsels meer verskans raak in hulle interne grammatika, hulle al meer die Wit Suid-Afrikaanse taalgebruikers begin akkommodeer deur daardie taalkundige verskynsels wat deur die Wit Suid-Afrikaanse sprekers (en moontlik ook die meer ervare tweedetaalsprekers) tot die poel van

taalverskynsel bygedra is, te selekteer. Hierdie bevindinge word toegeskryf aan die feit dat Engels nie deel vorm van die Swart Suid-Afrikaanse sprekers se identiteit nie en eerder die rol van sosiale mobiliteit vervul. Hierdie bevindinge stem ooreen met soortgelyke bevindinge van Mesthrie (2010) en Mesthrie, Chevalier en Dunne (2015) ten opsigte van fonetiese data. Met die minimale veranderinge wat die taalversorgers aan die Swart Suid-Afrikaanse taalgebruikers se tekste aangebring het in ag genome, wil dit voorkom asof die Swart Suid-Afrikaanse Engelse taalgebruikers se endonormatiewe gebruik van die taalverskynsels deur die taalversorgers in die publikasie-industrie aanvaar word.

Gegewe hierdie bevindinge, wil dit voorkom asof die norme (soos dit verband hou met die taalverskynsels wat in hierdie proefskrif ondersoek is) belyn is oor die STL- en IDG-komponente van die breër Suid-Afrikaanse Engelssprekende gemeenskap, en dat die aanleerderstaal van die Swart Suid-Afrikaanse Engelse taalgebruikers nie besig is om 'n radikale ander norm as die Wit Suid-Afrikaanse Engelse norm daar te stel nie. Dit beteken dat Swart Suid-Afrikaanse Engels hom steeds in Fase 3 van Schneider se Dinamiese Model vir die Evolusie van Postkoloniale Engels bevind.

Sleutelwoorde: Swart Suid-Afrikaanse Engels, Wit Suid-Afrikaanse Engels, taalaanleerders, ervare taalgebruikers, dubbel-oorganklike konstruksie, voorsetsel-datief-konstruksie, *that*-onderskikker, infinitiefmerker, Konstruksiegrammatika, korpuslinguistiek, Wêreldengels, taalverandering

Table of contents

Chapter 1	1
1.1 Contextualisation	1
1.2 Problem statement	7
1.3 Focus of this study	15
1.4 Research questions	16
1.5 Aims	16
1.6 Methodology	17
1.6.1 Theoretical approach	17
1.6.2 Corpus analysis	18
1.7 Structure of the thesis	21
Chapter 2	23
2.1 Introduction	23
2.2 The emergence of new varieties in language	25

2.2.1	The evolutionary theories of language change	30
2.2.1.1	<i>The theory of competition and selection</i>	34
2.2.1.2	<i>The ecology of language</i>	37
2.2.1.3	<i>Summary</i>	41
2.2.2	A description of language as an internal-external construct	41
2.2.2.1	<i>The relationship between the psycholinguistic and sociolinguistic dimensions of language</i>	42
2.2.2.2	<i>Investigating language change from a psycholinguistic perspective</i>	45
2.2.2.3	<i>Investigating language change from a sociolinguistic perspective</i>	56
2.2.2.4	<i>Summary</i>	63
2.3	The spread of New Englishes	64
2.3.1	Models describing the spread of New Englishes	64
2.3.1.1	<i>Strang's tripartite model of English</i>	64
2.3.1.2	<i>Quirk's views on the varieties of English</i>	65
2.3.1.3	<i>Kachru's three-circle model</i>	67
2.3.1.4	<i>Schneider's Dynamic Model of the Evolution of New Englishes</i>	72
2.3.1.5	<i>Synthesis</i>	79
2.3.2	Attitudes towards New Englishes	83
2.3.3	Synthesis and application to Black South African English	93
2.4	A grammatical description of the new varieties of English	102
2.4.1	Approaches to describing verb complementation	102
2.4.2	The Construction Grammar approach	102
2.4.2.1	<i>Defining Construction Grammar</i>	104
2.4.2.2	<i>Usage-based approaches to Construction Grammar</i>	111
2.4.3	Summary	125
2.5	Conclusion	125
Chapter 3	129
3.1	Contextualisation	129
3.2	Observational data	132

3.2.1	Corpus linguistics as research method	133
3.2.1.1	<i>Defining corpora</i>	138
3.2.1.2	<i>Description of the corpora used in this study</i>	143
3.2.1.2.1	<i>The Black South African English (BSAE) editing corpus</i>	144
3.2.1.2.2	<i>The White South African English (WSAE) editing corpus</i>	146
3.2.1.2.3	<i>The Tswana Learner English (TLE) corpus</i>	146
3.2.1.3	<i>Quantitative analysis of the corpora</i>	147
3.2.1.3.1	<i>Corpus analysis tools</i>	147
3.2.1.3.2	<i>Statistical measures</i>	151
3.3	Conclusion	157
Chapter 4	159
4.1	Contextualisation	159
4.2	A theoretical overview of the ditransitive construction	161
4.3	Factors driving dative alternation	168
4.4	Other factors influencing the structural and semantic realisation of the ditransitive	173
4.5	Results	177
4.5.1	General overview	177
4.5.2	Distinctive collexeme analysis	183
4.5.3	Conditional inference tree	185
4.5.4	Random forest	191
4.6	Conclusion	192
Chapter 5	195
5.1	Contextualisation	195
5.2	Theoretical overview of optionality	198
5.3	Factors driving <i>that</i> -omission	200
5.4	Results	204

5.4.1	General results	204
5.4.2	Conditional inference tree	210
5.4.3	Random forest	218
5.5	Conclusion	219
Chapter 6	223
6.1	Contextualisation	223
6.2	The <i>help</i> + infinitive constructeme in BSAE	228
6.2.1	Introduction	228
6.2.2	Factors driving infinitive <i>to</i> -omission: the case of <i>help</i> + infinitive ...	228
6.2.3	Results	233
6.2.3.1	<i>General results</i>	234
6.2.3.2	<i>Conditional inference tree</i>	236
6.2.3.3	<i>Random forest</i>	240
6.2.3.4	<i>Acceptability</i>	242
6.2.4	Summary	244
6.3	Undeletion of the infinitive <i>to</i> -marker	245
6.3.1	Contextualisation	245
6.3.2	Results	246
6.4	Deletion of the infinitive <i>to</i> -marker	249
6.4.1	Contextualisation	249
6.4.2	Results	250
6.5	Conclusion	251
Chapter 7	255
7.1	Introduction	255
7.2	The research questions and aims revisited	257
7.2.1	The nature of the verb-complementational profiles of ditransitive verbs in BSAE	257
7.2.2	The relative range and frequency of the anti-deletion profiles in BSAE	258
7.2.3	The extent to which the BSAE features have stabilised	262

7.2.4	The social acceptance of the BSAE features	263
7.2.5	The theoretical implications of the findings for the current scholarly discourse in New Englishes	263
7.3	Limitations and future research possibilities	265

List of tables

Table 2.1: Examples of constructions, varying in size and complexity (Goldberg, 2003:220)	106
Table 3.1: Summary of the details pertaining to the compilation of the BSAE editing corpus	145
Table 3.2: Summary of the details pertaining to the compilation of the WSAE editing corpus	146
Table 3.3: The frequency information required to execute a distinctive collexeme analysis (Stefanowitsch, 2013:296)	155
Table 4.1: The relative frequencies of alternating ditransitive constructions in BSAE, WSAE, Indian English and British English	179
Table 4.2: Collexemes distinguishing between the ditransitive and to-dative	184
Table 4.3: Summary of the variables expected to drive the dative alternation	186

Table 5.1: An overview of the percentage of omission of the <i>that</i> -complementiser in BSAE and WSAE across registers	205
Table 5.2: The raw frequencies of the complementation patterns of the verb <i>say</i> ...	207
Table 5.3: Overview of changes made in terms of the <i>that</i> -complementiser by the editor in the BSAE corpus	208
Table 5.4: An overview of the percentage of omission of the <i>that</i> -complementiser in the original BSAE texts and the edited BSAE across registers	208
Table 5.5: Normalised frequencies of the non-deletion of the <i>that</i> -complementiser in the direct speech construction of BSAE and WSAE	209
Table 5.6: Raw frequencies of the editorial intervention in the non-deletion of the <i>that</i> -complementiser in the direct speech construction in BSAE	210
Table 5.7: The hypothesised variables conditioning finite <i>that</i> -omission	211
Table 6.1: The various structural realisations of the <i>help</i> + infinitive constructeme	228
Table 6.2: Summary of the raw alternation frequency of the <i>help</i> + infinitive constructeme in the TLE, BSAE and WSAE corpora	234
Table 6.3: Results of the post-hoc pairwise Wilcoxon rank sum test for variable <i>to</i> -omission	236
Table 6.4: The hypothesised variables conditioning infinitive <i>to</i> -omission	237
Table 6.5: Summary of the <i>help</i> + NP + <i>to</i> allostruction and the <i>help</i> + NP + \emptyset allostruction	239
Table 6.6: A summary of the normalised frequencies of the <i>help</i> + <i>to</i> allostruction and the <i>help</i> + \emptyset allostruction across registers in BSAE and WSAE	242

Table 6.7: Raw frequencies of editors' changes made to the <i>help</i> + infinitive constructions in the BSAE	244
Table 6.8: A summary of the raw frequencies of the <i>help</i> + <i>to</i> allostruction and the <i>help</i> + \emptyset allostruction in the original BSAE corpus and the edited BSAE corpus	244
Table 6.9: A summary of the raw frequencies of the verb + NP + <i>to</i> allostruction and the verb + NP + <i>to</i> allostruction in the original BSAE corpus and the edited BSAE corpus	248

List of figures

Figure 2.1: A visual representation of Kachru's three-circle model (Crystal, 1997:54)	68
Figure 2.2: The various approaches to investigating new varieties of English (Schneider, 2007:15-16)	71
Figure 2.3: A visual representation of the lexicon-syntax continuum in CG (based on Hoffmann & Trousdale, 2013:1-2)	105
Figure 2.4: The symbolic structure of a construction (Croft, 2001:18)	107
Figure 2.5: The elements, components, and units of a construction (Croft, 2001:21)	108
Figure 2.6: A schematic representation of the grouping of exemplars (Bybee, 2010:27)	113

Figure 2.7: A schematic representation of the relation between a schema and its instances (Taylor, 2002:24, 125)	114
Figure 2.8: The related senses of the ditransitive construction (Goldberg, 1995:38)	117
Figure 2.9: The productivity cline (Bardhal, 2008:172)	119
Figure 2.10: The grammatical institutionalisation and conventionalisation of ditransitive verbs (Mukherjee, 2005:205)	122
Figure 3.1: The four stages of a research project (adapted from Graham, 2013:xi) .	130
Figure 3.2: A screenshot of a parallel concordance in Sketch Engine	149
Figure 3.3: A screenshot of a concordance in WordSmith 7	150
Figure 4.1: A representation of a schema and its instances (Taylor, 2002:125)	166
Figure 4.2: The ditransitive constructeme and its allostructions (Perek, 2015:156)	167
Figure 4.3: The frequency and distribution of the various allostructions of the ditransitive constructeme in BSAE and WSAE	178
Figure 4.4: A conditional inference tree indicating the variables that influence the ditransitive constructeme in BSAE and WSAE	187
Figure 4.5: Conditional importance of variables	191
Figure 5.1: Highlighting and greying in the joint attention frame (McGregor, 2013:1158)	199
Figure 5.2: <i>That</i> -omission ratio for BSAE and WSAE	204

Figure 5.3:	An overview of the proportional frequencies of the overt <i>that</i> -complementisers and the zero complementisers in WSAE and BSAE across four registers	206
Figure 5.4:	A conditional inference tree indicating the factors determining the retention and omission of the <i>that</i> -complementiser	212
Figure 5.5:	Conditional importance of the variables determining the retention and omission of the <i>that</i> -complementiser	218
Figure 6.1:	A diagrammatic representation of a temporal, finite clause (Langacker, 2008:117)	224
Figure 6.2:	A diagrammatic representation of an atemporal, infinitive clause (Langacker, 2008:117)	224
Figure 6.3:	Boxplot of the <i>to</i> -omission ratio in the <i>help</i> + infinitive constructeme for BSAE and WSAE	235
Figure 6.4:	A conditional inference tree of the variables predicting infinitive <i>to</i> -omission in the <i>help</i> + <i>to</i> constructeme in BSAE and WSAE	238
Figure 6.5:	Conditional importance of the variables believed to drive the omission of the infinitive marker in the <i>help</i> + infinitive constructeme	241
Figure 6.6:	Graph illustrating the number of changes the WSAE editors made to the <i>help</i> + infinitive constructeme.....	243
Figure 6.7:	The proportional frequencies of the overt infinitive <i>to</i> -complementiser and the bare infinitive <i>to</i> -complementiser with the lemmas <i>feel, hear, know, have, let, make, notice, observe, see, and watch</i>	247
Figure 6.8:	The proportional frequencies of the overt infinitive <i>to</i> -complementiser and the bare infinitive complementiser with the lemmas <i>force, allow, assist, cause, and enable</i>	250

Notational conventions

In the text

Lexical items are printed in italics, e.g. “*give*”.

References to notions and concepts are printed in small caps, e.g. “the concept SELECTION”.

Semantic attributes are provided in square brackets, e.g. “[X causes Y to receive Z]”.

When a technical term is preceded or followed by a definition, the term is printed in bold, e.g. “... **parole** relates to the social aspect of language”.

Syntactic abnormalities are indicated with a *, e.g. “*Nobody give”.

In the quoted examples

The verb of import is printed in **bold**.

The deletion of a linguistic item is indicated with a \emptyset .

Examples from the Black South African English editing corpus are followed with (BSAE).

Examples from the White South African English editing corpus are followed with (WSAE).

Examples from the Tswana Learner English corpus are followed with (TLE).

Examples from the Louvain Corpus of Native English essays are followed with (LOCNESS).

Examples from the British National Corpus Baby are followed with (BNC BABY).

In statistical tables

Following the conventions of Leech, Hundt, Mair and Smith (2009), the asterisks placed next to a numerical quantity in a statistical table serve as indicators of statistical significance:

- * indicates significance at the level $p < 0.05$
- ** indicates significance at the level $p < 0.01$
- *** indicates significance at the level $p < 0.001$
- **** indicates significance at the level $p < 0.0001$

Abbreviations

:	⇒	“realised by” (e.g. S:NP means that the clause element <i>subject</i> is realised by the syntactic form <i>noun phrase</i>)
AdjP	⇒	adjective phrase
AdvP	⇒	adverb phrase
A _{obl}	⇒	obligatory adverbial
A _{opt}	⇒	optional adverbial
BE	⇒	British English
BSAE	⇒	Black South African English
CG	⇒	Construction Grammar
DG	⇒	Descriptive Grammar
L1	⇒	first language
L2	⇒	second language
NP	⇒	noun phrase
O _d	⇒	direct object
O _i	⇒	indirect object
PP	⇒	prepositional phrase
S	⇒	subject
TLE	⇒	Tswana Learner English
V	⇒	verb
VP	⇒	verb phrase
WSAE	⇒	White South African English

Chapter 1

Introduction

1.1 Contextualisation

The linguistic situation in South Africa is both rich and complex. In English alone, four main ethnic Englishes can be distinguished, each with its own phonological and grammatical features. They are known as White South African English (WSAE), Black South African English (BSAE), Indian South African English (ISAE), and Coloured South African English (CSAE) (Branford, 1996:35; Mesthrie, 2010:3; Van Rooy, 2014b). Schneider (2007:174) ascribes this complexity not to multilingualism as such, but rather to the “comparatively high number of distinct, compartmentalized speech communities [that] have entered the arena at different points in time and have interacted with each other under varying social circumstances”.

In the last couple of decades, Black South African English (BSAE)¹ increasingly garnered the attention of researchers. BSAE is the variety of English spoken by the black people in South Africa who are mother-tongue speakers of the country's indigenous African languages (Buthelezi, 1995:242; De Klerk & Gough, 2002:356). Buthelezi (1995:242) attributes the emergence of this variety to a number of factors, amongst which "the educational experience of speakers, their culture and lifestyles which encourage code-mixing, a high degree of enclosure which encourages group cohesiveness, religious affiliations, and the overall black political experience in South Africa". This point of view on the emergence of BSAE corresponds with Schneider's (2003) view that New Englishes emerge as English is appropriated by the indigenous (IDG) strand and changes to reflect their own experiences and to accommodate their own identity.

The history of education in South Africa is most often cited as a principal factor in the emergence of the characteristic phonological and grammatical features evident in BSAE (Buthelezi, 1995:242; De Klerk, 1999:312; De Klerk & Gough, 2002:356). In the early 1900s, some black pupils were taught at missionary schools. The medium of instruction at these schools was English. Hirson (1981:220) points out that "[w]ith a few notable exceptions, the schools were poorly endowed, poorly equipped, overcrowded, and

¹ Although the term Black South African English used to be a contentious term, many researchers in this variety continue using this term (cf. Botha, 2012; De Klerk, 1999; De Klerk & Gough, 2002; Mesthrie, 2006; Minow, 2010; Van Rooy, 2006). In Supplement 38 of the South African Journal of Linguistics (a special issue dedicated to the topic of BSAE), Van Rooy (2000:ii) points out that there have been two major points of contention regarding this term. The first is that "[t]o label this form of English as a form of South African English hides the connection between BSAE and other forms of English in Africa". The second is that "[t]o label it Black is to reinforce racial or ethnic naming practices that have been such a common practice in South Africa, with such dire consequences".

With regard to the first point of criticism, Van Rooy (2000:ii) argues that it is important to regard BSAE as a variety of South African English, because not only do the different English varieties in South Africa influence one another, but the interaction between the varieties of English and other South African languages also influence one another. With regard to the second point of criticism, Van Rooy (2000:ii) makes it clear that the term "Black" is intended to point to the origin of the variety in township education as well as the fact that its speakers have already acquired at least one other language.

Coetzee-Van Rooy and Van Rooy (2005) support these arguments when they investigate the general labelling practices for the varieties of English in South Africa in order to determine what would be acceptable practice to society in general. In their research, they found that the participants in their study found the label Black South African English perfectly acceptable, and even assigned the label to the variety themselves.

incapable of providing more than a rudimentary education". In the years that followed, the population at schools increased dramatically, and state assistance became necessary. Consequently, the Bantu Education Act was implemented in 1953. During this time, the medium of instruction for black children was their mother-tongue language at primary school level and English at secondary school level (Hirson, 1981:221). After the Bantu Education Act was implemented, many competent black teachers withdrew from the system in protest against the act. Some parents also withdrew their children from the system in an attempt to boycott it, but most parents felt that it was necessary for their children to obtain an education. As a result, the number of students increased even more while the number of teachers who were proficient in English declined (De Klerk, 1999:311; Hirson, 1981:227).

Although the history of education (and the shortcomings thereof) in South Africa is a contributing factor in the emergence of BSAE, it is by no means the only factor that played a role. Croft (2000) argues that language change results from two independent processes namely linguistic innovation and linguistic conventionalisation. Generally, linguistic innovation is an individual, cognitive, and psycholinguistic process in which mapping from language structure to language function takes place. This process occurs even within monolingual native communities when speakers communicate with one another. However, in language contact situations where New Englishes tend to emerge, the potential for language change is greatly increased by means of factors such as performance errors due to transfer and overgeneralisation (Van Rooy, 2011:192).

As the grammatical and semantic features of BSAE emerged, linguists attempted to list and describe them. Buthelezi (1995), for example, lists some of the lexical, syntactic and semantic features of BSAE. Later, De Klerk and Gough (2002) and Gough (1996) also list phonological, grammatical, vocabulary, and discourse features that were regarded as characteristic of BSAE at the time. These studies, however, were based on the observations of the authors and no attempt was made to quantify the features by using data. As the occurrence of these features was not quantified, the extent to which they have conventionalised was not entirely clear. Furthermore, Mesthrie (2006:111) points out that while such research – which merely lists the features of a language variety – may have been a necessary first step in describing BSAE, “they are a long way off from being descriptively adequate”. Similarly, Minow (2010:1-2) points out that such research does

not provide much insight into how common specific features are or the extent to which they are truly characteristic of the variety as a whole.

In order to address this shortcoming in research on BSAE, scholars started collecting authentic data on BSAE. De Klerk compiled the corpus of spoken Xhosa English, Van Rooy compiled the Tswana Learner English corpus, and Meierkord compiled the Corpus of Black and Coloured South African English in Contact (Van Rooy, 2013:10-15). These corpora enrich scholars' understanding of the grammatical features of BSAE in three ways: (i) it empirically supports claims regarding features which were identified in less extensive data sets; (ii) it enables researchers to distinguish between performance errors and the actual grammatical features of BSAE; (iii) it enables researchers to investigate the variety more comprehensively and to discover new and unexpected features of the variety instead of regarding them as instances of deviance (Van Rooy, 2013:10, 15).

In spite of the strides that have been made in terms of using corpora to investigate features of BSAE, Van Rooy (2013:11) feels that current research on BSAE is still characterised by the 'comparative fallacy' which entails attempts to determine how BSAE differs from standard varieties of English instead of describing the features of BSAE in its own right. Recent publications on BSAE by Botha (2012), Mesthrie (2006), Van Rooy (2006, 2008a, 2008b, 2011) and Van Rooy and Terblanche (2006, 2009, 2010) are characteristic of research aimed at describing BSAE in its own right.²

With the 'comparative fallacy' preconditioning researchers to regard standard English as the only acceptable norm, researchers are inclined to regard innovative features of BSAE as instances of deviance and describe them as being "non-standard" rather than innovative. Bamgboṣe (1998:1) explains this by pointing out that innovations (i.e. acceptable variants) in New Englishes are often judged by contrasting them to the norms of L1 Englishes and not according to their function in the speech community. Van Rooy (2011:191) supports this standpoint when he states that "the distinction between error and conventionalized innovation is one of the crucial issues that researchers dealing with New Varieties struggle to come to terms with". Bamgboṣe (1998:1-2) argues that "[i]f

² It is important to recognise that even though some of the research listed here uses a first language (L1) corpus as a control corpus, it is inherently non-normative, descriptive, and data-driven instances of research done in BSAE.

innovations are seen as errors, a non-native variety can never receive any recognition". De Klerk (1999:315) ascribes this state of affairs to the fear that if deviations from the native English variety are allowed, all rules will be abandoned and chaos will ensue. She agrees with Bamgboṣe (1998:1-2) that the acceptability of New Englishes should not be determined by how it contrasts to L1 Englishes, but that the ultimate test for acceptability should be "the speakers' ability to deliver their message among themselves and to the outside world" (De Klerk, 1999:315).

Given the aforementioned argument, it is important to determine the point at which a speech community accepts linguistic innovations. In the literature, Mufwene's (2001b, 2008b) description of the linguistic feature pool plays an important role in determining the answer to this question. Mufwene (2001b, 2008b) argues that when two populations in the same geographical area coexist and interact with one another, features from non-native performance may feed into the linguistic feature pool where they are in competition with one another. Innovations are then selected from this feature pool and find their way into the output. At this point, it is important to recognise that these linguistic innovations may not yet have been accepted by the language community and that it is still only part of an individual, psycholinguistic process that takes place. As such, these innovations are initially only used occasionally. As the innovations become part of the linguistic feature pool and become part of the output, they are picked up by other members of the community who find them worthy of copying. The more frequently and continuously these innovations are used by the members of the community, the more entrenched they become in the minds of the members. Ultimately, through regular usage, the innovations become established patterns in the grammar and/or lexicon (Croft, 2000; Schneider, 2007:85, 98-99; Van Rooy, 2011:192). Therefore, linguistic conventionalisation is argued to be a social process (as opposed to linguistic innovation). Schneider (2007:99) succinctly summarises this relationship between linguistic innovation and linguistic conventionalisation when he argues "that innovations occur in the speech of individuals because of internal conditions, but they spread to the community for external reasons".

Mufwene (2008b:133-159) uses this description of the evolution of linguistic conventionalisation to illustrate how (under particular ecological conditions) competition is often (but not always) resolved in favour of substrate languages. Here, it is important to note that features from the local languages (also known as the substrate

languages) in the multilingual speech community may form part of the linguistic feature pool. Consequently, these features may find their way into the second language (L2) of these individuals. When other members of the speech community copy these transfers, and it becomes entrenched in the minds of the language users, it is referred to as the substrate influence. Therefore, it can be said that the “[s]ubstrate influence is the outcome of the cumulation of transfers in a communal variety that has developed its own separate norm” (Mufwene, 2008b:134). This substrate influence is often cited as an explanation for the characteristic features of New Englishes. In BSAE, for example, Buthelezi (1995:248) cites the substrate influence as a possible reason for the intransitive use of the verb *discuss* in BSAE. Similarly, Mesthrie (2006:142) concludes that “[m]esolectal BLSAfE shows a pendulum swing between L1 tendencies that favour undeleting and the influence of setting required by the standard form of the TL”.

In light of this discussion regarding linguistic innovation and the entrenchment of these innovations in the minds of language users, it is problematic to view the features of linguistic conventionalisation as “errors”. Van Rooy (2011:194-195) emphasises this fact when he demonstrates how an “(in origin) ‘erroneous’ form”, *can be able to* has become a conventionalised innovation that has found its way into other sub-varieties of South African English (SAE) in order to fill a gap in the system of English grammar.

The Dynamic Model of Postcolonial English – postulated by Schneider (2003, 2007, 2014) – provides a unified systematic approach to investigate the emergence of New Englishes, and allows researchers to consider the fundamentally uniform development processes evident in New Englishes despite the substantial differences among the indigenous languages and cultures that have come in contact with L1 varieties of English. Within the model, Schneider (2007:83) posits that “[i]nnovations and distinctive structural properties of PCEs [Postcolonial Englishes] are frequently positioned at the interface between lexis and grammar” in the sense that certain words of a word class (but not all of them) may prefer certain grammatical rules or patterns. This means that the syntactic behaviour of certain words is affected. Even though Schneider (2007:46) posits that “innovative assignments of verb complementation patterns to individual verbs” are likely to be found in New Englishes, this particular feature has not yet been investigated extensively in BSAE.

In the following section, an overview is provided of research done with regard to verb complementation in BSAE.

1.2 Problem statement

In the last 30 to 40 years, there has been an increase in scholarly investigation into the grammatical features of of New Englishes. These micro-linguistic descriptions aim to provide a foundation for the theoretical, applied, and political studies of New Englishes. They include descriptions of the phonology, lexis, and syntax of the new varieties.

While many of the micro-linguistic features have been investigated in other New Englishes, there are still certain gaps in terms of the micro-linguistic features described in BSAE, especially regarding innovative patterns of verb complementation. Aside from some research in the descriptions of the *that*-complementiser and the infinitive *to*-marker in BSAE (which are always discussed in conjunction with other grammatical and syntactic characteristics), not much has been said about the innovative patterns of verb complementation in the literature of BSAE or about the extent to which they are conventionalised. Conversely, this feature has enjoyed attention in the research of other New Englishes such as East African English (Schmied, 1991; Van Rooy, 2011), Indian English (Bernaisch, Gries & Mukherjee, 2014; Mukherjee, 2010; Mukherjee & Gries, 2009; Mukherjee & Hoffmann, 2006), New Zealand English (Hundt, 1998:109-112, 115-118) and Kenyan English (Buregeya, 2006). Schneider (2004) also compared the particle verbs in East Africa, Great Britain, India, the Philippines and Singapore while Mukherjee and Gries (2009) investigated verb-construction associations across Hong Kong English, Indian English and Singapore English.

Mukherjee and Gries's (2009) research regarding verb-construction associations provides useful insights into the issue of verb complementation. In their research, they find that the processes of structural nativisation of New Englishes can be observed at the level of verb-construction associations. Mukherjee and Gries (2009:27) investigate verb-construction associations in three varieties of New Englishes (i.e. Hong Kong English, Indian English, and Singapore English) which "represent markedly different stages in the process of the evolution of New Englishes". When considering the results from their investigation of these three varieties of New Englishes in terms of verb-construction

associations, they determine that the more advanced a New English variety is in the developmental cycle, the more dissimilar its collostructional³ preferences are to L1 English. This may be attributed to the fact that the typological factors and/or L2 influence trigger changes in a verb's preferences for certain constructions. An example of this phenomenon can be found in the corpus study conducted by Mukherjee and Hofmann (2006) in which they find that the verb *give* occurs most frequently in monotransitive patterns in Indian English, as opposed to *give* which occurs most frequently in ditransitive constructions in L1 English. Mukherjee and Gries (2009) point out that the strength of various collostructions should be investigated in other varieties of New English in order to confirm their finding, but thus far no similar study has been conducted in BSAE.

With regard to BSAE, De Klerk and Gough (2002:362) and Gough (1996:62) list "idiosyncratic" patterns of complementation as one of the grammatical features of BSAE. They provide the following examples from their data:⁴

(1) That thing **made** me **to** know God. (BSAE)

(2) I **tried that** I might see her. (BSAE)

De Klerk and Gough (2002:362) and Gough (1996:62) do not offer any explanation for the (idiosyncratic) patterns of complementation. This may partly be because they do not distinguish verb complementation from adjective complementation.

Buthelezi (1995:248) also mentions that *discuss* is used as an intransitive verb in BSAE. She provides the following example from her data:

(3) These markers show that he has three perspectives to **discuss about**.
(BSAE)

³ Mukherjee and Gries (2009:36) describe collostructional analysis as the investigation of "the attraction of each one of many words *W* to a particular syntactically defined slot (or repulsion of a word from the syntactic slot at hand) in a particular syntactic pattern *P* ... hence the name: a blend of collocation and construction".

⁴ Only examples of verb complementation are quoted here.

Buthelezi (1995:248) explains that because *discuss* tends to be used intransitively, it takes the preposition *about*. She goes on to say that the use of *discuss* as intransitive verb may be influenced by the verb *talk* (which mandatorily requires the preposition *about* in L1 varieties of English) or the noun *discussion* (which again mandatorily selects a preposition like *about* or *of*). This argument, however, is circular. In essence, Buthelezi argues that as the verb *discuss* is used intransitively, it follows that it will take the preposition *about*. At the same time, however, she posits that since the verb *discuss* takes the preposition *about*, it follows that *discuss* is used as an intransitive verb.

Another feature relating to verb complementation is new prepositional verb forms. De Klerk and Gough (2002:362) quote the following sentences as examples of this feature:

(4) He **explained about** the situation. (BSAE)

(5) They **were refusing with** my book. (BSAE)

These examples can be regarded as innovative patterns of verb complementation because it could be argued that the BSAE language users selected prepositional phrases as verb complementisers where noun phrases are conventionally used as verb complementisers in L1 English:

(6) He [**explained** [about [the situation]_{NP}]_{PP}]_{VP}. (BSAE)

(7) They [**were refusing** [with [my book]_{NP}]_{PP}]_{VP}. (BSAE)

Once again, theorists such as Buthelezi (1995), Gough (1996), De Klerk and Gough (2002) do not explain the presence of these features in much depth, besides attributing the occurrence of some of these characteristic features to the substrate influence.

Although these lists of features regarded as characteristic of BSAE do not provide much insight into why these features are present and to what extent they are used by the language community, they do provide researchers with a foundation to initiate more substantial research with regard to BSAE. Van der Walt and Van Rooy (2002) collected data from BSAE speakers to determine in which phase BSAE is in terms of Gill's (1999) model of three phases for the development of an indigenised norm. They analysed three

sets of data, taking into account the lists of features regarded as characteristic of BSAE. The first set comprised the attitudes of BSAE teachers towards the norm in the form of a questionnaire on English usage. The second set comprised the results of a correction exercise completed by BSAE teachers. The third set of data comprised the results of learner reception of the norm in the form of a grammaticality judgement task. The results indicated that while some features of BSAE are generally accepted, there is still widespread disagreement regarding the acceptability of certain features. These results situated BSAE in the liberation and expansion phase.

One of the characteristic features of BSAE discussed by Van der Walt and Van Rooy (2002:120-121) is preposition deletion and insertion (which links to the innovative use of prepositions in BSAE mentioned above). Although the examples below are discussed as examples of preposition deletion, they provide some insight into verb complementation in BSAE. In Example (8) below, the language users selected a noun phrase as complementiser where a prepositional phrase is conventionally used in L1 English. On the other hand, BSAE users selected a prepositional phrase as complementiser where a noun phrase is conventionally used in L1 English. These examples are indicative of innovative patterns of verb complementation in BSAE:

(8) [feel [∅ home again]_{NP}]_{VP} (BSAE)

(9) [discussed [about [their work]_{NP}]_{PP}]_{VP} (BSAE)

De Klerk (2003) reaches conclusions similar to those of Van der Walt and Van Rooy (2002) regarding the acceptability of features in BSAE. In her research, De Klerk (2003) investigates norms in Xhosa English (a subcategory of BSAE) by scrutinising some of the linguistic features regarded as characteristic of BSAE in the Xhosa English corpus. Her results confirm Van der Walt and Van Rooy's (2002) findings, and she concludes that varieties of BSAE are reaching a stage where local features are becoming increasingly widely used and accepted (though this point has not been reached yet). In her article, De Klerk (2003:474-475) also mentions the distinctive use of prepositions. As she does not provide complete example sentences, but only points to the distinctive use of the prepositions *about* (e.g. *discuss about, mention about*), *around* (e.g. *tell/say around*), *of* (e.g. *look of*), and *up* (e.g. *fill up the form*), not much can be gleaned from her research with regard to verb complementation.

Following the research of Van der Walt and Van Rooy (2002) and De Klerk (2003:474-475), Parkinson and Singh (2007) collected data from their BSAE students to determine the extent to which the features listed by Buthelezi (1995:245-249), De Klerk and Gough (2002:362-363) and Gough (1996:61-63) have stabilised. In order to achieve this goal, Parkinson and Singh (2007) stage an intervention to create awareness amongst BSAE students regarding the differences between BSAE and L1 English. After the intervention, they studied the writing of the BSAE students who took part in the intervention in order to determine which features conform more to the rules of the L1 variety. They argue that the rules which conform to the L1 variety after the intervention have not yet stabilised in the BSAE variety, while the rules which do not conform to the L1 variety after the intervention have stabilised in the BSAE variety. In the study, they conclude that the patterns of complementation have not stabilised as much in BSAE as some other features have. In this regard, it is also important to point out that students were sensitised to the fact that they are working in a formal context and that some of the forms they may have regarded as acceptable previously are not always found in formal written academic contexts (Parkinson & Singh, 2007:63).

Spencer (2011b) followed a different approach than Parkinson and Singh (2007) by obtaining feedback from L1 English-speaking teachers in her study to measure the acceptability ratings of the various features of BSAE. She finds that certain features in BSAE such as the question order which is retained in indirect questions and the extension of the progressive are more readily accepted by L1 speakers than for example gender conflation in pronouns and the use of *too* and *very much* as intensifiers. Spencer (2011b) also finds that L1 speakers are not always sure whether innovative patterns of complementation are acceptable or not. Based on these findings, she concludes that innovative patterns of complementation are one of the features that have not yet completely stabilised as a feature of BSAE.

In terms of finite complementation, Makalela (2013) and Mesthrie (2006) both investigate the *that*-complementiser in BSAE. Mesthrie (2006:117-118) points out that in L1 English there are situations in which the use of the *that*-complementiser is obligatory (also cf. Huddleston, 2002a:952), for example:

- (10) **That** she'd go for a walk was clear to us all. (* \emptyset for *that*) (L1 ENGLISH)

There are also situations where the *that*-complementiser should be omitted (also cf. Huddleston, 2002a:953), for example:

- (11) She said \emptyset , "I'll go for a walk". (**that* for \emptyset) (L1 ENGLISH)

Finally, there are situations in which the use of the *that*-complementiser is optional (also cf. Huddleston, 2002a:953-954):

- (12) She said **that** she'd go for a walk. (L1 ENGLISH)

- (13) She said \emptyset she'd go for a walk. (L1 ENGLISH)

In his study of the *that*-complementiser in BSAE, however, Mesthrie (2006) found that the BSAE speakers do not always adhere to the L1 rules. He quotes some examples from his corpus to show that the *that*-complementiser is used before direct quotes in BSAE, whereas he says it is obligatory to omit the *that*-complementiser in L1 English:

- (14) They'll just tell you **that**, "We have been using Fanakalo". (BSAE)

When having a look at other L1 corpora, though, it is worth noting that even L1 English speakers do not always view the omission of the *that*-complementiser before direct quotes as obligatory:

- (15) He also said **that**, "Florida State can't have the title because they lost to Notre Dame in a game in which they were supposed to prove that they were the best team in the country and that they played liked chunky soup in their bowl game". (LOCNESS)

- (16) Mpete Ketchapile (of the project) says **that** "[b]ecause they use piecing and shaving techniques in their healing methods, which could contribute to HIV transmission, they are now trained in hygienic practices." (ICE-SA)

As a result, Mesthrie's (2006:119) argument regarding the phenomenon where BSAE-speakers add the *that*-complementiser before direct quotes, does not hold much water. In this argument, he contends that the addition of the *that*-complementiser before direct

quotes in BSAE can be ascribed to two competing tendencies: the tendency to teach target language (TL) rules at schools and universities and the tendency to not omit the *that*-complementiser. This second tendency may, in turn, be attributed to three things: substrate influence, analogy or overgeneralisation in L2 acquisition (e.g. when speakers conflate direct and indirect speech), and the discourse tendencies in L1 English which allow *that* for purposes of clarification and so forth.

Considering the arguments above, it is clear that quantitative differences in usage, based on actual data, may provide valuable insights into the different ways in which the *that*-complementiser is used in BSAE and L1 English respectively (Mukherjee & Gries, 2009:28; Quirk, Greenbaum, Leech & Svartvik, 1985:16).

Makalela (2013:102-103) strongly argues for the substrate influence when he says that whereas the *that*-complementiser may fulfil a non-obligatory role in standard English, this is not the case for Sesotho (i.e. the equivalent of the *that*-complementiser in Sesotho, *gore*, is obligatory in cases where it is optional in L1 English). Lenake (2014) and Mathibela (2014) disagree with this point of view and argue that *gore* fulfils a non-obligatory role in Sesotho, similar to the non-obligatory role of the *that*-complementiser in standard English. In an attempt to explain phenomena pertaining to verb complementation in BSAE, it is important to carefully consider all the cognitive, psycholinguistic, and social factors which may result in a specific phenomenon, and not attribute all characteristics to transfer (cf. Mesthrie, 2003:452-453).

Mesthrie (2006:120) also observes that the *that*-complementiser appears in clefted *wh*-constructions in BSAE, whereas the *that*-complementiser must be omitted from clefted *wh*-constructions in L1 English (also cf. Huddleston, 2002a:953):

(17) **Who** does she think \emptyset is the ringleader? (*that for \emptyset) (L1 ENGLISH)

(18) So **what** I think **that** should be done ... is to ask students ... like “Up to this far, how do you feel?” (BSAE)

In terms of non-finite complementation Mesthrie (2006:121) investigates the infinitive marker *to*. He points out that the infinitive marker *to* is mandatory in L1 varieties of English after most verbs subcategorised to take on infinitive clauses. He points out that

most verbs in L1 varieties (such as *asked*) are predisposed to take on infinitive clauses (also cf. Huddleston, 2002b:1244):

(19) He asked me **to** go. (* \emptyset) (L1 ENGLISH)

On the other hand, there is a small set of verbs (e.g. *let*, *make*, *have*) that are predisposed to take bare infinitives:

(20) He made me \emptyset go. (**to*)

(21) He *let* me \emptyset go. (**to*)

(22) He had me \emptyset help with the errands. (**to*)

In BSAE, however, the infinitive *to* may be retained in cases where bare infinitive clauses are required in the L1 varieties (Mesthrie, 2006:122):

(23) My friends asked, "Why do you *let* your child **to** speak Zulu?"

(24) And even the teachers at school made us **to** hate the course.

Based on his findings regarding *that*-complementation and the infinitive *to*, Mesthrie (2006:129) states that grammatical features (such as the *that*-complementiser and the infinitive *to*) that are optional in L1 English, tend not to be omitted in BSAE. He then appends a corollary, which states that grammatical features (such as the *that*-complementiser and the infinitive *to*) that may be omitted in L1 English can also (variably) be omitted in BSAE, albeit at a lower rate of frequency.

Besides the characteristics listed in terms of patterns of complementation and the descriptions of the *that*-complementiser and the infinitive *to* in BSAE (which are always discussed in conjunction with other grammatical and syntactic characteristics), not much have been said about innovative patterns of verb complementation in the literature on BSAE or the extent to which they are conventionalised.

1.3 Focus of this study

In Section 1.1 and Section 1.2 above, it is indicated that patterns of verb complementation have proved to be a very fruitful area of investigation in New Englishes. Even though some dimensions of this area of research have been observed, it has not been exploited extensively in BSAE. The aim of this study is to investigate patterns of verb complementation more extensively.

In some cases where observations in this regard have been made, these observations were not quantified using data (e.g. Buthelezi, 1995; De Klerk & Gough, 2002; Gough, 1996). In this study, a concerted effort will be made to quantify observations using data. The results will provide invaluable insight in determining whether specific innovations have become entrenched and consequently conventionalised in BSAE, making the features truly characteristic of the variety as a whole.

Furthermore, the conventionalised innovations in BSAE will be described in their own right. Although the features will be contrasted to the patterns of verb complementation found in L1 English, they will be judged according to their function in the variety and not according to their relation to the norms of L1 English.

Considering the gaps in the research on verb complementation in BSAE, the overarching aim of this study is to identify the innovative patterns of verb complementation in BSAE and to determine the extent to which they have become conventionalised by looking at data. This will provide a platform from which patterns in BSAE can be compared to patterns in other New Englishes in order to contribute to the understanding of verb complementation patterns in New Englishes.

Finally, the implications of the findings above with regard to the process of stabilisation will be considered in order to make a theoretical contribution to the scholarly discourse in New Englishes. The discussion in this regard will take into account the cognitive perspectives on language evolution and language change as well as the sociolinguistic perspectives on the evolution of New Englishes.

1.4 Research questions

The discussion above results in the following research questions:⁵

- (i) What do the verb-complementational profiles of ditransitive verbs look like in BSAE, and what are the cross-varietal differences and similarities of these profiles between BSAE and WSAE?
- (ii) What is the relative range and frequency of the anti-deletion profiles of the finite *that*-complementiser and the non-finite infinitive *to*-marker in BSAE, and what are the cross-varietal differences and similarities of these profiles between BSAE and WSAE?
- (iii) To what extent have these constructions stabilised in BSAE?
- (iv) Have these constructions obtained social acceptance in the speech community (i.e. are these stabilised constructions accepted as conventionalised innovations)?
- (v) What are the theoretical implications of the findings with regard to the process of stabilisation for the current scholarly discourse in New Englishes?

1.5 Aims

The aims of this study are:

- (i) To establish the nature of the verb-complementational profiles of ditransitive verbs in BSAE, and to determine the cross-varietal differences and similarities of these profiles between BSAE and WSAE.

⁵ The first and second research questions are used to evaluate the frequencies of the various constructions in terms of a distributional statistic. The results obtained from these questions will enable the researcher to answer the third and fourth research questions relating to the conventionalisation of specific constructions.

- (ii) To establish the relative range and frequency of the anti-deletion profiles of the finite *that*-complementiser and the non-finite infinitive *to*-marker in BSAE, and to determine the cross-varietal differences and similarities of these profiles between BSAE and WSAE.
- (iii) To determine the extent to which these constructions have stabilised in BSAE.
- (iv) To determine whether the stabilised constructions have obtained social acceptance in the speech community (i.e. to determine whether the stabilised constructions are accepted as conventionalised innovation).
- (v) To infer the theoretical implications of the findings with regard to the process of stabilisation for the current scholarly discourse in New Englishes.

1.6 Methodology

Stefanowitsch and Gries (2005:2) point out that “collostructional analysis has grown out of a merger of two currents in modern linguistics”. The first current is the theoretical current and the second is methodological. The theoretical current will be discussed in Section 1.6.1, while the methodological current will be discussed in Section 1.6.2.

1.6.1 *Theoretical approach*

Verb complementation can be investigated from different perspectives, i.e. a formalist perspective (e.g. generative grammar), a descriptive perspective (descriptive grammar), a functionalist perspective (e.g. systemic functional grammar), a cognitive perspective (e.g. cognitive grammar) and a construction perspective (e.g. construction grammar). Stefanowitsch and Gries (2005:2), however, point out that it is important that the syntax be regarded as meaningful. Construction Grammar, in particular, provides linguists with the tools to view syntax as meaningful by extending Saussure’s (1916/1972) notion of a LINGUISTIC SIGN as arbitrary and conventional pairings of form and meaning to include all levels of grammar.

Mukherjee (2005:2) is of the opinion that “linguistic research is often too closely associated with a particular linguistic school of thought from the outset, whereas it would be useful to integrate concepts and suggestions made in other linguistic schools”. Based on this statement of principle, Mukherjee (2005:2) advocates the use of a pluralist theory in which existing frameworks are reviewed. The advantage of such an approach is twofold: on the one hand it allows researchers to identify useful points of reference, and on the other hand, it allows researchers to identify the crucial shortcomings of each approach. Considering the advantages of a pluralist approach, the study is based on a broadly functionalist approach. Functionalist theories explain linguistic behaviour in terms of the function of language and account for some basic facts of syntax in functional terms (Croft, 1995:490).

When following a pluralist approach, it is important to ensure that the methodology is relevant to and compatible with all the theories used. In principle, there are two kinds of data that can be used. On the one hand, there is the tradition of introspection and decontextualised sample sentences that are generally invented by linguists themselves (most often used by theorists in the generative framework). On the other hand, there is the usage-based tradition, which Langacker (1987b:6) describes as being a “bottom-up” approach to studying language. In usage-based theories, scholars assume that language users’ experience with language results in cognitive representations of language.

Within the usage-based framework, corpora are a useful tool. It allows linguists to determine the frequency of specific features and to consequently determine the extent to which specific innovative features have become conventionalised. Should introspection and made-up examples be used to answer these questions, the conclusions will be based on very insecure foundations (Egan, 2008:3). This – together with the fact that the majority of investigations into verb complementation in new varieties of English are investigated with the use of corpora – contributes to the decision to use corpora in this study.

1.6.2 *Corpus analysis*

The study will analyse the BSAE component of the corpus compiled within the Constrained Language project. This corpus is comprised of academic, instructional,

popular, and reportage texts. When compiling the corpus, it is important to keep in mind that L1 speakers of English most likely edit published articles. Scholars such as Mesthrie and Bhatt (2008:41), Gilquin and Granger (2011:59), and Götz and Schilk (2011:82) have pointed out that it is sometimes difficult to identify features of New Englishes varieties, since published texts tend to be edited. Furthermore, information about the editing process is rarely to be had. Kruger (2012), and more specifically Kruger (2017), have also argued that editorial intervention plays a greater role in the ultimate appearance of published translations than is accounted for. This is an important consideration because the process of editing entails the normalisation and conventionalisation of language structures that do not resemble L1 English so that it ultimately adheres to the characteristics of L1 English. Consequently, the results of research conducted in terms of the features of translated language may be misrepresented. The same may very well be true for texts written by BSAE English users.

Kruger and Van Rooy (2017:21) contend that scholars should leverage the edited texts to gain insight into the extent that innovative features of a variety have conventionalised. They argue that if the editors accept an innovative feature in a variety, and allow this feature to be disseminated in published writing, some degree of endonormative stabilisation has been attained. In this way, scholars can obtain useful information not only about the conventionalisation of features but also about the acceptability of these features.

Consequently, two versions of the texts are collected: the version written in BSAE, and the final version of the text after it was edited. This results in a parallel corpus consisting of an original BSAE subcorpus and an edited L1 English subcorpus. Gaining access to such a parallel corpus is quite useful, as the original texts written by BSAE users can be analysed in order to describe the verb complementation construction in BSAE while the edited texts may provide insight into the acceptability of these verb complementation constructions by L1 English users.

In order to determine which verb complementation constructions are truly characteristic of BSAE, the WSAE component of the corpus compiled within the Constrained Language project is employed as reference corpus. This corpus is also comprised of academic, instructional, popular, and reportage texts.

Finally, the study also analyses the use of verb complementation in the Tswana Learner English Corpus (also known as the TLE). This corpus is representative of the English written by L1 Setswana speakers from South Africa and Botswana (Granger, Dagneaux, Meunier & Paquot, 2009:38; Van Rooy, 2009:198-199). The TLE forms part of the International Corpus of Learner English (ICLE) and consists of 519 argumentative essays (totalling 203 965) words. This corpus is used to distinguish learner errors from established or emerging conventions in BSAE. If, for example, a feature occurs in the TLE corpus but less frequently (or not at all) in the BSAE corpus, it can be assumed that the feature is a learner error due to analogy or overextension. If, however, a feature occurs in the TLE corpus and the feature occurs as often in the BSAE corpus, it can be assumed that the feature has conventionalised in BSAE and is a feature of the variety (cf. Van Rooy, 2011).

The corpora are tagged using the Constituent Likelihood Automatic Word-tagging System (CLAWS4) with the C7-tagset in order to isolate the lexical verbs. It is important to isolate the lexical verbs, as the verb determines the clause elements that are likely to follow (Biber, Johansson, Leech, Conrad & Finegan, 1999:141). This is done by creating a concordance with the concordance tool in Wordsmith Tools 7. The concordance is then exported to Microsoft Excel where the features are analysed. The statistical tool R is used to determine the strength of the various features. Once all of the corpora have been analysed in terms of verb complementation, a primarily constructionist approach will be used to discuss and explain the results of the analysis.

In order to determine the extent to which specific features have become conventionalised, the original texts in the BSAE corpus are aligned manually to the edited versions of the same texts. Once the corpus is aligned and the necessary mark-up is done, the parallel corpus is uploaded to the online corpus-querying package, Sketch Engine. The corpus-specific linguistic features are investigated using the parallel concordance tool in Sketch Engine. These results are discussed and explained by referring to linguistic theories pertaining to language change and language contact.

1.7 Structure of the thesis

Chapter 2 provides an overview of some of the concepts in the evolutionary approaches to language change that are used in models explaining the spread of New Englishes. It is also argued that language exists not only in speech communities but also in the minds of speakers. The role and importance of linguistic analysis in the description of the linguistic features perceived to be characteristic of specific varieties are also discussed.

Chapter 3 explains the research methodology used in the study. The use of corpora is motivated, after which a description of the corpora is provided. The analysis tools and interpretative methods used to analyse the corpora are also described in this chapter.

The results of the methodological procedures set out in Chapter 3 are presented in Chapters 4, 5, and 6. Chapter 4 focuses on the BSAE language users' use of the ditransitive construction. Chapter 5 investigates the anti-deletion of the finite *that*-complementiser in BSAE, whereas Chapter 6 focuses on the anti-deletion of the infinitive *to*-marker. The findings in these chapters are expounded and elaborated on within the framework of the theories and concepts discussed in Chapter 2.

In the final chapter, Chapter 7, the findings are summarised, conclusions are presented, and possibilities for future research are provided.

Chapter 2

Theoretical framework

2.1 Introduction

Schneider (2007:173-174) describes the rich and complex linguistic situation in South Africa as follows:

The complexity of the sociolinguistic constellations in South Africa [...] probably exceeds that of any other country treated in this book. It is no coincidence that the country typically fails to show up in ENL – ESL or Inner – Outer Circle linguistics: South Africa would qualify for both of these categories at the same time, and also for neither of them, depending on which aspect of its language situation is emphasized.

Reference was made to this unique linguistic situation in South Africa in Chapter 1. Since this thesis aims to gain insight into verb complementation in BSAE – one of the English

varieties that forms part of the unique linguistic complexity which characterises South Africa – this chapter provides an overview of issues crucial to any investigation pertaining to the emergence and spread of New Englishes. The contents of this chapter contribute to the analysis and interpretation of the corpus data described in Chapter 3.

The first purpose of this chapter is to gain insight into the emergence of new varieties in language. Since the models describing the genesis and spread of New Englishes are founded on concepts from the evolutionary theories of language change, Section 2.2.1 focuses on two of the most important concepts in the evolutionary theories of language change: the theory of competition and selection (Section 2.2.1.1) and the ecology of language (Section 2.2.1.2). The theory of competition and selection allows scholars to determine whether the linguistic features encountered in varieties of language are conventionalised or whether they should be regarded as errors, while the ecology of language elucidates how language contact takes place (from both a psycholinguistic perspective and a sociolinguistic perspective) and how factors pertaining to ecology drives language change.

Section 2.2.2 builds on the idea that language exists in both internal ecologies (and external ecologies. The relationship between the psycholinguistic and sociolinguistic dimensions is explained in Section 2.2.2.1. Section 2.2.2.2 focuses on how language change is investigated from a psycholinguistic perspective, while Section 2.2.2.3 explains how language change is investigated from a sociolinguistic perspective.

Building on the foundation provided in Section 2.2, the spread of New Englishes is considered in Section 2.3. In Section 2.3.1, an overview is provided of some of the models used to describe the spread of New Englishes in the literature. These models include Strang's tripartite model of English (Section 2.3.1.1), Quirk's views on the varieties of English (Section 2.3.1.2), Kachru's three-circle model (Section 2.3.1.3), and Schneider's Dynamic Model of the Evolution of New Englishes (Section 2.3.1.4). This knowledge enables scholars to explain attitudes towards New Englishes (Section 2.3.2). In Section 2.3.3, the models describing the spread of New Englishes are used to provide an overview of the development of BSAE.

Section 2.4 provides the foundation for the linguistic analysis of the BSAE verb-complementational features. The features of BSAE are investigated utilising Construction

Grammar. This grammar is particularly well-suited to describe new linguistic features because it regards grammatical units as form-meaning mappings, and when language change takes place, it is realised in the reanalysis of the form-meaning mappings of the grammatical units. Furthermore, usage-based approaches to Construction Grammar equip scholars to observe how language is used in particular speech communities.

The chapter is concluded in Section 2.5.

2.2 The emergence of new varieties in language

When discussing the emergence of new varieties of English, it is important to be aware of the properties of language. On the one hand, language can be regarded as an internal system in the mind of the individual and may, therefore, be different from one individual to the next. In order for this construct to be of any use, the internalised representation of language in the mind of the individual must exhibit sufficient degrees of overlap with the internalised representation of language in the minds of other individuals in the speech community (Van Rooy, 2008c:9). Because individuals are enculturated to social institutions (Mufwene, 2001b:2), language forms part of a cultural evolution that shapes and regulates the internal language in the minds of the speakers who belong to the various speech communities. If there is no overlap in the internalised representation of language amongst individuals, individuals will not be able to cooperate and, as a result, there will be no speech community. As a point of departure, scholars can use this view to investigate language in terms of two factors: language as an internal construct in the minds of individuals, and language as a social and external construct used to facilitate cooperation between the individuals of a speech community.

The distinction between language as a mental construct and language as a social construct was first made by Saussure (1916/1972:9) in *Cours de Linguistique Générale*. In his description of this approach to language, he posits a theoretical notion of language that he calls *LANGAGE*. This concept can be described in terms of *langue* and *parole*. In this distinction, the *langue* relates to the variable, internalised mental representation of language in the minds of individuals (Saussure, 1916/1972:14), whereas the *parole* relates to the social aspect of language (Saussure, 1916/1972:13).

Several linguistic paradigms employ this notion of language as mental construct and language as social construct as a point of departure. In the generative approaches to language, internal language (I-language) is distinguished from external language (E-language). Chomsky (1986:22) defines **I-language** as internalised language: “some element of the mind of the person who knows the language, acquired by the learner, and used by the speaker-hearer”. On the other hand, Chomsky (1986:19) defines **E-language** as externalised language: all the technical concepts that are understood independently of the properties of the mind/brain, i.e. “where the language is ‘used by a population’ when certain regularities ‘in action or belief’ hold among the population with reference to the language, sustained by an interest in communication”.

Linguists who take an emergentist view of language make a similar distinction. Langacker (2008:216) points out that the process of communicating is a complex activity. This activity is two-fold in nature. On the one hand, the activity is controlled by neural processing, and as such can be regarded as a cognitive activity. On the other hand, language is acquired in order to enable participants to interact with one another in a social and cultural context, and as such can be regarded as sociocultural in nature.

Although it may seem as if the generativist and emergentist approaches make the same distinction, it is important to note that language as mental construct and language as social construct play a completely different role in the generativist and emergentist approaches.

The first issue on which the generativists and emergentists disagree is the issue of **the complexity of language**. The generativists consider the structure of language to be extremely complex (Pinker, 1994:117). In fact, Chomsky (1959:43-44) believes that the structure of language is so complex, that it is too difficult for children to acquire through exposure to the structures alone. Following this premise, he believes that there has to be some kind of inbuilt mechanism that enables children to acquire language. The idea is founded in the rationalist tradition and postulates that “there are innate ideas and principles of various kinds that determine the form of the acquired knowledge in what may be a rather restricted and organized way” (Chomsky, 1965:48). In this approach, a Platonist view is central in which “individual learning is largely a matter of *Wiedererzeugung*, that is, of drawing out what is innate in the mind” (Chomsky, 1965:51).

This complexity is attributed to properties of language: the first is the structure-dependency of language (Pinker, 1994:29-32) and the second is the recursivity of language (Pinker, 1999:8-10). Structure-dependency is related to the idea that even though sentences seem to be strings of words, speakers do not pick out words by way of their linear positions, but rather by way of phrases.⁶ These phrases, in turn, are grouped into even larger phrases, and each of these phrases are provided with a mental label, such as “subject noun phrase” or verb phrase (Pinker, 1994:29). These phrases are usually diagrammed as trees. Recursion relates to the idea that an infinite number of structures can be generated by a finite set of rules (Chomsky, 1957:13; Pinker, 1994:93). This recursion is achieved by a procedure that allows a structure that belongs to a category to invoke an instance of itself into a higher structure of the same kind (Pinker, 1994:93, 122-124, 1999:9-11).

The emergentists, on the other hand, hold the view that grammatical categories are not believed to exist in advance (i.e. there are no *a priori* categories in language) but are dependent on the speaker (Croft, 2001:362-368; Hopper, 1987:2-3). Langacker (2008:24) states that speakers abstract schemas⁷ (or generalisations) from occurring expressions, and once these schemas are established as units, they serve as “templates guiding the formation of new expressions on the same patterns” (also cf. Bybee, 2010; Lieven, Behrens, Speares & Tomasello, 2003). As a result, emergentists such as Labov (1994:600) do not believe that language is complex and innate, but that it is of a generalised nature. Hopper (1987:3) summarises this belief as follows:

The notion of Emergent Grammar is meant to suggest that structure, or regularity comes of discourse and is shaped by discourse as much as it shapes discourse in an on-going process. Grammar is hence not to be understood as a pre-requisite for discourse, a prior possession

⁶ Pinker (1994:512) explains that a phrase is “[a] group of words that behaves as a unit in a sentence and which typically has some coherent meaning”.

⁷ In this thesis, the notion of SCHEMATICITY is used as a synonym for Langacker’s (2008:244) notion GENERALITY. Traugott and Trousdale (2013:13) explain that schematicity is a feature of categorisation and involves abstraction. As such, they say that “[a] schema is a taxonomic generalisation of categories”. Similarly, Kemmer (2003:78) says that a schema is basically a cognitive representation of the perceived similarities across many instances of usage and refers to this concept as GENERALITY.

attributable in identical form to both speaker and hearer. Its forms are not fixed templates but are negotiable in face-to-face interaction in ways that reflect the individual speakers' past experience of these forms, and their assessments of the present context, including especially their interlocutors, whose experiences and assessments may be quite different. Moreover, the term Emergent Grammar points to a grammar which is not abstractly formulated and abstractly represented, but always anchored in the specific concrete form of an utterance.

Following the argument above, one can conclude that the generativists and the emergentists have inherently different views regarding the complexity of grammatical constructions. This fundamental difference in the approach to language results in a number of other differences. In the generativist framework, Chomsky (1965:3-5) distinguishes between the concepts *competence* and *performance*. **Competence** is regarded as the linguistic knowledge of the ideal listener speaker, whereas **performance** is regarded as the actual use of language.⁸ Chomsky (1986:10) argues that there is little reason to investigate performance, as it will provide insight only in as far as the understanding of the underlying competence permits. Consequently, the generativists see no purpose in investigating language as a social construct. Instead, they promote the investigation of language as an internal construct only:

Linguistic theory is concerned primarily with an ideal speaker-listener, **in a completely homogeneous speech community**, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest and errors (random or characteristic) in applying his knowledge of the language in actual performance.

(Chomsky, 1965:3; my emphasis)

⁸ Chomsky's distinction between *competence* and *performance* reminds strongly of Saussure's distinction between *langue* and *parole*. Saussure's *langue* and Chomsky's *competence* both point to the internalised mental representation of language (i.e. linguistic knowledge). On the other hand, Saussure's *parole* and Chomsky's *performance* both point to the use of language in social contexts.

From the emergentists' point of view, Croft (2000:3) criticises this approach when he says the following:

One of the problems with treating language as an idealized abstract system is that it makes language change into an inherent process: a single object – the abstract language system – changes over time. But the real, existing entities of linguistics are utterances and speakers' grammars. The evolution of both of those entities, particularly utterances, occurs through replication, not inherent change.

This belief that language is a structure that emerges from repetitive use in social events stands in direct contrast to the generativist approach. Langacker (2008:222) expounds on this view when he states that he regards language “as a structured inventory of conventional linguistic units” and goes on to say that this structure is closely related to language use. This idea is also emphasised by Bybee (2006:712) who argues that both grammatical form and grammatical meaning “come into being through repeated instances of language use”.

Another substantial difference between the generativist approaches and the emergentist approaches relates to the nature of language as internal construct. Chomsky (1986:23) believes that everyone has a blueprint of language known as the **Universal Grammar** (UG). He construes UG as “the theory of human I-languages, a system of conditions deriving from the human biological endowment that identifies the I-languages that are humanly accessible under normal conditions”.

Conversely, Langacker (2008:217) emphasises the importance of regarding language as an activity, i.e. viewing language as “something people do” as opposed to “something people have”. As such, one can say that the mental grammars of individual language users are not identical. Langacker (2008:217) explains that individuals have their own distinct linguistic system (also known as an “idiolect”). Although these systems exhibit a strong family resemblance, some systems resemble one another closely while others resemble one another more distantly.

For the purposes of this study, it is equally important to consider language as an internal construct and language as a social construct. This view is supported by Labov (1994:26)

when he says that it is essential to consider both the variation due to internal factors and the variation due to social factors when investigating language change. The importance of acknowledging that language is both an internal and a social construct is also evident when considering Croft's distinction between innovations and conventionalisations. Croft (2000:59-60) contends that when an individual devises an innovation, it will not be conventionalised⁹ unless a large group of individuals within the speech community appropriate the innovation. This idea is also reflected in Mufwene's (2001b, 2008b) theory of the linguistic feature pool, where language change will not take place if linguistic innovations are not added to the linguistic feature pool and conventionalised by the members of the speech community. Therefore, both the role of individual speakers and the role of language communities should be investigated in studies involving language change.

2.2.1 *The evolutionary theories of language change*

In this study, a generalised theory of evolution is used to understand issues pertaining to language change. This theory originated in biology, but Dawkins (1989:190) points out that the arguments relating to biological evolution could, in essence, be applied analogously to cultural¹⁰ evolution (Dawkins, 1989:189-190).

This idea originated in studies where metaphors are used to analogise biological evolution to scientific progress by means of natural selection (Hull, 1988; Popper, 1979). Darwin (1871/1981:59) was the first scholar to make this analogy when he said that "[t]he formation of different languages and distinct species, and the proofs that both have been developed through a gradual process, are curiously the same." Dediu, Cysouw, Levinson, Baronchelli, Christiansen, Croft, Evans, Garrod, Gray, Kandler and Lieven

⁹ The concept *conventionalisation* allows linguists to distinguish between utterances that are grammatical and utterances that are ungrammatical. Langacker (2008:227) points out that the boundary between grammatical and ungrammatical utterances are not clear-cut and is continually adjusted by the speech community. This idea of conventionalisation is usually applied in L1 contexts, but here it is extended to L2 contexts. The issue of *conventionalisation* is discussed in more detail in Section 2.2.1.1.

¹⁰ Here, the concept *culture* is used to point to the ideas, customs, and social behaviour of a particular society.

(2013:307) also argue that “[b]oth biology and culture are sources of variation and constancies, and change is ongoing in both.” Therefore, linguists see the potential of using similar metaphors to gain insight into language¹¹ change (cf. Croft, 2000; Mufwene, 2001a, 2001b, 2008b). Croft (2000:10) explains this evolutionary approach as follows:

Evolution is recognized as a process that occurs with certain types of entities. The process is probably best understood as it occurs with populations of biological organisms; that is evolutionary biology. The hypothesis is that language change is an example of the same process, or a similar process, occurring with a different type of entity, namely language.

In this approach, Croft (2000) and Mufwene (2001a, 2001b, 2008b) use the language-as-species metaphor¹² to explain how certain issues that play an integral role in the evolution of biological species can be applied to language in order to gain insight into issues pertaining to language change.

When we investigate language evolution by means of the language-as-species metaphor, McMahon (1994:334) warns that it is important “to lay out the basis of comparison [between biological evolution and language change] carefully and that we need not equate to compare”. A similar approach is followed in theories pertaining to cultural evolution and scientific progress. Both Hull (1988:402) and Dawkins (1989:191), for example, unequivocally argue that when attempting to describe scientific progress and cultural evolution by means of biological evolution, it cannot be done by following the principles of biological evolution too strictly. Hull (1988:402) explains it as follows:

I define the entities that function in the evolutionary process in terms of the process itself, without referring to any particular level of organization. Any entities that turn out to have the relevant characteristics belong to the

¹¹ Language is only one factor amongst many other factors (such as dress, diet, customs, art, architecture, engineering, and technology) that comprise culture.

¹² Lakoff and Johnson (1999:45) argue that metaphors allow us to apply a conventional mental image from a source domain to a target domain. This application is made possible by the abstract schematic similarities that exist between the domains.

evolutionary kind. Entities that perform the same function in the evolutionary process are treated as being the same, regardless of the level of organization they happen to exhibit. Generalizations about the evolutionary process are then couched in terms of these kinds. The result is increased simplicity and coherence.

With specific reference to language evolution, Croft (2000:10-12) points out that the literal approach claims a biological basis for the universal properties of languages, while the analogical approach claims analogies between biological processes and processes of language change.¹³ The generalised approach, however, proposes that languages are not biological systems. Mufwene (2001b:145) makes the same point when he says:

However, I also argue that the linguistic species need not be a clone of any biological species. In fact, the proposed population genetics of language evolution is more than an analog of population genetics, although its heuristics has been very much inspired by the latter.

A linguistic species must be defined on its own independent terms and its evolutionary properties hypothesized according to its own combination of ontological properties.

As languages are not regarded as biological systems, it evolves by non-genetic means and, therefore, may take place at a much faster rate than is the case in instances of biological evolution (Dawkins, 1989:189).

Croft (2000:4) makes another important point with regard to the evolutionary approaches to language change when he says that it is important to acknowledge that languages in themselves do not change. Instead, people change language by means of their actions. This means that language change is a theory “that [is] thoroughly based on what happens to language in use, from the origin of an innovation to its adoption as a convention of the speech community, and which unifies internal and external causes of

¹³ Christiansen and Kirby (2003:305) explain that language “is never stationary, changing over time and within populations, which themselves are dynamic”.

language change” (Croft, 2000:12). These changes can be of a structural nature¹⁴ or of a pragmatic nature¹⁵ and can only be investigated when actual instances of language use are considered. This point of view contrasts quite strongly with the generativist view in which language is treated as an idealised abstract system, which in turn relegates language change to being an inherent process (cf. Section 2.2.2.2). When we investigate language change by considering how language is used, we are able to investigate language change from both a sociolinguistic perspective as well as a structural/functional perspective (Croft, 2000:5-6). This explains why evolutionary approaches to language change subscribe to emergentist approaches to language.

Another integral aspect of any theory of language change pertains to the acknowledgement that language change takes place by means of two processes (Croft, 2000:4-5). The first process is the process of innovation, which entails the creation of new forms in language. New forms are created when speakers alter linguistic features in the duplication process. These altered linguistic features are contributed to the linguistic feature pool, where they compete with the features that are typically replicated. The second process is the process of propagation, which entails the diffusion or loss of forms in language. If the altered replications are selected from the linguistic feature pool, they are propagated and diffused in the speech community. If they are not selected, though, the altered features are lost.

As pointed out earlier in Chapter 1, the linguistic feature pool plays a pivotal role in the theory of competition and selection. In the following section, the theory of competition and selection is discussed. This theory provides useful insight into the process of innovation and the process of propagation.

¹⁴ Structural changes refer to changes of morphosyntactic or lexical nature, or changes in encoding meaning which leads to more or less structural complexity (Mufwene, 2001a:145-146).

¹⁵ Pragmatic changes refer to changes of sociological constraints that regulate the usage of expressions within a speech community (Mufwene, 2001a:145-146).

2.2.1.1 *The theory of competition and selection*

When investigating language change in terms of the theory of competition and selection, it is important to point out that language is regarded as a species and not an organism (Croft, 2000; Mufwene, 2001a). According to this analogy, language is comprised of a population of utterances (Croft, 2000:7). **Utterances** – which can be defined as “the product of human behavior in communicative interaction” (in other words, the actual pieces of language) – are likened to the notion of DNA (Croft, 2000:26). In turn, each utterance is comprised of **linguemes**, which are the linguistic patterns¹⁶ that are inherited in replication. These linguemes are the linguistic equivalent of the notion of GENE in biology (Croft, 2000:28). Mufwene (2008b:11) argues that language change is largely brought about by the individual communicative acts (i.e. utterances) of its speakers. These utterances populate the **feature pool**¹⁷, which is an analogue to the notion of GENE POOL in biology (Mufwene, 2001a:46). In biology, genes can replicate themselves. In his discussion on biological replicators, Dawkins (1989:12-20) points out that the copying process is not always perfect. In fact, it is essential that imperfect replication takes place so that progressive evolution can happen (Dawkins, 1989:16). Similarly, the transmission of linguistic features usually takes place with modification (Mufwene, 2001a:46). This modification can be attributed not only to performance errors¹⁸ due to transfer, simplification and overgeneralisation (Van Rooy, 2011:192) but also to “speakers’ constant search for regular patterns” (Deutscher, 2005:208).¹⁹

Mufwene (2001b:4-6) explains the competition and selection that take place within the feature pool using a three-tier model. In this model, the upper tier represents the linguistic input. At this level, language contact plays an important role. As the speakers of

¹⁶ Croft (2000:28) points out that these linguistic patterns could be anything “from a phoneme to a morpheme to a word to a syntactic construction” as well as “their conventional semantic/discourse-functional (information-structural) values”.

¹⁷ Croft (2000) uses the term *lingueme pool*, whereas Mufwene (2001a, 2001b, 2008b) uses the term *feature pool*. For reasons of consistency, I use the term *feature pool* throughout.

¹⁸ The issue of performance errors is discussed more comprehensively in Section 2.2.2.

¹⁹ Deutscher (2005:171-209) argues that changes in linguistic structures interact with speakers’ craving for order. As new linguistic structures evolve, speakers notice coincidental patterns and misconstrue them as meaningful and “extend them by analogy to anything else that seems to fit” (Deutscher, 2005:208).

different languages²⁰ come into contact with one another, they contribute the linguistic features of the respective languages they speak to the feature pool by means of their utterances.

The middle tier represents the feature pool. In the feature pool, linguistic features can be similar to or different from one another (Mufwene, 2001b:5). Features that are similar reinforce one another. This process of reinforcement is known as **congruence** (Mufwene, 2001a:46). However, where features differ from one another, there is room for **competition** (Mufwene, 2001a:46). Furthermore, some features that differ from one another can be associated with the same or similar grammatical functions. These features also compete with one another (Mufwene, 2001b:4).

Thomason (2001:86) points out that the competition and selection can be influenced by either **internally-motivated choices** or **externally-motivated choices**. Although it is difficult to distinguish between these two kinds of choices, they can be distinguished from one another by considering the kind of trigger that gives rise to the change. To distinguish between the two kinds of triggers, Thomason (2001:88-89) considers the lexical items *photocopy* and *animal*. Approximately 40 years ago, the most common word for a photocopy was a *xerox copy* (derived from the name of the large company that produces photocopy machines, Xerox Corporation). After a while, the meaning of the lexical item *xerox copy* was limited to refer only to refer to photocopies made on a Xerox photocopy machine.²¹ Consequently, the word *photocopy* replaced the word *xerox copy*. On the other hand, the word *animal* was borrowed from the Latin *animalis*. It is presumed that the word was initially used to signify a deer. As time passed, the words *animal* and *deer* competed with one another for the general meaning [animal]. The word *deer* was not lost completely but continued to exist with a more restricted meaning. Thomason

²⁰ Here it is important to note that competition and selection do not only take place when the speakers of different languages come into contact with one other, but also when the speakers of a monolingual community come into contact with other monolingual speakers. In this regard, it can be said that each speaker has her own idiolect, and the features of these idiolects contribute to the features in the feature pool. As this study focuses on a new variety of English that emerged from a multilingual context, the explanation of competition and selection focuses on the emergence of new varieties by means of language contact in multilingual situations.

²¹ Thomason (2001:88) speculates that the Xerox Corporation took exception to the general use of the word *xerox copy* and that this may be the reason why the word *photocopy* was coined.

(2001:88-89) points out that the transition period during which *xerox copy* and *photocopy* competed for the same semantic space is basically identical to the transition period during which the word *animal* and *deer* competed for the same semantic space. However, these lexical items differ from one another in that *photocopy* was invented in the language, whereas *animal* was loaned from another language. Therefore, it is said that *photocopy* is representative of an internally-sourced change, whereas *animal* is representative of a contact-induced change.²²

The bottom tier represents the output varieties, i.e. the emergent varieties (Mufwene, 2001b:4-5). These varieties come about when the speakers of the new varieties select modified features from the feature pool and combine them with one another to create a new system. In biological evolutionary terms, it is said that the genes that are selected and replicated are fit; from there the notion of SURVIVAL OF THE FITTEST (Omoto & Lurquin, 2012:153). Here it is important to note that the fittest genes are not necessarily the strongest, the smartest or the healthiest. Instead, **fitness** is determined by means of selection or, as Dawkins (1989:136) phrases it, “[f]itness’ has the special technical meaning of reproductive success”. This means that a gene is considered fit if it is deemed to be the best fit in the (socially determined) communicative ecology.

With regard to the replication of linguistic features, it is important to distinguish between the two kinds of replication. They are **normal** (i.e. identical) **replication** and **altered replication**. When normal replication takes place, it is said that the features in the utterance conform to the linguistic conventions of the speech community (Croft, 2000:30). This links to the notion of **CONVENTION**. Convention can be explained by considering the individuals in a speech community and the speech community as a whole. Each speaker has a grammar, i.e. “the cognitive structure in a speaker’s mind that contains her knowledge about her language” (Croft, 2000:26). Each speaker gains her knowledge from the other speakers’ use of the language within the speech community.²³ Therefore, no two speakers in a speech community will ever have exactly the same grammar (Croft,

²² The differences in terms of internally-motivated changes and contact-induced changes are discussed in Section 2.2.2).

²³ Croft (2000:8) states that “[a]ll people in a society are members of multiple speech communities”. (Here the concept **SPEECH COMMUNITY** does not refer to a collection of speakers, but rather to the domains of use). Speech communities are addressed in more detail in Section 2.2.2.3.

2000:26). In spite of this, they do share common ground, i.e. mutual knowledge. Convention is a property of this mutual knowledge (Croft, 2000:7).

On the other hand, when the altered replication of linguistic features takes place, it is said that the utterance does not conform to the linguistic conventions of the speech community (Croft, 2000:8, 30). This links to the notion of **INNOVATION**. Such an altered replication will only become conventionalised when it is repeatedly selected from the feature pool and used frequently by the other speakers of the output variety (Croft, 2000:38).

Croft (2000:8, 38, 166) maintains that there are distinct causal mechanisms that bring about the innovation and the propagation of language change. Innovation is brought about by means of functional factors (i.e. a linguistic-functional process in which disruptions take place in the mappings from language form to language function), whereas propagation is brought about by means of social factors (i.e. the relationship between the speaker and the speech community).

Furthermore, Croft (2000:30) emphasises that the replication of linguistic features is fundamentally a social process, i.e. “[i]f a speaker doesn’t speak, she will not replicate any linguemes”. Therefore, Croft (2000:32) argues that the selection and replication of features are influenced by social factors, just as the selection and replication of genes are influenced by ecological factors in biology. As such, Croft (2000:27) describes the environment (i.e. the ecology) as “the other members of the speech community, the social context of the speech event, and the goals of the speech event itself”. Considering the central role that ecology plays in language change, the issue of language ecology is addressed in the following section.

2.2.1.2 *The ecology of language*

Omoto and Lurquin (2012:157) point out that ecology plays a central role in the biological evolutionary process. They explain that populations need to adapt to their environment in order to survive. Consequently, fitness also “depends on the environment or circumstances in which an individual lives” (Omoto & Lurquin, 2012:157).

In evolutionary linguistics, the sociohistorical contexts (as well as the social contexts of speech events) are known as the **ecology** in which language change takes place. Therefore, when addressing the emergence of new varieties of English, Schneider (2007:4) argues that it is necessary to consider the sociohistorical contexts of their emergence. In terms of the sociohistorical contexts and the social contexts of speech events, a discussion of the ecology should consider issues pertaining to language contact and accommodation.

Thomason (2001:1) defines the notion of **LANGUAGE CONTACT** as “the use of more than one language in the same place at the same time”. Mufwene (2008b:17), in turn, points out that it is not sufficient that two populations coexist with one another in the same place; they have to interact with one another. With regard to the emergence of New Englishes, the most important form of language contact took place during colonisation. **Colonisation** entailed the migration of a population of its own free will from one territory to settle in another territory (Mufwene, 2001a:153). There the population controlled much of the fate of the territory and its inhabitants (Mufwene, 2001b:106). Schneider (2003:242) describes the ecology in which New Englishes emerge in terms of the perspectives of these two communities, i.e. the settlers (known as the **STL strand**) and the indigenous populations (known as the **IDG strand**). Both the STL speakers and the IDG speakers created and added new linguistic features to the feature pool (discussed in Section 2.3.1.4).

Several of the driving factors that played an important role in colonisation form part of the ecology in which English was used. Schneider (2003:235) points out that some of these factors were of a political, economic, military and religious nature. These factors determined the kind colonisation that took place. Mufwene (2001a:204-206) distinguishes four kinds of colonisation styles. They are: trade colonisation (which entailed the exchange of commodities and led to the development of pidgins); settlement colonies (where contact took place under conditions of socioeconomic integration and where the lexifier²⁴ was the only significant contributor in terms of language); social and/or regional segregation (where the new varieties had a more divergent structure

²⁴ The notion **LEXIFIER LANGUAGE** here refers to the dominant language from which a pidgin or a creole draws most of its vocabulary (Mufwene, 2001b:3).

reflective of the non-European languages due to social and/or regional segregation); and exploitation colonisation (which took place where segregation went hand-in-hand with power stratification and where the lexifier was introduced in an artificial form through education). The different colonisation styles are indicative of the kinds of contact that took place between the populations. Mufwene (2001a:206) states that in places where different forms of colonisation coexisted (e.g. South Africa), the New Englishes need to be studied very carefully.

Although there is extensive variation in the colonisation styles, Croft (2000:221) observes that the linguistic patterns found in New Englishes show remarkable similarity. Schneider (2003:234) makes a similar observation when he states that the different varieties of New Englishes display similar linguistic characteristics despite the many differences between the various indigenous languages and cultures that influenced their development. Schneider (2003) ascribes this phenomenon to the influence of the **external ecologies** (i.e. the social factors that are characteristic of language contact) on the **internal ecologies** (i.e. the linguistic features of the New Englishes).

In terms of the external ecology, scholars believe that variants may diminish under conditions of mutual accommodation (Schneider, 2003:234; Trudgill, 2004:27-28).²⁵ But whereas Trudgill (2008:244) argues against any relationship between accommodation and identity in colonial new-dialect formation,²⁶ Schneider (2003, 2008) argues that identity plays an integral role in the process of accommodation and, consequently, the emergence of new varieties of English.

Several theorists do not subscribe to Trudgill's point of view. Croft (2000:73), for example, argues that accommodation is inherently a social phenomenon and is directly linked to the notion of IDENTITY, as accommodation takes place because the speaker wishes to identify with the interlocutors in the social group. So, although the changes are

²⁵ Croft (2000:73) defines the concept ACCOMMODATION as the process in which speakers alter their utterances to resemble the utterances of the group in which they find themselves in order to enable social relational interaction.

²⁶ Here it is important to note that Trudgill (2008:244) does not deny that identity factors are present in new-dialect formations, nor does he deny that it is a viable sociolinguistic concept. Instead, he simply does not believe that identity is a contributing factor in new-dialect formation.

not intentional, they are driven by the speaker's desire to achieve particular social goals. Identity also plays a central role in Schneider's (2003) model. He argues that:

The individual parties who come into cultural and linguistic contact with each other needed to define and redefine themselves and their social roles in the light of the presence of other groups, of their own historical roots and cultural traditions, and in their relationship to territories and distant centers of political and military power. As these relationships changed over time, so did their identities, their images of themselves in relation to others and the world, and, in turn, their language usage as an expression of these changing identities.

(Schneider, 2003:240)

Thus far, the external ecologies of language were considered.²⁷ However, Haugen (1972/2001:57) points out that even though the "true environment of a language is the society that uses it", language also exists in the minds of the speakers who use it. Therefore, one could say that part of the ecology of language is also psychological, meaning that it interacts with other languages in the minds of bi- and multilingual speakers. As such, the behaviour of language is dual: "it is outward action, performance, but it is also inward potential, competence, which we infer from the performance and in turn use to explain the performance" (Haugen, 1972/2001:58).

Kruger and Van Rooy (2016a) use the term **CONSTRAINED LANGUAGE** to refer to language produced in situations where constraints emanating from the cognitive²⁸ and social environment affect communication. In this regard, they posit that sociolinguistic constraints (imposed by the conscious construal of context and audience needs) together with certain psycholinguistic processing constraints (imposed by bilingual language activation) contribute to the bilingual processing constraints in L2 varieties (Kruger & Van Rooy, 2016a:27). It is important to distinguish between these external and internal ecologies, as the study of the interaction between the different kinds of ecologies provide

²⁷ The issues of identity and accommodation are addressed in more detail in Section 2.2.2.3.

²⁸ The cognitive ecology of language is discussed in more detail in Section 2.2.2.2.

linguists with crucial insights on the causes of language change and how these changes spread (Mufwene, 2008a:23).

2.2.1.3 *Summary*

Some of the most pertinent concepts pertaining to the evolutionary theories of language change have been discussed in this section. The theory of competition and selection provides insight into the conventionalisation of linguistic features, while the theory on the ecology of language explains how language exists in different settings and how these settings influence the changes that take place in language.

Since the importance of describing language as an internal and external construct has been mentioned several times so far, it is discussed in the following section. The discussion of the ecology of language in this section set the groundwork for describing language as an internal and an external construct.

2.2.2 *A description of language as an internal-external construct*

In Section 2.2.1.2, it was said that language exists in internal and external ecologies. On the one hand, language exists in the minds of the speakers who use it. Our knowledge concerning the inward potential of language is primarily derived from studies conducted in the field of psycholinguistics. At the same time, it is important to remember that language also exists in the speech communities who use it. Our knowledge of the existence of language in social contexts is primarily derived from studies conducted in the field of sociolinguistics. To date, scholars have for the most part approached research on language change from either a psycholinguistic point of view or a sociolinguistic point of view. Yet, Mufwene (2008b:23) submits that scholars may gain more insight into language change by considering the interaction between these two ecologies, as opposed to merely studying the one or the other.

This section provides an overview of the physiognomies of language. Language change is described from both a psycholinguistic perspective (Section 2.2.2.2) and a sociolinguistic perspective (Section 2.2.2.3). However, the complex interaction between language as an

internal construct and language as an external construct is considered first in Section 2.2.2.1.

2.2.2.1 The relationship between the psycholinguistic and sociolinguistic dimensions of language

Kruger and Van Rooy (2017:24) acknowledge the importance of both the psycholinguistic dimension and the sociolinguistic dimension in language change. As such, they posit that scholars can perceive language change as innovations in the process of being conventionalised (referred to as **BOTTOM-UP CONVENTIONALISATION**), or as errors in relation to some exonormative standard (referred to as **TOP-DOWN STANDARDISATION**). In order to determine the origin of novel linguistic features, Van Rooy (2011:197) proposes two sets of criteria against which these features can be evaluated. They are **grammatical systematicity** and **acceptability** (discussed in more detail in Section 2.3.2).

Grammatical systematicity involves **form-function reanalysis**, i.e. when speakers and listeners reanalyse the form-meaning mapping in a grammatical construction. When language change takes place by means of form-function reanalysis, the language user is regarded as the agent of language change (Croft, 2000:117; Deutscher, 2005:61; McMahon, 1994:8; Mufwene, 2008b:64). In this regard, Croft (2000:117) explains that speakers and hearers possess a cache of internalised knowledge of form-meaning constructions. These constructions are stored together with information about the contexts in which they were used. Consequently, when speakers want to communicate, they retrieve a number of form-meaning pairings and combine them in such a way that their intended meaning is conveyed to the hearer. However, the repeated use of similar constructions across various contexts of use is inevitable. As such, there may be a certain degree of indeterminacy or ambiguity in the attribution of the semantic components to the syntactic elements in an utterance. In other words, there may be a possibility that the speaker or the hearer may not attribute the semantic components to the syntactic elements in precisely the same way. This phenomenon is known as **FORM-FUNCTION REANALYSIS**. Croft (2000:118) points out that form-function reanalysis is a non-intentional mechanism for innovation and that these innovations are the result of speakers attempting to conform to convention. As other speakers repeatedly start attributing the same function to the same syntactic components, the innovation is propagated and

eventually conventionalised. In emergentist usage-based studies, increased frequency plays an important role since it contributes to the cognitive entrenchment of the features (cf. Section 2.4.2.2). This process is known as **BOTTOM-UP CONVENTIONALISATION** (Kruger & Van Rooy, 2017:23-27).

At the same time, however, Croft (2000:88) points out that language use encompasses more than the attribution of semantic components to syntactic elements; it also entails interaction between speakers and hearers (also cf. McMahon, 1994:17; Mufwene, 2008b:11). Similarly, Kruger and Van Rooy (2017:28) point out that the legitimisation processes that impact on diffusion and conventionalisation in more complex ways also need to be considered. An important issue pertaining to the processes of legitimisation is the issue of **acceptability**. If speakers do not deem the novel construction to be conventional and appropriate, it will not be propagated and eventually conventionalised. As discussed in Section 2.3.2, the acceptability of a construction is primarily determined by the authoritative measure (i.e. the extent to which editors, publishing houses, and other authorities accept the innovation) and the codification measure (i.e. whether innovations are codified in resources such as style guides and dictionaries (Bamgboṣe, 1998:4). Consequently, it is posited that acceptability is determined by the extent to which speakers and hearers deem the novel construction to be conventional and appropriate in various contexts. Kruger and Van Rooy (2017:23-27) refer to this process as **TOP-DOWN STANDARDISATION**.

Kruger and Van Rooy (2017:24) argue that these two processes, bottom-up conventionalisation (which can be investigated from a psycholinguistic perspective) and top-down standardisation (which can be investigated from a sociolinguistic perspective) are the fundamental cornerstones for any investigation pertaining to language change. The relationship between these two dimensions, however, are quite complex. Kruger and Van Rooy (2017:24) attribute this complexity to the fact that “much of the social diffusion and conventionalisation of innovations are, in fact, driven by psycholinguistic processes based on frequency effects associated with input and exposure, which, in turn, are dependent on a variety of other interwoven psycholinguistic and social factors”.

The complexity of the relationship between these dimensions can be illustrated quite clearly by means of the process through which new words are admitted to dictionaries. The process starts when an individual attributes a semantic component to a syntactic

element in a new and innovative way (i.e. when form-function reanalysis takes place). If the speech event fails, the word will not be used again. If, however, the speech event is successful and other members of the speech community continue to employ the new word, it is added to the linguistic feature pool where it enters into competition with other linguistic features. As more speakers find the new word acceptable and employ it in their utterances, the word is propagated and becomes more entrenched in the minds of the speakers. At the same time, the increased usage of this new word comes to the attention of lexicographers during their perusal of large bodies of (electronic) texts. Initially, lexicographers flag the feature and keep monitoring its use (Oxford English Dictionary, 2016b). The more frequently the lexicographers encounter the word, the more likely it becomes that it will be included in the dictionary since its usage signals not only its success in communicative events but also its acceptance by various speakers in the speech community. Once the word reaches “a level of general currency where it is unselfconsciously used with the expectation of being understood” by a number of speakers, it is put forward for inclusion (Oxford English Dictionary, 2016b). Finally, these words are reviewed by “a range of experts” before being included in the dictionary (Oxford English Dictionary, 2016a).²⁹

In order to provide a nuanced account of the integrative interaction of the cognitive and social factors that are inherent in language change, Kruger and Van Rooy (2017:25) propose the use of the notions CONVENTIONALISATION and LEGITIMISATION to refer to the processes of bottom-up conventionalisation and top-down standardisation respectively.

²⁹ Here it is important to emphasise that subjective linguistic ideologies do not play a role in either the lexicographers’ decision to include a word in the dictionary or in the language experts’ description of the word. As explained earlier, the lexicographers rely on instances of actual language use to determine the extent to which speakers deem the word to be acceptable and the extent to which the word has consequently been conventionalised, while the language experts provide insight in issues such as word etymology based on diachronic corpus studies. Therefore, the codification of language norms in these dictionaries follow a historical and sociolinguistic approach to language standardisation as opposed to an ideological approach (Peters, 2014:582). Haugen (1966) and Milroy (2001) provide additional insight into these two approaches to standardisation.

In Section 2.2.2.2, language change is scrutinised from a psycholinguistic perspective. In Section 2.2.2.3, on the other hand, language change is scrutinised from a sociolinguistic perspective.

2.2.2.2 Investigating language change from a psycholinguistic perspective

In this section, the aim is to describe language as an internal construct and consider the psycholinguistic factors that contribute to the emergence of new varieties of English.

When investigating language, it exhibits regularity of patterning despite variability. In studies pertaining to the emergence of new varieties of English (e.g. Mesthrie, 2004; Mesthrie & Bhatt, 2008:39-155; Schneider, 2003:234; Sebba, 1997:70), many regularities have been observed in the structural patterning of the various varieties, in spite of the many variables (e.g. differences in the indigenous languages and cultures) that factored into their emergence. An investigation into language as an internal construct provides insight into the regularity of patterning exhibited in the structure of language, and to understand how language variation originates within individual speakers while they attempt to negotiate the social structures in which they find themselves (Bybee, 2010:1; Eckert, 2000:3-4; Hickey, 2012:390). Therefore, some scholars have argued that language change should first be studied from an internal point of view before considering the external factors that may have contributed to the change (Thomason, 2006:339).

In the paragraphs that follow, some of the psycholinguistic factors that play a role in the emergence of New Englishes will be considered. These include issues regarding speakers' individual grammars, second-language acquisition (SLA), bilingual activation, as well as constrained language.

Whereas the speech community is central to a sociolinguistic investigation into language change (cf. Section 2.2.2.3), the **individual grammar** is central to a psycholinguistic investigation into language change. Each individual possesses her own internal grammar.

This internal grammar comprises the grammatical knowledge³⁰ of a speaker. It is “made up of a system of grammatical variables with associated social values as well as semantic values” that allow the individual to produce meaningful utterances that can be understood by other members of the speech community (Croft, 2000:26, 52). Individuals acquire their individual grammars through contact with the other individuals of the speech community and through their exposure to the linguistic structures embodied in the other members’ utterances. Since the contact patterns differ from one individual to the next, there will be differences in the individual grammars of speakers (Croft, 2000:29). The differences in individual grammars can also be ascribed to the fact that individuals cognitively organise their experiences with language in different ways (Bybee, 2006:711).

Bybee (2015:9-10) attributes language change to three cognitive factors. The first pertains to cognitive access (i.e. “recalling words and constructions from memory”) and the motor routines of production (i.e. “articulation”) (Bybee, 2015:9). The more we access words and constructions, the more entrenched their memory representations become and the easier they are accessed (Bybee, 2006:715, 2015:9; Bybee & Thompson, 1997:378-380). At the same time, the repetition of the motor routines of production leads to a more fluent execution of the routine. As the routine becomes more fluent, the routine articulatory gestures undergo reduction and overlap, resulting in the establishment of a new routine of articulatory gestures, e.g. when a group of words are processed as a single unit (e.g. *do not* → *don’t*) (Bybee, 2006:714-715, 2015:9; Bybee & Scheibman, 1999). This process is known as the **REDUCTION EFFECT**.

The second factor that contributes to language change is the human penchant for forming patterns from experience and applying these patterns to new experiences (Bybee, 2015:10). When speakers use language, they are continually finding patterns and reinforcing them through repetition (Lieven *et al.*, 2003:361-362). It allows speakers to gain faster access to lexical items and is known as the **CONSERVING EFFECT** (Bybee &

³⁰ Croft (2000:29) explains that the knowledge of language essentially comprises the individual’s ability to replicate the linguistic features of a language in the appropriate social-communicative context. Bybee (2010:14) refers to this knowledge as **EXEMPLAR REPRESENTATIONS**. She explains that they consist of “phonetic detail, including redundant and variable features, the lexical items and constructions used, the meaning, inferences made from this meaning and from the context, and properties of the social, physical linguistic context” (Bybee, 2010:14).

Thompson, 1997:380-381). The conserving effect can be explained by considering noun phrases and pronouns. Although pronouns diachronically derive from nouns and nouns and pronouns occupy the same positions syntactically, pronouns are used much more frequently than noun phrases. Bybee and Thompson (1997:380-381) argue that the frequent use of pronouns made them resistant to change, and, consequently, they preserved their distinct forms for the nominative and accusative case, while nouns have lost their case distinctions in English.

The third and final factor is related to the conserving effect and is known as **AUTONOMY**. Bybee (2006:715) explains that autonomy “refers to the fact that morphologically complex forms (or strings of words) of high frequency can lose their internal structure as they become autonomous from etymologically related forms”. An example, here, is the process of grammaticalisation that *be going to* underwent. Deutscher (2005:147-155) explains that *going to* originally meant [walking somewhere] or [travelling somewhere]. The phrase *going to do something* only emerged later in the 15th century. It was used as a kind of shorthand for [going somewhere in order to do something]. Gradually other verbs were used in the expression, e.g. *going to bring*, signalling a move towards abstraction. During the 17th century, the [future] meaning was included more frequently in expressions containing *going to*, e.g. *is going to bring* and *is going to deliver*. By this time, *going to* could be used as a future marker, and need not have included any residual meaning of [movement]. This process led to the grammaticalisation of the semi-auxiliary verb *going to*. Since *going to* lost its independent content and was used more frequently, the erosion in sound became possible (i.e. the increase in the use of the construction minimised the risk of misunderstanding when speakers used shortcuts in reducing *going to* to *gonna*) (Deutscher, 2005:154-155).

Bybee (2006:712-713) argues for a similar process with conventionalised word sequences, i.e. prepositional verbs (e.g. *look at*), phrasal verbs (e.g. *shut up*), collocations (e.g. *run a business*), fixed expressions (e.g. *as a matter of fact*), and idioms (e.g. *the long and the short of it*). These word sequences are used so frequently that they have become conventionalised and obtained extended meanings (also cf. Section 2.4.2.1). Both Bybee (2015:10) and Deutscher (2005:154-155) point out that the manner in which language users employ the words or patterns of a language in context may lead to changes in meaning in any language beyond the compositional meanings that these sequences had

at earlier stages of use. As a result, language change is inevitable in all languages and language change takes place in similar ways across languages.

An important aspect to consider when investigating language change from a psycholinguistic perspective – especially regarding the emergence of the new varieties of English – is the issue of **language acquisition**. Sridhar and Sridhar (1992) – first published in *World Englishes* in 1986 – pointed towards what they called the “paradigm gap” between research on SLA and research on the new varieties of English and called for an integrated approach in order to bridge this gap. Such an integrated approach is necessary since insights gained in the field of SLA can also provide insights into new varieties of English. This is the case because both learner Englishes and new varieties of English emerge in multilingual ecologies where language contact takes place (Hundt & Mukherjee, 2011a:1; Mair, 2013:254-255; Mesthrie & Bhatt, 2008:156).³¹ In this regard, Kachru (1992a:54-55) argues that while learner Englishes (which are performance varieties) and New Englishes (which are institutionalised varieties) are distinct from one another, they overlap with each another, because “[a]n institutionalized variety always starts as a performance variety, with various characteristics slowly giving it a different status” (Kachru, 1992a:55). Therefore, it comes as no surprise that a large overlap in the sources of new features that arise in learner varieties of English and New Englishes can be observed, including language learning strategies, learner error, processing difficulty, and substrate transfer (Gilquin, 2015:97; Kruger & Van Rooy, 2017:21; Schneider, 2007:99-109). Unfortunately, as Hundt and Mukherjee (2011a:1-2) point out, “these two objects of inquiry have continued to be treated as fundamentally different and unrelated areas of research”. They go on to say that it is imperative that these two fields “were described and compared on an empirical basis in order to draw conceptual and theoretical conclusions with regard to their form, function and acquisition” (also cf. Mufwene, 2008b:149-159). Lately, however, scholars have made a concerted effort to bridge this gap (cf. Davydova, 2012; Deshors, 2014; Gilquin, 2015).

The acquisition of an L1 and the acquisition of an L2 take place by decidedly different means. It can be attributed to the fact that first language acquisition (FLA) coincides with

³¹ In this study, the notion of MULTILINGUALISM includes the notion of BILINGUALISM. Multilingualism is discussed from sociolinguistic perspective in Section 2.2.2.3.

cognitive development in children. SLA, on the other hand, typically occurs at a later stage in life when circumstances in the speaker's ecology necessitate the maturation of other cognitive abilities – most likely at the expense of the language acquisition ability (Van Rooy, 2010:15). Therefore, it comes as no surprise that SLA takes place by different cognitive means than FLA. FLA takes place through the imitation and repetition of high-frequency items and is entrenched directly into the internal grammar of the individual speaker, which gives rise to schema-like generalisations (Bybee, 2006:714, 2010:16-17; Lieven *et al.*, 2003). As a result, L1 speakers do the bulk of the work with their procedural memory (Paradis, 2004:9-12, 57-61).³² Items that occur less frequently, are accounted for by more general rules (only acquired later on) and resides in the L1 speakers' declarative memory (Paradis, 2004:9-12, 57-61; Van Rooy, 2010:15).³³ SLA, on the other hand, tends to take place when the learners are cognitively more mature (Corder, 1967:163). When language learning takes place at a later age, the learning burden is shifted to the declarative memory, and the procedural memory no longer plays such a large role (Paradis, 2004:11, 57-61). Therefore, L2 learners' general schemas are less instance-based than the schemas of L1 speakers, whose cognitive and linguistic development takes place concurrently (Van Rooy, 2010:15). It explains why speakers find it harder to acquire an L2 later in life than acquiring an L1.

Since speakers experience a decline in procedural memory for language acquisition, scholars have proposed **the critical age hypothesis** (cf. Pinker, 1994:295-301). According to this hypothesis, language development is “on a maturation timetable, like teeth” and prohibits native-like acquisition from around adolescence (Pinker, 1994:294). As a result, L2 “development often ‘fossilizes’ into permanent error patterns that no teaching or correction can undo” when it is acquired after the critical age (Pinker, 1994:295). Several scholars, such as Aitchison (1991:167-168), Croft (2000:57-58), and Ortega (2009:14-17) have criticised this hypothesis and argue that there is no sudden cut-

³² Paradis (2004:9) explains that **procedural memory** encompasses “the implicit competence which underlies the performance of motor and cognitive skills” and ultimately contribute to the automatic performance of a task. Procedural memory contrasts with declarative memory.

³³ According to Paradis (2004:9), **declarative memory** subserves “everything that can be represented at the conscious level”. It basically encompasses the speaker's encyclopedic knowledge (i.e. knowledge of the world) episodic memory (i.e. the speaker's recollection of past experiences).

off of language learning when speakers reach adolescence. Croft (2000:57-58) also calls into question the privileged position assigned to the native speaker. In this regard, he argues that even adults' "linguistic ability can evolve beyond the so-called critical period" (Croft, 2000:57). Croft's view is couched in the fact that the lexicon of adult speakers continues to change throughout adulthood and that they even acquire additional social and stylistic registers.³⁴ Croft (2000:57-58) goes on to say that even native speakers lose some of their native ability when it is not reinforced by language use in the community around them (also cf. Grosjean, 2013:13). Considering the aforementioned, Croft (2000:58) calls for the abandonment of the distinction made between native speakers and non-native speakers.³⁵

Underlying any discussion pertaining to SLA is the issue of **LANGUAGE CONTACT**. Scholars largely tend to discuss language contact in sociolinguistic terms, but Weinreich (1953:1) points out that when a speaker is bi- or multilingual (i.e. when two or more languages are present in the speaker's individual grammar), it is inevitable that language contact will take place in the speaker's individual grammar.³⁶ Therefore, he regards "[t]he language-using individuals" as "the locus of the contact" (Weinreich, 1953:1). Matras (2009:2-3) explains this psycholinguistic approach to language contact as follows:

[S]peakers' communicative goals and intentions, their discourse strategies, and their language processing capacities are at the core of any speech production and so also of the structural innovations that constitute

³⁴ Paradis (2004:59) attributes this phenomenon to the fact that the vocabulary (i.e. the sound-meaning pairing of words) – unlike prosody, phonology, morphology, and syntax – is supported by the declarative memory. Consequently, the vocabulary of a speaker is not susceptible to the critical period. Here it is important to note that Paradis (2004:60) acknowledges that L2 speakers can achieve native-like performance, but points out that these speakers use "speeded-up control rather than automatic processing". This means that their knowledge is still not of an implicit (i.e. procedural) nature, but of an explicit (i.e. declarative) nature. Interestingly, Givón (2005:95) points out that although "a well-coded lexicon can be acquired by many non-human species [...] [t]he communicative natural use of grammar in non-human species has never been attested" (also cf. Hurford, 2012:515).

³⁵ The problems associated with distinguishing native speakers from non-native speakers are addressed in more detail in Section 2.3.2.

³⁶ Matras (2009:2-4) argues that it is important to investigate language contact from both a psycholinguistic and sociolinguistic perspective. Therefore, language contact is discussed from a sociolinguistic perspective in Section 2.2.2.3.

the seeds of potential language change. They therefore merit consideration when we set out to interpret processes of contact-induced change. 'Contact' is, of course, a metaphor; language 'systems' do not genuinely touch or influence one another. The relevant locus of contact is the language processing apparatus of the individual multilingual speaker and the employment of this apparatus in communicative interaction. It is therefore the multilingual speaker's interaction and the factors and motivations that shape it that deserve our attention in the study of language contact.

According to Matras (2009:2), the psycholinguistic aspects of language contact have not enjoyed much attention in the literature since Weinreich's (1953) work. The last decade, however, saw scholars increasingly considering the psycholinguistic aspects of language contact at work (e.g. Grosjean, 2011:11-12; Kruger & Van Rooy, 2016a:27, 2016b:118-119; Lanstyák & Heltai, 2012:99; Matras, 1998:281-282, 2009; Schneider, 2013:143).

When language contact takes place in speakers' grammars, speakers are often said to be bi- or multilingual. It is important, though, to note that "bilinguals are not simply the sum of two monolinguals in one" (Matras, 2009:61). Matras (2009:61) explains his statement on the basis of two factors. In the first place, he maintains that bilinguals may prefer certain language patterns or patterns of dominance of one language over another. In the second place, bilinguals are not only able to sustain monolingual conversation; they also have the ability to contrast languages in conversation. Consequently, he argues, multilingual speakers demonstrate unique psycholinguistic processing abilities. Upon studying these abilities, scholars recognised that in situations where language contact takes place within speakers' individual grammars, language production in specific

communicative contexts might be influenced by the constraint of **bilingual activation**.³⁷ Treffers-Daller and Sakel (2012:3) explain that when bilingual activation occurs, speakers are juggling their languages. In other words, they are continuously involved in a process of attentional control in which attention and cognitive resources are directed away from one language to another language. This juggling process makes it impossible for L2 speakers to keep their languages separate at all times. As a result, schemas, word-forms, and constructions from the language that is not in current overt production use are often transferred to the active language during communicative events. It is important to be cognisant of the fact that the multilingual constraints are at work to some degree, even when the speaker is in monolingual mode (De Groot & Christoffels, 2006:189; Kruger & Van Rooy, 2016b:119). Therefore, multilingual speakers are said to activate the languages at their disposal to varying degrees in varying contexts, ranging from a wholly monolingual mode to a wholly multilingual mode (Grosjean, 2013:15).

The constraint of bilingual activation often governs the phenomenon known as **CROSS-LINGUISTIC INFLUENCE (CLI)**.³⁸ CLI refers to the influence that the speaker's L1 (i.e. the speaker's deactivated language) exerts on the speaker's L2 (i.e. the speaker's active language) in the form of interferences (Grosjean, 2013:20). Weinreich (1953) was the

³⁷ When language production takes place under certain constraints, it is said that the constraints result in unique language forms known as **CONSTRAINED LANGUAGE** in the literature (Kruger & Van Rooy, 2016a; Lanstyák & Heltai, 2012:100). Lanstyák and Heltai (2012:100) point out that all communicative events are subjected to certain constraints. They identify several kinds of constraints, including physical constraints (e.g. noise, limitations of time and space), physiological constraints (e.g. stress, fatigue, brain lesions), psychological constraints (e.g. emotional state), cognitive constraints (e.g. limitations of memory capacity and attention), linguistic constraints (e.g. the resources available in a given language), and social constraints (e.g. pressure on the individual to observe norms and the requirements of maintaining face). The bilingual activation constraint, then, is a specific kind of constraint that play a role in the language production of multilinguals. It is important to note that there are other constraints that also tend to accompany the constraint of bilingual activation. Kruger and Van Rooy (2016a:27), for example, point out that written text production in non-native varieties of English also constrained by "overly strict conformance to perceived standard language norms". This constraint is listed by Lanstyák and Heltai (2012:100) as a social constraint. A discussion on language norms can be found in Section 2.3.2.

³⁸ In the literature, scholars use different terminology to refer to this phenomenon. They include L1 INFLUENCE, TRANSFER, and INTERFERENCE (cf. Gut, 2011:104). Since the terms NATIVE and NON-NATIVE SPEAKERS are surrounded by so much controversy, and since there are so many indigenous languages in South Africa, it seems prudent to use the term CROSS-LINGUISTIC INFLUENCE (CLI) in this study.

first scholar to analyse this phenomenon systematically. In his analysis, he posits that adult learners initially process their L2 vis-à-vis their knowledge of their L1 (Weinreich, 1953:7-8). This leads Matras (2007:34) to attribute the likelihood of CLI to three factors: (i) the intensity of exposure to the contact language; (ii) the structural similarities and differences (congruence) among the languages concerned; and (iii) the inherent semantic-pragmatic or structural properties of the affected categories. CLI can occur at all levels of language and is exacerbated in the presence of other constraints in the communicative event (Grosjean, 2013:21; Siegel, 2003:187).

At this stage, it is expedient to point out that some scholars have argued that the speakers' L2 influences their L1. In this regard, Cook (2003:1) points out that while scholars readily recognise the influence that speakers' L1 has on their L2, they tend to neglect the fact that the L2 has a similar effect on the L1. In this regard, she states the following:

As well as the first language influencing the second, the second language influences the first. Perhaps this effect is less detectable in our everyday experience: only complex instrumental analysis of a Spanish speaker's accent in Spanish will reveal whether the speaker also knows English. It becomes blatant only when the first language starts to disappear, for instance when a speaker brings more and more L2 words into his or her first language.

(Cook, 2003)

In the first compilation of its kind, authors such as Cook, Iarossi, Stellakis and Tokumaru (2003), Jarvis (2003), Murphy and Pine (2003), and Pavlenko (2003) illustrate the influence that an L2 can exert on an L1. Despite the fact that these scholars have demonstrated that an L2 can influence an L1, this topic has remained under-researched with only a few scholars (e.g. Ulbrich & Ordin, 2014) investigating the phenomenon until recently. In the following paragraphs on CLI, all references to the influence that the L1 exerts on the L2 subsumes the influence that the L2 exerts on the L1.

According to Grosjean (2011:14-17, 2013:21) and Paradis (2004:214-216), two kinds of CLI can be distinguished. The first is **transfer** which refers to static phenomena. It is reflected in the permanent traces of the L1 (also known as the **substrate language**) on

the L2 (also known as the **superstrate language**).³⁹ These phenomena are linked to the speaker's L2 competence. Therefore, Corder (1967:166-167) refers to this phenomenon as **errors of competence** (i.e. L2 learners' systematic errors). The second is **interference** which refers to dynamic phenomena. These phenomena are linked to "the ephemeral intrusions of the other language" (Grosjean, 2011:15). They are regarded as accidental and are linked to processing, and are therefore related to the speaker's performance. Therefore, Corder (1967:166-167) referred to this phenomenon as **errors of performance** (i.e. L2 learners' random mistakes).

Transfer is of particular importance here. Weinreich (1953) distinguishes between two kinds of transfer. The first kind of transfer takes place when the multilingual speaker models an L2 structure on an L1 structure in order to facilitate the correct acquisition of the corresponding L2 structure. In the literature, this kind of transfer is known as **positive transfer**. The second kind of transfer takes place when the multilingual speaker erroneously models an L2 structure on an L1 structure. In the literature, this kind of transfer is known as **negative transfer**. Siegel (2003:187) argues that language features exhibiting transfer from the L1 do not always immediately appear as part of the grammar of the L2. He argues that when speakers attempt to speak a common L2, they transfer features from their L1 to their L2. These L2 forms with L1 properties are added to the linguistic feature pool used in the contact situation (cf. Section 2.2.1.1). When the speech community starts to shift from their L1 to the contact variety, **levelling** (when some features are selected while other features are discarded) occurs (also cf. Mufwene, 2008b:133-159). Therefore, certain features of CLI are transitional in nature and disappear after speakers have achieved a certain level of proficiency, whereas other features of CLI remain as persistent features of a stabilised L2 variety (Gut, 2011:102-103). CLI (more specifically transfer) is often cited as the reason for the

³⁹ Within the New Englishes paradigm, the indigenous languages tend to be the substrate whereas the European languages tend to be the superstrate. Sebba (1997:25) points out that the vocabulary of the superstrate resting on the grammar of the substrate "bears more than a passing resemblance to the colonialists' model of a (European) social elite resting (literally!) on a social inferior indigenous underclass". He points out that this situation paints an unfortunate picture of the social relations between the STL strand and the IDG strand in the postcolonial setting. Therefore, Sebba (1997:25-26) cautions against the use of the term *substrate*, saying that it "suggests that indigenous speakers somehow played a secondary or inferior role" in the emergence of the new varieties of English. The social relationships between the STL strand and the IDG strand and the resulting attitudes towards the new varieties of English are discussed in more detail in Section 2.3.

emergence of certain features observed in new varieties of English, but Mufwene (2008b:158) cautions that scholars should ensure that sufficient evidence exists before attributing features to transfer. Therefore, other aspects – such as the cognitively demanding environment of language acquisition which leads to analogy or overgeneralisation in L2 acquisition (McWhorter, 2001; Trudgill, 2001) and contextual and historical factors (Schneider, 2007:3) – should also be considered.

The feature of **simplification**, for example, is said to be a coping mechanism employed by adult L2 learners when attempting to deal with the cognitive difficulty of acquiring an additional language (Kortmann & Szmrecsanyi, 2009:21-26; McWhorter, 2011; Trudgill, 2001:372). This feature is said to be present when a variety of language displays a simpler grammar than other varieties of a language. The hypothesis is that creoles tend to be grammatically simpler than standard varieties (Kortmann & Szmrecsanyi, 2009:266; McWhorter, 2001:162). Trudgill (2001:372) explains this phenomenon by saying that “the longer a language exists, the more ‘historical baggage’ it acquires”.

Another feature of non-native varieties of English, which is linked to the feature of simplification, is the feature of analyticity (Kortmann & Szmrecsanyi, 2009). Kruger and Van Rooy (2016a:31) define **analyticity** as “the more transparent and regular mapping of form and function”. This feature leads to constructions that are more explicit. L2 speakers employ these strategies to ensure that their intended meaning is transparent and to avoid misunderstanding when communicating with other members of the speech community (Kruger & Van Rooy, 2016a:31; Rohdenburg, 1996:151-152). In this regard, Van Rooy, Terblanche, Haase and Schmied (2010) have found that East African speakers demonstrate a preference for more explicit referencing, explicit constructions, and increased repetition. In turn, Mesthrie (2006) has investigated the preference for the anti-deletion of the infinitive *to*-complementisers and the finite *that*-complementisers in BSAE.⁴⁰ Similarly, Kruger and Van Rooy (2016b) have found a consistent preference for the use of the explicit *that*-complementiser in the reported speech of L2 varieties of English influenced by Afrikaans. This phenomenon can be explained with Hawkins’s (2003, 2014:34-38) theory of processing efficiency. The theory posits that there is a

⁴⁰ Mesthrie (2006) attributes the anti-deletions of these complementisers not only to a need to be as explicit as possible in order to avoid any potential misunderstanding, but also to overgeneralisation during the acquisition of the L2.

systematic preference for the more explicit variant containing the overt complementiser – even when there is no danger of ambiguity (Hawkins, 2003:200). This is in line with Rohdenburg's (1996:151-152) **complexity/transparency principle** (on the basis of Hawkins, 1990), which posits that “[i]n the case of more or less explicit grammatical options, the more explicit one(s) will tend to be favored in cognitively more complex environments”. In addition, Kruger and Van Rooy (2016b:129) argue that the explicit use of the *that*-complementiser is the result of cognitive entrenchment since the L2 speakers are more likely to encounter formal texts in the L2 context.

Mesthrie and Bhatt (2008:162) point out that some features of the L2 varieties of English are indicative of more conservative, standard-like choices. At this point, it is useful to consider the emergence of the L2 varieties from a sociolinguistic perspective. Therefore, the social contexts that play a role in the emergence of the new varieties of English are discussed in more detail in the following section.

2.2.2.3 *Investigating language change from a sociolinguistic perspective*

As mentioned earlier, the social contexts of speech events play an integral role in language. Kachru (1996:135) points out that the social nature of language gives rise to issues such as diversification, codification, identity, creativity, cross-cultural intelligibility, power, and ideology and these issues play an important role in language change. Furthermore, the investigation of the social nature of language provides valuable insights into descriptions pertaining to the linguistic features of the variety used (cf. Croft, 2000:90; Hudson, 1996:2-3; Schneider, 2007:14-16). In terms of the overarching aim of this dissertation, the investigation of the social nature of language will help determine the extent to which specific verb complementation patterns have conventionalised in BSAE.

Croft (2000:90) states that there would not be language without people. He goes on to say that “[w]e cannot talk about language use – speech acts in conversation and the non-linguistic actions they serve – without talking about the people that use them” (Croft, 2000:90). Sociolinguistics is typically the field in which we study language in relation to society – and by implication, people (Hudson, 1996:1). Croft (2000:92) and Tomasello (2008:72-73) define a **society** as a group of speakers who interact with one another socially. Within the society, individuals interact with one another in order to achieve some

shared goal by utilising processes of shared intentionality. When the members of the society achieve their shared goal, they have achieved **social success** (Croft, 2000:88-90). Tomasello (2008:171) explains it as follows:

[I]ndividuals care about the reputation they have among others in the social group, since having a reputation as a good helper and co-operator contributes in important ways to social success.

Every society consists of multiple **speech communities** (Croft, 2000:92). These speech communities play a central role in understanding the social nature of language. Although individuals belong to multiple speech communities, each individual can only identify with one specific speech community at any given time. The community with which the individual identifies depends on what is relevant and important in the circumstances. Wardhaugh (1992:122) explains it as follows:

You are a member of one speech community by virtue of the fact that on a particular occasion you identify with X rather than Y when apparently X and Y contrast in a single dimension.

Wardhaugh's point above is reiterated by Croft (2000:181) when he states that speakers either may wish to identify with a more powerful group or they may wish to identify with a less powerful group. Usually, the variant used by the more powerful group forms the **standard variant**, whereas the variant used by the less powerful group form the **vernacular**.

Each of these speech communities uses a specific **code** – a language (or a variety of a language) – that two or more people employ for the purposes of communication in a specific community or domain (Wardhaugh, 1992:89). Croft (2000:92) emphasises the fact that “every language is a multiplicity of codes”. It means that English, for example, is regarded as a set of codes (e.g. the language of sociolinguistics, the language of classical music, and so on). Each of these codes is specialised to varying degrees. Individuals within the English speech community will only be familiar with these codes to the extent that they are members of the relevant communities. Furthermore, Croft (2000:92) explains that there is a substantial overlap in the codes used in the communities of a single society.

As a result of this overlap, the internal-external boundary in the theory of language change is broken down, and this is one of the reasons why languages are fundamentally variable.

An issue that relates quite strongly to the concepts *SPEECH COMMUNITY* and *CODES* is the issue of **identity**. Jenkins (2008:11-12) defines identity as follows:

[T]he human capacity – rooted in language – to know ‘who’s who’ (and hence ‘what’s what’). This involves knowing who we are, knowing who others are, then knowing who we are, us knowing who they think we are, and so on: a multi-dimensional classification or mapping of the human world and our places in it, as individuals and as members of collectives.

This definition considers not only individual identity, but also social identity. According to Hudson (1996:10-11), it is important that the individual be considered (even in sociolinguistics) as no two speakers experience language in the same way.⁴¹ At the same time, social identity also plays a crucial role, as language users use language to assume a particular social identity (Deumert, 2010a:245; Labov, 2001:xv). They achieve this goal by using certain linguistic characteristics – together with a number of other characteristics, such as social, cultural and political characteristics – to achieve group identity with other language users (Wardhaugh, 1992:118). Mesthrie, Swann, Deumert and Leap (2000:148) regard language varieties as meaningful, as “they are indexical of a speaker’s origin or of aspects of their social identity [...] but they also carry certain social values related to the speakers who use them and the contexts in which they are habitually used.” Therefore, sociolinguistically speaking, it is useful to group language users belonging to specific groups according to dimensions of age, region of origin, social class and gender (Hudson, 1996:11-12).

In his Dynamic Model for the description of New Englishes (cf. Section 2.3.1.4), Schneider (2003:239) explicitly focuses on this notion of *SOCIAL IDENTITY* by pointing towards its

⁴¹ The fact that individuals have their own idiolects has already been mentioned earlier in Section 2.2.2.2. There the point was made that the mental grammars of individual language users are not identical and, as such, individual language users have their own distinct linguistic systems. From a sociolinguistic perspective, Hudson (1996:10-11) makes the same point when he says that speakers experience reality differently.

“construction and reconstruction by symbolic linguistic means”. It involves drawing a line between individuals who form part of the speech community (also known as *us* in postcolonial literature) and individuals who do not form part of the speech community⁴² (also known as *them* in postcolonial literature). This distinction is based not only on regional and social histories but also on value orientation and custom modes of behaviour. Schneider (2003:239) posits that the expression of this distinction tends to be realised more readily and rapidly by means of linguistic variability within the speech community than by any other means. He attributes this phenomenon to the fact that it is usually more costly, difficult, and sometimes impossible to try and express identity boundaries and solidarity by other means (Schneider, 2003:240).

Furthermore, it is important to note that linguistic and social identities are dynamic. This means that “speakers keep redefining and expressing their linguistic and social identities, constantly aligning themselves with other individuals and thereby accommodating their speech behavior to those they wish to associate and be associated with” (Schneider, 2007:21). This idea is central to theories about the notion of **ACCOMMODATION**⁴³. The theory of accommodation plays an integral role in sociolinguistics and is based on theories in social psychology (Croft, 2000:181). The theory is used to argue that speakers alter their speech in order to accommodate to the hearers (Giles, 1973:89-90). Boylan (2009) points out that accommodation is not a mere mimicry of the participants in the social interaction, but rather an identification with their way of “*seeing and saying things*” (Boylan, 2009:288; his emphasis).

Two kinds of accommodation may take place. The first kind of accommodation takes place when speakers recognise that the hearers are from a different speech community and therefore make adjustments to their speech to accommodate them. These adjustments may include slower speech rates, elaboration of content (to allow for the lack of common ground), and the simplification of grammatical constructions (Croft, 2000:182; Giles, 1973:90). The second kind of accommodation takes place when speakers make

⁴² In other words, those with whom the members of the speech community do not share these social and historical qualities.

⁴³ Some scholars (e.g. Thomason, 2001:142-146) use the term **NEGOTIATION** to refer to instances where speakers change their language to approximate the patterns of another language or language variety.

adjustments to their speech to shift towards the conventions of the language of the speech community to which the hearers belong. These adjustments may include changes in pronunciation, pause and utterance lengths and vocal intensity. In cases such as these, the speakers adjust their speech in order to identify with the community of the hearer (Croft, 2000:182; Giles, 1973:90). Consequently, accommodation is said to affect language at any linguistic level (Tuten, 2003:29). It is important to be cognisant of the fact that accommodation is not a conscious action (Thomason, 2001:42; Trudgill, 2008:243; Tuten, 2008:260).

Scholars, such as Schneider (2003, 2007, 2008) and Tuten (2003, 2008), argue that accommodation and identity play a pivotal role in the emergence of new varieties of English. Tuten (2003:29) formulates this point of view as follows:

Given that most contributing varieties in a prekoine linguistic pool are mutually intelligible [...] many of the alterations in speech that take place are not strictly speaking necessary to fulfil communicative needs [...] Rather, *speakers accommodate to the speech of their interlocutors in order to promote a sense of common identity.*⁴⁴

Similarly, Croft (2000:73) states that accommodation is inherently a social phenomenon. Since accommodation takes place when the speaker wants to identify with the other members of a speech community, identity is said to be directly related to accommodation. Therefore, the linguistic changes wrought during accommodation are driven by the speaker's desire to achieve certain social success.

Considering the aforementioned, there is an interesting relationship between identity and accommodation. Schneider (2008:264) explains this relationship by saying that the notions of IDENTITY and ACCOMMODATION are the flip sides of the same coin. On the one hand, accommodation is a process in which solidarity is realised by means of linguistic approximation in order to attain group cohesiveness. On the other hand, identity is

⁴⁴ Tuten (2008:260) acknowledges Trudgill's (2008:243) view that Tuten's use of the phrase "in order to" suggests that accommodation is a conscious action, and clearly states that he does not regard accommodation as a conscious action. This is in line with the statement made earlier on that accommodation is not a conscious action.

representative of the individual's stance concerning the social structures present in the community. Consequently, Schneider (2008:264) regards accommodation as a mechanism used to express one's identity choices. Tuten (2008:259-260), in turn, argues that accommodation allows the members of different communities to cooperate and negotiate common practices which may result in a new rudimental community identity.

When considering issues concerning social identity and accommodation, it is also important to consider issues concerning **multilingualism**, as it is a contributing factor in language change (Bybee, 2015:248; Croft, 2000:92).⁴⁵ Mair (2013:260) posits that monolingualism is "a luxury which few speakers of a peripheral language can afford". Thomason (2001:31) goes further and states that "[t]he idea that monolingualism is the human norm is a myth". Deumert (2011:262) proves this point when she states that even Iceland, traditionally described as being monolingual, is, in fact, multilingual as Icelanders use three different languages on a regular basis. They are Icelandic, Danish, and Icelandic Sign Language. Despite the common occurrence of multilingual communities, Croft (2000:92) points out that linguistic theories are based on the assumption that speech communities are monolingual. Kachru (1996:252) attributes this state affairs to the following:

We are still hesitant to cross the threshold and face the complexities of multilinguals' language behavior and the impact of that language data on our hypotheses and our attitudes. We are reluctant to modify, reformulate, revisit and reassess our favorite paradigms.

It is critical to account for the multilingual nature of speech communities since the presence of multilingual speakers can give rise to language change (Bybee, 2015:248). According to Lewis, Simons and Fennig (2016), there are currently 7 097 known languages in the world. More specifically, in South Africa alone, there are 30 living

⁴⁵ Multilingualism can be discussed on two levels, namely **individual multilingualism** and **societal multilingualism** (cf. Deumert, 2011). Individuals are said to be multilingual when they speak more than one language. Societies are regarded as multilingual when more than one language is spoken by the majority of people (Mesthrie *et al.*, 2000:37). In both individual and societal multilingualism, multilingualism can be regarded as a question of degree (Bybee, 2015:248; Deumert, 2011:264). This means that it can range from knowledge of a few words in a language other than the mother tongue to competency in more than one language.

languages – 20 of which are indigenous and 10 of which are non-indigenous (Lewis *et al.*, 2016). Due to the multilingual nature of many speech communities, language contact is inevitable. Thomason (2001:1) defines the notion of **LANGUAGE CONTACT** as “the use of more than one language in the same place at the same time”. Language contact played a pivotal role in colonialism since the STL strand and the IDG strand came into contact with one another which led to the emergence of the new varieties of English. Deumert (2011:263) also states that the increasing pace of international migration and human mobility has further contributed to multilingualism (and language contact) since the beginning of the twentieth century.

Thomason (2001:86-88) lists a number of changes that can be observed when languages come into contact with one another. They are feature loss (e.g. the loss of phonological, morphological, lexical, and syntactic features), feature addition (e.g. the addition of phonological, morphological, lexical, and syntactic features primarily by means of borrowing), the replacement of old native linguistic features with new interference features. Here it is imperative to note that Thomason (2001:88) attributes the effects of contact-induced changes to the competition between innovative features and conventionalised features (cf. Section 2.2.1). The intensity of the language contact determines the kind of features that change (Thomason, 2001:69). For example, when low-intensity contact takes place, elements of the other language’s lexicon are more likely to be borrowed. This situation can be attributed to the fact that speakers need not be fluent in a language to borrow lexical items from that language. However, the more intense the language contact becomes and the more multilingualism improves, the more likely it becomes that speakers will borrow structural features in addition to the lexical items. It means that speakers will not borrow structural features before lexical items.

Thomason (2001:69-71) proposes that the different levels of contact can be illustrated with a borrowing scale. She identifies four points on this scale: (i) casual contact; (ii) slightly more intense contact; (iii) more intense contact; (iv) intense contact. Scholars apply this scale in the colonial context. When the IDG strand came into contact with the STL strand for the first time, the IDG strand was not fluent in English. Nor was the STL strand fluent in the indigenous languages. Therefore, casual contact took place, and the STL strand only borrowed content words (mostly nouns used to signify fauna, flora, and places) from the indigenous languages (Thomason, 2001:70). As the IDG strand and the STL strand started to interact more frequently, the second stage of the borrowing scale

was reached. As the STL strand occupied the position of power, the IDG strand became more fluent in English. The IDG strand started borrowing not only function words, but also structural features from their mother tongue (Thomason, 2001:70). Gradually, the contact became more intense (i.e. more members of the community became bilingual, and certain social factors favoured more intense borrowing). This situation contributed to the borrowing of more function words as well as the borrowing of more significant structural features (Thomason, 2001:70). Finally, when intense contact was reached, the majority of the population was bilingual, and the social factors favoured borrowing (Thomason, 2001:70-71).

Elements of Thomason's (2001:86-88) borrowing scale is evident in models describing the emergence of new English varieties, such as Schneider's (2003) Dynamic Model of the Evolution of New Englishes discussed in Section 2.3.1.4.

2.2.2.4 *Summary*

The interaction between language as an internal construct and language as an external construct has been considered in Section 2.2.2.1. It is pointed out that it is crucial to consider the interaction between these two ecologies to gain a full understanding of how languages evolve.

Psycholinguistic investigations enable linguists to determine how language variation originates from within individual speakers. Individual grammars, SLA, bilingual activation, and CLA all explain how language evolves in internal ecologies.

Sociolinguistic investigations, on the other hand, allow linguists to determine the role that language plays in the identity constructions of speakers, and how speakers adjust their language to accommodate hearers.

2.3 The spread of New Englishes

2.3.1 *Models describing the spread of New Englishes*

This section is concerned with gaining insight into the emergence of new varieties of English. First, an overview is provided of the models found in the literature on New Englishes to describe the spread of New Englishes. These models include Strang's tripartite model of English (cf. Section 2.3.1.1), Quirk's views on the varieties of English (cf. Section 2.3.1.2), Kachru's three-circle model (cf. Section 2.3.1.3), and Schneider's Dynamic Model of the Evolution of New Englishes (cf. Section 2.3.1.4).

While some linguists advocate assigning status to new varieties of English, other linguists (specifically those who advocate for the prescriptive use of standard English) oppose the legitimisation of these varieties. Consequently, Section 2.3.2 focuses on the attitudes towards New Englishes. The knowledge obtained from Section 2.3.1 and Section 2.3.2 is synthesised in Section 2.3.3 to outline the position BSAE occupies in the South African landscape.

2.3.1.1 *Strang's tripartite model of English*

Strang (1970:17) subscribes to the view that language is both internal (in the sense that acquisition takes place in the mind of the individual) and social (in the sense that learning takes place due to exposure to the social use of language). She points out that should individuals wish to be understood and accepted by the speech community, they cannot be too idiosyncratic. Taking the argument above as a point of departure, Strang (1970:17) states that whereas "the governing conditions come from society [...] executive language acts (speaking, writing, etc.) are made by individuals".

In terms of language as a social construct, Strang (1970:17-18) distinguishes between A-speakers, B-speakers, and C-speakers. She maintains that the spread of English has necessitated a model to account for all the different varieties of English. She defines **A-speakers** as those speakers who speak English as their mother tongue and comprises speakers from countries such as the UK, the USA, Canada, Australia, New Zealand, and

South Africa. **B-speakers**, on the other hand, are defined as speakers for whom English is not necessarily a mother tongue, but who learnt English in early childhood. She points out that English in these communities tends to have a special status. English in these communities may or may not be regarded as an official national language and is regarded as the language to be used when dealing with advanced academic work and for participation in national and international matters. B-speakers are found in Asia (especially India) and Africa (especially the former colonial territories). Finally, there are **C-speakers**. These speech communities employ English as a first foreign language as part of the country's educational curriculum. In these communities, English has no official or traditional standing in the country.

Strang (1970:18-19) emphasises the fact that even though countries in which the speech communities are regarded as A-speakers of English for the most part, not all members of the speech community have mastered the language equally. As such, "[t]he equation of 'population of English-speaking country' with 'size of English-speaking community' cannot be made without important qualifications" (Strang, 1970:18). Furthermore, it is also important to recognise that there is considerable variety amongst, for example, the A-speakers, not only in terms of individual, idiolectal variations but also in terms of the distinct institutionalised varieties.

Approximately a decade later, Quirk adopted this model in describing the spread of English.

2.3.1.2 *Quirk's views on the varieties of English*

Quirk *et al.* (1985) describe the spread of English in terms of Strang's (1970) model. The English of the A-speakers are called **English as Native Language** (ENL), the English of B-speakers are called **English as Second Language** (ESL), and the English of C-speakers are called **English as Foreign Language** (EFL).

Just as Strang (1970) expounds on the various functions that English has for the speakers of English, Quirk *et al.* (1985) maintain that ESL has five functions:

(1) *instrumental*, for education; (2) *regulative*, for government administration and the law courts; (3) *communicative*, for interpersonal communication between individuals speaking different native languages; (4) *occupational*, both intranationally and internationally for commerce and for science and technology; (5) *creative*, for nontechnical writings, such as fiction and political works.

(Quirk *et al.*, 1985:4)

Another aspect to consider is that whereas ESL and EFL are described in a seemingly objective manner in Quirk *et al.* (1985), Quirk has very particular ideas of what language variation should entail. He believes that variation can only be attributed to “our social and regional backgrounds, our professional careers, and indeed our creative urges as individuals” (Quirk, 1990/2003:98). He postulates that deviations from the standard should not be allowed for any other reason. The implication is that whereas varieties such as Legal English (related to the legal profession) and Scientific English (related to the sciences) are perfectly acceptable, varieties that do not resemble L1 English linguistically (e.g. South Asian English and African English) are not acceptable at all. He claims that the “profusion and [...] confusion of *types* of linguistic variety that are freely referred to in educational, linguistic, sociolinguistic, and literary critical discussion” is a serious issue that requires careful consideration (Quirk, 1990/2003:98; his emphasis). Furthermore, he is also of the opinion that ESL and EFL speakers should not be allowed to teach English:

Not merely may their own English be far from standard but they may have little respect for it and may well have absorbed (at second or third hand) the linguistic ethos that is simplified into the tenet that any English is as good any other.

(Quirk, 1990/2003:401)

In regarding any variety of English as good as the other, Quirk (1990/2003:101-102) believes that educators are trapping students in their social and ethnic circumstances by creating barriers to their educational progress, their career prospects and their social and geographical mobility. Kachru (1991), however, points out that what Quirk fails to consider is the issue of limited resources: there simply are not enough able and willing L1 English-speaking teachers to provide in the demand across the world.

Not all linguists agree with Quirk's views on the importance of standard English. One such linguist is Kachru. An overview of Kachru's views follows, after which an explanation is provided on how Kachru's views differ from Quirk's.

2.3.1.3 *Kachru's three-circle model*

Kachru's (1996:135-136) model is underpinned by the view that English is a **pluricentric** language.⁴⁶ He attributes the pluricentric nature of English to the diaspora of English. He classifies the diaspora into three phases: the first phase involves the spread of English to Wales and Scotland; the second phase involves the spread of English to North America, Canada, Australia, and New Zealand; and the third phase (what Kachru calls the RAJ PHASE) involves the spread of English to Asia, Africa and the Philippines. Kachru (1996:136-137) especially attributes the pluricentric nature of English to the third phase during which English was influenced by the non-Western cultures and their sociolinguistic contexts. The pluricentric nature of English is accounted for in three concentric circles in Kachru's (1992c) model: the "Inner Circle", the "Outer Circle" and the "Expanding Circle". The circles "represent the types of spread, the patterns of acquisition, and the functional allocation of English in diverse cultural contexts" (Kachru, 1992c:356-357). Upon considering the three circles, it becomes clear that they broadly correspond to the ENL/ESL/EFL distinction.

The **Inner Circle** comprises countries such as the United Kingdom (UK) and the United States of America (USA) where English is spoken as a native variety (Kachru, 1992c:356). The varieties of English in these countries are regarded as norm-providing, as they are considered representative of the traditional cultural and linguistic bases of English (Kachru, 1985:16, 1996:138). English is the vernacular language of a significant portion of the population in these countries.

The **Outer Circle** comprises countries such as India and Singapore, which are said to be ESL countries. The varieties of English in these countries are regarded as norm-

⁴⁶ Kachru (1996:135) defines pluricentric languages as languages that "have multilinguistic identities, multiplicity of norms, both endocentric and exocentric, and distinct sociolinguistic histories".

developing (Kachru, 1985:17, 1992a, 1996:138). This means that even though English coexists with strong indigenous languages, non-native varieties of English are spoken widely and are often used in an official capacity. As a result, these non-native varieties of English have become institutionalised and are developing their own standards.

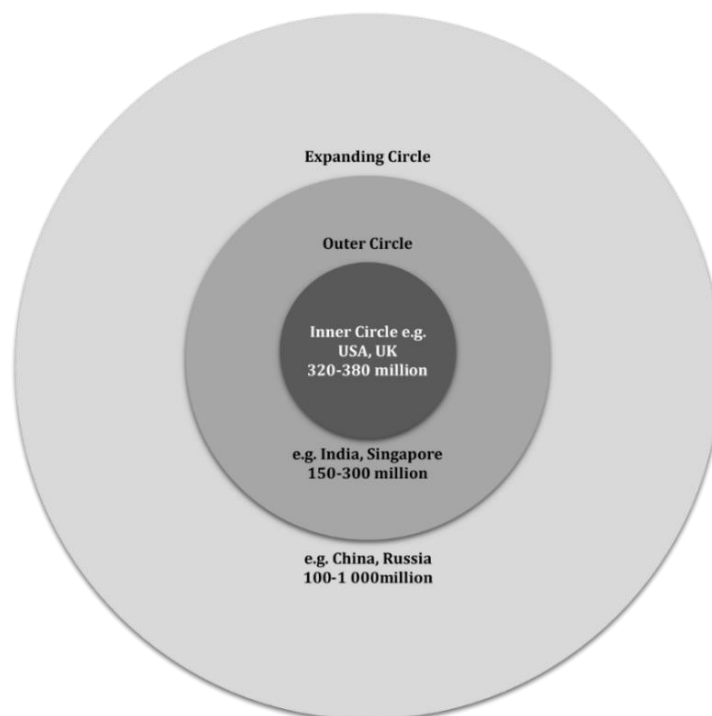


Figure 2.1: A visual representation of Kachru's three-circle model (Crystal, 1997:54)

The most prominent feature of an institutionalised variety is a large geographical spread (Bamgboṣe, 1998:4). It indicates the large-scale selection and propagation of the features that are characteristic of the variety. In this regard, Bamgboṣe (1998:4) states that “[t]he greater the geographical spread [...] the higher is its acceptance as a standard form”. Furthermore, Kachru (1992a:55) points out that an institutionalised variety always develops from a performance variety. As the performance variety acquires stable characteristics, it gains the status of being institutionalised. Some of these features include its extension of use, the emotional attachment that the L2 speakers have with the variety, its functional importance, and its sociolinguistic status (Kachru, 1992a:55). Moag’s (1992:241-245) view of the institutionalisation of new varieties overlaps with Kachru’s point when he says that a new variety can be considered institutionalised when authors write literary works in the variety, when the new variety is used in education, and when the variety is used in print media.

Finally, the **Expanding Circle** comprises countries such as China and Russia, which are said to be EFL countries. The varieties of English in these countries are regarded as norm-dependent, i.e. these varieties are regarded as performance varieties which have no official status and are therefore dependent on the standards provided by the L1 English speakers in the inner circle (Kachru, 1985:17, 1992c:356-357, 1996:138, 147).

Kachru (1991) unequivocally disagrees with Quirk's (1990/2003) view that native varieties of English are superior to second-language varieties. He argues that linguists should be more concerned with the non-native varieties of the Outer Circle and the Expanding Circle, as these are the circles in which the most linguistic development can be observed (Kachru, 1992c:357-358). Therefore, he argues that the Inner Circle countries should no longer determine the norms for the Outer Circle. Instead, the norms should be determined by the Outer Circle countries for themselves.

Despite the fact that many linguists (especially applied linguists) have adopted Kachru's model, the model has certain shortcomings. Kachru (1996:137) himself points out that his model is lacking in that it is difficult to account for the complex linguistic situations in countries such as South Africa, Ireland, and Jamaica. This is the same problem the ENL/ESL/EFL model faced (cf. Section 2.3.1.1).

Bekker (2009:70-77, 2012a:128-134) describes the development of English in South Africa. His description demonstrates that the linguistic situation in South Africa is complex due to the different contact situations between different groups, at different times and different places. To complicate matters even more, Van Rooy (2014a:22) points out that the history of segregation during the 20th century in South Africa influenced the nature of the contact in that the various groups each experienced the contact very differently. South Africa is also a multilingual country with eleven official languages. Despite South Africa's multilingual nature, many speech communities adopted English as their L1 and, therefore, it becomes difficult to classify South Africa as an ENL country or an ESL country (Schneider, 2007:12).

Furthermore, Kachru (1992a) is of the opinion that it is difficult to investigate micro-linguistic and descriptive issues, and therefore takes a more macro-linguistic approach when addressing negative attitudes towards the non-native varieties of English. Some scholars regard both the ENL/ESL/EFL model and Kachru's three-circle model as "rather

superficial and fuzzy in their capacity for establishing categories of linguistic description and classification” (Schneider, 2007:14).⁴⁷ The importance of micro-linguistic analysis is demonstrated in Labov’s (1994, 2001, 2010) series, *Principles of linguistic change*. In this series, Labov proves that the variation found in language is not necessarily free and chaotic, but inherently structured. Similarly, micro-linguistic studies in New Englishes have revealed remarkably similar characteristics despite the many variables in terms of the substantial difference between the indigenous languages and cultures that played a pivotal role in their development (Schneider, 2003:234).⁴⁸ Based on these analyses, scholars such as Chambers (2003:22-23) argue that language variation cannot be dismissed as irrelevant.

In many recent studies, scholars have worked towards bridging the gap between macro-linguistic and micro-linguistic descriptions in their work on the new varieties of English. These studies have demonstrated that more insight can be gained in the macro-linguistic aspects by considering micro-linguistic aspects. Examples of such studies are the work of Van Rooy (2006, 2008a, 2014b) and Kruger and Van Rooy (2017) on BSAE; the work of Schmied (1991) and Van Rooy (2011) on East African English; and the work of Bernaisch *et al.* (2014), Mukherjee (2010), and Mukherjee and Gries (2009) on Indian English.

Melchers and Shaw (2003:29-41) provide a more flexible framework for the classification of New Englishes. Their framework not only allows for a macro-linguistic approach but also for a micro-linguistic approach. It considers the following aspects: (i) the political classification of the ideological stances (i.e. conservative, liberal, and radical approaches); (ii) the issue of standardisation and codification and language use in writing; (iii) the type of prestige the variety gives its speakers; (iv) the issue of standardisation and the degree of editorial intervention; (v) the domains of English use in various countries (following

⁴⁷ Schneider (2003:239) also points out that although Kachru does not regard the inner circle varieties as superior to the outer and expanding circle varieties, the fact that the ENL countries form the “centre” of the three-circle model may inadvertently place the inner circle varieties in a position where it is “entitled to establish norms of correctness”.

⁴⁸ With regard to the similarities of characteristics in New Englishes, consider, for example, Mesthrie (2004).

Kachru's three-circle model); (vi) and the English proficiency of speakers in various countries.

Schneider (2007), Van Rooy (2013), and Bolton and Botha (2015) provide overviews of the kinds of research that have been conducted in terms of the description and classification of New Englishes. These overviews consolidate the direction of scholarship to more general approaches in linguistic enquiry (Schneider, 2007:15-16). The overview arranges the kind of research that has been done along two dimensions: "attention to linguistic structure" and "level of generality". The first dimension allows linguists to describe the structure of new varieties by investigating their phonology, lexis, and grammar, whereas the second dimension allows linguists to both observe new varieties by means of case studies, and to make generalisations of some kind regarding all the new varieties of English. (These categories are illustrated in Figure 2.2.)

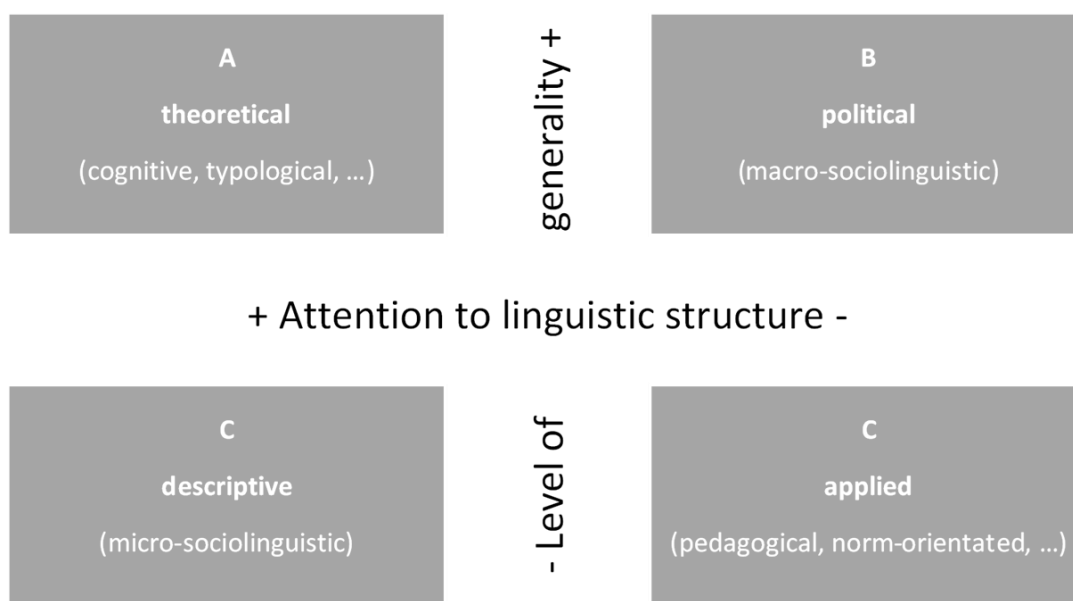


Figure 2.2: *The various approaches to investigating new varieties of English (Schneider, 2007:15-16)*

In terms of the kind of research being done in New Englishes, Schneider (2007:16) observes that linguists from developing countries tend to focus more on questions of an applied and political nature as they are more exposed to the immediate practical needs of society. Linguists from developed countries, on the other hand, tend to ask general, comparative, and theoretical questions to provide a more objective description. Despite these differences, he feels that both positions are legitimate and need to complement one

another to be effective. Based on this belief, Schneider (2003, 2007, 2014) proposes a model for investigating New Englishes (known as the Dynamic Model of the Evolution of New Englishes) that not only takes into account issues of applied and political nature, but also allows linguists to ask theoretical and descriptive questions in order to focus more on linguistic structure. An overview of this model is provided in Section 2.3.1.4.

2.3.1.4 *Schneider's Dynamic Model of the Evolution of New Englishes*

Schneider's (2003) Dynamic Model of the Evolution of New Englishes posits that a fairly uniform process underlies the emergence of New Englishes, specifically Postcolonial Englishes (Schneider, 2003:241, 2014:10).⁴⁹ This uniform process accounts for the similarities that can be observed between the different varieties of New Englishes. The similarities are grounded in the sociolinguistic conditions that are characteristic of colonisation. Based on this view, Schneider (2003:241) formulates the focal point of his theory as follows:

But the essential point of my model of two intertwined strands is that both groups who share a piece of land increasingly share a common language experience and communication ethnography, and thus the forces of accommodation are effective in both directions and in both communities, and result in dialect convergence and increasingly large shared sets of linguistic features and conventions. The end result is thus the emergence of an overarching language community with a set of shared norms.

Several theories inform the Dynamic Model. They are theories of language contact, sociolinguistics, social identity and language evolution (Schneider, 2014:11). In terms of these theories, two concepts play a central role in Schneider's (2003) model. The first is

⁴⁹ Schneider (2007:55) cautions that models are meant to "provide a uniform description of a set of processes that have occurred independently of each other in reality", and that it should not be confused with reality itself. Therefore, he argues, his model by no means claim to account for every complex reality (Schneider, 2007:55): it simply attempts to account for the similarities that might have gone unnoticed. Considering the aforementioned, he goes on to say that his Dynamic Model can be improved upon, modified and developed further to "provide an ever-closer match to reality" (Schneider, 2007:55).

the concept of ECOLOGY (cf. Section 2.2.1.2) and the second is the concept of SOCIAL IDENTITY (cf. Section 2.2.2.3). Using these concepts as a point of departure ensures that the focus is on the different speech communities and the social environment in which they are situated instead of generalising the speech communities based on the country in which they are situated.

Identity pertains to how individuals and communities define who they are and whom they wish to be. Schneider (2003:239) explains that:

Based on regional and social histories but also value orientations and customary modes of behaviour, a line is drawn between ‘us’ (those who share essential parts of that history and orientation, those we wish to socialize and be associated with) and ‘others’ (who don’t share these qualities), and these attitudes and socialization patterns usually find symbolic expression—including (and perhaps most readily and rapidly) by means of linguistic variability.

In terms of colonisation, Schneider (2003:240) points out that when the STL strand and the IDG strand came into cultural and linguistic contact with one another, they needed to “define and redefine themselves and their social roles in light of the presence of the other groups, of their own historical roots and cultural traditions, and their relationship to territories and distant centers of political and military power”.⁵⁰ These changes resulted in language changes which were an expression of their changing identities.

The Dynamic Model comprises a progression of five distinct stages. The first stage is known as the **foundation phase** (Schneider, 2003:245-246). This stage is characterised by the use of English on a regular basis in a country that has never had an English-speaking community before. It is the result of the migration and settlement of a significant group of English speakers in a new country due to certain political or economic motivations in the homeland. (Section 2.2.1.2 explains how these motivations resulted in

⁵⁰ Schneider’s references to the STL strand and the IDG are reminiscent of Moag’s (1992:135) discussion of the colonials (speaking the imported native English variety) and the natives (speaking the local indigenous variety). In fact, Schneider (2003) references Moag’s (1992) work on several occasions in his description of the Dynamic Model.

various types of colonisation.) As indigenous communities already live in these countries, complex contact situations arise between the STL strand and the IDG strand.

It is important to consider that not all the English speakers in the STL strand emigrate from the same regional backgrounds (nor are they necessarily from the same social backgrounds), and therefore all the speakers in the original English speech community do not behave in a linguistically homogenous way (Schneider, 2003:244). In order to facilitate optimal communication, the linguistic forms that are widely used are used more often, whereas the linguistic forms that are not used often (and consequently result in communication failure) are avoided (Schneider, 2003:244). This process is generally known as **koinéisation**.

However, the STL strand also interacts with the IDG strand and this interaction results in a more complex contact situation (Schneider, 2003:244-245). Initially, the contact serves utilitarian purposes, as there is no lingua franca to facilitate communication. This means that the STL strand and the IDG strand continue to communicate predominantly within the confines of their respective communities, and only communicate outside these confines when they needed to trade and negotiate (Schneider, 2003:245). As a result of this restricted contact, a reduced code is used to communicate. This reduced code is generally known as a **pidgin**. In this phase, the settler language is not greatly influenced by the indigenous languages. The only notable exception in this regard is the names of places.

The second stage is known as the **exonormative phase** (Schneider, 2003:245-247). During this phase, the settlers are governed by the homeland and, as a result, English is regarded as the language of administration. The result is that English is spoken regularly in the new environment. Some members of the indigenous populations have regular contact with the settlers, and they acquire language in an attempt to gain economic prosperity. Gradually, the interaction between the STL strand and the IDG strand leads to the reconstruction of identity in both the STL strand and the IDG strand. Schneider (2003:246) refers to the reconstructed STL strand identity as the “English-plus” identity, and the reconstructed IDG strand identity as the “English-cum-local” identity.

The new environment in which the members of the STL strand find themselves gradually have an impact on the English they speak (Schneider, 2003:246). In this stage, the settlers

no longer only borrow the names of places from the indigenous languages, but also other meaningful words. These words usually refer to new objects they encounter for the first time in the new environment, such as animals and plants. At the same time, structural nativisation takes place when the members of the indigenous communities start using English (Schneider, 2003:246). This can be attributed to the phenomena of transfer from the indigenous languages to English in terms of phonology and grammatical structures. Schneider (2003:246) points out that it is at this point where expanded pidgins develop and **creolisation** occurs.

The third stage is known as the **nativisation phase**. Schneider (2003:247) regards this phase as “the central phase of both cultural and linguistic transformation”. This is the stage at which the members of the STL strand and the IDG strand realise that their identities have changed permanently. This realisation has an immediate linguistic impact in that plenty of innovation takes place, together with some gradual diffusion.

During this stage, the STL strand gradually starts to transition “from the acceptance of a distant mother country as the source of both political power and linguistic and cultural guidance to gradual independence” (Schneider, 2003:247). For the first time, the STL strand and IDG strand become more intertwined with one another. Here, it is important to notice that although the STL strand adopt some elements of the local culture, the burden of accommodation falls to the members of the IDG strand who are expected to assimilate into the culture and of the IDG strand (Schneider, 2003:247).

This state of affairs results in interesting linguistic changes, because the members of both strands now interact with one another on a daily basis and have to accommodate one another (Schneider, 2003:248). Within the STL strand, some members reject any form of linguistic innovation and regard the original linguistic norms as the only acceptable norm. Consequently, differences of opinion arise as to which form of language is to be regarded as the norm. Although the language of the STL strand is influenced by the indigenous languages, it is not influenced to the same extent as the languages of the IDG strand is influenced by the language of the settlers (Schneider, 2003:248).

In linguistic terms, the changes of the language of the STL strand now transcends the borrowing of lexical items from the IDG strand: new word-formation products evolve from both STL and IDG stocks; local collocations and set phrases develop; varying

prepositional usage develop; innovative verb complementation patterns develop; and alternative morphosyntactic behaviour arises (Schneider, 2003:249). Linguistic phenomena – such as code-switching – also emerge at this stage.

The fourth stage is known as the **endonormative stabilisation phase** (Schneider, 2003:249-253). This phase generally follows political independence. During this phase, the members of the STL strand community perceive that they are no longer members of the homeland, but rather members of a new community that includes the members of the IDG strand. According to Schneider (2003:250), this transition is usually the result of “some exceptional, quasi-catastrophic political event which ultimately causes the identity alignment of STL-strand speakers to switch from a self-association with the former mother country, however distant, to a truly independent identity”. Schneider (2003:250) calls this **Event X**.

Linguistically, Schneider (2003:253) points out that the difference between the third and fourth phases of the Dynamic Model is expressed by the labels “English in X” (e.g. English in Hong Kong), which is characteristic of the nativisation phase, and “X English” (e.g. Hong Kong English), which is characteristic of the endonormative stabilisation phase (Schneider, 2003:253). Whereas the first label indicates that the variant has no discrete character of its own, the second label indicates that the variant has its own discrete character with its own distinct type. This means that the use of the label “X English” indicates a variety that is on equal terms with other varieties of English.

Linguistic homogenisation is characteristic of this phase (Schneider, 2003:251). This homogenisation is to a certain extent the result of actual usage⁵¹ but is also an identity-driven discourse construct. In cases where linguistic distinctiveness can be observed, the members of the speech community tend to ignore it. During this stage, dictionaries are published. It is important to note that grammars are only published later. This is attributed to the fact that “the number of grammatically divergent patterns is smaller than the number of local words, and in the light of the assumption of an internationally

⁵¹ This aspect of the endonormative phase points back to the theory of competition and selection (cf. Section 2.2.1.1). The more frequently new, innovative features are replicated by members of the speech community, the more conventionalised these features become, resulting in linguistic homogenisation and the stabilisation of the variety.

homogeneous ‘common core’ of English grammar they are more difficult to accept as correct, it appears” (Schneider, 2007:52).

The fifth stage is known as the **differentiation phase** (Schneider, 2003:253-254). During this phase, the independence of the community mentioned in the previous stage has solidified. Due to the political, economic, and social stability, there is room for internal differentiation:

[I]n the absence of an external challenge or need to demarcate a community as against some outside entity, differences within a society and between individuals with respect to their economic status, social categories, and personal predilections come to light and can be given greater prominence.

(Schneider, 2007:53)

Schneider (2007:54) points out that even though a new language variety has emerged, it is not indicative of the end of linguistic evolution. Now, new varieties that are characteristic of new group identities can emerge from the formerly new variety.

Although Schneider’s (2003) Dynamic Model is widely applied to describe the emergence of new varieties of English, there are some elements of the model that have been criticised. In fact, Schneider (2014:17) himself says that the model has its own limitations. One of the major issues of contention is the extent to which the model can be used to investigate both Inner Circle varieties and Outer Circle varieties using the same framework. Mesthrie and Bhatt (2008:35) argue that the model cannot be used to investigate both Inner Circle varieties and Outer Circle varieties, as different conditions of ‘installation’ apply. Van Rooy and Terblanche’s (2010:358) view, on the other hand, contrasts strongly with this criticism when they say that “[o]ne of the most important virtues of this model is that it incorporates both native and non-native varieties in one coherent account”. Schneider (2014:16) responds to Mesthrie and Bhatt’s (2008) criticism by saying that he has intentionally built the joint perspective into the Dynamic Model. He goes on to say:

It depends on the granularity of one’s perspective, I think, on how finely graded an analysis is desired. Zooming in more closely, for example, by

looking at the Outer Circle varieties, might bring out even more similarities, but the very fact that to some extent there is a coherent underlying pattern that can be identified across both variety (and Circle) types is a significant claim in itself.

(Schneider, 2014:16)

Scholars have also called for more careful consideration when applying the model to New Englishes. Van Rooy (2014a:22) argues that although Schneider's Dynamic Model "breaks with the simplistic characterisation of an entire country belonging to a category like Kachru's Inner or Outer Circle", scholars should rather investigate "the contact between the STL strand and the IDG strand [...] in terms of the local settings within a country, rather than to an entire country". Van Rooy (2014a) motivates his argument by providing evidence from the postcolonial English varieties in two countries (i.e. the United States of America and South Africa) that have complex postcolonial population demographics. This point of view is also motivated by Bekker's (2009:432-433) research in which he investigates the vowels of South African English (SAE). He concludes that the various contact situations in the colonial history of South Africa have resulted in "successive 'waves' of nativisation" (i.e. whereas WSAE has already reached the phase of endonormative stabilisation in Schneider's (2003) model, BSAE has thus far only reached the nativisation phase).

Furthermore, Van Rooy (2014a) argues that the notions ENDONORMATIVITY and HOMOGENEITY, two features central to the endonormative stabilisation phase of the Dynamic Model, must be disentangled. He demonstrates that the process of identity rewriting leads to endonormativity, whereas the process of extended contact leads to homogeneity.⁵² Although these processes may often run concurrently, it is important not to conflate them into a single process (Van Rooy, 2014a:35-36). If these notions (i.e. ENDONORMATIVITY and HOMOGENEITY) are disentangled, it explains why the new varieties of

⁵² Here Van Rooy (2014a:35) points to the fact that BSAE never really stabilised and may be in danger of being absorbed into the STL strand. By contrast, Coloured South African English (CSAE) and Indian South African English (ISAE) have stabilised to a certain extent and are therefore less likely to be absorbed into the STL strand.

English have achieved some semblance of endonormativity but not homogeneity (cf. Section 2.3.3 for a full explanation).

2.3.1.5 *Synthesis*

This section provided an overview of the models proposed by scholars in the literature to describe the spread of English. The section started with Strang's tripartite model of English in which she distinguishes between A-speakers (i.e. mother-tongue speakers of English), B speakers (i.e. speakers who acquired language in their early childhood), and C-speakers (i.e. foreign speakers of English).

Quirk's views on the varieties of English was considered next. He distinguishes between English as Native Language (ENL), English as Second Language (ESL), and English as Foreign Language (EFL). A recurring theme in Quirk's work is the idea that varieties that do not resemble standard English linguistically are not acceptable.

Quirk's views were followed by Kachru's three-circle model. This model comprises the Inner circle (which is representative of the native speakers of English), the Outer Circle, (which is representative of the L2 speakers of English) and the Expanding Circle (which is representative of the foreign speakers of English). Whereas Quirk views L2 varieties of English as not being acceptable, Kachru argues that linguists should be more concerned with studying L2 varieties because this is the circle in which most of the linguistic development occurs.

Schneider regards Kachru's model as "rather superficial" in its capacity for establishing categories of linguistic description and proposes a different model for investigating New Englishes. This model is known as the Dynamic Model of the Evolution of New Englishes. In his model, Schneider contends that a fairly uniform process underlies the emergence of New Englishes which accounts for the similarities between the varieties of New Englishes. He distinguishes five stages in this uniform process: the foundation phase, the exonormative phase, the nativisation phase, the endonormative stabilisation phase, and the differentiation phase.

Of all the models, Kachru's (1992) three-circle model and Schneider's (2003, 2007) Dynamic Model of the Evolution of New Englishes are perhaps the most venerated by linguists today.

Hundt and Mukherjee (2011b:210), for example, approach their discussion of the distinction between English as native, second, and foreign language from the perspective of Kachru's (1992) three-circle model. Initially, the purpose of Kachru's model was to compile a sociolinguistic profile of English in which each of the circles "represent[s] the types of spread, the patterns of acquisition, and the functional allocation of English in diverse cultural contexts" (Kachru, 1992c:356). With this model, he hoped to create an awareness of the implications of the multicultural nature of English and how it is used across cultures (Kachru, 1992c:357). Overall, scholars believe that Kachru succeeded in his endeavour. Bruthiaux (2003:160), for example, states:

Introduced at a time when the duopoly of American and British English was unquestioned and metropolitan attitudes to postcolonial variants often ranged from amused condescension to racist stereotyping [...] the model broke new ground in raising awareness of the very existence of dynamic varieties of English with growing populations of speakers and increasingly vibrant media, literatures, and popular cultures.

However, the model is not without its limitations. Bruthiaux (2003:160), for example, also points out that since the model provides no account of dialectal variation within the varieties which are representative of the inner circle, English is perpetuated as "monolithic and standardised". Similarly, Seargeant & Tagg (2011:498) believe that Kachru's model overlooks "much of the variation that occurs within countries or regions". This view was compounded by the fact that Kachru's model abstracts away from countries (such as South Africa) with complex postcolonial multilingual realities where ENL, ESL, and EFL speakers live alongside one another (Hundt & Mukherjee, 2011b:210). Consequently, some scholars have come to regard the concentric circles as hermetically sealed units that overlook the linguistic variation that exist in each of the circles.

If the entire body of Kachru's work is considered, though, it becomes clear that Kachru never intended for the circles to be viewed in this way. In fact, Kachru (1982:49) explicitly states that "speech communities – monolingual or multilingual – are not homogeneous

users of single codes". He supports this statement by demonstrating how different codes influence one another. He takes the argument a step further by saying that, taken together (from the perspective of the language users and the community to which they belong), all the codes can be viewed as a single system. In light of this argument, he makes it clear that any description of New Englishes should account for this linguistic and cultural pluralism (Kachru, 1982:51). Although Kachru certainly does not believe English to be a monolithic and standardised language with no variety in each of the circles, the fact remains that it is still difficult to situate South Africa within the three-circle model.

In light of these criticisms, many scholars prefer Schneider's (2003; 2007) Dynamic Model of the Evolution of New Englishes. Van Rooy (quoted by Hundt & Mukherjee, 2014:211), for example, states that he finds Schneider's model more helpful "because it considers the (changing) sociolinguistic relationship between speaker groups and the different acquisition histories within a community rather than grouping them all together into either an undifferentiated ENL, ESL or EFL community".

Explaining the development of the linguistic situation in South Africa using the Dynamic Model, however, is also not without problems. In this regard, Van Rooy (2014:22) suggests that the contact between the STL and IDG strand should first be applied to local settings within a country and not the country in its entirety. By approaching the development of the new varieties of Englishes in this manner, scholars can account for the development of all the varieties of English in South Africa. This can be seen by contrasting the development of WSAE with the development of BSAE in South Africa. Schneider's (2007:173-188) description of the development of English in South Africa mostly pertains to WSAE. In the paragraphs that follow, a short description of the development of BSAE is provided.

The period during which the first contact between the English STL strand and the indigenous African IDG strand took place can be traced back to the establishment of British sovereignty in 1815 and the consequent British settlement in 1820. During this time, there was significant warfare between the British and the Xhosa and other tribes and the majority of the indigenous African population continued to follow their traditional lifestyles (Hodgson, 1997). Since there was a large contingent STL population and many of the intermediary roles were played by other IDG strand, there was not an immediate need to foster a class of indigenous intermediaries amongst the black African

population. This resulted in an extended time-frame for the foundation phase in BSAE (Van Rooy, 2014a:30).

By the latter half of the 19th century, a number of missionary schools were established. Some of these schools are said to have been staffed by British teachers during the first half of the 20th century. It is during this time that “[t]he formation of an educated black elite began in earnest” (Hodgson, 1997:80). It resulted in the emergence of a generation of university lecturers, newspaper editors, and novelists. Although some scholars (e.g. De Klerk, 2006:7-8) maintain that the standard of education in these schools was generally of a high standard, other authors (e.g. Hirson, 1981:220) caution that while there were indeed one or two schools that provided quality education, the majority of the mission schools were in fact “poorly endowed, poorly equipped, overcrowded, and incapable of providing more than rudimentary education.” On the grounds of his analysis of enrolment numbers, he states that very few black children had access to education before the apartheid government implemented Bantu education in the middle of the 20th century. In light of this information, it can be said that the foundation phase of BSAE lasted from approximately 1820 to 1950.

Hirson (1981:227) points out that the implementation of Bantu education in 1953 resulted in a dramatic increase in the number of children attending school. At the same time, however, many of the competent black teachers left the school system, and the children were educated by teachers who had limited language proficiency in English. This resulted in more black children receiving a poorer quality of education.

Aside from the education policies, the labour policies implemented by the apartheid government further limited contact between black labourers and white English speakers, especially since the majority of the white supervisors were Afrikaans-speaking (Lanham, 1996:27). Also, the policy of homelands and decentralisation also played a role in restricting the evolution of BSAE (Van Rooy, 2014a:31). In light of this information, the period of exonormative stabilisation for BSAE can be said to have lasted for most of the latter part of the 20th century, from approximately 1950 to 1994.

Although Schneider (2007:185) regards the 1994 political transition as “Event X” in the South African context, signifying the phase of endonormative stabilisation for South Africa in general, Van Rooy (2014a:32) is hesitant to say that BSAE has reached the stage of

exonormative stabilisation. In fact, he points out that aside from “the Black Consciousness Movement in literature” which led to the production of poems in a distinctive voice, any form of endonormativity beyond the field of literary writing is all but absent” in BSAE before 1994 (Van Rooy, 2014a:31). In this regard, Van Rooy (2014a:32) states that while there are indications for greater tolerance towards a range of pronunciations, the acceptance of non-standard grammatical features remains ambiguous. Consequently, linguists are hesitant to say that BSAE has progressed to the phase of endonormative stabilisation and do not foresee that it will progress beyond the phase of nativisation in the near future. The development of BSAE (and WSAE) is discussed in more detail in Section 2.3.3.

A recurring notion in all of these models is the attitudes that speakers have toward those varieties that are not L1 varieties of English. This issue is addressed in more detail in the section that follows.

2.3.2 *Attitudes towards New Englishes*

Scholars have been debating the legitimacy of New Englishes, and according to Van Rooy (2008a:35), the debate is yet to be settled. Whereas some scholars (e.g. Bamgboṣe, 1992, 1998; Kachru, 1991, 1992b; Schneider, 2003, 2007) are in favour of assigning status to non-native varieties of English similar to the status assigned to native varieties of English, other scholars (e.g. Quirk, 1990/2003) oppose this view and regard non-native varieties in a negative light. In the paragraphs that follow, a brief overview of this debate is provided.

Standard English is the variety of English that has been codified in dictionaries and grammars and is often proffered as the **norm (of correctness)**.⁵³ There are a number of local standards of English (e.g. UK English, US English, Australian English and South African English).⁵⁴ These standards have been adopted by most of the major publishing houses in the world, resulting in a high degree of uniformity with regard to the language

⁵³ Melchers and Shaw (2003:30) explain that the term **NORM** is used to refer to the implicit set of rules that is appropriate to use in a specific grammatical or social context.

⁵⁴ Currently, there is no international standard. This means that publishers use the local standard.

used in texts published world-wide (Biber *et al.*, 1999:18; Melchers & Shaw, 2003:31). Biber *et al.* (1999:18), however, warn that the term STANDARD ENGLISH can be misleading, because (in most contexts) it is assumed that the variety is uniform and unchanging. In this regard, scholars such as Leech *et al.* (2009) and Mair (2006) have proven that standard varieties of English indeed change. In large part, Leech *et al.* (2009:237) attribute these changes to language-internal processes, discourse-based phenomena, and language-external phenomena (also cf. Section 2.2.2).

The flipside of standard English is **non-standard English** (i.e. English that is not regarded as suitable or adequate in formal situations). Whereas standard English is often linked to the issue of NORMS, non-standard English is often linked to the issue of DEFICIT LINGUISTICS. Kachru (1991:207) explains that the term **DEFICIT LINGUISTICS** is used in the context of language learners with inadequate linguistic competence (also cf. Milroy & Milroy, 2012:69). As such, discussions regarding the deficit concern are often characterised by words such as *deficient*, *error*, and *deviation*. These words signal how discussions pertaining to standard and non-standard varieties of English are often characterised by inequalities and power differentials.⁵⁵

The deficit concern often emphasises the idea of **NATIVE-LIKE PROFICIENCY**. Mesthrie and Bhatt (2008:36) point out that the distinction between native and non-native speakers of English seems like a clearly discernible one:

Traditionally a native speaker is assumed to be one who has learnt a language from birth without formal instruction. By contrast a non-native speaker of a language has learnt it as a second (or later) language some time after being initiated into his/her native language as in the native language.

⁵⁵ Mair (2013:257-258) demonstrates that power differentials can be observed not only between standard English and non-standard English, but also amongst natively spoken educated standards. The social power that a variety enjoys is predominantly determined by demographic weight and institutional support. Although a variety should at least be codified in order to enjoy institutional support, institutional support tends to be realised when the variety enjoys official status and is used in prestigious domains of communication. These aspects are addressed later in this section.

Following this seemingly clear distinction between native and non-native speakers, some scholars (e.g. Quirk, 1990/2003:100) assume that native-like proficiency can only be achieved by native speakers of English, as they have proper intuitions regarding the structural properties of the language. However, Schneider (2007:17) states that this is a very simplistic view and does not allow for the complexities of reality. This view is simplistic in two regards. First, it assumes that all speech communities are monolingual when most speech communities are in fact multilingual (Kachru, 1985:11; Mair, 2013:254-255; Mesthrie & Bhatt, 2008:36). Second, it denies that competence in a language is tied to its constant use (Croft, 2000:57; Schneider, 2007:17). This means that the L2 English speakers in the traditional Outer Circle who shifted to using English only (or even shifted to using English predominantly) should be regarded as native speakers of English. Kachru refers to this kind of nativeness as **functional nativeness** (Yano, 2001:122-123). He contrasts functional nativeness with **genetic nativeness**, i.e. the nativeness of the L1 English speakers who find themselves in the Inner Circle (Yano, 2001:122-123). In light of these arguments, it could be reasoned that native-like proficiency is a rather “elusive yardstick” and that it is more important to consider that English is used in various linguistically and culturally pluralistic societies and a variety of situations (Kachru, 1992a:52). Therefore, Croft (2000:58) calls for the abandonment of the native speaker and non-native speaker distinction:

We must abandon the assumption that there are ‘native’ speakers whose grammars are fixed and always determine the conventions of the speech community as opposed to ‘nonnative’ speakers, whose grammatical knowledge is ‘imperfect’ and variable and who only try to approximate the conventions established by the ‘native’ speakers.

Traditionally, New Englishes tend to be characterised by a degree of variability. Mair (2013:254-255) and Schneider (2007:17) ascribe this to the global spread of English and its consequent contact with not only indigenous languages, but also other non-standard varieties of English. In this regard, Mair (2013:255) says that “[t]he more English spreads globally, the more heterogeneous it becomes internally”. Schneider (2014:9) points out that whereas the growing heterogeneity of English could primarily be observed in the Outer Circle countries during the 20th century, the expansion of English dramatically increased in the Expanding Circle countries during the last couple of years.

Given this variability, Kachru (1992a:53) suggests that English should be regarded as a spectrum of varieties, “ranging from standard native varieties to standard non-native varieties”. This position, however, gives rise to another problem, namely the problem of determining whether a linguistic feature is an innovation (i.e. a new acceptable variant) or simply an error (i.e. a mistake or an uneducated guess). Gut (2011:120) maintains that the distinction is made on attitudinal and political grounds rather than linguistic grounds. The difficulty in distinguishing innovations from errors is a recurrent problem in the field of New Englishes and has been vocalised by several scholars (cf. Bamgboṣe, 1998; De Klerk, 1999:315; Gut, 2011; Kachru, 1992a:61-62; Van Rooy, 2011). The importance of this issue should not be underestimated, because the recognition of non-native varieties is dependent on the recognition that innovations are not errors (Bamgboṣe, 1998:2).

Bamgboṣe (1998:3) suggests several measures to assist in this regard. These measures are: (i) the number of people using the innovation (i.e. the demographic measure); (ii) the area across which the innovation is dispersed (i.e. the geographical measure); (iii) the intelligibility of the innovation (i.e. the intelligibility measure);⁵⁶ (iv) the people using the innovation (i.e. the authoritative measure); (v) the areas where the use of the innovation are sanctioned (i.e. the codification measure); and (vi) the attitude of users and non-users to it (i.e. the acceptability measure). Each of these issues is addressed in the following paragraphs.

Bamgboṣe (1998:3) posits that the **demographic measure** concerns numbers. It also links strongly to the concepts INNOVATION and CONVENTIONALISATION which form an integral part of the theory of competition and selection (cf. Section 2.2.1.1). When an innovative linguistic feature is added to the linguistic feature pool but is never replicated by other members of the speech community (i.e. if it is never diffused), it will remain an innovation (regarded as an error) and not become conventionalised as a generally accepted innovation. On the other hand, the larger the number of speakers in the speech community that select and replicate the innovation, the more likely it is that the

⁵⁶ The intelligibility measure is not one of the measures proposed by Bamgboṣe (1998), but it enjoys so much attention in the literature, that it has been included in the list of measures.

innovation will become entrenched (in individual grammars) and conventionalised (in the language used by society).

The **geographical measure** relates to the area across which the diffusion of the linguistic feature takes place (Bamgboṣe, 1998:3). Within the South African context, this measure can be explained by considering BSAE. In Section 2.3.3, BSAE is described as the English spoken by the indigenous African IDG strand who speak a number of Bantu languages. Traditionally, the speakers of these Bantu languages were settled in specific regions in South Africa, e.g. the Xhosa people were mainly settled in south-east South Africa, the Zulu people were mainly settled in the province of Kwa-Zulu Natal, while the Sotho people were mainly settled in central and north-east South Africa and Lesotho. The English spoken by the members of each of the Bantu languages in South Africa share many linguistic features. So, when scholars investigate BSAE, they are essentially investigating the linguistic features that cut across the aforementioned language backgrounds and geographical areas.

Another measure not considered by Bamgboṣe (1998), is the measure of **intelligibility**. Nelson (1984:17) describes the notion as follows:

Intelligibility is a characterization of effective conveyance of information appropriate to the context through language. Given that the producer knows what he wants to say, the message is intelligible to the receiver when it is couched in such linguistic and non-linguistic form and context that the receiver apprehends what the producer means.

Considering Nelson's (1984) definition of intelligibility, it is clear that it can be regarded as a feature of a communicative interchange and is dependent on the sociolinguistic background of the participants. Therefore, it is not possible to quantify intelligibility. In other words, it is not an absolute criterion, but rather participant- and context-dependent and should be considered a cline as opposed to a clear-cut dichotomy (Nelson, 1984:15-17). The notion of INTELLIGIBILITY is also captured in one of Keller's (1994:94) maxims of linguistic actions: "[t]alk in such a way that you are understood". Based on these premises, some scholars (e.g. De Klerk, 1999:315) argue that speech events can only be successful when speakers are able to deliver their messages in such a way that it is understood by the hearers. While this approach may not be problematic when speakers

and hearers belong to the same speech community, the picture changes when they do not belong to the same speech community. Following this argument, Svartvik (1985:33; my emphasis) says:

The main reason for the majority of people wanting to acquire proficiency in English as a second or foreign language is to use it for some general or specific purpose: to do business, administer, read text books or detective stories, attend conferences, travel, and what not [...] *The strong argument in favour of English as an international medium is that it is the most widely used language, but will remain useful only as long as it remains intercomprehensible.*

Thus, my defence of the native-speaker norm, or rather a native-speaker norm, is based on functional criteria. They are not those of 'correctness' (whatever that may be) or social status (we know what that is) or even admiration of the political systems of the English-speaking nations. Nor is my recommendation based on fear of 'deviations' (which will occur anyway, even when there is a 'home-based' norm, as testified by the popularity of usage-books for native consumption).

Although Svartvik (1985) makes a valid point, Kachru (1985:23) points out that even among the native varieties of English, the speakers are not always intelligible to one another:

True, educated (standard) Indian English, Singapore English, Nigerian English, or Kenyan English is not identical to RP or GA (General American). It is different; it *should* be different. Do such educated varieties of non-native Englishes create more problems of intelligibility than does, for example, a New Zealander when he or she talks to a midwestern American?

(Kachru, 1985:24-25; his emphasis)

Consequently, Kachru (1985:23) sees "no reason to expect homogeneity in the multi-ethnic and multilingual societies of Africa, South Asia, Southeast Asia or the Philippines".

When a feature is used by influential groups, e.g. authors, teachers, media practitioners, examination bodies and publishing houses, it can be assumed that they have accepted and legitimised the feature. This is known as the **authoritative measure** (Bamgboṣe, 1998:4; Melchers & Shaw, 2003:32). Some authors (e.g. Gilquin & Granger, 2011:59; Götz & Schilk, 2011:82; Mesthrie & Bhatt, 2008:41) regard edited data as contaminated by the process of editorial intervention when investigating errors and conventionalisations in New Englishes. Kruger and Van Rooy (2017:21), however, contend that “editorial processes may provide important information about the degree to which a variety has progressed towards endonormative stabilisation”. They argue that the use of features that are characteristic of non-native varieties of English in published written texts signals and propagates the conventionalised status of these features. They explain their position as follows:

[I]f innovative features of a variety are accepted and allowed to disseminate in published writing by the linguistic gatekeepers of the publishing industry of a particular community, some degree of endonormativity has been attained – even in the absence of overt codification of the variety in question. Furthermore, once this point of tacit legitimisation is reached, the frequency of input of the innovative feature in published media rises, initiating a feedback loop that further contributes to conventionalisation. We therefore propose that the acceptance of innovative features by editors in the context of the variety in question is a significant marker of and contributing force in the endonormative stabilisation of the variety.

(Kruger & Van Rooy, 2017:21-22)

The authoritative measure gives rise to the codification measure since editors make use of style guides and reference manuals prescribed by the publishing houses. The **codification measure** refers to the practice of putting the acceptable features of a variety into a written form in dictionaries, grammars, style guides or any other kind of reference manual. (Bamgboṣe, 1998:4). Kachru (1985:18) extends this idea a bit more when he says that “codification implies determining the bounds of [...] innovations or creativity – in other words, ‘allowable’ deviation from the native norms”. Since a feature can only be prescribed once it is codified, codification is regarded as an important measure (Melchers

& Shaw, 2003:32). This means that “[c]orrect’ can now mean ‘according to an abstract principle of correctness’ rather than ‘conforming to an observed norm’” (Melchers & Shaw, 2003:32). Bamgboṣe (1998:5) explains it as follows:

[A]s long as non-native English norms remain uncoded, they cannot become a point of reference for usage and acceptance. By default, the only codified norms available (which are based on native varieties) will continue to license what is acceptable and what is not, even when there is a desire to encourage and institutionalize non-native English norms.

Finally, codification gives rise to the **acceptability measure**. Bamgboṣe (1998:4) regards this measure as “the ultimate test of admission of an innovation”. Once an innovation has been accepted, it is likely to have a reasonable lifespan (Bamgboṣe, 1998:4).⁵⁷ The acceptability measure is also regarded as important by other scholars. Kachru (1992a:56), for example, agrees that a non-native variety will only be regarded as the norm once it has been recognised and accepted.

To a large extent, acceptability is determined by the notion of **ATTITUDE**. Attitude not only includes the attitudes of the speakers of the native varieties of English but also the attitudes of the speakers of the non-native varieties of English. Generally, the acceptability of non-native varieties can be associated with the reliance on standard English. Mair (2006:7) believes that there is no need for influential groups to be unduly conservative by relying on standard English and that there is no reason why non-native varieties should not be regarded as “legitimate new standards of English”. Even in communities consisting primarily of native speakers of English, the speakers who habitually use only standard English are in the minority (Kachru, 1992a:51-52).

Yet, native speakers have traditionally not been very accepting of the non-native varieties of English, and their non-acceptance of these varieties often gives rise to what is known as the **COMPLAINT TRADITION** in which the “falling standards” of English is lamented. Schneider (2007:43) draws on Milroy and Milroy’s (2012:32) description of the

⁵⁷ Bamgboṣe (1998:4) is careful to point out that the lifespan of a linguistic feature is subject to the normal processes of language change.

complaint tradition within the native-speaker community, where native speakers complain about non-standard (but native) language use, to describe how native speakers complain about non-standard, non-native language use. In the complaint tradition of the native-speaker communities, Milroy and Milroy (2012:32) point out that native speakers' complaints about the non-standard use of language by other native speakers in the community carry with it "the implication that general standards of conduct and morality in society are also in decline". They go on to say it is inevitable that the stigmatisation of non-standard forms will have social consequences since the stigmatised forms are more commonly used by speakers in lower social groups (Milroy & Milroy, 2012:69). They attribute this to the symbolic roles language play in society. They posit that "[...]language attitudes stand proxy for a much more comprehensive set of social and political attitudes, including stances strongly tinged with authoritarianism, but often presented as 'common sense'" (Milroy & Milroy, 2012:45-46). From a New Englishes perspective, Schneider (2007:43) points out that the complaint tradition is usually associated with members of the STL strand, who reject the idea of linguistic innovation and still regard themselves as an extension of their country of origin. In both the complaints traditions, "the parameters for making judgments on the formal and functional uses of English continue to be culturally and linguistically ethnocentric", with the main difference being the identity of the people subjected to the "othering" (Kachru, 1992a:59-60).

Unfortunately, the full recognition of non-native varieties is often not only hindered by the negative attitudes of speakers of the native variety, but also by the linguistic insecurity of the speakers of the non-native variety. He explains the problem as follows:

Speakers of these post-colonial non-native Englishes are often caught in a double bind. A too-perfect approximation to the former colonial norm is socially undesirable [...] but many of the stable phonetic and grammatical features that have emerged still tend to be seen as interference-caused errors rather than potential harbingers of a new and legitimate local norm of English usage.

(Mair, 2006:7)

Kachru (1992a:60) makes a similar point when he says that although the non-native speakers do not want to emulate the native varieties, they are often not able to recognise

the “ecological validity” of their non-native variety. According to him, this gives rise to a phenomenon he calls **LINGUISTIC SCHIZOPHRENIA**. Despite this phenomenon, many non-native English speakers continue using English, since they feel that access to English is advantageous in that it contributes to the upward mobility of its speakers in terms of economic and educational success (Thomason, 2001:35).

It is important to consider that the success of (standard) English can primarily be ascribed to the fact that its (native) speakers are often the ones who are in power (Chambers, 2003:233). This has led some scholars to describe (standard) English in decidedly negative terms. Skutnabb-Kangas (2003:33), for example, describes English as a “killer language” guilty of “linguistic genocide”. Similarly, Swales (1997:374) describes it as a *Tyrannosaurus rex*, i.e. “a powerful carnivore gobbling up the other denizens of the academic linguistic grazing grounds”. Therefore, Swales (1997) feels that English is becoming too successful and that speakers should be more resistant to its “triumphalism”. Similarly, Seidlhofer and Jenkins (2003) encourage linguistic scholars to describe and codify the non-native varieties to allow these varieties to “emancipate” themselves of the imperialist variety.

The emancipation of the non-native varieties links strongly to the notion of **EXPERIENCE**, which in turn contributes to social identity which plays a pivotal role in postcolonial Englishes (cf. Section 2.3.1.4). The idea that language is used to construe experience is explained theoretically by Halliday and Matthiessen (2004:30):

It is clear that language does – as we put it – construe human experience [...] [T]here is no facet of human experience that cannot be transformed into meaning. In other words, language provides a theory of human experience, and certain of the resources of the lexicogrammar of every language are dedicated to that function.

Considering the role of language in the construal of experience, scholars are obligated to not simply regard the innovations in non-native varieties as deviant, but rather as an expression of a unique set of experiences. Van Rooy (2008a:353) expresses this idea succinctly when he states that “we need to start looking at Outer Circle Englishes in a very different way – not in terms of failed attempts to internalize some a priori Inner Circle grammatical system, but in terms of unique kinds of consistency that reveal themselves

in textual development". The important role of language in the expression of experiences is also vocalised by the African-American novelist, James Baldwin (cited by Achebe, 2003:64-65; my emphasis):

My quarrel with the English language has been that the language reflected none of my experience. But now I began to see the matter another way ... *Perhaps the language was not my own because I had never attempted to use it, had only learned to imitate it.* If this were so, then *it might be made to bear the burden of my experience if I could find the stamina to challenge it, and me, to such a test.*

Achebe (2003:65) makes a similar point when he says:

I feel that the English language will be able to carry the weight of my African experience. But it will have to be a new English, still in full communion with its ancestral home, but altered to suit its new African surroundings.

The discussion in this section has demonstrated that New Englishes are central to the expression of the experiences of the non-native speakers of English. The attitudes of both the native and the non-native speakers have a significant impact on the use and acceptance of these varieties. In the section that follows, the theory encapsulated in Section 2.3.1 and Section 2.3.2 is applied to the development of (Black) South African English specifically and its role in the speech communities of South Africa.

2.3.3 Synthesis and application to Black South African English

Currently, there are 30 living languages in South Africa (Lewis *et al.*, 2016). Of these 30 languages, 20 are indigenous, and ten are non-indigenous (Lewis *et al.*, 2016). Considering this number of languages, it is inevitable that the majority of speech communities in South Africa are multilingual. One of the largest non-indigenous languages in South Africa is English. Kamwangamalu (2002:161) points out that English plays a special role in South Africa, as is evident from its use in higher domains:

This is evident from the language practices in the higher domains such as the media, the legislature, education, the army and correctional services. In each of these domains English has become increasingly more dominant [...] The hegemony of English is also evident from the language practices for political events, such as the inauguration in 1994 of Nelson Mandela as the first democratically elected president of South Africa, the annual openings of Parliament, the signing ceremony of 10 December 1996 of South Africa's new constitution, and various official announcements or press releases.

Given the multiple functions English has in South African society, English is perceived as an instrument of “upward economic mobility” and is synonymous with “being educated” (Coetzee-Van Rooy & Verhoef, 2000:164; Kamwangamalu, 2002:162), even though less than 10% of the South African population are native speakers of English (StatsSA, 2012:22-23).

Although English is regarded as the language to be used in government, education, and economy, official bodies (such as the Pan South African Language Board (PANSALB))⁵⁸ promote functional multilingualism:

The national and provincial governments may use any particular official languages for the purpose of government, taking into account usage, practicality, expense, regional circumstances and the balance of the needs and preferences of the population as a whole or in the province concerned; but the national government and each provincial government must use at least two official languages.

(PanSALB, 1999:1)

Despite the promotion of functional multilingualism, English is the only language in South Africa that is represented in all nine provinces (De Klerk, 1997:98). In this regard,

⁵⁸ PanSALB was established in terms of the Pan South African Language Board, Act 59 of 1995 (amended as the PanSALB Amendment Act of 1999) to develop the 11 official languages of South Africa and to promote multilingualism in South Africa (PanSALB, 2015).

Kamwangamalu (2002:162) states that “[m]ultilingualism in the new South Africa in practice means English plus any other languages, and not the use of any two languages without English”. Consequently, some have argued that English is replacing the African languages in South Africa (e.g. Kamwangamalu, 2003:74-75). Deumert (2010b) disagrees with this statement, arguing that there is not sufficient proof that this is the case. Coetzee-Van Rooy (2012) and Morreira (2012:197-198) find that the use of English is confined to specific domains (e.g. as the language of media and education) and that it is not replacing the home languages of the African speakers in interactions with family and friends.

Silva (1997:4) points out that even though some white native speakers of English in South Africa are critical of the local varieties of English, South Africans in general are not aware of the extent to which their varieties of English differ from other world varieties. Although several South African dictionaries have seen the light since 1997 – e.g. *South African Concise Oxford English Dictionary* (2002) and *A Dictionary of South African English on Historical Principles* (2010) – it is mainly WSAE that has been codified. The preface of the *South African Concise Oxford English Dictionary* (2002) clearly states that the material used in compiling the dictionary include “news material, academic, legal, medical, and scientific papers, web sites marketing companies and their products, [and] travel and tourism information”. These kinds of material tend to be written by native speakers of English, and should they be written by non-native speakers of English, they are most likely edited by a native speaker of English. Therefore, the *South African Concise Oxford Dictionary* – like many other dictionaries of SAE – are predisposed to codifying WSAE.

As pointed out earlier in Chapter 1, four main ethnic Englishes can be distinguished in South Africa. They are WSAE, BSAE, CSAE, and ISAE. These varieties have their roots in the multiple contact situations that occurred during the colonisation of South Africa. The single largest ethnolect in South Africa is BSAE, spoken by the indigenous African IDG strand who also speak a number of Bantu languages (Van Rooy, 2014a:30).

In this section, a very brief overview is given of the arrival of English in South Africa, which resulted in WSAE. A more detailed overview is provided for the emergence of BSAE. The emergence of this variety is mapped using Schneider’s (2003) Dynamic Model of the Evolution of New Englishes (cf. Section 2.3.1.4). In order to gain a full understanding of the emergence of BSAE, the various contact situations that arose from the arrival of English in South Africa are considered.

Bekker (2012a) distinguishes three contact situations in South Africa that contributed to the emergence of WSAE. The first contact situation arose at the turn of the 18th century with the first British occupying the Cape in 1795 and the subsequent establishment of British sovereignty over the Cape in 1815 (cf. Giliomee & Mbenga, 2007:85). In 1795 and 1806 a large number of British officials and soldiers, together with some traders and farmers arrived in the (Eastern) Cape. By 1820 a more permanent settlement was established with the arrival of some 4 000 immigrants – generally of some financial means but lacking large capital – from London and the Home Counties (Giliomee & Mbenga, 2007:86). As there was no existing English-speaking population at that time in the Cape, it can be said that this constituted a *tabula rasa* context for dialect mixing, koinéisation and new-dialect formation, even though it was somewhat complicated by the prior presence of the Afrikaans population (Bekker, 2012a:129).

The second contact situation arose between 1848 and 1862 when a new settlement of English-speaking settlers was established in Natal (now known as Kwa-Zulu Natal) with a large number of middle- to upper-class individuals immigrating from Britain (Bekker, 2012a:130). Bekker (2012a:131) argues that “there is evidence to suggest that the Founder Effect [...] was minimal in Natal, and that the context was effectively *tabula rasa*”. Unlike the situation in the Cape, there was not as much contact with the Afrikaans population.

The final contact situation arose in the fourth quarter of the 19th century in Johannesburg, where there was “a major influx of British subjects, as part of a more general influx of immigrants to South Africa” (Bekker, 2012a:131). Just as was the case in the Cape, there were also Afrikaans speakers present in the Johannesburg context (cf. Bekker, 2012a, 2012b). It is important to be cognisant of the fact that WSAE is the result of these multiple contact settings.

With regard to contact between English and the African languages, Schneider (2007:175-181) points out that there appears to be little evidence to suggest that there was much contact between the English STL strand and the African IDG strand initially. Bilingualism in English amongst the African IDG group primarily spread by means of missionary schools. The first missionary school was established in 1799 by Dr van der Kemp of the London Missionary Society (Hirson, 1981:219). By the middle of the 19th century, several missionary schools had been established with the assistance of the

colonial government. However, the establishment of the missionary schools was a slow process and, consequently, the missionary education was only available to “the more talented (or fortunate) pupils” (Hirson, 1981:219). This resulted in an extended time-frame for the foundation phase in BSAE (Van Rooy, 2014a:30).

During the second half of the 19th century, several excellent missionary schools developed and produced a generation of university lecturers, newspaper editors and novelists (Peterson, 2012). These missionary schools, however, were the exception and many schools only provided a rudimentary education. In this regard, Hirson (1981:219) cites R Hunt Davis as saying that the intent of the schools was to “civilise races emerging from barbarism”:

Thus, a suitable education would be one which inculcated Christianity, taught habits of self-control and moral discipline, imparted a knowledge of English, provided training in proper diet, cleanliness, and other aspects of personal health, and gave lessons in household care, agriculture, and handicrafts. Instructions in reading, writing and arithmetic would be sufficient to meet the needs of the working class.

Hirson (1981:220) goes on to say that, in general, the mission schools were overcrowded, received minimal government backing in terms of educational provisions and, therefore, were not in a position to provide more than a basic education. Given these educational circumstances, Van Rooy (2014a:30) postulates that something similar to Phase 2 of the Dynamic model “was achieved by the end of the mission school phase, but for a relatively small elite of the black population”. This can be attributed to the regular contact that these students had with the STL strand within an educational context. Given these circumstances, Van Rooy (2014a:31) warns that “any analysis of the history of BSAfE should be cautious to attribute widespread use and extensive stabilisation to BSAfE prior to the middle of the 20th century”.

With the implementation of apartheid in 1948, the situation started to deteriorate rapidly. The situation was worsened even more with the implementation of the Bantu Education Act in 1953 (De Klerk, 1999:312). This act instituted mother-tongue instruction at primary school level and English at secondary school level (Hirson, 1981:221). During this time, the Congress Youth League (CYL) took a stand against mother-tongue

instruction on two main grounds. The first was that they believed they would receive inferior education, and the second was that they believed the system was designed to divide the African people tribally (Hirson, 1981:227). As a result, many of the teachers left the education system, and large numbers of pupils were taught by teachers who had limited language proficiency in English (Gough, 1996:54). Despite the CYL's call for parents to keep their children from attending school and the parents' disapproval of the new system, they wanted to secure an education for their children and continued sending them to school (Hirson, 1981:227). This resulted in a situation where more students were accepted into the schools, but there were not enough teachers proficient in English to educate the pupils (De Klerk, 1999:312). As a result, the education provided was inferior to the education provided by the independent schools (Hirson, 1981:227).

Besides the restrictive nature of the aforementioned educational policies, the apartheid government also implemented restrictive labour laws. Although employment opportunities increased greatly for black South Africans in the mining industries, they were channelled into unskilled occupations in order to limit contact between the black labourers and their white supervisors, who were primarily Afrikaans (Lanham, 1996:26). Van Rooy (2014a:31) points out that the separation effect of these labour laws was further strengthened by a policy of homelands and decentralisation. This policy forced black South Africans to relocate to the homeland territories "where they were expected to 'develop separately' in the ideology of the ruling National Party" (Van Rooy, 2014a:31).

The restrictive nature of the educational policies, the restrictive labour laws, and the policy of homelands and decentralisation prevented the possibility of stabilisation from 1950 to 1990 (Van Rooy, 2014a:31). Van Rooy (2014a:31) states that there is hardly any evidence of endonormativity beyond the distinctive voice found in the poems of the Black Consciousness Movement in the field of literary writing. As a result, "attitudes towards an exonormative English remained extremely favourable".

The socio-political situation in South Africa changed drastically in 1994. In this year, the first fully democratic elections were held. Not only were all adults allowed to vote, but all enforced racial segregation came to an end and was replaced by black empowerment and employment equity policies. Schneider (2007:185) regards this landmark in South African history as an Event X, which constitutes the point at which SAE moved into Phase 4. However, Schneider (2007:185-188) observes certain aspects of SAE that point

towards what Van Rooy (2014a:32) calls the “atypical status” of SAE. Two of the observations that are pertinent to this thesis are that: (i) there is no local variant of English that acts as carrier of the new national identity; and (ii) SAE is slowly moving toward endonormativity, though not homogeneity. In the paragraphs that follow, some of these observations will be expanded and commented upon.

Coetzee-Van Rooy (2010:3) confirms Schneider’s (2007:185) view that there is no single local variety of English that acts as carrier of the new national identity; just as there is not any single “dominant” language. This situation, she says, gives rise to a distinctive expression of language and identity (Coetzee-Van Rooy, 2010:3). In her research, Coetzee-Van Rooy (2014) investigates the role of language in the identity construction of individuals in South Africa. One of her findings is that the majority of the respondents in her study feel that their identity is primarily determined by their home language rather than English (Coetzee-Van Rooy, 2014:52). This finding is also reflected in Mesthrie’s (2015:83) observation that young black university students organise their speech repertoires, in order of preference, as follows: (i) oral proficiency in the language(s) of the home and neighbourhood; (ii) oral proficiency in the street language in urban areas; and (iii) oral and written proficiency in English, as promoted in classrooms. Even though English is last on the university students’ list of language preferences, they still feel that English plays an important role in their identity construction; their home language and English are simply used to perform different identity constructions. The use of language to negotiate social networks, is also pointed out by Wilmot (2011:86-87). She says that the speakers demonstrate “flexibility [...] with their language” in that “they use it strategically to suit the different identity positions they wish to take up in different contexts” (Wilmot, 2011:87). Considering the manner in which language is used in the identity construction of South Africans, Coetzee-Van Rooy (2014) argues that English in South Africa is still in Phase 3 of the Dynamic Model and will most likely not progress to Phase 4 in the near future.

Widespread multilingualism may be one the reasons why English is not that central to the process of identity construction South Africa (Coetzee-Van Rooy, 2014). Coetzee-Van Rooy (2014) also cautions that “English is not seen as a symbol of South African social

integration” and as such does not entirely comply to the prerequisites⁵⁹ postulated for Phase 4 of the Dynamic Model. Although English is not central to the process of identity construction, South Africans still regard English as beneficial. Since English played a key role in the fight against apartheid, it also represents positive connotations as the language of liberation and resistance (Coetzee-Van Rooy & Verhoef, 2000:164; De Klerk, 1997:98).

The fact that English is not an “identity carrier” in South Africa is also one of the reasons why English may be moving towards endonormativity. Van Rooy (2014a:35) argues that identity rewriting is still underway in South Africa, as reflected in the social tension that still permeates South Africa. However, on a symbolic level, common identity is established during major events such as the rugby and soccer world cups (Spencer, 2011a; Van Rooy, 2014a:35). Considering this state of affairs, endonormativity is only achieved to “the extent that a measure of national unity has developed since 1994” (Van Rooy, 2014a:33). According to Van Rooy (2014a), the lack of complete endonormativity is illustrated by the fact that “[e]vidence for the acceptance of non-standard grammatical features remains ambiguous at best (Spencer, 2011b; Van der Walt & Van Rooy, 2002; Van Rooy & Terblanche, 2010)”, in spite of “indications of much more tolerance for a range of pronunciations (Coetzee-Van Rooy & Van Rooy, 2005) and lexical features (Van Rooy & Terblanche, 2010) than before”.

The issue of homogeneity, however, is linked to the issue of extended language contact. Here, it is important to note that the multiple contact situations in South Africa (described earlier in this section), together with South Africa’s history of social segregation, are at the root of the unique linguistic situation in South Africa. Van Rooy (2014a:34) argues that each of the multiple contact situations found in the history of South Africa may set the phases of the Dynamic Model in motion. This explains why other, smaller varieties such as CSAE and ISAE are more established than BSAE (also cf. Mesthrie, 2010, 2015:84-85; Mesthrie, Chevalier & Dunne, 2015). It also explains why SAE is not moving towards homogeneity. If provision is made for multiple contact situations in the history

⁵⁹ In this regard, Schneider (2003:250-251) argues that “the newly achieved psychological independence and the acceptance of a new indigenous identity result in the acceptance of local forms of English as a means of expressing identity”. In South Africa, however, there is evidence that multilingualism is a salient linguistic marker of South African identity (Coetzee-Van Rooy, 2010:3-4; Coetzee-Van Rooy & Verhoef, 2000:178).

of a country, the Dynamic Model can be used more effectively, and account for the ‘waves’ of nativisation found in the English varieties of South Africa (cf. Bekker, 2009:432-433).

In terms of WSAE, scholars seem to be in agreement that the variety has progressed to Phase 4 of the Dynamic Model (Bekker, 2009:185; Schneider, 2007; Van Rooy & Terblanche, 2010). More recent research conducted by Du Plessis (2015:188) has found that WSAE may already have entered Phase 5. Considering the discussion above, however, it is clear that WSAE is not moving towards homogeneity. So, although WSAE has already reached Phase 4 and Phase 5 of the Dynamic Model, scholars are more hesitant to say that BSAE has progressed beyond Phase 3 (e.g. Bekker, 2009:432-433; Du Plessis, 2015:192).

In terms of BSAE, scholars have found that native English speakers are ambivalent in their acceptance of BSAE features. While the native speakers demonstrate tolerance towards innovations in the pronunciation of BSAE speakers (especially the acrolectal form)⁶⁰ and innovative lexical features contributed by BSAE speakers, the acceptance of non-standard grammatical features remains ambiguous (Coetzee-Van Rooy & Van Rooy, 2005; Spencer, 2011b; Van der Walt & Van Rooy, 2002; Van Rooy & Terblanche, 2010; Van Rooy, Van Rooyen & Van Wyk, 2000). Considering the aforementioned, scholars are arguing that BSAE is gaining momentum in progressing to Phase 4 but may remain in Phase 3 for the foreseeable future.

⁶⁰ According to Van Rooy (2004:944) the mesolectal form of BSAE “is spoken fluently by educated speakers, but because of salient features of pronunciation (like vowel mergers) and certain features of grammar, it would not be judged as overtly prestigious by speakers of the variety of other South Africans”. On the other hand, the acrolectal form is “closer to native varieties of SAfE in many respects, but at the same time, it is characterised by more variability than less” (Van Rooy, 2004:947).

2.4 A grammatical description of the new varieties of English

2.4.1 *Approaches to describing verb complementation*

Earlier, it was mentioned that Schneider (2003:234) attributes the similarities observed between the different varieties of New Englishes to the uniform process that underlies the development of these varieties. One of the similarities that Schneider (2003:249) observes is the innovative assignment of verb complementation patterns that characterise new varieties of English when they are in Phase 3 of his model. Since scholars situate BSAE in Phase 3 of Schneider's model, and innovative verb complementation patterns are cited as one of the features of BSAE, this study focuses on the verb complementation patterns in BSAE. This section is focussed on the grammatical theories that underpin the analysis of the verb-complementational patterns in this study.

Descriptions of verb complementation may be approached from a number of perspectives. They include formalist perspectives (e.g. Radford, 1997), functionalist perspectives (e.g. Halliday & Matthiessen, 2014), cognitive perspectives (e.g. Langacker, 1987a, 2008), and constructional perspectives (e.g. Goldberg, 1995). In turn, these perspectives may include studies of a lexical nature and studies of a syntactic nature.

In this study, a primarily constructional (and therefore syntactic) approach will be followed. This approach allows for the investigation of contextual, authentic language data.⁶¹ In the sections that follow, important elements of this approach will be addressed. In Section 2.4.2.1, a definition of Construction Grammar (CG) is provided. This section is followed by Section 2.4.2.2 in which the usage-based approaches to CG are considered.

2.4.2 *The Construction Grammar approach*

The CG approach derives from Saussure's (1916/1972) notion of the LINGUISTIC SIGN, which is an arbitrary and conventional pairing of form (i.e. the signifier) and meaning (i.e. the signified). For example, the English sign *tree*, the Afrikaans sign *boom*, and the Sotho

⁶¹ Motivations for the use of authentic data are provided in Section 2.4.

sign *sefate* all signify the underlying meaning [tree]. Since the 1980s, scholars have been exploring the idea that these arbitrary⁶² form-meaning pairs might not only be useful in the description of words and morphemes but that they might also be useful in the description of all levels of grammar. This extended notion of the Saussurean sign has come to be known as **CONSTRUCTIONS** (Hoffmann & Trousdale, 2013:1). These constructions are central to CG.

There are a number of tenets that underlie CG. One of the first tenets is that it does not assume that “a particular subset of the data is part of a privileged ‘core’” (Goldberg, 2003:219). As such, it sets out to account for all the characteristics of language – even problematic or unusual linguistic phenomena. This makes the approach particularly useful in explaining the grammar of the new varieties of English (cf. Bernaisch *et al.*, 2014; Mukherjee, 2010; Mukherjee & Gries, 2009; Mukherjee & Hoffmann, 2006). Goldberg (2003:219) also posits that by considering unusual linguistic phenomena, general linguistic issues are illuminated, providing scholars with a more complete account of language.

Another tenet of CG is that the structure of language is shaped by discourse (Goldberg, 2003:219; Hopper, 1987:3). In this regard, it is posited that input and general cognitive mechanisms – together with pragmatic and processing constraints – play an important role in the acquisition of constructions (Goldberg, 2006:3). This means that linguistic forms are not fixed templates, but are negotiable in face-to-face interaction. This contrasts with the tenet of generative grammar in which it is argued that language is too complex for humans to acquire without the existence of *a-priori* categories in their mental grammar.⁶³

⁶² Although the cognitive process of categorisation plays an important role in CG, scholars acknowledge the arbitrariness of language. Croft (2001:7) explains that not everything in language can (or should) be explained in terms of (amongst other things) abstract generalisation or by means of formal or functional general principles. If this were to be the case, “all languages would be alike, all languages would be internally invariant, and no languages would change” (Croft, 2001:7).

⁶³ A more detailed discussion of the differences between the emergentist and generativist views are provided in Section 2.2.

These tenets have certain important implications for the study of linguistic phenomena. Since *a-priori* categories are taken to be absent in the mental grammars of speakers, it is argued that “virtually all aspects of the formal representation of grammatical structure are language-particular” (Croft, 2001:4).⁶⁴ Consequently, there is no need for constructionists to learn a complex technical language in order to describe human language. Instead, constructions are the basic units of representation.

Furthermore, the use of corpora is particularly useful in CG – especially since constructionists acknowledge that language is shaped by discourse and that linguistic forms are negotiable in face-to-face interaction. Gries (2013a:107) points out that many methodological innovations and techniques have resulted from developments in corpus linguistics and that these innovations and techniques are particularly relevant to CG in particular. Therefore, constructionists will benefit from keeping up-to-date with developments in the field of corpus linguistics. This issue is addressed in more detail in Chapter 3.

The aforementioned basic tenets are the tenets that are of particular importance to this thesis. The following paragraphs provide a more detailed definition of the CG and demonstrate its usefulness in usage-based approaches.

2.4.2.1 *Defining Construction Grammar*

As mentioned earlier, the notion of **CONSTRUCTIONS** is at the heart of CG. Constructions consist of form-meaning⁶⁵ pairings which are partially arbitrary (Croft, 2001:4, 18; Goldberg, 1995:1). Goldberg (2006:5) explains that any linguistic pattern can be recognised as a construction if it fulfils one of two conditions. The first condition is that the form and function should not be “strictly predictable from its component parts or from

⁶⁴ In this regard, Croft (2001:134-147) argues for language universals without universal syntactic roles.

⁶⁵ The concept **MEANING** depicts all of the conventionalised aspects of a construction’s function (Croft, 2001:19). This includes the properties of the situation described in the utterance, the properties of the discourse in which the utterance is found, as well as the pragmatic situation of the interlocutors. As such, the concepts **MEANING** and **SEMANTIC** are used to point towards “any conventionalized feature of a construction’s function” (Croft, 2001:19).

other constructions recognised to exist” (Goldberg, 2006:5). The second condition is that the linguistic pattern occurs with sufficient frequency – even if it is predictable – to be stored as a construction. It is argued that a construction can be anything from a morpheme to a word, to an idiom, to an abstract phrasal pattern. Therefore, it is posited that language is made up of constructions and that these constructions exist at all levels of language. The constructions existing at the various levels of language can be placed on a lexicon-syntactic continuum (Croft, 2001:16-17; Goldberg, 2003:219-220, 2006:5; Hoffmann & Trousdale, 2013:1-2). The continuum is illustrated in Figure 2.3, while the range of constructions is represented in Table 2.1 on p. 106.

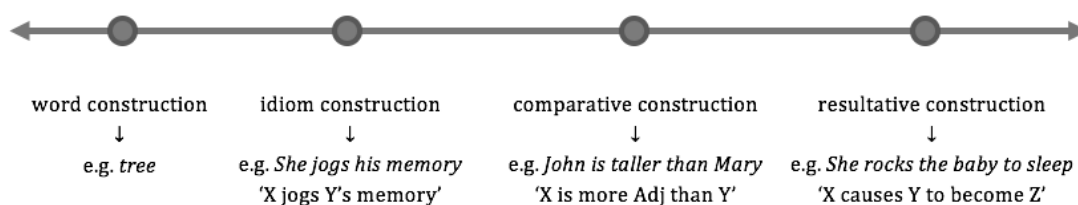


Figure 2.3: A visual representation of the lexicon-syntax continuum in CG (based on Hoffmann & Trousdale, 2013:1-2)

It is important to be aware that constructions are fundamentally symbolic units⁶⁶ in which the form-meaning pairs cannot be separated from one another (Croft, 2001:18; Traugott & Trousdale, 2013:1). The form and the meaning are connected by means of a symbolic link which is internal to the construction (Croft, 2001:19). The construction as a symbolic unit is represented visually in Figure 2.4 on (p. 107).

⁶⁶ Langacker (1987a:60) states that “[a] conceptual unit becomes a proper candidate for linguistic description only when it functions in a symbolic unit, as either its entire semantic structure or a significant component thereof”. Croft (2001) derives his notion of a construction being a symbolic unit from this statement.

Table 2.1: Examples of constructions, varying in size and complexity (Goldberg, 2003:220)

Construction	Form/Example	Function
morpheme	e.g. <i>anti-, pre-, -ing</i>	
word	e.g. <i>avocado, anaconda, and</i>	
complex word	e.g. <i>daredevil, shoo-in</i>	
idiom (filled)	e.g. <i>going great guns</i>	
idiom (partially filled)	e.g. <i>jog (someone's) memory</i>	
covariational-conditional construction	form: the Xer the Yer e.g. <i>The more you think about it, the less you understand</i>	Meaning: linked independent and dependent variables
ditransitive construction	form: Subj [V Obj1 Obj2] e.g. <i>He gave her a Coke; He baked her a muffin</i>	Meaning: transfer (intended or actual)
passive	form: Subj auxVpp (PP _{by}) e.g. <i>The armadillo was hit by a car</i>	Discourse function: to make undergoer topical and/or actor non-topical

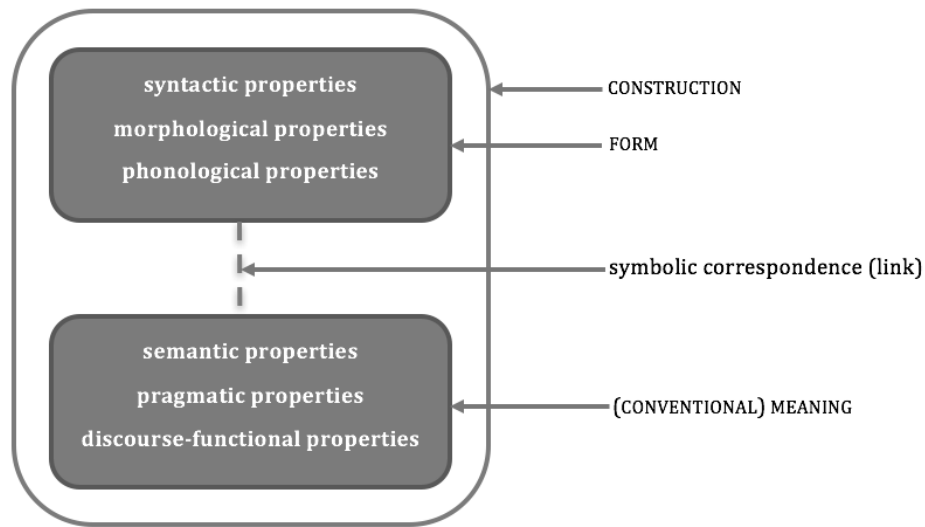


Figure 2.4: *The symbolic structure of a construction (Croft, 2001:18)*

Traugott and Trousdale (2013:8) represent the symbolic structure of a construction as follows:

$$[[F] \leftrightarrow [M]]$$

In this template, F is short for FORM (i.e. the syntactic, morphological, and phonological properties), whereas M is short for MEANING (i.e. the semantic, pragmatic, and discourse-functional properties). The double-headed arrow signifies the symbolic link between form and meaning. Finally, the external brackets indicate that the form-meaning pair is a conventionalised unit and is therefore regarded as a construction. This template will be used in this study to represent the symbolic structure of constructions.

Constructions consist of various parts (Croft, 2001:18-21). These parts are explained by means of the following sentence:

- (1) It rains.

Example (1) is an instance of an intransitive construction.⁶⁷ The construction consists of elements, components, and symbolic units. The **ELEMENTS** consist of the parts of the syntactic structure, whereas the parts of the semantic structure are known as **COMPONENTS**. **SYMBOLIC LINKS** (as illustrated by the broken lines in Figure 2.4 and Figure 2.5) connect the elements of the syntactic structure to its corresponding components of the semantic structure. Another symbolic link (as illustrated by the middle link in Figure 2.5) connects the entire syntactic structure to the entire semantic structure. An element and its corresponding component are known as a **SYMBOLIC UNIT**. Finally, the symbolic units as a whole make up the **CONSTRUCTION**. This description of constructions demonstrates that constructions specify not only syntactic information, but also lexical, semantic and pragmatic information (Fillmore, Kay & O'Connor, 1988:501).

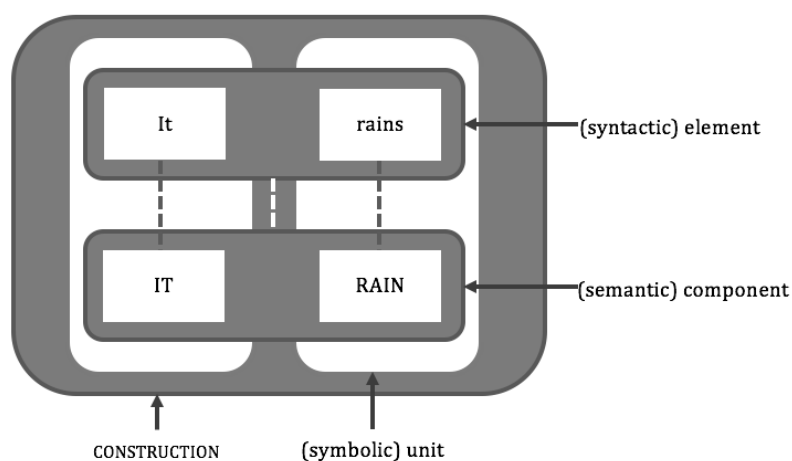


Figure 2.5: The elements, components, and units of a construction (Croft, 2001:21)

In Section 2.2.2, it is pointed out that speakers use language to construe their experiences (cf. Halliday & Matthiessen, 2014:30) and that constructions provide linguists with insight into the manner in which speakers use language to construe the world (Goldberg,

⁶⁷ At this point, it is important to take note that there are two kinds of syntactic constructions in CG (Bardhal, 2008:5). The first is **sentence-type constructions** (e.g. active and passive sentences, questions, and imperatives). The second is **argument structure constructions** (e.g. intransitives, transitives, and ditransitives). Since the research questions of this thesis centre around argument structure constructions, the examples used to illustrate important concepts in the following paragraphs are of the second kind (i.e. argument structure constructions).

2006:9). The careful attention given to the subtle ways in which language is used to construe meaning in CG is illustrated in [2] and Example [3] below:

(2) Sally bought Tumi a scarf.

(3) Sally bought a scarf for Tumi.

Example (2) and Example (3) are examples of a dative ditransitive alternation. Haspelmath (2015:25) explains that ditransitive alternations are “competing ditransitive constructions that can be used with the same verb and roughly the same meaning”. Example (2) exhibits a single syntactic structure with a single meaning. Its symbolic structure can be transcribed as $[[S V O_i O_d] \leftrightarrow [X \text{ transfers } Y \text{ to } Z]]$ where X is the agent, Y is the patient and Z is the beneficiary. The alternation in Example (3), however, exhibits a single syntactic structure with multiple meanings. On the one hand, it could mean that the scarf is intended for Tumi (i.e. $[[S V O_d PP] \leftrightarrow [X \text{ intends to transfer } Y \text{ to } Z]]$). On the other hand, it could mean that Sally bought the scarf for Tumi – since Tumi was not in a position to do so herself – so that Tumi could, in turn, give it to a third party (i.e. $[[S V O_d PP] \leftrightarrow [X \text{ intends to transfer } Y \text{ to } Z \text{ so } Z \text{ can transfer it to } A]]$ where A is another beneficiary).

The examples above provide some initial insight into a very important principle in Radical Construction Grammar (RCG) specifically. This principle is that there are no syntactic relations in constructions. In this regard, Croft (2001:5) argues that “[t]he only internal syntactic structure to constructions is their meronomic structure: the syntactic structure of constructions consists only of their elements [...] and the roles that they fulfil in the construction”. This principle can be considered more fully by considering the notion of **SELECTIONAL RESTRICTIONS**⁶⁸ demonstrated in (4) and (5):

(4) They buried the treasure.

(4a) The snow buried his body.

⁶⁸ The notion of **SELECTIONAL RESTRICTIONS** points towards the presupposed meanings attached to lexical items which places certain restrictions on the co-occurrence of words.

(5) The snow entombed his body.

(5a) *They entombed the treasure.

Both (4) and (5) have the same form, i.e. [S V O]. The meanings of the verbs, however, are restrictive. Both the lexical items *bury* and *entomb* mean [to put or hide an object underground or underneath something]. But while *bury* can be used with both animate and inanimate objects, *entomb* can only be used with animate (human) objects. Therefore, (5a) is not regarded as semantically compositional, while (4a) is regarded as semantically compositional.⁶⁹

These examples demonstrate that the restrictions placed on *bury* and *entomb* are not dependent on the syntactic elements. Instead, the meaning of the entire construction is a function of the meaning of the elements of the construction (Croft, 2001:180). This means that the constructions are semantically compositional. Since constructionists posit that syntactic structures are only relevant in terms of the semantic roles they fulfil, it allows them to provide uniform descriptions of not only 'core' (or high-frequency) linguistic features, but also so-called 'peripheral' (or low-frequency) linguistic features⁷⁰ (Croft, 2001:15; Fillmore *et al.*, 1988:503-504). Goldberg (2006:12) posits that the focus on the

⁶⁹ Bybee (2006:716) explains that collocations (also known as prefabs) are word sequences that are conventionalised and semantically predictable. Since these conventionalised sequences occur repeatedly, the form-meaning mappings of the individual words that occur in sequence become partially assembled (Barlow, 2000:381-319). When this happens, it is said that the individual words are **semantically compositional**. It is important to keep in mind, though, that this process is not only applicable to strong lexical collocations, but also to looser collocational links between words, e.g. [*it is one thing to X, it is another thing to Y*] (Barlow, 2000:319).

⁷⁰ An example of a so-called 'peripheral' linguistic feature is idioms, which are syntactically and semantically irregular by definition.

semantic aspects of linguistic features explains why constructionists make every effort to provide **general-functional explanations** when describing language.⁷¹

The cognitive-functional CG approach followed in this thesis subscribes to usage-based research methods. Consequently, frequency counts are essential for grammatical description. Type frequency counts, for example, enable linguists to determine how argument structure constructions are organised in a schematicity hierarchy (Bardhal, 2008:5). Token frequency counts, in turn, enable linguists to determine whether patterns are sufficiently frequent to be stored in the speakers' mind – even when they are fully regular instances of other constructions and therefore predictable (Goldberg, 2006:12-13). The most important issues pertaining to such a usage-based description of language are therefore addressed in the following section.

2.4.2.2 *Usage-based approaches to Construction Grammar*

Bybee (2006:711) posits that all linguists are likely to argue that “grammar is the cognitive organisation of language”, but that only usage-based scholars will argue that “grammar is the cognitive organisation of one’s experience with language”. From this perspective, language cannot be regarded as an object apart from the speaker.

Since the speaker’s experience with language is represented directly in cognition, Bybee (2013:51) argues that the direct pairing of form to meaning that characterises constructions in CG is particularly appropriate for usage-based models. The constructions are stored as a structured inventory according to frequency counts and are represented by scholars as taxonomic networks (Hoffmann & Trousdale, 2011:4). Considering the aforementioned importance of frequency and cognition, constructionists argue that constructions are acquired on the basis of input and general cognitive mechanisms (cf. for

⁷¹ Goldberg (2006:11) explains that there are different general types of explanation that are relevant to language. The first type of explanation is known as **general-formal explanation**, which entails the demonstration that a feature is an instance of a larger generalisation. The second type of explanation is known as **general-functional explanation**, which entails the identification of the purpose or function of the linguistic feature. The third type of explanation is known as the **general-historical explanation**, which entails the identification of general constraints on historical change.

example Bybee, 2006; Bybee & Scheibman, 1999; Bybee & Thompson, 1997; Culicover & Jackendoff, 1999; Lieven *et al.*, 2003). In the following paragraphs, an overview is provided of the most important issues related to input and the general cognitive mechanisms in the usage-based approaches in CG. These insights will help bridge the theory discussed in Section 2.2 and Section 2.3 and explain how concepts such as *error*, *innovation*, and *conventionalisation* can be operationalised in the verb-complementational patterns that are investigated in Chapters 4, 5, and 6.

An important notion in usage-based approaches is the notion of EXEMPLARS. Bybee (2010:7-8) argues that exemplars are important since they represent speakers' memory for linguistic forms. Each token of language experience is regarded as an exemplar. When speakers encounter an exemplar, a new "trace" is produced in their memory (Hintzman, 1986:411). These traces coexist with all the other traces of the same exemplar, together with its own relevant context and meaning. The traces enable speakers to organise and store exemplars that are deemed identical together. As such, it is posited that each exemplar has some effect on the speakers' memory storage and organisation of linguistic items (Bybee, 2010:18). After repeated encounters with various exemplars, speakers group the exemplars together and map the identical parts on one another (Bybee, 2010:27). During this process, categories are schematised from these exemplars.

Bybee (2010:26) explains the notion of EXEMPLARS with the following examples taken from the British National Corpus (BNC):

- (6) It drives me crazy.
- (7) they drive you mad
- (8) that drives me mad
- (9) The death of his wife the following year drove him mad.
- (10) A slow-witted girl drove him mad.
- (11) It drove the producer mad.

(12) A couple of channels that used to drive her up the wall.

(13) This room drives me up the wall.

Bybee (2010:27) argues that each of the eight utterances above could be regarded as an exemplar. As the speakers encounter the exemplars, they group them together and map their identical parts onto one another. The categories that are schematised from these exemplars are illustrated in Figure 2.6 below:

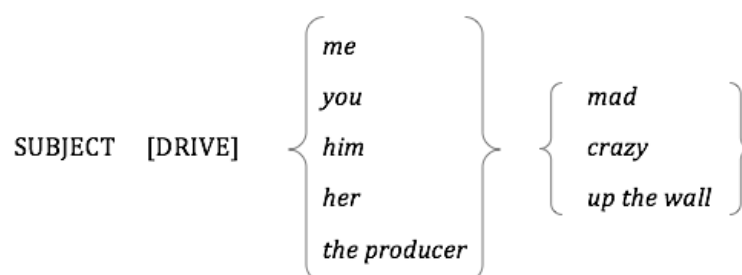


Figure 2.6: A schematic representation of the grouping of exemplars (Bybee, 2010:27)

Typically, the category SUBJECT can be filled by any noun phrase (Bybee, 2010:27). The verb *drive* is represented by the category DRIVE.⁷² The experiencer slot is filled most frequently by a pronoun that refers to an animate (and typically human) experiencer. The final slot can be filled by any adjective or prepositional phrase that is of a strong semantic character (Bybee, 2010:27). Subconsciously, speakers are aware of the semantic restrictions placed on each category, since each exemplar is stored with its own relevant context and meaning. Theoretically, each of the exemplar categories illustrated in Figure 2.6 represents a position within the resultative construction. The following template can be used to represent the resultative construction:

[[S V O_a O_p] ↔ [[X causes Y to become Z]]

Nosofsky (1988:54) demonstrates that each of these exemplar categories is structured not only by means of similarity but also by means of frequency. Exemplars that are the

⁷² The category is placed in square brackets to indicate that any form of the lemma *drive* can be used in this slot.

result of a large number of tokens will be represented more strongly than exemplars that are the result of a smaller number of tokens (Bybee, 2013:53; Nosofsky, 1988).

An important notion that arises from exemplars is the notion of SCHEMATICITY. Goldberg (2006:47) says that evidence has shown that our knowledge of instances leads to generalisations. As such, Traugott and Trousdale (2013:13-14) postulate that schematicity is a property of categorisation, which crucially involves abstraction. In other words, when speakers group exemplars together and map them on one another, they tend to abstract more general linguistic patterns from the series of more specific patterns (Bybee, 2010:27, 80; Croft, 2001:4, 27; Traugott & Trousdale, 2013:14). These abstractions are known as **SCHEMAS**.⁷³

Scholars have long posited that schemas are representative of speakers' knowledge of both item-specific instances as well as schematised instances (Langacker, 1987a:132; Traugott & Trousdale, 2013:14). As such, the main interest in terms of schematisation resides in the abstractness of a schema in relation to its instantiations. In Cognitive Grammar, schematicity is illustrated as follows:

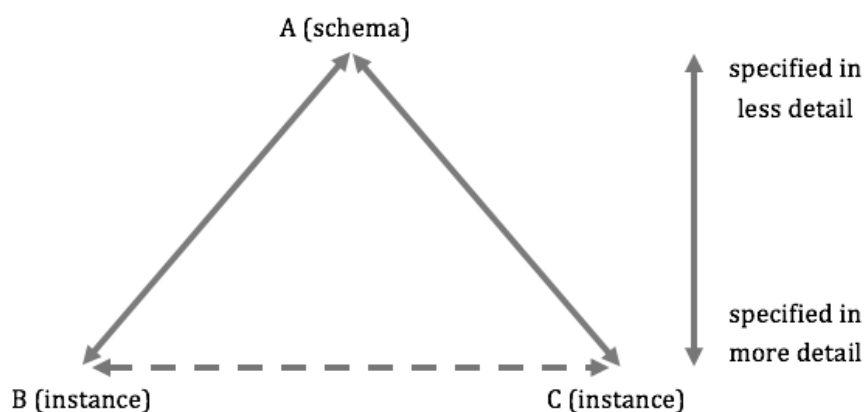


Figure 2.7: A schematic representation of the relation between a schema and its instances (Taylor, 2002:24, 125)

⁷³ Traugott and Trousdale (2013:14) are careful to point out that schemas (and subschemas) are not necessarily mental representations, but rather categories linguists use to systematise language in order to analyse and discuss linguistic systems. There may, however, be an overlap between the mental representations and the linguists' categories for description.

In Figure 2.7, it can be said that A is schematised from the more specific instances B and C. It can also be said that B and C are instantiations of the schema A. Taylor (2002:24) points out that although B and C INHERIT particular specifications from A, they flesh out the schema in more detail and in contrasting ways. Furthermore, it is also said that the wider the range of instantiations that instantiate the schema, the more schematic the schema becomes (Bybee, 2010:67).

In terms of constructions specifically, it can be said that the more instantiations there are of a construction, the more schematic it will be. Here one can consider the difference between the word *green* (as a lexical construction) and *Sally gives Tumi a cake* (as a ditransitive construction). Hoffmann and Trousdale (2011:4) point out that *green* is fully specified in terms of its phonology, with the specified form /gri:n/. In contrast, the ditransitive construction with the specified form [S V O_i O_d] can occur with many expressions, for example, *Sally gives Tumi a cake* and *Sally baked Tumi a cake*. Consequently, the ditransitive construction is said to be more schematic.

Another important concept that plays a role in schematicity is the notion of **SEMANTIC COHERENCE**.⁷⁴ Goldberg (2006:39-40) posits that the Semantic Coherence Principle ensures that the participant role of the verb and the argument role of the construction are semantically compatible. In this regard, she states that:

[T]he more specific participant role of the verb must be construable as an instance of the more general argument role. General categorization processes are responsible for this categorization task and it is always operative. This principle follows from the idea that argument structure constructions are learned by generalizing over the semantics of instances of the pattern used with particular verbs.

(Goldberg, 2006:40)

⁷⁴ Croft (2001:179-185) makes a similar point when he discusses the notions **SELECTIONAL RESTRICTIONS** and **COLLOCATIONAL RESTRICTIONS**. These restrictions are restrictions that are placed on possible combinations of words. These restrictions are determined only by the semantics denoted by the word.

This principle can be illustrated by considering the systematic variation that exists in ditransitive constructions. These systematic variations depend on the verb class with which the constructions interact (cf. Goldberg, 1995:35, 2003:220). The verb *gives*, for example, is a three-argument verb: it requires an agent (i.e. a giver), a recipient and an object that is transferred between the agent and the recipient. Example (14) below can be regarded as a prototypical instance of the ditransitive construction:

(14) Sally gives Tumi a cake.

Example (14) exhibits a perfect match between the lexical semantics of the verb and the constructional semantics since the agent causes the recipient to receive the object being transferred (i.e. Sally transfers the cake to Tumi). The symbolic structure of this construction can be represented as follows: $[[S\ V\ O_i\ O_d] \leftrightarrow [[X\ \text{transfers}\ Y\ \text{to}\ Z]]$.

Over time, however, the ditransitive construction has been extended to appear with additional verbs. An example of such an instance is where the ditransitive construction has been extended to appear with verbs such as *bake*, as in Example (15) below:

(15) Sally bakes Tumi a cake.

In this instance, the meaning of the verb and the meaning of the construction may not always correspond with one another, i.e. they are not always fully semantically coherent. Therefore, additional clusters of constructions (known as SUBSCHEMAS) may exist. These subschemas are linked to a central sense, creating a network. Consequently, Goldberg (1995:35) posits that constructional semantics are highly polysemous (also cf. Croft, 2001:27; Traugott & Trousdale, 2013:14-15). (A representation of such a network is illustrated in Figure 2.8. on p. 117) Example (15) is an example of an instance belonging to such a subschema, where the meaning of the verb does not correspond entirely to the meaning of the construction. In other words, the verb *bake* does not independently encode transfer semantics. Instead, the meaning is derived from the ditransitive construction. In this regard, Boas (2013:236) explains that “the verbs are associated with specific lexical semantic information that allows them to integrate (or ‘fuse’) with the semantics of an Argument Structure construction”. Whereas the meaning of Example (14) is more prototypical of the ditransitive construction, the meaning of Example (15) is more peripheral in that the agent intends to transfer the object to the recipient (i.e. Sally intends

to transfer the cake to Tumi). The symbolic structure of this construction can be represented as follows: $[[S V O_i O_d] \leftrightarrow [[X \text{ intends to transfer } Y \text{ to } Z]]]$. Furthermore, it is said that the schema instantiated by Example (14) (i.e. $[[S V O_i O_d] \leftrightarrow [X \text{ transfers } Y \text{ to } Z]]]$) is more schematised than the schema instantiated by Example (15) (i.e. $[[S V O_i O_d] \leftrightarrow [[X \text{ intends to transfer } Y \text{ to } Z]]]$), since it abstracts over more instances of use.

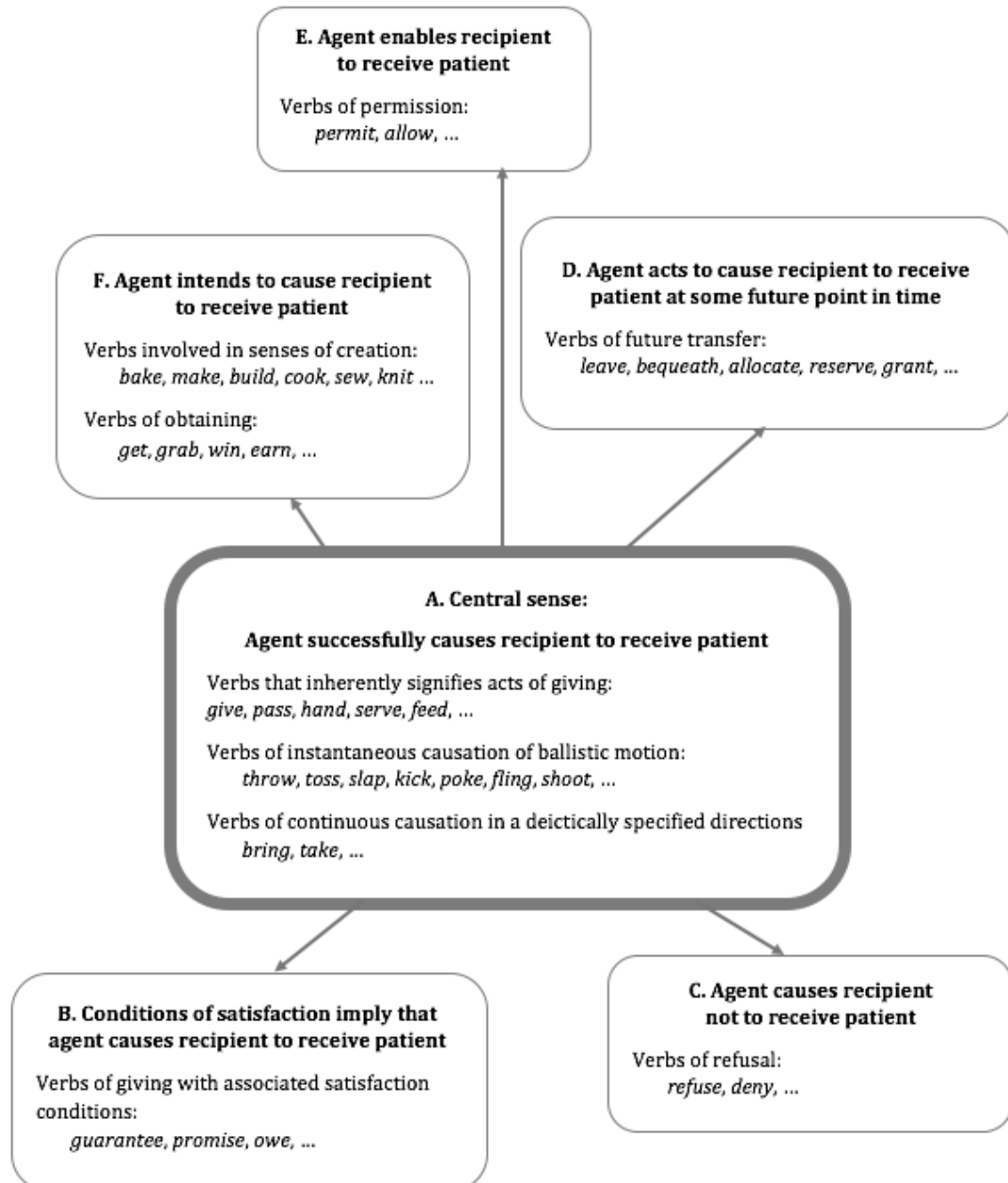


Figure 2.8: *The related senses of the ditransitive construction (Goldberg, 1995:38)*

Another important issue that arises from the notion SCHEMATICITY is the notion of PRODUCTIVITY. Langacker (2000:26) explains that **productivity** indicates the availability of a pattern for the sanctioning of novel expressions. According to Croft (2001:28), a pattern will only be productive if it is highly schematised. However, the productivity of a pattern is determined not only by the degree of schematicity but also by factors such as type frequency, semantic coherence, and token frequency. In the following paragraphs, these factors are discussed together with the notion DEGREE OF SCHEMATICITY.

Boas (2013:247) explains that **type frequency** is measured “to determine how many different items occur in the various schematic slots of a construction”. It also correlates with the construction’s ability to occur with novel items. For example, if a construction can occur with a wide range of verbs in its verb slot, it is said that the construction is productive. Goldberg (1995:98-99) avers that “constructions that have been witnessed with many different verbs are more likely to be extended to appear with additional verbs” (also cf. Boas, 2013:247; Bybee, 2010:67).⁷⁵ The caused-motion construction (which have been witnessed with many different verbs) can be extended to occur with a verb such as *sneeze*, for example, *He sneezed his tooth right across town* (example quoted from Goldberg, 2003:220).

However, Bybee (2010:67) is careful to point out that type frequency interacts with other factors such as the degree of schematicity (as mentioned earlier). When high schematicity is combined with high type frequency, a highly productive construction arises. When high schematicity is combined with low type frequency, the construction becomes less productive. When low schematicity is combined with a high type frequency, however, productivity will be limited since the number of instantiations are limited. Bardhal (2008:172-173) also emphasises that type frequency should not be considered on its own, but in relation to semantic coherence (also cf. Goldberg, 1995:129-138). Based on

⁷⁵ Here it is important to be cognisant of the fact that the construction (i.e. its particular semantic structure in combination with its associated syntactic structure) should be considered independent of the lexical items which instantiate them (Goldberg, 1995:1). The reason for this approach is to “avoid the claim that the syntax and semantics of the clause is projected exclusively from the specifications of the main verb” (Goldberg, 1995:224). Therefore, it could rather be said that the semantics of verbs predispose them to occur within certain clauses (cf. for example Levin, 1993). Boas (2013:236) argues that this approach prevents scholars from positing “implausible verb senses” in cases where verbs occur in an “unusual environment”.

her considerations of historical and psycholinguistic data, Bardhal (2008:172-173) posits that productivity is a function of type frequency, semantic coherence, and the inverse correlation between them. As such, she places the different degrees of productivity on a cline. This cline is illustrated in Figure 2.9. Constructions located at the top of the cline exhibit high type frequency, but a low degree of semantic coherence. Constructions located at the bottom of the cline exhibit low type frequency with a high or low degree of semantic coherence.

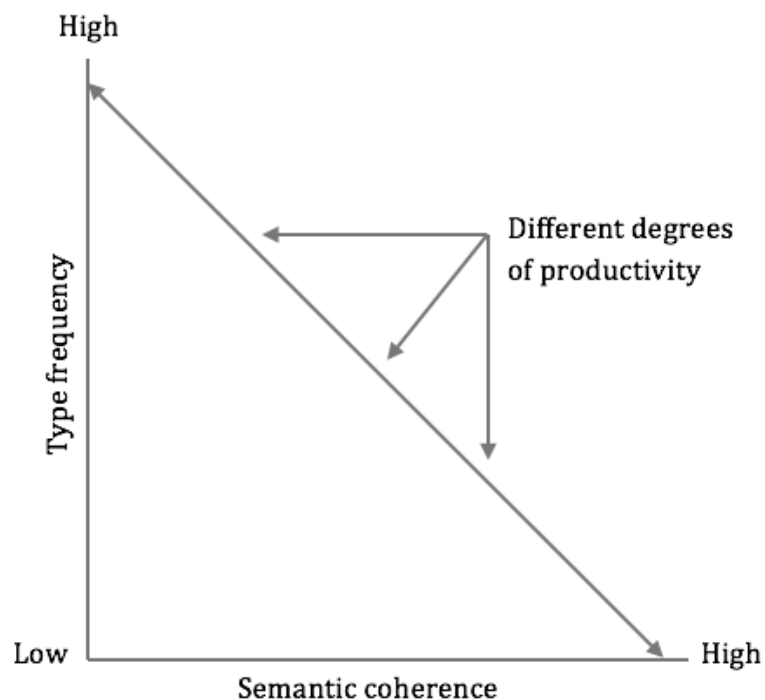


Figure 2.9: The productivity cline (Bardhal, 2008:172)

Although the importance of type frequency is emphasised in the literature, it is important not to neglect token frequency. Croft (2001:28) states that **token frequency** refers to the number of times a constructional schema is activated. If a constructional schema is activated frequently, it is said to have a high token frequency. If, on the other hand, it is activated infrequently, it is said to have a low token frequency. These frequency counts are so important because they are linked to Langacker's (1987a) notion of **ENTRENCHMENT**:

Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, *units are variably entrenched depending on the frequency of their occurrence.*

(Langacker, 1987a:59; my emphasis)

Bybee (2006:711) posits that “the frequency of use of particular instances of constructions have an impact on representation that we can see evidenced in various ways, for example, in speakers’ recognition of what is conventionalised and what is not, and even more strikingly in the nature of language change”. It is said that frequency strengthens the memory representations of constructions (Bybee, 2006:715). Therefore, the more entrenched a construction becomes, the lower the processing costs when speakers retrieve them from memory (Hoffmann, 2004:186).

The concepts described in the paragraphs above, i.e. SCHEMATICITY and PRODUCTIVITY are used by constructionists to explain language change, and more specifically, grammaticalisation. Bybee (2006:719) defines the notion **GRAMMATICALISATION** (also known as GRAMMATICISATION) as the creation of “a new construction out of a particular instance of an old construction”. In turn, Hopper and Traugott (2003:1) provide a more extensive definition of grammaticalisation:

“[G]rammaticalization” refers to that part of the study of language that is concerned with such questions as how lexical items and constructions come in certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions.

During this process, an existing construction will be used more frequently and change in various ways to become a new construction (cf. for example the grammaticalisation of *(be) going to* described in Section 2.2.2.2).

This view of grammaticalisation, however, is extremely restrictive and of little use in terms of argument constructions. Therefore, Mukherjee (2005:198) derives the notion **GRAMMATICAL INSTITUTIONALISATION** from the notion of **GRAMMATICALISATION**. He points out

that grammatical institutionalisation may refer to one of two processes. In the first instance, it is concerned with the process by which a new construction enters the language system. In the second instance, it is concerned with the process by which an existing construction becomes admissible in new contexts. This distinction corresponds with Traugott and Trousdale's (2013) distinction between the notions CONSTRUCTIONALISATION and CONSTRUCTIONAL CHANGE. Traugott and Trousdale (2013:22) define the notion **CONSTRUCTIONALISATION** in its most basic form as "the creation of a form_{new}-meaning_{new} pairing", i.e. the development of a new sign. Should only the form or the meaning of a sign change, it would not constitute constructionalisation, but rather constructional change. Traugott and Trousdale (2013:26) define the notion of **CONSTRUCTIONAL CHANGE** as "a change affecting one internal dimension of a construction". **Gradual constructionalisation** typically entails constructionalisation that is preceded and followed by a number of constructional changes (Traugott & Trousdale, 2013:22). It is also important to keep in mind that these changes are often accompanied by "changes in degree of schematicity, productivity and compositionality" (Traugott & Trousdale, 2013:27).

Mukherjee (2005:205) avers that it is important to distinguish between the notions GRAMMATICAL INSTITUTIONALISATION and CONVENTIONALISATION.⁷⁶ He says that distinguishing these notions from each other enables linguists to explain two important aspects of language use. First, the distinction enables linguists to gain insight into language users' understanding of the novel structures that may have started the process of institutionalisation but have not (yet) been conventionalised. Second, the distinction enables linguists to understand "the familiarity of linguistic forms that occur fairly frequently" and have therefore already been conventionalised (Mukherjee, 2005:205). The relationship between grammatical institutionalisation and conventionalisation can be described by considering Figure 2.10.

⁷⁶ The notion of CONVENTIONALISATION has been discussed in Section 2.2.2. Within CG, the term can be operationalised by saying that conventionalisation has taken place when possible forms have been turned into probable forms (Mukherjee, 2005:204). This can be determined by means of frequency counts.

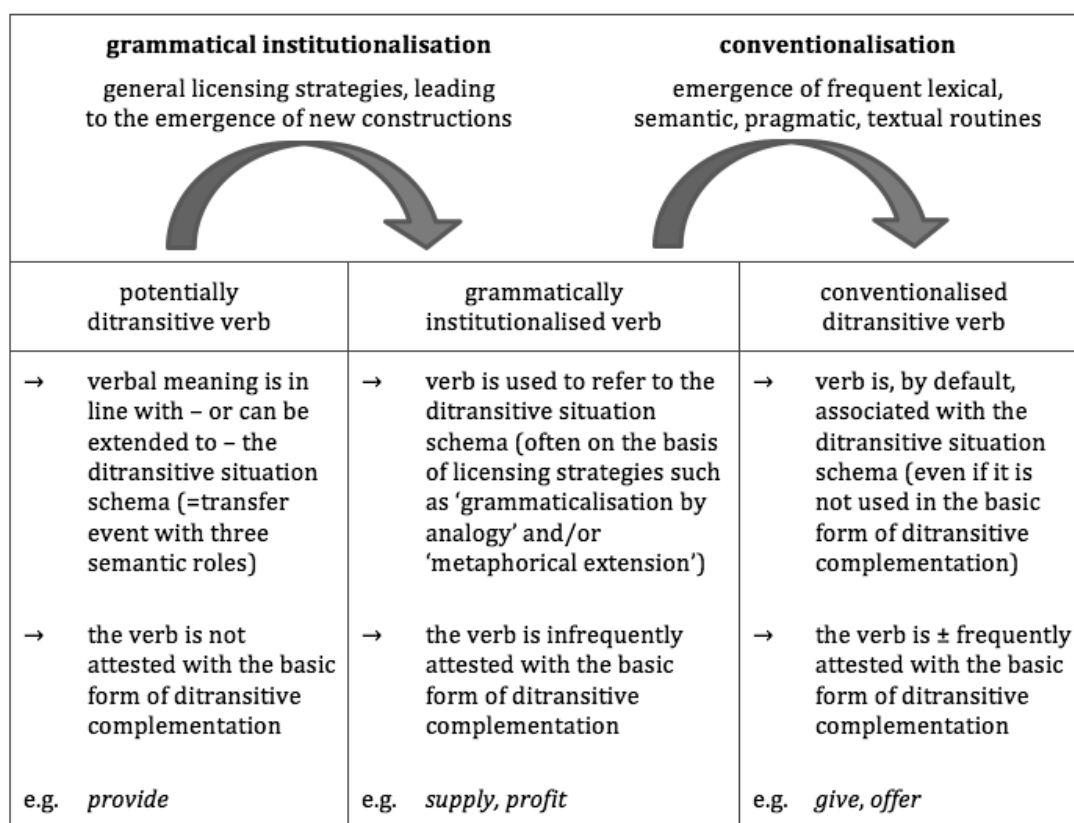


Figure 2.10: *The grammatical institutionalisation and conventionalisation of ditransitive verbs (Mukherjee, 2005:205)*

The first column is dedicated to verbs that have the potential to be used as ditransitive verbs. These are typically the verbs that are on the periphery of ditransitivity. Mukherjee (2005:205-206) attributes their peripheral status to the fact that they have been grammatically institutionalised as ditransitive verbs, but that speakers do not use them with sufficient frequency to be considered institutionalised. He goes on to explain that speakers start using these verbs in ditransitive constructions primarily because their verbal meaning is closely related to the verbal meaning of verbs typically used in ditransitive constructions (Mukherjee, 2005:206-207). The second column is dedicated to verbs that have been grammatically institutionalised as ditransitive verbs (Mukherjee, 2005:204). These are typically verbs that occur infrequently in ditransitive constructions. Finally, the third column is dedicated to verbs that are regarded as prototypes of ditransitivity. These are the verbs that are typically associated with the ditransitive construction.

From a methodological perspective, the quantitative data obtained from corpus analysis enables researchers to determine whether a verb is grammatically institutionalised or whether it is conventionalised. However, Mukherjee (2005:204-205) is careful to emphasise that simple frequency counts are not sufficient when investigating grammatical institutionalisation and conventionalisation. Hoffmann (2004:190) makes a similar point when he warns that frequency only “becomes a meaningful diagnostic tool when if it is compared with the frequency of occurrence of related linguistic phenomena”. Therefore, he advocates an analysis where both quantitative data and qualitative data are considered:

[A] frequency-based analysis not only needs to determine how often a particular item is actually found but also how often it *could* have occurred but in fact didn't because the concept was expressed differently. Such an approach represents a more differentiated, two-layered view of frequency phenomena.

(Hoffmann, 2004:190; his emphasis)

Considering Hoffmann's (2004) view above, it is useful to consider **COLLOSTRUCTIONAL ANALYSES**⁷⁷ when investigating grammatical institutionalisation and conventionalisation. This approach is particularly useful, not only because it transcends frequency counts but also because it increases the adequacy of grammatical description (Stefanowitsch & Gries, 2003:210). Stefanowitsch and Gries (2003:211) explain that collostructional analyses allow linguists to determine “the extent to which particular slots in a grammatical structure prefer, or are restricted to, a particular set or semantic class of lexical items”. In other words, it enables linguists to determine whether there are significant associations between lexemes and the grammatical structures in which they occur.

Following the issue of grammatical institutionalisation and conventionalisation, Mukherjee (2005:204-205) emphasises the importance of describing the mechanisms that underlie the process of grammatical institutionalisation. He calls these mechanisms **LICENSING STRATEGIES**. Two important licensing strategies are distinguished, namely

⁷⁷ The word *collostruction* is a blend of the notions COLLOCATION and CONSTRUCTION.

metaphorical extension (Goldberg, 1995) and **grammaticalisation by analogy** (Hoffmann, 2004). These licensing strategies are considered in the following paragraphs.

Metaphorical extension is based on the categories of related meanings. Goldberg (1995:33) explains that the semantics associated with a construction tends to be associated with “a set of systematically related senses”.⁷⁸ Goldberg (1995:33) argues that this **constructional polysemy** (i.e. where the same form is paired with different but related senses) enables linguists to capture the relationships between the different senses in a natural way. To illustrate, the constructional polysemy of the ditransitive construction can be considered. The central meaning of the ditransitive construction involves “the successful transfer of an object to a recipient, with the referent of the subject agentively causing the transfer” (Goldberg, 1995:33). This meaning can be extended metaphorically to point to a situation where the intended transfer of an object to a recipient takes place. The verbs used with the central meaning are regarded as more salient and therefore categorised as conventionalised ditransitive verbs, whereas the verbs used with the peripheral meanings occur less frequently and are regarded as grammatically institutionalised verbs.

Hoffmann (2004:194-195) proposes another important licensing strategy. This strategy is known as **grammatical analogy**. This strategy licenses low-frequency combinations to be grammatically institutionalised by means of analogy to their more frequent ‘structural relatives’ (Hoffmann, 2004:195). This means that the salience of a particular construction can at least partly be ascribed to the extent to which the form of a construction is similar to the form of constructions that occur more frequently. In his description of ditransitive constructions, Mukherjee (2005:206) demonstrates that while the notion of grammatical analogy is quite useful, it must be complemented with a semantic component in the case of argument structure. In other words, there should be some kind of “semantic analogy that can be drawn between the verbal meaning of a potentially ditransitive verb and a typical ditransitive verb”. This idea was touched upon earlier in this section while

⁷⁸ This issue has been addressed to a certain extent earlier in this section in the discussion pertaining to Example 15 and Example 16.

discussing Goldberg's notion of SEMANTIC COHERENCE, i.e. the notion that the participant role of the verb and the argument role of the construction should be compatible.

2.4.3 Summary

Section 2.4 provides the tools for the grammatical analysis of the verb-complementational features in BSAE. The realisation of the form-meaning pairing in constructions has been considered, together with the role that frequency plays in the schematisation and ultimately the conventionalisation of innovative grammatical features.

2.5 Conclusion

The first part of this chapter focussed on the emergence of new varieties of English. The section commenced by providing an overview of the models used to describe the emergence and spread of New Englishes. Four models were addressed in this chapter. They are Strang's tripartite model of English, Quirk's views on the varieties of English, Kachru's three-circle model, and Schneider's Dynamic Model of the Evolution of New Englishes. The first three models address the spread of New Englishes from a primarily sociolinguistic perspective. The last model, however, integrate sociolinguistic perspectives with psycholinguistic perspectives, providing scholars with more insight into the spread of New Englishes.

The overview of the models used in the description of the spread of Englishes was followed by a section in which the attitudes towards New Englishes were considered. Here, it was shown that the attitudes towards New Englishes are often verbalised in terms of standard English and non-standard English. Whereas standard English is often discussed in terms of norms (of correctness), non-standard English is often discussed in terms of deficit linguistics. In this regard, native-like proficiency is often considered to be the yardstick against which language competence is measured. However, since native-like proficiency is difficult to define, the yardstick has proven to be elusive. In part, the difficulty in defining native-like proficiency was attributed to the heterogeneous nature of English. The heterogeneous nature of English has also given rise to the problem of

distinguishing errors from innovations. In this regard, it was posited that attitudes towards varieties play an important role with regard to acceptability. This includes not only the attitudes of L1 speakers, which tend to be negative, but also the attitudes of non-native speakers, which tend to be characterised by linguistic insecurity.

This section was followed by a section in which the theory from all of the aforementioned sections was synthesised and applied to (Black) South African English. In the section, it was posited that South Africa poses certain challenges to the models used for the description of the spread of New Englishes. This is not only due to South Africa's complex and unique history of colonisation, but also due to the multilingual nature of South Africa's speech communities.

Since language change plays a fundamental role in the emergence of new varieties of language, some of the most pertinent aspects regarding the evolutionary approaches to language change were discussed. They include the theory of competition and selection, the ecology of language change, and the founder effect. These aspects play an important role in theories geared towards describing and explaining the emergence of new varieties of English.

An exploration of language as an internal construct and language as an external construct was also undertaken. Typically, language as an internal construct is investigated in the field of psycholinguistics, whereas language as an external construct is investigated in the field of sociolinguistics. It has been demonstrated, however, that the psycholinguistic aspects and the sociolinguistic aspects of language are closely intertwined. This complicates investigations into the relationship between psycholinguistics and sociolinguistics. From a psycholinguistic perspective, research is conducted to determine how language variation originates from within individual speakers. In this regard, individual grammars, SLA, bilingual activation and CLA were shown to be of particular importance. The issues pertaining to bilingual activation and CLI are subsumed under the notion of constrained language. From a sociolinguistic perspective, language is typically studied in relation to society. As such, speech communities play a pivotal role in sociolinguistics. Two important aspects in this regard are identity and accommodation. Since no two speakers experience language in exactly the same way, language becomes indexical of a speaker's origins or of a speaker's social identity. It is important, though, to

keep in mind that linguistic and social identities are dynamic and that speakers tend to adjust their language to accommodate hearers.

The second part of the chapter started with discussion aimed at contextualising the emergence of new varieties of English and the role of linguistic analysis in this regard. This discussion has favoured the emergentist perspective. The underlying tenets of the emergentist perspective make it particularly suitable for studies investigating language change. The emergentist perspective posits that *a-priori* categories do not exist in language. In other words, language is “something speakers do” as opposed to “something speakers have”. As such, language performance is not disregarded in favour of language competence. This allows scholars to investigate not only language change, but also language variety. Furthermore, it was posited that language does not exist in completely homogeneous speech communities.

Several approaches in the description of verb complementation are mentioned. CG was shown to be particularly suitable for the investigation of language phenomena from a usage-based perspective. Since the results of this thesis are primarily investigated, described, and interpreted from a CG perspective, a definition of CG was provided. The definition was followed by an overview of the most significant usage-based approaches that underlie this thesis.

The methodological aspects regarding the analysis of authentic BSAE data are described in the next chapter.

Chapter 3

Methodology

3.1 Contextualisation

Graham (2013:x-xi) suggests that researchers approach their research in stages when they wish to launch an investigation into any phenomenon. The first, most critical stage, is the stage in which researchers pose their research questions. This stage is followed by the stage in which data is collected. The third stage entails the analysis of the data. Finally, the results are interpreted in the fourth stage. Graham (2013:xi) illustrates these four stages as follows:

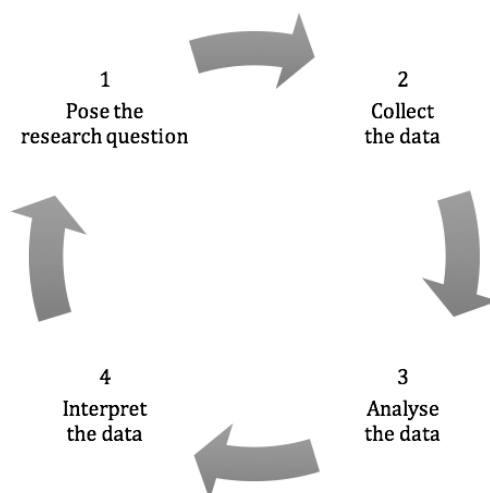


Figure 3.1: The four stages of a research project (adapted from Graham, 2013:xi)

The first stage of this research was addressed in Chapter 1. In this chapter, the research problem was contextualised and the research questions were formulated accordingly. These research questions are geared towards gaining insight into the verb complementation patterns observed in BSAE. The current chapter focuses on the data collection, as well as the methods employed to analyse the data. The interpretation of the data takes place in Chapter 4 onwards.

In studies of a linguistic nature, scholars can collect various kinds of data. Scholars such as Gilquin and Gries (2009:1-2) and Gries (2013a:94) distinguish two major kinds of linguistic data. They are introspective data and observational data. Traditionally, **introspective data** abandoned the use of collected data. The abandonment of data collection is accompanied by the introduction of “very informally collected linguistic acceptability judgements (primarily by the analyst him/herself) as the primary source of data” (Gries, 2003b:94). This approach results in situations in which analysts simply discard any data that conflict with their own judgements. This approach becomes problematic since linguistic competence and performance are too complex to be described by introspective acceptability judgements alone (Svartvik, 1992). To complicate matters even more, Kirsner (2014:2) finds that native speaker intuitions are inconsistent. In spite of the aforementioned problems associated with introspective data, Gilquin and Gries (2009:3) and McEnery and Wilson (2001:12) point out that introspective data should not be discarded completely since it is a form of data, and as

such it has the potential to be a valuable tool – especially when collected with the necessary care and precaution.

Introspective data is increasingly used in conjunction with **observational data**. In the category of observational data, Gries (2006b:3-4) notes that corpus linguistics has rapidly become an autonomous methodological paradigm within linguistics during the past few decades. He goes on to say that many studies (especially studies conducted from a usage-based perspective) are now based on electronic corpora which consist of naturally-occurring language. Gries (2006c:191-192) emphasises some advantages that can be attributed to the use of corpus data. One of the most important advantages is that corpus data enable researchers to quantify the use of specific linguistic features, allowing “for a rather objective identification of what may be considered important and what may be considered rather marginal” (Gries, 2006c:191). By quantifying features, the researcher also has the means to determine the extent to which these features have conventionalised, resulting in discussions that are descriptively more adequate. In this regard, Bybee (2010:12) argues that the use of quantifiable data provided by corpus data is particularly useful in the CG and usage-based framework.

At this stage, it should be pointed out that several scholars advocate the combined use of introspective data and observational data. Croft (1998:152), for example, explains that “[i]ntrospective linguistic evidence can limit the range of alternative mental representations to a set of possibilities”, while observational evidence (e.g. corpus data) enables linguists “to narrow this set of possibilities to a single plausible model”. Fillmore (1992:35), in turn, expresses the view that introspective data and observational data complement one another:

The first [observation] is that I don't think there can be any corpora, however large, that contain information about all the areas of English lexicon and grammar that I want to explore; all that I have seen are inadequate. The second observation is that every corpus that I've had the chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way. My conclusion is that the two kinds of linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body.

Gries (2006a:87) is also very emphatic in stating that he does not advocate the use of corpus data on its own since the results obtained from the different sources of data will not remain constant. Therefore, linguists should always check their findings against different sources of data.

In this thesis, corpora provide data that allows the researcher to transcend (subjective) observations by quantifying innovative linguistic features. In Chapter 1, it has been pointed out that the nature of the research aims of this study does not allow for the use of decontextualised sample sentences. Instead, a usage-based approach that favours the use of corpora is followed. Consequently, corpora are particularly suitable for the purpose of revealing innovative verb complementation patterns in BSAE. The principal corpus data used in this study are comprised of texts written by BSAE language users. Each text is aligned with the edited version of the same text. This enables the researcher to check whether specific features are accepted and legitimised by the authoritative measures, i.e. the editor in this case (cf. Section 2.3.2).

This chapter sets out to describe the two kinds of data that are employed in this study. The corpus data is described in Section 3.2.1.2. This section is followed by Section 3.2.1.3 in which the analyses of the data are addressed. The chapter is concluded in Section 3.3.

3.2 Observational data

As pointed out in the previous section, two main kinds of data can be distinguished in linguistic studies. They are observational data and introspective data. This section focuses on observational data and (more specifically) corpus data. In Section 3.2.1, corpus linguistics as a research method is considered. It is followed by Section 3.2.1.2 in which the corpora used in this study are described. Finally, Section 3.2.1.3 focuses on the quantitative analysis of the corpus data.

3.2.1 *Corpus linguistics as research method*

Corpus linguistics is an approach or research method for studying language use and is regarded as a complementary approach to some of the more traditional approaches (Biber, Conrad & Reppen, 1998:9-10; Bowker & Pearson, 2002:9; Leech, 1992:105-106; Thompson & Hunston, 2006:8). In this regard, Leech (1992:105-106) draws a clear distinction between corpus linguistics and other branches of linguistics such as psycholinguistics and sociolinguistics. He argues that whereas theories in psycholinguistics and sociolinguistics are domains of study within linguistics, corpus linguistics provides “a methodological basis for pursuing linguistic research” (Leech, 1992:105).

At this point, it should be mentioned that not all linguists exploit corpus data in the same way. In fact, there are many different methods that can be used to analyse corpus data (Teubert, 2005:1-2). Taylor (2008:182-183) points out that these methods are often reflected in the terminology used by scholars (even though they may not always consciously be aware of the scientific context in which they work). Therefore, she investigates the different ways in which scholars refer to the notion of CORPUS LINGUISTICS in their research. She demonstrates that scholars rarely use the term CORPUS LINGUISTICS but rather tend to use terms such as CORPUS, CORPUS-BASED, CORPUS-DRIVEN, CORPUS-ASSISTED, CORPUS ANALYSIS, CORPUS APPROACH and CORPUS STUDY. Gilquin and Gries (2009), in turn, specifically investigate how linguists exploit their corpus data. They have found that studies in which corpora are the only source of data, scholars tend to approach the corpus data in an exploratory fashion, i.e. there is no rigorously formulated hypothesis as the data is expected to speak for itself (Gilquin & Gries, 2009:10). On the other hand, they have found that in studies in which corpora are not the only source of data, scholars tend to explicitly formulate a hypothesis (Gilquin & Gries, 2009:11-12).

The aforementioned distinction points to a distinction that is often mentioned in approaches to corpus linguistic approaches. It is the distinction between corpus-based approaches and corpus-driven approaches. Biber *et al.* (1998) declare that they are adherents of the corpus-based approach. They demonstrate that the corpus-based approach enables linguists to conduct research of a scope and reliability of analysis that is robust and consistent (also cf. Gries, 2012:478). This can be attributed to the fact that

this methodology is empirical, uses computers for analysis and requires both quantitative and qualitative analytical techniques (Biber *et al.*, 1998:4). By following a corpus-based approach, Biber *et al.* (1998:5) argue that linguists can “identify and analyze complex ‘association patterns’”.⁷⁹ They point out that corpus analyses allow linguists not only to analyse a particular linguistic feature but also to consider their systematic associations with other features.

Tognini-Bonelli (2001), however, have a stricter view of corpus-based linguistics. She states that the term **CORPUS-BASED RESEARCH** is used to refer to a methodology in which the main purpose of the corpus is to “expound, test or exemplify existing theories and descriptions” (Tognini-Bonelli, 2001:65). Tognini-Bonelli (2001:66) criticises the corpus-based approach primarily on the basis that the approach is used to validate existing theoretical models:

We could say, therefore, that corpus-based linguists adopt a ‘confident’ stand with respect to the relationship between theory and data in that they bring with them models of language, and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve the data accordingly [...] corpus evidence is brought in as an extra bonus rather than a determining factor with respect to the analysis, which is still carried out according to pre-existing categories [...] [the corpus evidence] is never really in a position to challenge [such categories] as there is no claim made that they arise directly from the data.

Based on the criticism above, Tognini-Bonelli (2001:84) advocates a corpus-driven approach in corpus analysis. She avers that the term **CORPUS-DRIVEN** research is used to refer to a methodological approach in which “the commitment of the linguist is to the

⁷⁹ Biber *et al.* (1998:5) explain that the notion of **ASSOCIATION PATTERNS** refers to “the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features”.

integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus-evidence” (Tognini-Bonelli, 2001:84).

Some scholars, however, question the viability of Tognini-Bonelli’s (2001) strict view of corpus-based and corpus-driven studies. Mukherjee (2005:72), for example, finds Tognini-Bonelli’s (2001) view of corpus-driven studies implausible and unrealistic. First, he states that her view is implausible “since any corpus is compiled on the grounds of linguists’ informed intuitions about language in the first place” (Mukherjee, 2005:72). He goes on to explain that even during the process of corpus compilation, linguists’ intuitions about language play a role when decisions are made regarding size and representativeness.⁸⁰ In fact, Tognini-Bonelli (2001:178) herself states that it is impossible for scholars to have a “theory-neutral stance”. Second, Mukherjee (2005:72) states that Tognini-Bonelli’s (2001) view of corpus-driven linguistics is unrealistic because “any linguistic research activity stems from some sort of initial intuitions about language”. He provides the following example:

[A] study of ditransitive verbs can hardly ever be corpus-driven because the concept of ditransitivity itself is a received idea that has a well-established tradition in linguistics [...] Such a received idea is always at the basis of corpus-linguistic research into any linguistic phenomenon because the phenomenon at hand has to be defined (or at least circumscribed) initially before corpus data are scrutinised. So, there always is some sort of theoretical preconception involved, and, what is more, even the avoidance of an *a priori* theory is a theoretical preconception.

(Mukherjee, 2005:72)

⁸⁰ The characteristics of corpora are discussed in more detail in Section 3.2.1.1.

Mahlberg (2005:18) has similar concerns regarding Tognini-Bonelli's (2001) view of corpus-driven linguistics and questions whether a corpus-driven approach is executable in practice.⁸¹ In this regard, she asks:

How far do we have to go in an attempt to let the data come first? In the search for a description of general nouns can we take for granted that we know what a noun is, or would we have to start even earlier with a description of all words before we can form classes such as nouns?

Other linguists' views of corpus-based research also do not agree wholly with Tognini-Bonelli's (2001) view thereof. For example, Biber *et al.*'s (1998:4-5) view of corpus-based research mentioned earlier transcends the expounding and exemplifying of "theories and descriptions that were formulated before large corpora became available to inform language study" (Tognini-Bonelli, 2001:65) in that they do not model their research on existing theories and descriptions. Instead, they set out to describe language by "studying the use of language characteristics by considering the relevant 'association patterns'" (Biber *et al.*, 1998:5). They explain that with the use of association patterns, they measure the extent to which linguistic features and variants are associated with contextual factors.

In light of the aforementioned difficulties in defining the notions of CORPUS-BASED RESEARCH and CORPUS-DRIVEN RESEARCH, some linguists – such as Rayson (2008:521) – use the notion of DATA-DRIVEN RESEARCH to draw attention to the fact that while he uses a corpus-driven approach, he still relies on categories established by certain linguistic theories (i.e. he relies on part-of-speech tagging which is based on pre-determined linguistic categories).

In this study, a primarily corpus-based approach is followed. This approach does not preclude the adjustment of grammatical categories as suggested by Tognini-Bonelli (2001). It means that the grammatical categories described in CG – more specifically the grammatical categories pertaining to argument structure – are used to provide a starting point for the analysis of corpus data. These grammatical categories may need to be

⁸¹ Here it is important to note that even though Mahlberg (2005:18) questions this strict view, she acknowledges that researchers must be cognisant of the fact that the terminology used to describe data should be treated with caution since it may impact the results.

adjusted, since the experiences of the speakers in different speech communities may not be expressed in the same way and these differences may be reflected in the speakers' language. This approach is in line with the aim of the study to not merely find out more about what is already known of verb complementation in BSAE, but also to gain insight into the patterns of verb complementation in BSAE that was never available before.

Furthermore, it has been pointed out in Chapter 2 that language change primarily takes place during language use, and that it is largely influenced by repetition and frequency of use (also cf. Bybee, 2010:11). Since corpora are reflections of language in use, it is a useful tool in obtaining empirical data. For example, frequency effects may raise many interesting questions (Bybee, 2013:68), such as "questions concerning the interaction of token frequency with type frequency, the gradual development of autonomy, the interaction of schematicity and type frequency in determining productivity, and the question of the effects of different levels of token frequency" (also cf. Teubert, 2005).

Gries (2010:198), however, points out that scholars cannot rely on frequencies of occurrence to predict aspects of processing perfectly. He lists two reasons for this fact. In the first place, he states that frequency of occurrence is noise-ridden due to its intercorrelation with factors such as the salience of words or concepts, the recency of occurrence, concreteness/manipulability, and so on. In the second place, he argues that frequency of occurrence does not "enter into a straightforward one-to-one relationship with aspects of processing because any particular frequency of occurrence can arise from very different distributional patterns" (Gries, 2010:198). He explains the problem as follows:

[A] word *w* may occur 18-20 times in each of ten very different registers, or it may occur 190 times in only one of the ten registers. While these two results look the same in a frequency list of the complete corpus of ten registers, it is obvious that these results would not be the same: they would not be the same for the corpus linguist who may be interested in register-dependent vocabulary differences, and they would not be the

same for the psycholinguist or language acquisition researcher who knows that learning processes in general exhibit a distributed learning or spacing effect.

(Gries, 2010:198)

In this study, the aim is to overcome these biases that may arise from the measurement of the frequency of occurrence by employing statistical tools such as the Chi-squared test for independence, the Fisher exact test, collocation analysis, inferential decision trees, and random forests.

Since corpora will allow the researcher to obtain the frequencies of occurrence, the general concept **CORPUS** is defined in Section 3.2.1.1. This section is followed by Section 3.2.1.2, which provides a description of the corpora used in this study. In Section 3.2.1.3.1, an overview is provided of the corpus analysis tools used to analyse the corpora. Finally, Section 3.2.1.3.2 provides an overview of the statistical measures that are employed to gain insight into the various linguistic features.

3.2.1.1 *Defining corpora*

Teubert (2005:1) points out that working with real language data is not a new approach to studying language:

When, two hundred years ago, the philologists embraced the philosophy of the enlightenment and set off to find the laws that make language work, they quickly found out that first of all, they needed reliable language data. And while some of them never gave up the dream of uncovering the secret eternal laws that make language work, most philologists were happy to make tentative generalisations about the multitude of data they had to deal with. Yet the more language data they had, the more complicated these generalisations became.

(Teubert, 2005:1)

In the last few decades, however, scholars came to realise that – in the field of corpus linguistics, at least – not any collection of texts can be regarded as a corpus. Before a collection of texts can be regarded as a corpus, it must fulfil certain requirements. These requirements have been reiterated by many corpus linguists. The first requirement is that the corpus must be **machine-readable**. When corpora are machine-readable (i.e. computerised), scholars can use specialised corpus analysis tools to analyse data in a fraction of the time it would take humans to do the analyses by hand. Not all analyses can be done automatically by computers, though, and many analyses still require interactive techniques which require the input of the linguist (Biber *et al.*, 1998:4; Bowker & Pearson, 2002:10).

The second requirement pertains to the **size** of the corpus. The general consensus is that the corpus should be large. However, Bowker and Pearson (2002:45) point out that “the adjective ‘large’ is rather vague” and that there are no hard and fast rules to determine the ideal size of a corpus. However, one can always assume that bigger is always better. In this regard, McEnery and Wilson (2001:30-31) suggest that when linguists start the process of compiling corpora, it is useful to decide the number of samples and the number of words that are to be collected. This kind of corpus is known as a finite corpus. The number of texts from different registers, the number of samples from each text, as well as the number of words in each sample, should also be considered (Biber *et al.*, 1998:248-249). By considering these issues, linguists are well on their way to create a corpus that is not only large but also representative and balanced.

Biber (1993:243) defines **representativeness** as “the extent to which a sample includes the full range of variability in a population”. Although the concept REPRESENTATIVENESS is elusive, i.e. it cannot reasonably be obtained⁸² (Zanettin, 2012:46), it is something that all linguists attempt to obtain by considering variability from both a situational and a linguistic perspective (Biber, 1993:243). As such, linguists consider the extent to which the range of text types are represented in the corpus together with the range of the linguistic distributions in the language. Given the aforementioned considerations, Zanettin (2012:45-46) argues that it is important to ensure the size of the (sub)corpus is

⁸² Also cf. Atkins, Clear and Ostler (1992) who state that any attempt to achieve “a perfectly ‘balanced’ corpus” is “theoretically suspect”.

“proportional to the relative frequency of occurrence of these texts in the textual categories represented by it”. In BSAE, for example, language users do not often write in instructional and popular registers. Therefore, when the BSAE corpus is compiled, the instructional and popular registers will not be as large as the academic or reportage registers.

Finally, it is important that the texts included in the corpus be produced in a natural communicative setting (Gilquin & Gries, 2009:6). In other words, the texts must be **authentic**. This is a particularly important requirement since linguists need to analyse authentic texts produced in natural environments to gain a full understanding of the usage and function of utterances. As pointed out earlier, the usage and function of utterances provide the linguist with unique insights into the experiences that speakers have with language.

Biber *et al.* (1998:4) point out that, taken together, the aforementioned characteristics attributed to corpora result in a scope and reliability of analysis that would not have been possible otherwise.

There are also several advantages attributed to the use of corpora in linguistic research. The first advantage of using corpora is that the statements linguists make about a language variety are grounded in the actual use of the language (Egan, 2008:3). When scholars use their intuition (especially when investigating new varieties of English), they are more likely to spot unusual linguistic features than more typical features (Biber *et al.*, 1998:3). Once they have spotted an unusual linguistic feature, they are primed to find more instances of the features. Consequently, the frequency with which the feature occurs is overestimated. This phenomenon is known as **CONFIRMATION BIAS** (Kahneman, 2011:81).

Croft (1998:168-169) points out that another advantage of the use of corpora is that scholars can make statements regarding the distinctness of use of a linguistic feature (Altenberg & Granger, 2001), evidence for the conventionality of use (cf. Van Rooy, 2011), as well as evidence for the relatedness of uses (Pedersen, Pakhomov, Patwardhan & Chute, 2007).

A third advantage of studying corpora of naturally occurring language is that linguists are better able to understand how speakers experience language. In this regard, Van Rooy (2013:10-11) says that since linguists can retrieve many instances of a linguistic feature, they are in a position to obtain a comprehensive overview of the functional possibilities. Consequently, they are in a better position to understand the function of language. This is particularly the case when linguists follow a corpus-driven approach since this approach enables linguists to spot linguistic features and patterns that “occur too far apart for the human analyst to spot the correspondences in active memory” (Van Rooy, 2013:11).

Corpus linguistics is a particularly useful method for the investigation of language change. In Chapter 2, it has been pointed out that language change primarily takes place during language use and that it is largely influenced by repetition and frequency of use (also cf. Bybee, 2010:11). Since corpora are reflections of language in use, it is a particularly useful tool in obtaining empirical data. For example, frequency effects may raise many interesting questions (Bybee, 2013:68), such as “questions concerning the interaction of token frequency with type frequency, the gradual development of autonomy, the interaction of schematicity and type frequency in determining productivity, and the question of the effects of different levels of token frequency” (Teubert, 2005). Gries (2013b:4) points out that quantitative results allow linguists “to identify what merits a qualitative discussion in the first place”.

In this study, the distinction between the notion of ERROR and the notion of CONVENTIONALISATION is an important one and plays a pivotal role in the analyses of the corpora (cf. Section 2.2.1.1 and Section 2.3.2). However, Hundt and Mukherjee (2011b:215) point out that it is difficult to operationalise the labelling of a linguistic pattern that diverges from the ENL norm as an “error” or an “innovation” because the distinction is not based on linguistic criteria.

In the first instance, it is important to keep in mind that what counts as an error in standard English may not necessarily count as an error in the institutionalised variety of BSAE. This means that even though an error-tag analysis seems to be the obvious solution to identifying errors in the corpora, it is not appropriate in studies such as this one, because an error-tag analysis typically takes place within the paradigm of second-language acquisition, while this study deals with BSAE as an institutionalised variety of English.

Since errors typically arise from a slip of the tongue in institutionalised varieties (cf. Mesthrie & Bhatt, 2008:45-46), one can assume that an error will not occur as consistently and frequently as a conventionalisation. A conventionalisation, on the other hand, can be regarded as a linguistic form that occurs regularly and widely in a variety (cf. Mesthrie & Bhatt, 2008:47). Van Rooy (2011) argues that a linguistic form becomes a feature when it becomes entrenched in the individual grammars of the members of the speech community. This entrenchment points to the grammatical stabilisation of the linguistic form. Typically, grammatical stabilisation is measured by means of frequency counts.

Using frequency counts to distinguish errors from features is not without problems, though. Frequency as a measure of conventionalisation raises the question of how often a specific linguistic form should occur in a corpus before it can be regarded as a feature. In light of this problem, another form of measurement is required. Gut (2011) and Van Rooy (2011) suggest that if a linguistic form is accepted by the members of a speech community, it points to the fact that the linguistic form is accepted and should be considered a feature of the variety (cf. Section 2.3.2). Since a high frequency count of the linguistic form in the language of a speech community already signals the community's acceptance of the form, it may be more useful to consider the linguistic gatekeepers' acceptance of the form. This is also known as the authoritative measure. Kruger and Van Rooy (2017:21) view the authoritative measure as a contributing force in the endonormative stabilisation of a variety.

From a corpus linguistic perspective, the authoritative measure can be gauged by aligning the source texts originally written by the BSAE speakers with the edited versions of the same texts. If a feature is largely left unchanged by the editors, it points to the linguistic

gatekeepers' acceptance of the linguistic form of as a feature of the variety and the potential endonormative stabilisation of the variety. If, however, the linguistic gatekeepers remove the linguistic form from the texts they edit, it suggests that the linguistic gatekeepers do not find the linguistic form acceptable.

In the section that follows, the corpora used in this study are described.

3.2.1.2 Description of the corpora used in this study

In this study, three corpora are employed to answer the research questions posed in Chapter 1. The first corpus is the BSAE editing corpus. It is a parallel corpus of original texts written by experienced BSAE writers and their edited counterparts. The original texts are used to describe the verb complementation patterns of experienced⁸³ BSAE writers, while the edited counterparts are used to determine the acceptance of innovative verb complementation patterns amongst editors. Kachru (1992a:56-57) explains that “a variety may exist, but unless it is recognised and accepted as a model it does not acquire a status”. This approach allows the researcher to measure the extent to which the linguistic gatekeepers of the publishing industry accept the features used by the BSAE language users, and so gain insight into the dynamics of the interaction between conventionalisation and legitimisation.

The second corpus employed in this study is the WSAE editing corpus and serves as a control corpus. As such, the results from the analysis of this corpus are used not only to identify the innovative verb complementation patterns in the BSAE editing corpus but also to determine the extent to which these innovative patterns have stabilised.

The third corpus employed in this study is the Tswana Learner English (TLE) corpus. In some earlier research, for example, observed learner errors were sometimes presented

⁸³ Here, the fact that the BSAE writers are experienced writers (as opposed to learner writers) is important since they are regarded as “the norm-setting segment” of the BSAE population (Van Rooy, 2011:204).

as features of BSAE. An analysis of the TLE corpus will assist in the contextualisation of some of the findings in this study that contradict these observations.

In the sections that follow, a comprehensive description of the corpora employed in this study is provided.

3.2.1.2.1. The Black South African English (BSAE) editing corpus

The first corpus used in this study is the BSAE editing corpus. This corpus was compiled in the Constrained Language project. The BSAE editing corpus is a parallel corpus comprising 244 texts (totalling 279 740 words). Of the 244 texts, 122 texts (totalling 137 262 words) are original texts written by 35 BSAE speakers. Each of these texts is aligned to its edited counterpart (totalling 142 550 words). The texts were edited by professional South African editors who are either L1 English speakers or balanced English/Afrikaans bilinguals.⁸⁴ The details pertaining to the compilation of this corpus are summarised in Table 3.1.

Four registers are represented in this corpus. They are academic, instructional, popular, and reportage. The academic register is comprised of academic articles, research proposals, master's dissertations, and doctoral theses. The instructional register is comprised of learning material and material for government departments. The popular register is comprised of two religious texts. Finally, the reportage register is primarily comprised of newspaper articles from two community newspapers. All of these texts were written for a South African audience, although Kruger and Van Rooy (2017:32-33) point out that the audience for the academic texts may be more diffuse, even if all of the texts were sourced from locally produced publications.

⁸⁴ Kruger and Van Rooy (2017:27) point out that language editing in South Africa is "predominantly done by either native English speakers (users from the STL-strand), or by balanced English/Afrikaans bilinguals, while BSAE users are underrepresented in the industry". Consequently, the acceptance and legitimisation of innovative features in BSAE are constrained by linguistic gatekeepers who are exposed to BSAE, but are not users of the same variety (Kruger & Van Rooy, 2017:27).

Table 3.1: Summary of the details pertaining to the compilation of the BSAE editing corpus

Register	Number of texts	Word count (original texts)	Word count (edited texts)
Academic	21	80 682	81 629
Instructional	5	23 291	30 346
Popular	2	4 741	4 596
Reportage	94	28 548	25 979
Total	122	137 262	142 550

The texts in the academic, instructional, and popular registers were primarily sourced from South African editors who freelance for language services companies and private clients. The editors collected the relevant biographic information from the authors and obtained permission for their inclusion in the corpus. To protect the anonymity of both the authors and the editors, all references to the authors and editors were removed from the texts. However, the relevant biographical information was relayed by means of corpus mark-up. The texts in the reportage register were sourced directly from two community newspapers. The one newspaper is distributed in an urban area in Gauteng, whereas the other newspaper is distributed in a semi-urban area in Mpumalanga.

Initially, the corpus was skewed towards academic texts. Two factors contributed to this situation. The first is that the academic register contains several master's dissertations and doctoral theses. These texts tend to be substantially longer than the average research proposal and research article. In the interest of making the corpus more representative and balanced, smaller sections were extracted from the master's dissertations and doctoral theses. The second is that texts from the other registers are more difficult to source (Kruger & Van Rooy, 2017:32). Although quite a large number of texts was sourced in the reportage register, these texts were considerably shorter than the average academic text. So, even though the number of texts and the word count of these texts may not seem balanced across registers, attempts have been made to ensure that the size of the various registers is proportional to the frequency of occurrence of these texts in real life (cf. Section 3.2.1.1).

3.2.1.2.2. The White South African English (WSAE) editing corpus

The WSAE editing corpus is used as a basis for comparison. It was also compiled in the Constrained Language project. Consequently, the academic, instructional, popular and reportage registers are also represented in this corpus. The registers in this corpus are comprised of the same kinds of texts used in the BSAE editing corpus and were sourced in the same way as the BSAE texts. This is an important factor, especially if it is to be used as a control corpus. However, the reportage register did not contain any texts from WSAE authors. Therefore, texts from the reportage register of ICE-SA was included. Given the purpose of this corpus, the edited versions of these texts were disregarded. This means that the WSAE editing corpus is comprised of 163 texts, totalling 235 355 words. The details regarding the compilation of this corpus are summarised in Table 3.2:

Table 3.2: Summary of the details pertaining to the compilation of the WSAE editing corpus

Register	Number of texts	Word count (original texts)
Academic	11	46 227
Instructional	28	93 528
Popular	3	54 802
Reportage	121	40 798
Total	163	235 355

Initially, the corpus was skewed towards academic texts and instructional texts. The academic register once again contained master's dissertations and doctoral theses, while the instructional register contained several textbooks. Consequently, smaller sections were extracted from the master's dissertations, doctoral theses, and textbooks in this corpus to ensure a more balanced corpus.

3.2.1.2.3. The Tswana Learner English (TLE) corpus

The Tswana Learner English (TLE) corpus is a BSAE learner corpus consisting of 519 essays (totalling 201 244 words) written by L1 Setswana speakers from South Africa and Botswana. The texts were sourced within the framework of the International Corpus of

Learner English (Granger *et al.*, 2009) from four South African tertiary institutions where the speakers were enrolled as undergraduate students. These students were enrolled for courses in English which formed part of their humanities or social degree programmes. According to Van Rooy (2008b:273), the students' demographic data indicate that they had been educated in township schools situated in towns and small cities in the North-West and Northern Cape provinces of South Africa.

3.2.1.3 *Quantitative analysis of the corpora*

This section provides an overview of the tools used to analyse the corpora (Section 3.2.1.3.1) and the statistical measures used to interpret the data (Section 3.2.1.3.2).

3.2.1.3.1. Corpus analysis tools

Following the compilation of the corpus, the texts were marked up. Bowker and Pearson (2002:80-81) point out that it is not only expensive to compile a corpus, but also time-consuming. Therefore, linguists aim for reusability when compiling a corpus. In this regard, corpus mark-up plays a central role in the compilation of corpora.

First, the corpora were marked up with the biographic details of the authors and the editors. In this study, the biographic details of the authors had to be in place to ensure that the BSAE corpus is comprised of texts written by BSAE speakers and that the WSAE corpus is comprised of texts written by WSAE authors. Next, the corpus was marked up in terms of its primary data (i.e. footnotes, references, quotes, and spelling errors) to facilitate the corpus searches. Texts quoted from other sources (e.g. interviews and books) were also removed at this stage.

After the mark-up was completed, the original texts in the BSAE editing corpus were manually aligned in Excel with their corresponding edited counterparts at sentence level. This resulted in a parallel corpus which was uploaded to the online corpus-querying package, Sketch Engine (Lexical Computing, 2018). When a corpus is uploaded to Sketch Engine, it is automatically part-of-speech tagged and lemmatised using the English Penn Treebank part-of-speech tag set which contains modifications made by Sketch Engine

(cf. Infogistics, 2001). The parallel concordance tool in Sketch Engine allowed the researcher to draw parallel concordances to not only investigate the features of interest but also to see whether the editors made any changes to the way in which the feature was used by the BSAE authors (cf. Figure 3.2 on p. 149).

The WSAE editing corpus and the TLE corpus were part-of-speech tagged using the Constituent Likelihood Automatic Word-tagging System (CLAWS4) with the C7-tagset in order to isolate the lexical verbs (UCREL, 2018). Isolating the lexical verbs is an important step in the analysis of the corpora as the potential of the verb determines the clause elements that are likely to follow (Biber *et al.*, 1999:141). The part-of-speech tagged WSAE editing corpus and the part-of-speech tagged TLE corpus were then analysed using the concordance tool in Wordsmith Tools 7 developed by Scott (2018) (cf. Figure 3.3 on p. 150).

The concordances created in both Sketch Engine and WordSmith Tools were downloaded and converted to xlsx-format. This allowed the researcher to analyse the relevant linguistic features in Microsoft Excel.

Once the analyses were completed, the results were analysed using the statistical software, R (R Foundation, 2018). Section 3.2.1.3.2 provides an overview of the statistical measures employed to interpret the results.

PARALLEL CONCORDANCE BSAE editing corpus FULL (original), English 1

simple help, help 153

BSAE editing corpus FULL (edited), English 2

h.file_id=A-006-O	The leaders/ managers are the most qualified people to help them with their development. They are familiar with performance levels and skills gap.	The leaders or managers are the people most qualified to help them with their development since they are familiar with performance levels and the skills gap.
h.file_id=A-006-O	This can help the employees develop and grow (Blake, 2006).	This can help employees to develop and grow (Blake, 2006).
h.file_id=A-006-O	* Are energised, committed and work hard to help the company succeed.	* They are energised, committed and work hard to help the company succeed.
h.file_id=A-006-O	help them to see that their roles have meaning and add value to the organisation;	help them to see that their roles have meaning and add value to the organisation;
h.file_id=A-006-O	help them to see the importance of their jobs;	help them to see the importance of their jobs;
h.file_id=A-006-O	help them to feel proud of their achievements;	help them to feel proud of their achievements;
h.file_id=A-006-O	This will also help them grow, all employees want to feel like they are benefiting from their work.	This will also help them grow; all employees want to feel that they are benefiting from their work.
h.file_id=A-006-O	In addition, where the employees feel that they are being cared for and supported they feel an obligation to put more effort in helping the organisation reach its goal.	In addition, where employees feel that they are being cared for and supported, they feel an obligation to put more effort into helping the organisation to reach its goal.
h.file_id=A-006-O	This helps employers keep the organisational culture churning (Shahnawaz & Jafri, 2009).	This helps employers keep the organisational culture churning (Shahnawaz & Jafri, 2009).
h.file_id=A-010-O	The History curriculum in general is meant to promote acquisition of knowledge and understanding of human activity in the past, linking it with the present, helping learners to understand issues of cause and consequence, continuity and change and the general evolution of society over time to where it is today. (Footreference 7)	The history curriculum in general is intended to promote the acquisition of knowledge and the understanding of human activity in the past and to link it with the present so as to help learners to understand causes and consequences, continuity and change, and the general evolution of society over time to become what it is today. (Footreference 7)

Back to the original interface

Figure 3.2: A screenshot of a parallel concordance in Sketch Engine

N	Concordance	Tag	Word #	Sent #	SentPara #	Para #	H...	H...	Sect	Sect	File	Date	%
1	is_VBZ making_VVG in_II achieving_VVG results_NN2 and_CC help_VV0 collect_VVI evidence_NN1 about_II what_DDQ work_VV0	..*	1 276	33	17	0	1		0	1	tagged_I-071-O.	2017-Feb-06 00:	24%
2	and_CC overwhelmed_VVN by_II the_AT operation_NN1 and_CC help_VV0 she_PP1S1 received_VVD from_II the_AT doctor_NN1	..*	268	8	14	0	267		0	267	tagged_R-092-O.	2017-Feb-06 00:	74%
3	encompassing_VVG 90_MC plus_II variables_NN2 that_CST help_VV0 researchers_NN2 and_CC practitioners_NN2 to_II	..*	479	15	34	0	478		0	478	tagged_A-121-O.	2017-Feb-06 00:	8%
4	meet_VV0 _get_VV0 to_TO know_VVI _learn_VV0 from_II _help_VV0 _support_VV0 _and_CC are_VBR kept_VVN	..*	3 644	92	28	0	3		0	3	tagged_A-047-O.	2017-Feb-06 00:	66%
5	_the_AT report_NN1 indicates_VVZ that_CST communities_NN2 help_VV0 mobilise_VVI sp=mobilize_NULL resources_NN2	..*	969	33	16	0	968		0	968	tagged_A-117-O.	2017-Feb-06 00:	18%
6	see_VVI the_AT importance_NN1 of_IO their_APPGE jobs_NN2 ;_help_VV0 them_PP1O2 to_TO feel_VVI proud_JJ of_IO their_APPGE	..*	2 516	85	81	0	2		0	2	tagged_A-006-O.	2017-Feb-09 00:	49%
7	development_NN1 ;_listen_VV0 to_II their_APPGE views_NN2 ;_help_VV0 them_PP1O2 to_TO see_VVI the_AT importance_NN1 of_IO	..*	2 506	85	71	0	2		0	2	tagged_A-006-O.	2017-Feb-09 00:	49%
8	possible_JJ the_AT community_NN1 and_CC teachers_NN2 help_VV0 to_TO build_VVI the_AT education_NN1 system_NN1	..*	86	3	20	0	85		0	85	tagged_R-106-O.	2017-Feb-06 00:	27%
9	them_PP1O2 tick_VVI) ;_care_VV0 about_II them_PP1O2 ;_help_VV0 them_PP1O2 to_TO see_VVI that_CST their_APPGE	..*	2 477	85	42	0	2		0	2	tagged_A-006-O.	2017-Feb-09 00:	48%
10	keep_VVI them_PP1O2 interested_JJ _This_DD1 will_VM also_RR help_VVI them_PP1O2 grow_VVI _all_DB employees_NN2	..*	3 388	135	7	0	3		0	3	tagged_A-006-O.	2017-Feb-09 00:	66%
11	are_VBR not_XX proud_JJ of_IO _punct=NULL it_PP11 also_RR help_VV0 us_PP1O2 not_XX reconvinced_VVN _"/quote_NULL	..*	286	11	62	0	285		0	285	tagged_R-078-O.	2017-Feb-06 00:	92%
12	_FU Proportion_NN1 of_IO ODA_NP1 provided_VVD to_TO help_VVI build_VVI trade_NN1 capacity_NN1 *_FU Debt_NN1	..*	2 165	37	152	0	2		0	2	tagged_I-069-O.	2017-Feb-06 00:	35%
13	traditional_JJ religions_NN2 and_CC sought_VVD to_TO help_VVI Christians_NN2 to_TO understand_VVI Congolese_JJ	..*	4 908	152	17	0	4		0	4	tagged_A-120-O.	2017-Feb-06 00:	92%
14	environmentally-friendly_JJ method_NN1 of_IO farming_NN1 to_TO help_VVI combat_NN1 hunger_NN1 among_II communities_NN2 in_II	..*	203	6	49	0	202		0	202	tagged_R-051-O.	2017-Feb-06 00:	54%
15	of_IO Code_NN1 of_IO Ethics_NN is_VBZ practically_RR to_TO help_VVI employees_NN2 do_VDI what_DDQ is_VBZ expected_VVN	..*	1 241	41	14	0	1		0	1	tagged_A-015-O.	2017-Feb-06 00:	26%
16	in_II the_AT class_NN1 _not_XX knowing_VVG how_RRQ to_TO help_VVI him/her_PP1O1 _We_PP1S2 have_VH0 learners_NN2	..*	4 236	135	41	0	4		0	4	tagged_A-090-O.	2017-Feb-06 00:	80%
17	what_DDQ work_VV0 and_CC what_DDQ did_VDD not_XX _to_TO help_VVI improve_VVI planning_NN1 and_CC implementation_NN1	..*	1 288	33	29	0	1		0	1	tagged_I-071-O.	2017-Feb-06 00:	24%
18	feeding_VVG projects_NN2 in_II South_NP1 Africa_NP1 to_TO help_VVI improve_VVI the_AT nutritional_JJ status_NN1 of_IO the_AT	..*	956	28	37	0	955		0	955	tagged_A-048-O.	2017-Feb-06 00:	36%
19	_Mzayifani_NP1 Nosenga_NP1 _we_PP1S2 vow_VV0 to_TO help_VVI _knowing_VVG that_CST if_CS Nosenga_NP1 was_VBDZ	..*	146	4	45	0	145		0	145	tagged_R-092-O.	2017-Feb-06 00:	40%
20	_can_VM you_PPY really_RR expect_VVI us_PP1O2 to_TO help_VVI our_APPGE learners_NN2 with_IW grade_NN1 9_MC	..*	5 024	167	49	0	5		0	5	tagged_A-090-O.	2017-Feb-06 00:	94%
21	club_NN1 made_VVD amends_NN2 and_CC vow_VV0 to_TO help_VVI the_AT hopeless_JJ woman_NN1 _Speaking_VVG to_II	..*	99	3	12	0	98		0	98	tagged_R-092-O.	2017-Feb-06 00:	27%
22	foundation_NN1 years_NN2 back_RP with_IW hope_NN1 to_TO help_VVI the_AT community_NN1 _and_CC today_RT all_DB	..*	77	3	22	0	76		0	76	tagged_R-099-O.	2017-Feb-06 00:	23%
23	energised_VVN _committed_JJ and_CC work_VV0 hard_RR to_TO help_VVI the_AT company_NN1 succeed_VV0 *_FU Utilise_VV0	..*	1 845	65	30	0	1		0	1	tagged_A-006-O.	2017-Feb-09 00:	36%
24	in_II the_AT gift_NN1 of_IO the_AT Holy_JJ Spirit_NN1 and_CC to_TO help_VVI them_PP1O2 to_TO be_VBI transformed_VVN into_II	..*	3 751	143	92	0	3		0	3	tagged_P-007-O.	2017-Feb-06 00:	73%
25	raise_VVI the_AT income_NN1 of_IO poor_JJ beneficiaries_NN2 to_TO help_VVI them_PP1O2 to_TO smooth_VVI their_APPGE	..*	2 622	88	16	0	2		0	2	tagged_I-071-O.	2017-Feb-06 00:	49%
26	traders_NN2 _and_CC find_VV0 sp=fine_NULL away_RL to_TO help_VVI traders_NN2 in_II Sedibeng_NP1 _N1 reac	..*	294	8	39	0	293		0	293	tagged_R-100-O.	2017-Feb-06 00:	98%
27	_The_AT discussions_NN2 are_VBR designed_VVN to_TO help_VVI you_PPY think_VVI about_II how_RRQ the_AT	..*	381	7	9	0	380		0	380	tagged_I-069-O.	2017-Feb-06 00:	6%
28	page_NN1 _The_AT activities_NN2 are_VBR designed_VVN to_TO help_VVI you_PPY think_VVI about_II how_RRQ the_AT	..*	276	3	9	0	275		0	275	tagged_I-069-O.	2017-Feb-06 00:	4%
29	page_NN1 _The_AT activities_NN2 are_VBR designed_VVN to_TO help_VVI you_PPY think_VVI about_II how_RRQ the_AT	..*	178	3	9	0	177		0	177	tagged_I-068-O.	2017-Feb-06 00:	3%
30	_The_AT discussions_NN2 are_VBR designed_VVN to_TO help_VVI you_PPY think_VVI about_II how_RRQ the_AT	..*	283	7	9	0	282		0	282	tagged_I-068-O.	2017-Feb-06 00:	5%
31	that_CST if_CS Nosenga_NP1 was_VBDZ alive_JJ would_VM help_VVI _According_IJ21 to_IJ22 the_AT doctor_NN1 's_GE	..*	157	4	66	0	156		0	156	tagged_R-092-O.	2017-Feb-06 00:	43%
32	better_JJR and_CC that_CST we_PP1S2 as_II learn_VV0 but_CCB I_PP1S1 am_VBM also_RR sick_JJ _so_CS	..*	5 110	170	36	0	5		0	5	tagged_A-090-O.	2017-Feb-06 00:	95%
33	process_NN1 _a_AT1 theory_NN1 of_IO measurement_NN1 can_VM help_VVI clarify_VVI whether_CSW a_AT1 particular_JJ practice_NN1	..*	272	7	49	0	271		0	271	tagged_A-022-O.	2017-Feb-06 00:	5%
34	are_VBR a_AT1 lonely_JJ road_NN1 _no-one_PN1 can_VM help_VVI him_PP1O1 at_II home_NN1 _I_PP1S1 know_VV0	..*	5 072	169	17	0	5		0	5	tagged_A-090-O.	2017-Feb-06 00:	95%
35	also_RR identifies_VVZ key_JJ issues_NN2 that_CST will_VM help_VVI in_II developing_JJ recommendations_NN2 and_CC the_AT	..*	4 654	145	80	0	4		0	4	tagged_A-015-O.	2017-Feb-06 00:	100%
36	or_CC is_VBZ supposed_JJ to_TO do_VDI) _This_DD1 will_VM help_VVI in_II affirming_VVG the_AT value_NN1 and_CC worth_NN1	..*	275	7	6	0	274		0	274	tagged_A-047-O.	2017-Feb-06 00:	5%
37	_Police_NN2 hope_VV0 the_AT community_NN1 could_VM help_VVI in_II finding_VVG their_APPGE grandfather_NN1 _T1 miss	..*	318	10	24	0	317		0	317	tagged_R-102-O.	2017-Feb-06 00:	98%
38	during_II the_AT year_NN1 xD1_FO which_DDQ would_VM help_VVI in_II promoting_VVG transparency_NN1 ;_and_CC can_VM	..*	4 541	121	29	0	4		0	4	tagged_A-119-O.	2017-Feb-06 00:	95%
39	to_TO drink_VVI them_PP1O2 _the_AT teacher_NN1 can_VM help_VVI in_II this_DD1 regard_NN1 to_TO remind_VVI them_PP1O2	..*	2 232	76	47	0	2		0	2	tagged_A-090-O.	2017-Feb-06 00:	42%
40	and_CC see_VVI how_RRQ best_VV0 we_PP1S2 could_VM help_VVI _"/quote_NULL he_PP1S1 said_VVD _will_VM	..*	373	13	30	0	372		0	372	tagged_R-057-O.	2017-Feb-06 00:	98%
41	the_AT highly-sophisticated_JJ technology_NN1 that_CST will_VM help_VVI researchers_NN2 to_TO identify_VVI new_JJ topics_NN2	..*	204	7	12	0	203		0	203	tagged_R-025-O.	2017-Feb-06 00:	64%
42	training_NN1 for_IF the_AT EPWPs_NP2 that_CST will_VM help_VVI the_AT success_NN1 of_IO the_AT program_NN1 _"/	..*	343	10	54	0	342		0	342	tagged_R-090-O.	2017-Feb-06 00:	93%
43	to_II narrowly_RR defined_VVN political_JJ ideologies_NN2 can_VM help_VVI them_PP1O2 to_TO be_VBI critical_JJ of_IO their_APPGE	..*	3 107	81	28	0	3		0	3	tagged_A-010-O.	2017-Feb-06 00:	96%
44	of_IO water_NN1 _and_CC hot_JJ tea_NN1 _This_DD1 will_VM help_VVI to_TO clear_VVI mucus_NN1 _Avoid_VV0 drinking_VVG	..*	317	7	6	0	316		0	316	tagged_R-125-O.	2017-Feb-06 00:	67%
45	of_IO public_JJ works_NN programmes_NN2 should_VM help_VVI to_TO determine_VVI the_AT impact_NN1 of_IO a_AT1	..*	2 595	87	14	0	2		0	2	tagged_I-071-O.	2017-Feb-06 00:	48%
46	point_NN1 because_CS it_PP11 sp=becauseit_NULL can_VM help_VVI venture_NN1 managers_NN2 understand_VV0 and_CC	..*	747	21	20	0	746		0	746	tagged_A-034-O.	2017-Feb-06 00:	10%
47	your_APPGE work_NN1 environment_NN1 _These_DD2 will_VM help_VVI you_PPY to_TO learn_VVI from_II your_APPGE peers_NN2	..*	303	8	6	0	302		0	302	tagged_I-068-O.	2017-Feb-06 00:	5%

Figure 3.3: A screenshot of a concordance in WordSmith 7

3.2.1.3.2. Statistical measures

In quantitative corpus-linguistic methods, frequency counts are often employed to answer research questions (Gries, 2012:479). As such, corpus linguists are essentially concerned with whether a linguistic feature occurs (i.e. $n>0$) or not (i.e. $n=0$). If a feature does occur, we may ask how often it occurs. These quantitative results allow us “to identify what merits a qualitative discussion in the first place” (Gries, 2013b:4). It is important to be cognisant of the fact that linguists are not interested in the frequencies of a linguistic feature *per se*, but rather in what the frequencies and the distributional characteristics of the feature reveal about its structure, semantics/meaning, and pragmatics/function. Biber *et al.* (1998:5) make a similar point when they say that “[t]he goal of corpus-based investigations is not simply to report quantitative findings, but to explore the importance of these findings for learning about the patterns of language use”. Insights into these aspects of language are gained by using statistical tools which enable linguists to not only evaluate these frequencies but also to gain insight into the direction of certain effects and whether these effects are significantly different from chance (Gries, 2006c:198).

Distributional analysis

The data in this study are not distributed normally. Since normal distribution is an assumption of parametric tests, parametric tests cannot be used to compute the probabilities for ranges. Therefore, non-parametric tests are harnessed to describe the distribution of features.

In order to describe the distribution of certain features, descriptive statistics are reported with a boxplot illustrating median values and the interquartile range of a set of values for a specific group. The median values measure the value that is most typical of the set of values (i.e. central tendency) for the group, while the interquartile range measures the dispersion of the set of values in the group (Levshina, 2015:43-50). To determine whether the differences in the measures of central tendency between the groups are significant, a Kruskal-Wallis rank test is performed. A Kruskal-Wallis rank test ranks the values of the entire data set before comparing the ranks across the different groups (Levshina,

2015:178-179). This test is used to determine whether the differences between the groups under investigation are statistically significant.

The Chi-squared (χ^2) test for independence

The χ^2 -test helps the researcher determine whether the differences that are observed in the distribution of a set of categories are statistically significant, and is based on a comparison of the observed and expected frequencies (Levshina, 2015:212). There are only two assumptions for the χ^2 -test. The first is that the sample is randomly selected and that the observations are independent (Conover, 1999:204). The second is that every observation can be classified into exactly one category according to the criterion (Conover, 1999:205). When one or more values in the contingency table are smaller than five (5), though, the χ^2 -test is not considered to be robust (Levshina, 2015:29). In those instances, the Fisher exact test is used.

The Fisher exact test

Similar to the χ^2 -test, the Fisher exact test helps the researcher establish whether the differences that are observed in the distribution of a set of categories are statistically significant. However, the Fisher exact test is preferred to the χ^2 -test when one or more values in the contingency table are smaller than five. Similar to the χ^2 -test, the Fisher exact test is also based on the assumption that the sample is randomly selected and that each observation is independent (Conover, 1999:188).

Collostructional analysis

Methods pertaining to **collostructional analysis** are one of the measures employed in this study. Collostructional analysis is a method of analysis used to investigate the co-occurrence of words and is based on inferential statistics (Stefanowitsch & Gries, 2003). Stefanowitsch and Gries (2003:210-211) point out that collostructional analysis methods have two main applications. The first is to improve strategies of grammatical description. The second is to provide a foundation for linguistic theorising and model-building.

Stefanowitsch and Gries (2003) developed collocation analysis methods to overcome the fact that co-occurrence frequencies are not an ideal measure of association strength – not only for empirical reasons but also for theoretical reasons. They explain that these shortcomings can be attributed to the fact that “raw frequency counts do not take into account the overall frequencies of a given word in the corpus, and therefore the most frequent collocates of any given word are typically function words, which are often of little use” (Stefanowitsch & Gries, 2003:216). They go on to say that many of the measures that have been proposed in the context of collocation analysis are problematic in at least one of two ways. First, many of the proposed statistical measures that have been put forward involve distributional assumptions that are not justified. These include Berry-Rogghe’s (1974) *z*-score and Church, Gale, Hanks & Hindle’s (1991) *t*-score. Second, some of the statistical measures tend to overestimate association strengths and underestimate the probability of errors. These tests include non-parametric improvements such as the χ^2 -statistic or Dunning’s (1993) log-likelihood coefficient which are dependent on the chi-square distribution for significance testing. Stefanowitsch and Gries (2003:217) explain that these measures are problematic since “the vast majority of collexemes occurring within any given construction have a very low frequency in that construction”.⁸⁵

According to Stefanowitsch and Gries (2003:217), there is only one statistical measure that is not subject to the aforementioned shortcomings. This measure is known as the **Fisher exact test** (cf. Pedersen, 1996). The Fisher exact test is a statistical test used to determine if there are non-random associations between two categorical variables. The

⁸⁵ This phenomenon is represented in Zipf’s law. The law is attributed to George Kingsley Zipf (1935:v-vi) who theorised that in any large body of language (e.g. a corpus), the frequency of each word is close to inversely proportional to its rank in the frequency table. This means that the most frequent word in a frequency list will occur twice as often as the second most frequent word; the second most frequent word will occur three times as often as the third most frequent word; and so on. This theory is expressed by the following formula:

$$P_n \propto 1/n^a$$

test can be performed on a two-by-two table that represents the single and joint frequencies of the **collexeme**⁸⁶ (L) and the **collostruct**⁸⁷ (C) found in the corpus.

At this stage, it is important to note that three variants of collostructional analysis can be distinguished. They are (i) simple collexeme analysis; (ii) distinctive collexeme analysis; and (iii) covarying collexeme analysis. The underlying method used in each of the aforementioned variants are the same, but the nature of the linguist's research question will determine the kind of analysis to be used (Stefanowitsch, 2013:291). For example, should the linguist want to determine which verbs are strongly attracted to or repelled by the ditransitive construction, they would execute a **simple collexeme analysis** (Stefanowitsch & Gries, 2003). To determine the association strength of all the words that occur in a specific slot of two semantically or functionally similar constructions across the two constructions, the linguist would execute a **distinctive collexeme analysis** (Stefanowitsch, 2013:295-296). If a linguist wants to identify pairs of words that occur with each other in two slots of the same construction, they would execute a **covarying collexeme analysis** (Stefanowitsch, 2013:298).

In this study, a distinctive collexeme analysis is used to investigate two semantically similar allostructions, the double object allostruction and the prepositional dative allostruction, and the lexemes that occur in them. The analysis determines which lexemes favours which allostruction (cf. Chapter 4).⁸⁸

Stefanowitsch (2013:296) explains what frequencies are needed to execute a distinctive collexeme analysis:

⁸⁶ The concept *collexeme* refers to lexemes that are attracted to particular constructions. As such, a collexeme analysis investigates the strength of association between a construction and its constituent lexemes (Stefanowitsch & Gries, 2003:215).

⁸⁷ The concept *collostruct* refers to constructions that are associated with particular lexemes. As such, a collostructional analysis investigates the strength of association between a lexeme and its constituent constructions (Stefanowitsch & Gries, 2003:215).

⁸⁸ The distinctive collexeme analysis in this study was executed using Gries's (2007) Coll.analysis 3.2a script in R.

[I]n order to determine the association of a particular lexical construction l_i belonging to a word class L to two grammatical constructions c_1 and c_2 , both belonging to a class of construction S , the following frequencies must be determined: the frequency l_i in c_1 , the frequency of l_i in c_2 , the frequency of other members of L in c_1 , and the frequency of other members of L in c_2 .

These frequencies are then entered into a two-by-two contingency table such as the one illustrated in Table 3.3:

Table 3.3: *The frequency information required to execute a distinctive collexeme analysis (Stefanowitsch, 2013:296)*

	Word l_i of Class L	Other words of Class L	Total
construction c_1 of Class C	Frequency of $L(l_i)$ in $C(c_1)$	Frequency of $L(\neg l_i)$ in $C(c_1)$	Total frequency of $C(c_1)$
construction c_2 of Class C	Frequency of $L(l_i)$ in $C(c_2)$	Frequency of $L(\neg l_i)$ in $C(c_2)$	Total frequency of $C(c_2)$
Total	Total frequency of $L(l_i)$ in $C(c_1, c_2)$	Total frequency of $L(\neg l_i)$ in $C(c_1, c_2)$	Total frequency of $C(c_1, c_2)$

Inferential decision trees

As mentioned earlier in Chapter 2, different linguistic forms sometimes have the same function. The ditransitive constructeme, for example, can be realised by the double-object allostruction or the prepositional dative allostruction. Variation analyses allow linguists to determine how these allostructions vary and whether there are factors that drive speakers to choose one variant over another.

In the literature, for example, it has been found that certain variables (some related to form and others related to function) determine whether speakers choose to employ the double-object allostruction or the prepositional dative allostruction. In terms of form, for example, linguists have found that if the recipient in the construction is realised by a pronoun, it is more likely that the speaker will choose to use the double-object construction (cf. Chapter 4). In terms of function, on the other hand, linguists have found that when verbs denote the transfer of abstract things in situations where the agent and

the recipient are not required to be in close proximity to one another, speakers are more likely to use the prepositional dative construction.

In this regard, Labov (1969) observed that choices such as these vary in systematic ways, and therefore should be modelled quantitatively. Sankoff (1988:151) makes a similar point:

Now, whenever a choice can be perceived as having been made in the course of linguistic performance, and where this choice may have been influenced by factors such as the nature of the grammatical context, discursive function of the utterance, topic, style, interactional context or personal or sociodemographic characteristics of the speaker or other participants, then it is difficult to avoid invoking notions and methods of statistical inference, if only as a heuristic tool to attempt to grasp the interaction of the various components in a complex situation.

In this study, conditioning variables are identified that have been determined to predict language users' choice of construction. The factor that is of utmost importance in this study is the factor **VARIETY**. The statistical analyses will highlight all the instances where differences in the use of constructions between the BSAE and WSAE can be attributed to the fact that they are two different varieties of English,

The complex interaction between these conditioning variables is visualised with conditional inference trees (Tagliamonte & Baayen, 2012:135). Bernaisch *et al.* (2014:14) and Strobl, Malley and Tutz (2009b:4) explain that decision trees are basically a recursive partitioning approach towards classification and regression.

In this study, the conditional inference trees are drawn by means of the `ctree` algorithm in the `partykit` R package (Hothorn, Hornik & Zeileis, 2015). The `ctree` algorithm recursively performs univariate splits of the dependent variable based on a set of covariate independent variables. These univariate splits predict the best-known outcomes of the dependent variable.

However, Bernaisch *et al.* (2014:14) and Tagliamonte and Baayen (2012:163) point out that decision trees sometimes fail to identify predictive interactions. Therefore, random forests are also fitted to the data in this study to ensure that the analyses account for all the predictive variables.

Random forests

Random forests are also based on classification trees, but they provide additional layers of randomness to the analysis (Bernaisch *et al.*, 2014:14). They provide information about the importance of multiple predictors, both factorial and continuous (Tagliamonte & Baayen, 2012:135). When random forests are generated, many different trees are built on random bootstrapped samples of the data. Each split in a tree chooses from a randomly restricted subset of available predictors, and the overall result amalgamates the multitude of trees that has been generated (Bernaisch *et al.*, 2014:14-15; Strobl *et al.*, 2009b:11).

In this study, the random forests are drawn by means of the `cforest` algorithm in the `party` package in R (Hothorn *et al.*, 2015). The `cforest` algorithm recursively performs univariate splits of the dependent variable based on a set of covariate independent variables. These univariate splits predict the best-known outcomes of the dependent variable.

3.3 Conclusion

This study is a corpus-based study in which corpora are leveraged to describe how language is used by language users in “real life”. The results obtained from the corpus linguistic investigations are analysed quantitatively and qualitatively. This approach is particularly appropriate in light of the fact that this study approaches the description of specific linguistic features from a usage-based perspective.

Chapter 4

The ditransitive construction

4.1 Contextualisation

Language change in verb complementation profiles is investigated increasingly not only in the Inner Circle varieties such as British English, American English, and New Zealand English (cf. Bresnan & Ford, 2010; Bresnan & Hay, 2008; Egan, 2008; Mair, 1990, 1995, 2002; Wolk, Bresnan, Rosenbach & Szmrecsanyi, 2013), but also in the Outer Circle varieties such as Indian English and Hong Kong English (Bernaisch *et al.*, 2014; Mukherjee, 2010; Mukherjee & Gries, 2009; Mukherjee & Hoffmann, 2006; Olavarría de Ersson & Shaw, 2003). More often than not, linguists tend to focus on the changes that take place on the level of the lexis. But Olavarría de Ersson and Shaw (2003:138) point out that “[v]erb complementation is an all-pervading structural feature of language and thus likely to be more significant in giving a variety character than, for example, lexis”.

One of the features that have garnered a lot of attention in the new varieties of English is the ditransitive construction (Bernaisch *et al.*, 2014; Mukherjee, 2005; Olavarría de Ersson & Shaw, 2003). In their investigation of three varieties of Asian Englishes, Mukherjee and Gries (2009:47-48) find that these three new varieties can be positioned on a cline of collostructional nativisation. They base this statement on the fact that the preferences of a verb “may change in the course of the evolution of a new variety”. They point out that the ditransitive verb, in particular, is correlated with the variety’s evolutionary stage in Schneider’s (2003, 2007) Dynamic Model.

Following Mukherjee and Gries’s (2009:47-48) investigation, Bernaisch *et al.* (2014) investigate the ditransitive construction in terms of a number of processing-related factors that are said to drive the language user’s choice in ditransitive allostruction (in both L1 and L2 varieties of English). They find that the systematic lexico-constructional differences that are observed in six varieties of Asian Englishes develop independently of the processing-related factors that have been determined to influence the allostructional choices made by language users. This means that the processing-related factors known to drive the allostructional choices made by L1 speakers also drive the allostructional choices made by L2 speakers. Bernaisch *et al.*’s (2014) findings support Hoffmann’s (2011:278) finding that input and processing factors seem to influence the construction networks of both British and Kenyan English.

However, Bernaisch *et al.* (2014) are careful to emphasise that these findings do not mean that cognitive processing efforts are irrelevant to the process of structural nativisation in postcolonial Englishes. In this regard, they point to research conducted by Mukherjee and Hoffmann (2006) and Koch and Bernaisch (2013) in which sets of “new ditransitives” are identified in South Asian Englishes that are not present in British English, their historical input variety. It is argued that these novel verb-construction associations have been derived by means of “semantico-structural analogy” (Mukherjee, 2007:175). This means that the speakers of the South Asian English varieties are competent L2 English speakers who are able to “introduce new forms and structures into the English language on grounds of semantic and formal templates that already exist in the English language system” (Mukherjee, 2007:176).

These studies provide evidence that the verb-complementational profiles are realised differently in L2 varieties of English than in L1 varieties of English. This chapter sets out to analyse the verb-complementational profiles of the ditransitive verbs in BSAE. In Section 4.2, a theoretical overview is provided of the ditransitive construction. The factors that are said to drive the dative alternation in both standard Englishes and new varieties of Englishes are discussed in Section 4.3, while some additional factors that influence the structural and semantic realisation of the ditransitive are addressed in Section 4.4. The results of the analysis are presented in Section 4.5. Concluding remarks can be found in Section 4.6.

4.2 A theoretical overview of the ditransitive construction

Scholars agree that the ditransitive construction denotes some kind of transfer and is prototypically realised by the double-object structure⁸⁹, exemplified in (16) below (cf. Biber *et al.*, 1999:150; Mair, 1990:94; Mukherjee, 2005:65; Quirk *et al.*, 1985:1208):

- (1) [The government]_{S:NP} ... [must **give**]_{V:VP} [learners]_{IO:NP} [healthy food]_{DO:NP...} (BSAE)

The construction is comprised of three semantic roles: the **AGENT** (i.e. the acting entity) who transfers a **PATIENT** (i.e. the transferred entity) to a **RECIPIENT** (i.e. the affected entity). The semantic component that expresses such a situation type of transfer – i.e. [X transfers Y to Z] – is not limited to the syntactic structure [S:NP V:VP DO:NP IO:NP], though.⁹⁰ In fact, it may also be linked to the syntactic structure [S:NP V:VP DO:NP IO:PP] as exemplified in (1a) below:

⁸⁹ The ditransitive construction is the only construction in which two noun phrases co-occur without there being a copular relationship between the entities they realise (i.e. the receiver and the patient).

⁹⁰ There are several instances where alternations arise due to “the ubiquitous fact that words – foremost verbs – turn up in multiple structural patterns” (Cappelle, 2006:1). Examples include the dative alternation (as illustrated here) and the particle placement alternation (e.g. Cappelle, 2006; Gries, 2003a).

(1a) [The government]_{S:NP} [must give]_{V:VP} [healthy food]_{DO:NP} [to learners]_{IO:PP}.

Instances such as these give rise to the debate in which constructionists argue whether constructions such as [[X transfers Y to Z] ↔ [S:NP V:VP IO:NP DO:NP]] and [[X transfers Y to Z] ↔ [S:NP V:VP DO:NP IO:PP]] should be regarded as two independent constructions or as two instances of the same schematic construction.

Goldberg (1995:89-95) posits that there is a semantic difference between (1) and (1a), even though these two structures may seem synonymous. She argues that the *to*-variant is a metaphorical extension of the caused-motion construction by means of semantic extension via metaphor.⁹¹ In this metaphorical extension, possession is to be understood “as the ‘possessed’ being located next to the ‘possessor’, transferring an entity to a recipient as causing the entity to move to that recipient, and transferring ownership away from a possessor as taking that entity away from the possessor”. She goes on to argue that the metaphor is “motivated by the fact that giving prototypically correlates with movement from a possessor to a recipient” (Goldberg, 1995:89). In the literature, this metaphorical extension is often referred to as the **MEANING-TO-STRUCTURE HYPOTHESIS**.

Goldberg’s (1995:89-95) argument seems to be motivated by the use of the locative *to*-preposition which suggests that the patient is transferred along an abstract path from the agent to the receiver who is construed as the goal of a path. In constructional terms, it means that these two constructions are actually not synonymous and that the double-object construction should be represented as [[X causes Y to have Z] ↔ [S:NP V:VP IO:NP DO:NP]] and the *to*-dative construction should be represented as [[X causes Z to go to Y] ↔ [S:NP V:VP DO:NP IO:PP]] (cf. Krifka, 1999).⁹² Since these two

⁹¹ Goldberg (1995:91) points out that although the ditransitive construction is not related to the caused-motion construction on a syntactic level, they are synonymous on a semantic level.

⁹² In the literature, dative alternation is most often attributed to semantics. Some scholars, e.g. Oehrle (1976) and Green (1974), argue for lexical semantic differences. Lexical semantic differences also underpin the work of scholars such as Gropen, Pinker, Hollander, Goldberg and Wilson (1989) and Pinker (1989:110-111) on language learnability and the dative alternation. Other scholars, in turn, argue for lexical equivalence and constructional semantic differences (Goldberg, 1995; Gries & Stefanowitsch, 2004). Be that as it may, semantics plays a central role in accounting for the dative alternation.

syntactic realisations of the ditransitive are linked to different meanings, it is argued that these two constructions cannot be regarded as alternations.

Rappaport Hovan and Levin (2008), however, point out that evidence to support Goldberg's (1995) hypothesis that the *to*-dative actually depicts an abstract path component is scarce. They argue that the *to*-phrase in caused-motion constructions take distinct spatial goals, whereas the dative *to*-phrase in ditransitive constructions take distinct possessional goals. Rappaport Hovan and Levin (2008:137-142) support their argument by illustrating that the *to*-phrase in ditransitive constructions – such as the ditransitive construction depicted in (1a) above – cannot be questioned by the locative *wh*-word *where* – as illustrated in (1b) below:

(1b) ***Where** must the government give healthy food?

The *to*-phrase in ditransitive constructions can, however, be questioned by *wh*-word *who(m)* (Rappaport Hovan & Levin, 2008:142-144):

(1c) **To whom** must the government give healthy food?

Furthermore, the preposition *to* cannot be substituted by other locative prepositions:

(1d) *The government must give healthy food **at learners**.

Nor can the dative *to*-phrase be combined with other locative prepositional phrases to denote a complex path:

(1e) *The government must give healthy food **off the shelf to learners**.

Similarly, Bresnan, Cueni, Nikitina and Baayen (2007:5-6) quote a number of instances they observe in which language users do not adhere to the meaning-to-structure hypothesis:

- (2) Karen spoke with Gretchen about the procedure for registering a complaint, and [**hand-carried**]_{V:VP} [*her*]_{IO:NP} [*a form*]_{DO:NP}, but Gretchen never completed it.
- (3) Therefore, when he got to purgatory, Buddha [**lowered**]_{V:VP} [*him*]_{IO:NP} [*the silver thread of a spider*]_{DO:NP} as his last chance for salvation.
- (4) This story is designed to [**give**]_{V:VP} [*the creeps*]_{DO:NP} [*to people who hate spiders*]_{IO:PP}, but is not true.
- (5) Stories like these [**must give**]_{V:VP} [*the creeps*]_{DO:NP} [*to people whose idea of heaven is a world without religion*]_{IO:PP}.

The examples in (2) and (3) above (as quoted by Bresnan *et al.*, 2007:6) should not be grammatically possible, since the use of the verbs *hand-carried* and *lowered* imply that the patients – *a form* and *the silver thread of a spider* – is moved from one place to another. Despite the distinct allative meaning (i.e. the meaning of moving from one place to another) apparent in these examples, the *to*-dative is not employed.

By the same token, the examples in (4) and (5) above (also as quoted by Bresnan *et al.*, 2007:5) should not be grammatically possible either, since the phrase *giving someone the creeps* does not possess an allative meaning. Yet, language users do employ the prepositional dative in these instances.

In spite of these shortcomings associated with the meaning-to-structure hypothesis, Goldberg (2002) argues that the different syntactic realisations of the ditransitive construction, i.e. [S:NP V:VP IO:NP DO:NP] and [S:NP V:VP DO:NP IO:PP] should be regarded as two independent constructions. To support her standpoint, she goes on to posit the **SURFACE GENERALISATION HYPOTHESIS**:

There are typically broader syntactic and semantic generalizations associated with a surface structure form than exist between the same surface form and a distinct form that it is hypothesized to be syntactically or semantically derived from.

(Goldberg, 2002:329)

In light of the fact that each “surface pattern” should be described in terms of its own syntactic, semantic, and pragmatic properties, Goldberg posits that “the syntax of argument structure should be represented without recourse to derivations” (Goldberg, 2002:330).

In a similar vein, Gries (2003a:139-140) argues that even though certain constructions can often be used interchangeably, they “exist independently because of the cognitive-functional properties manifesting themselves in the variables’ values/levels”. This means that although constructions such as the dative alternation are often used as if they are semantically synonymous, they are not pragmatically synonymous.

Other constructionists, however, do not share this view. Although Cappelle (2006:12) agrees with Gries (2003a:139-140) that each construction is associated with its own distinct discourse-functional properties, he contends that “[j]ust because the two orderings are not linked by a truly Chomskyan transformation, it does not necessarily mean that language users are not aware of their relatedness” (Cappelle, 2006:12).⁹³ In fact, language users’ awareness of these relationships has long been promulgated in cognitive grammar and has been expressed by means of schemas and their elaborative relationships (cf. Langacker, 2008:55-57; Taylor, 2002:123-126). These relationships were mentioned earlier in Section 2.3.1.3 and illustrated in Figure 7, but are illustrated here again for ease of reference:

⁹³ Some constructionists (e.g. Götzsche, 2013) deny that there is any relationship between the two alternations on the grounds that alternations are descendant of lexical rules and transformations (cf. the notion of ‘deep structures’ in Generative Grammar), and must therefore be regarded with suspicion (also cf. Leclercq, 2015).

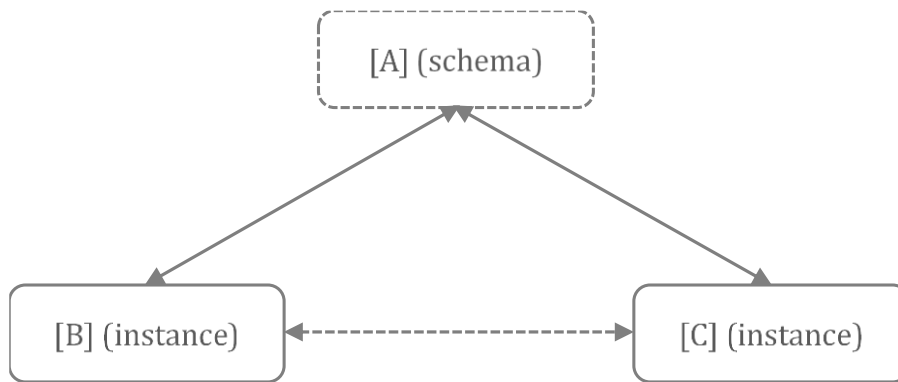


Figure 4.1: A representation of a schema and its instances (Taylor, 2002:125)

In Figure 4.1, the instantiating items [B] and [C] are said to flesh out the schema [A] in contrasting ways, while the schema abstracts the commonalities of the instances. Langacker (2008:56-57) explains it as follows:

A schema should therefore be seen as immanent in its varied instantiations, not as separate and distinct (even if shown individually for analytical purposes). By its very nature, a schema serves a categorizing function: capturing what is common to certain previous experiences, it can be applied to any new experience exhibiting the same configuration.

(Langacker, 2008:56-57; his emphasis)

The notions of SCHEMA and INSTANTIATION are applicable to every aspect of language structure, including phonology, morphology, syntax, and semantics. These notions have also been adopted by many usage-based construction theorists, and it is generally accepted that language users are able to schematise categories from exemplars and that these schematised networks form part of language users' linguistic knowledge. Therefore, to deny the relationships between constructions is to deny an important component of

the language users' linguistic knowledge (Cappelle, 2006:149; Kay & Fillmore, 1999:1; Perek, 2015).⁹⁴

On the basis of the notions of SCHEMAS and INSTANTIATIONS, Cappelle (2006) proposes the notion of **ALLOSTRUCTION**. Allostructions are “variant structural realisations of a construction that is left partially underspecified” (Cappelle, 2006:18). The partially underspecified construction is called a **CONSTRUCTEME** and conveys the idea that the two allostructions are linked by a common generalised supercategory in the minds of the language users and not by a generativist derivational rule. The inheritance links between the constructeme and the allostructions illustrate the similarity of the allostructions and represent the language users' shared knowledge of the different allostructions and the manner in which they are linked to one another. Concomitantly, it also considers the individual semantic and pragmatic details which distinguish the allostructions from one another. Cappelle (2006:19) emphasises that the two allostructions are not in complementary distribution. In other words, the allostructions may substitute one another without changing the (truth-conditional) meaning. This approach is illustrated in Figure 4.2.

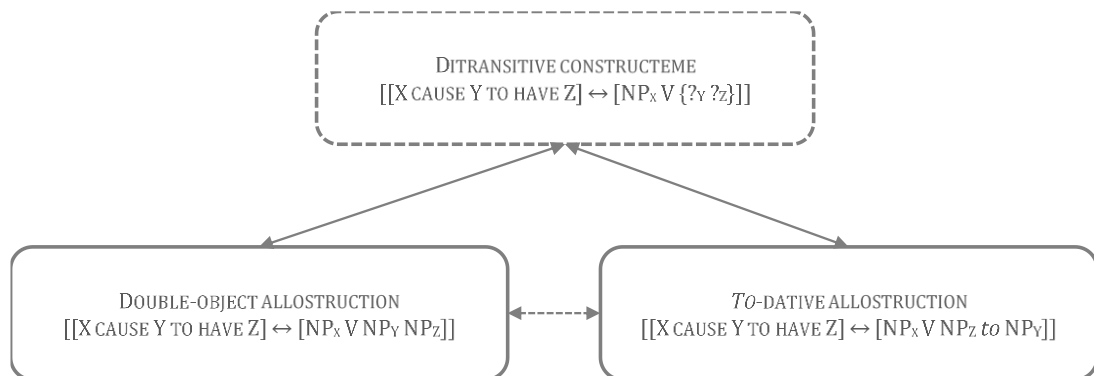


Figure 4.2: *The ditransitive constructeme and its allostructions (Perek, 2015:156)*

⁹⁴ It is interesting to note that although Goldberg (2002:329) promulgates the surface generalisation hypothesis and denies any link between alternations, she admits that possible paraphrase relations do in fact play a role in “the learning, processing or representation of language” (Goldberg, 2002:349).

In light of the discussion above, the analysis of the ditransitive construction in this thesis is grounded in the standpoint that the double-object construction and the *to*-dative construction are both allostructions of the ditransitive constructeme, and can be used alternatively by language users.

From this standpoint, though, the question of how language users choose which allostruction to use remains. Consequently, the factors driving dative alternation are considered in the next section.

4.3 Factors driving dative alternation

While scholars such as Goldberg (1995, 2002) argue that language users base their choice of dative on semantic grounds, other scholars have determined that fine-grained lexical semantic differences between the allostructions tend to be inconsistent, as mentioned earlier in Section 4.1 (cf. Bresnan *et al.*, 2007; Bresnan & Nikitina, 2007). Instead, the choice between the allostructions can often be predicted reliably and systematically by considering multidimensional information (cf. Bernaisch *et al.*, 2014; Bresnan *et al.*, 2007; Bresnan & Ford, 2010; Bresnan & Nikitina, 2007; De Cuypere & Verbeke, 2013; Thompson, 1990).

In the literature, several scholars have pointed out that the **verb lemma** governs the syntax of clauses (Bresnan & Ford, 2010:173; e.g. Levin, 1993:45-49). In (1), for example, the verb *give* can be replaced with the verb *offer* without changing the syntax or affecting the semantic equivalence of the clause. However, it cannot be replaced with the verb *say*:

(1) The government ... must **give** learners healthy food... (BSAE)

(1f) The government ... must **offer** learners healthy food...

(1g) *The government ... must **say** learners healthy food...

Given the way that verbs govern the syntactic elements in a clause, the factor LEMMA is included as a predictor variable in the analysis.

Other variables that have been determined to influence language users' choice in dative alternation are the **length of the recipient** (RECLENGTH) and **the length of the patient** (PATLENGTH). The length of these structures can be measured efficiently by counting the number of graphemic words, making these variables ratio variables (Arnold, Losongco, Wasow & Ginstrom, 2000:29; Bernaisch *et al.*, 2014:13; Bresnan & Hay, 2008:249; Szmrecsányi, 2004:1037; Wasow, 1997:102).

Here, it is important to note that increasing size often implies increased syntactic complexity. For example, pronouns tend to be shorter than noun phrases, while noun phrases tend to be shorter clauses. This is important because scholars sometimes include **the relative syntactic complexity of the recipient** (RECSYNTAX) and **the relative syntactic complexity of the patient** (PATSYNTAX) as factors that drive dative alternation (e.g. Bernaisch *et al.*, 2014:13). These variables are concerned with the syntactic structure by which the recipient and the patient are realised. The recipient, for example, can be realised by a pronoun or a noun phrase, whereas the patient can be realised by a pronoun, a noun phrase, or a clause. It is important to note that increased syntactic complexity is often linked to an increase in syntactic length. As a result, the variables RECLENGTH and PATLENGTH are said to be epiphenomena of the variables RECSYNTAX and PATSYNTAX. When interpreting the results of the analysis, this should be kept in mind.

By juxtaposing linguistic theories and psycholinguistic language processing insights with language data, Hawkins (1990, 1994) determines that language users base their syntactic choices on the notion of minimal complexity. In this regard, it is determined that the longer and the more syntactically complex a constituent is, the more difficult it becomes to process. The implication is that language users will choose to use the syntactic structures that are easy to process more often than not since these structures are recognised rapidly on-line by both speakers and hearers and so reduces cognitive effort. This is one of the reasons why structures that frequently occur in language are not very complex, while more complex structures occur less frequently, all other factors being equal (also cf. the explanation of exemplar theory in Section 2.1.3.2).

However, Hawkins (1990, 1994) posits that syntactic length is not the only factor that determines processing complexity. He finds that processing complexity is minimised when 'lighter' constituents (e.g. NPs that are shorter and less complex) are placed before

the ‘heavier’ constituents (e.g. NPs that are longer and more complex). This principle has come to be known as the **END-WEIGHT PRINCIPLE** (Wasow, 1997:81).

The principle can be explained by means of **VIEWING WINDOWS**: the shorter the viewing window, the easier the language user can recognise and process the constituents; the longer the viewing window, the harder it becomes for the language user to recognise and process the constituents (Hawkins, 1990, 1994). The viewing windows for (3) and (4), together with their transformed counterparts (3a) and (4a) are visualised below:

(3) ...Buddha [lowered]_{VP} [him]_{NP} [the silver thread of a spider]_{NP}...

1
2
3

(3a) ...Buddha [lowered]_{VP} [the silver thread of a spider]_{NP} [to him]_{PP}...

1
2
3
4
5
6
7
8

(4) This story ... [gives]_{VP} [the creeps]_{NP} [to people who hate spiders]_{PP} ...

1
2
3
4

(4a) This story ... [gives]_{VP} [people who hate spiders]_{NP} [the creeps]_{NP} ...

1
2
3
4
5
6

As seen earlier, within the framework of the meaning-to-structure hypothesis, the transformed versions of (3) and (4) above – i.e. (3a) and (4a) – would be regarded as being grammatically “more acceptable” since they are more congruent with the hypothesis. Yet, language users choose to use the structures that are less congruent with the hypothesis. Their decisions can be accounted for by considering the size of the viewing windows. In the original versions, (3) and (4), the viewing windows are relatively small indicating minimal complexity in terms of cognitive processing, whereas the viewing windows of their transformed counterparts in (3a) and (4a) are larger, indicating more complexity in terms of cognitive processing. Since many scholars have observed similar examples, it has been said that the psycholinguistic processing constraints supersede the semantic constraints proposed in the meaning-to-structure hypothesis (cf. Bresnan *et al.*, 2007;

In terms of animacy, **animacy of the recipient** (RECANIM) and **animacy of the patient** (PATANIM) are considered. Animacy is a semantic feature that is used to indicate how alive or sentient an entity is. Entities that are alive, such as humans and animals, are regarded as animate, while all other entities are regarded as inanimate. Consequently, the variables RECANIM and PATANIM take the value 'animate' or the value 'inanimate'. In previous research, Bresnan and Hay (2008) have found that all else being equal, New Zealand English speakers are more likely to produce inanimate recipients in the double-object construction than American speakers.

The **definiteness of the recipient** (RECDEF) and **definiteness of the patient** (PATDEF) are also considered in this analysis. As with animacy, definiteness is a semantic feature. If an entity is classified as being definite, it means that it is easily identifiable and distinguishable from other referents. Definite noun phrases are typically preceded by the definite article *the* or another kind of determiner such as *that* and *her*. If, however, an entity is classified as being indefinite, it means that it is not easily distinguishable from other entities. Indefinite noun phrases are typically preceded by the indefinite article *a*. In this analysis, both the recipients and the patients are classified as being either definite or indefinite. Scholars such as Thompson (1990) and Collins (1995) have found that whereas recipients tend to be more definite, patients tend to be more indefinite.

Another factor to consider is the **person of the recipient** (RECLocal). Scholars tend to assign values to the variable PERSON: 'local' and 'non-local'. In this analysis, the same values are used. First person (i.e. the speaker) and second person (i.e. the addressee) are typically regarded as being 'local' to the speech act, whereas third person (i.e. all the others) are regarded as being 'non-local'. PERSON is said to influence syntactic alternations categorically in some languages, but variably in English. It is expected that local recipients/patients will precede non-local recipients or patients.

The **concreteness of the patient** (PATCONCRETE) is also coded. Concrete patients are seen as concrete entities that can be perceived through one of the five senses. If an entity cannot be perceived by means of one of the five senses, it is classified as being abstract. In this analysis, the values 'concrete' and 'abstract' are used to code the concreteness of the patient. Concrete patients are expected to occur in the prepositional dative constructions.

Finally, **variety** (VARIETY) is coded with the values ‘BSAE’ and ‘WSAE’. This variable is included to determine the extent to which the choice in allostruction can be attributed to the variety that is used.

A summary of the variables expected to drive the dative alternation is provided in Table 4.3 on p. 186.

Thus far, only two of the syntactic structures said to realise the ditransitive construction have been mentioned. The other syntactic realisations of the ditransitive construction are addressed in Section 4.4.

4.4 Other factors influencing the structural and semantic realisation of the ditransitive

Aside from the double-object allostruction and the prepositional dative allostruction discussed in Section 4.1 and Section 4.3, Mukherjee (2005) and Mukherjee and Hoffmann (2006) attest the ditransitive construction in three more syntactic patterns. These patterns are exemplified in (7)-(9) below:

- (7) [Ikenga-Metuh (1987:1-12)]_{S:NP} [**gave**]_{V:VP} [∅]_{IO:∅} [several reasons for the different approaches scholars have taken in the study of ATR]_{DO:NP}. (BSAE)
- (8) The other major point he raises is in addressing the question of “why [**give**]_{V:VP} [∅]_{IO:∅} [∅]_{DO:∅} in the first place?” (as quoted by Mukherjee & Hoffmann, 2006:151 from ICE-GB)
- (9) The church vowed to reach out to the community as to [**give**]_{V:VP} [∅]_{DO:∅} [to those who are sick at Tshepiso Clinic]_{IO:PP}. (BSAE)

In all of these examples, the reader can retrieve the “omitted” object from the context of the text. For example, (7) is quoted from an academic text. Therefore, the reader can assume that the intended audience of Ikenga-Metuh’s text is the recipient in the construction. Therefore, the sentence could be reformulated as follows:

- (7a) [Ikenga-Metuh (1987:1-12)]_{S:NP} [**gave**]_{V:VP} [his audience]_{IO:NP} [several reasons for the different approaches scholars have taken in the study of ATR]_{DO:NP}.

In (9), the transferred entity (i.e. the patient) can also be inferred from the text. In the text, the author quotes an interviewee as saying:

We are showing our part of social responsibility to the community by **giving back** to the community. As the church, it's our responsibility to serve to community. It is hard for a theme to wait for medication with an empty stomach... (BSAE)

In view of this context, (9) can be reformulated as follows:

- (9a) The church vowed to reach out to the community as to [**give**]_{V:VP} [food and other necessities]_{DO:NP} [to those who are sick at Tshepiso Clinic]_{IO:PP}.

Despite the omitted syntactic elements in the aforementioned examples, the meaning [X transfers Y to Z] is undisputed.

In some instances, though, the meaning [X transfers Y to Z] becomes fuzzy, even in the presence of all the syntactic elements. The fuzziness is demonstrated in (10) and (11) below with the verbs *blame* and *remind*:

- (10) ... [the host population]_{S:NP} [**blames**]_{V:VP} [tourism]_{IO:NP} [for all negative aspects of life in the area]_{DO:PP}. (WSAE)
- (11) [Key words]_{S:NP} [**remind**]_{V:VP} [you]_{IO:NP} [of the main idea in the sentence]_{DO:PP}. (WSAE)

In these sentences, the meaning [X transfers Y to Z] is not perspicuous. In (10), for example, one cannot say that *the host population* (the agent) transfers *all negative aspects of life in the area* (the object) *to tourism* (the patient). Similarly, in (11), one cannot say

that *key words* (the agent) transfer *the main idea in the sentence* (the object) to *you* (the patient). Yet, scholars argue that these verbs are to be regarded as ditransitives.

In this regard, Goldberg (1995:31-39) explains that the instances where the agent causes the transfer of a patient to a recipient are the prototypical realisations of the ditransitive. When other instances of transfer are realised, though, it does not necessarily mean that the structure cannot be considered a ditransitive (cf. Section 2.3.1.2). Consequently, one can make a case for expressions involving verbs of judgement, whether negative or positive, (e.g. *blame, condemn, praise, compliment*) occurring in the ditransitive construction as it may denote a metaphorical or abstract process of transfer. In (10), for example, the agent *the host population* transfers an abstract patient to an inanimate recipient that does not have the ability or volition to accept the patient. However, this does not mean that a transparent interpretation of the sentence is not possible. In (10), then, it can be said that the agent *the host populations* transfers the abstract patient (*blame for*) *all negative aspects of life in the area* to the inanimate abstract receiver, *tourism* – i.e. [X ACTS TO VERBALLY TRANSFER Z TO Y] ↔ [S:NP V:VP IO:NP DO:PP]. Consequently, one could argue that expressions involving verbs of judgement are related to prototypical ditransitive constructions involving prototypical ditransitive verbs and as such extrude from the central sense of the ditransitive.

This approach allows for the metaphorical use of constructions. In the case of the ditransitive construction, the metaphorical use of the ditransitive accounts for those instances where the agent is not an animate being exhibiting volition and also for those instances where the recipient is not an animate being able to accept and possess the patient willingly (Goldberg, 1995:31-30).

The fuzziness of the meaning of transfer in (10) can in part be attributed to the inanimacy of the recipient. It is difficult to argue that an inanimate recipient such as *tourism* has the ability and willingness to take possession of the patient since it is strictly speaking not able to accept the blame for anything, neither is it possible to demonstrate the quality of willingness. Similarly, the fuzziness of the meaning of transfer in (11) can be attributed to the inanimacy of the agent.

Goldberg (1995:31-39) argues for a polysemous semantic approach in which related senses of the ditransitive are also viewed as ditransitives. This means that while (10) and (11) do not imply that an agent literally transfers a patient to a recipient (i.e. the recipient does not take possession of an object), they are systemic metaphors of causal events as transfers (cf. Goldberg, 1995:148) extending the prototypical meaning. In (10), then, it can be said that the agent *the host populations* transfers the abstract patient (*blame for*) *all negative aspects of life in the area* to the abstract receiver, *tourism*. Similarly, in (11), it can be said that the agent *key words* transfers the abstract patient, (*a reminder of*) *the main idea of the sentence*, to the recipient *you*.

Furthermore, Quirk *et al.* (1985:1208-1210) put forward syntactic reasons for classifying (10) and (11) as ditransitive constructions. They argue that these verbs are to be regarded as ditransitive on the grounds that should these sentences be transformed to the passive voice – cf. (11a) and (12a) – the meaning will remain the same as in (10) and (11).

(11a) The next stage, antagonism, is an extreme point where all negative aspects of life in the area are blamed on tourism (by the host population). (WSAE)

(12a) You are reminded of the main idea in the sentence (by key words). (WSAE)

As pointed out earlier, though, syntactic characteristics should not be the only factor to play a role in classifying the transitivity of a verb. In a similar train of thought, Mukherjee (2005:12) argues that the overextension of systematic correspondences results in the inclusion of a very heterogeneous group of verbs and that this approach is “rather extreme and not very useful”. Therefore, he argues that the ditransitive verb must be attested in the prototypical form of the ditransitive pattern – i.e. [S:NP V:VP IO:NP DO:NP] – for it to be classified a ditransitive verb (also cf. Biber *et al.*, 1999:150-151). Following this argument, the verb *express* in (12) and (13) cannot be classified as a ditransitive verb — read with (12a) and (13a) below:

(12) Once more, the management, staff and students of the university wish to [express]_{V:VP} [their condolences]_{DO:NP} [to the parents and friends of the deceased]_{IO:PP}. (BSAE)

- (12a) *Once more, the management, staff and students of the university wish to [express]_{V:VP} [the parents and friends of the deceased]_{IO:NP} [their condolences]_{DO:NP}.
- (13) Government is committed to [providing]_{V:VP} [food]_{DO:NP} [to poor households]_{IO:PP} ... (BSAE)
- (13a) *Government is committed to [providing]_{V:VP} [poor households]_{IO:NP} [food]_{DO:NP}.

Mukherjee's (2005) approach, however, is not without its own limitations. As he himself points out, there are some verbs such as *express* and *provide* in (12) and (13) that are not attested in the basic ditransitive pattern – as illustrated in (12a) and (13a) – yet, these verbs do convey the meaning [X transfers Y to Z] (Mukherjee, 2005:12).

Taking the aforementioned into account, constructions in this study are only classified as being ditransitive when they express a situation type of transfer. This means that verbs such as *blame* in (10), *remind* in (11) *express* in (12) and *provide* in (13) are classified as ditransitive.

4.5 Results

4.5.1 General overview

Figure 4.3 provides an overview of the frequency and distribution of the syntactic patterns associated with the ditransitive construction in the BSAE corpus and the WSAE corpus.

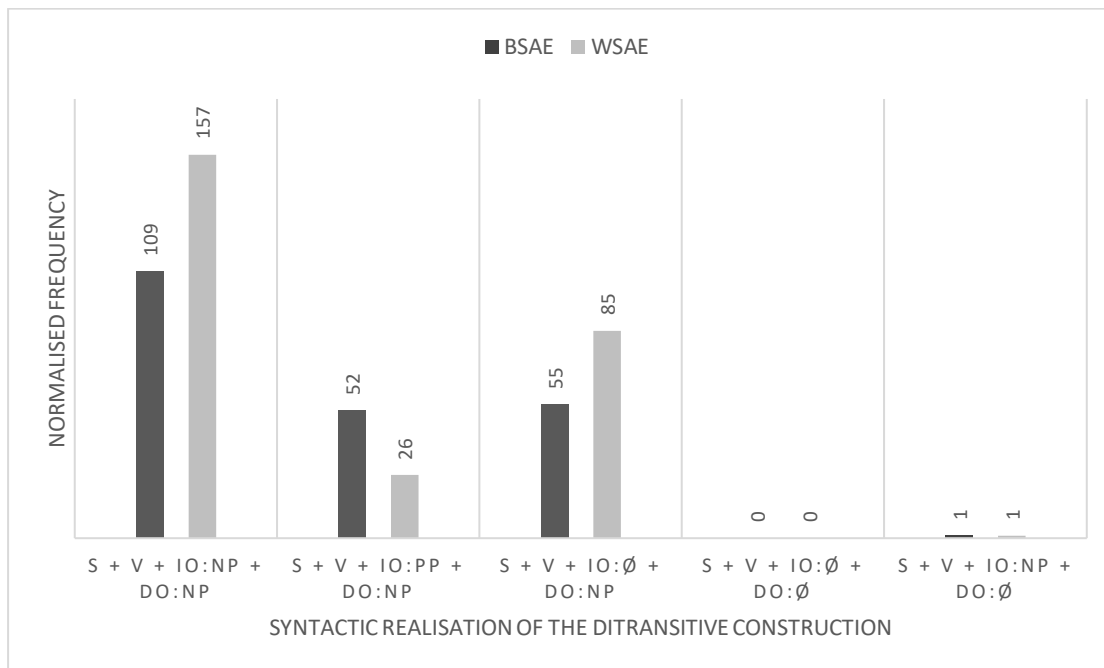


Figure 4.3: *The frequency and distribution of the various allostructions of the ditransitive constructeme in BSAE and WSAE*

At first glance, there seems to be quite a large divergence between the BSAE corpus and the WSAE corpus in terms of the relative frequencies of the first two syntactic patterns. A chi-squared test indicates that the difference in the use of the double-object allostruction between BSAE and WSAE is statistically significant ($\chi^2 = 13.91$ with $p = 0.0002$). The difference in the use of the prepositional dative allostruction between BSAE and WSAE is also statistically significant ($\chi^2 = 16.36$ with $p = 0.00005$). Proportionally, the prepositional dative comprises 32% of the ditransitive constructeme in BSAE, while the prepositional dative comprises a mere 13% of the ditransitive construction in WSAE. These results are broadly similar to the results of De Cuypere and Verbeke (2013) and Mukherjee and Hoffmann (2006). They found that the prepositional dative proportionally occurs more frequently in Indian English (i.e. in the ICE-India corpus and the Kolhapur corpus) than in British English (i.e. in the ICE-GB corpus). Their results (together with the results of this analysis) are provided in Table 4.1.

Table 4.1: *The relative frequencies of alternating ditransitive constructions in BSAE, WSAE, Indian English and British English*

	BSAE	ICE-India [†]	Kolhapur [‡]	WSAE	ICE-GB [†]
Double-object construction	68%	56%	56%	87%	69%
Prepositional dative construction	32%	44%	44%	13%	31%

[†] (Mukherjee & Hoffmann, 2006) [‡](De Cuypere & Verbeke, 2013)

When WSAE and British English are compared to one another, though, the prepositional dative seems to occur half as often in WSAE than it does in British English. Considered in this light, the prepositional dative seems to occur only slightly more often in BSAE than it does in British English.

In the literature, scholars have put forward several hypotheses (of both a sociolinguistic nature and a psycholinguistic nature) to account for this phenomenon in Indian English. Olavarría de Ersson and Shaw (2003:159) argue that culture (a sociolinguistic factor) may be a driving force in the grammatical choices made by Indian English speakers in terms of their allostruction choice in the ditransitive constructeme. They argue that Northern European cultures tend to view the individual as being at the centre of the world, while South Asian cultures tend to view the individual as “a part or a small object in a larger whole” (Olavarría de Ersson & Shaw, 2003:159). In his 6-D model of national culture, Hofstede (2011:11-12) describes this issue in the INDIVIDUALISM VERSUS COLLECTIVISM dimension which deals with “the degree to which people in a society are integrated into groups”. Hofstede (2011:11) describes this dimension as follows:

On the individualist side we find cultures in which the ties between individuals are loose: everyone is expected to look after him/herself and his/her immediate family. On the collectivist side we find cultures in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families (with uncles, aunts and grandparents) that continue protecting them in exchange for unquestioning loyalty, and oppose other in-groups.

In the individualism category, India scores an intermediate score of 48, while Britain scores a high score of 89, which confirms Olavarría de Ersson and Shaw's (2003:159) observation that South Asian cultures have more collectivist traits while British cultures have more individualist traits (Hofstede Insights, 2018). This is realised in the grammar of the ditransitive constructeme as follows:

[W]hen offered the choice between two syntactic structures that focus either on what is provided or on the recipient to express more or less the same thing, BrE speakers would be likely to profile the recipient more frequently in their use of language than their Indian counterparts do, whereas IndE speakers will be more likely to use the structure which profiles what is provided rather than the structure where the recipient is profiled. This seems to make an extraordinarily – some might say disturbingly – direct connection between grammar and “culture”, and it would be very interesting if corroborative evidence could be found.

(Olavarría de Ersson & Shaw, 2003:159)

Hofstede's 6-D model of national culture assigns quite a high score of 65 in the individualism category for South Africa but states that the scores are for the white population of South Africa and does not account for the black population (Hofstede Insights, 2018). But scholars characterise indigenous African cultures as being collectivist (cf. Eaton & Louw, 2000; Triandis, McCusker & Hui, 1990). As such, it could be inferred that the BSAE speakers are characterised as being more collectivist than the WSAE speakers.

This cultural hypothesis, however, is not supported by the realisation of the ditransitive in the Bantu languages. In Sesotho (one of the indigenous Bantu languages spoken by BSAE speakers in South Africa), Demuth, Odat, Machobane and Mloi (2014:9) explain that the recipient is typically realised as a preverbal clitic (cf. (14)) rather than a postverbal noun phrase (cf. (15)) in everyday Sesotho discourse:

(14) *Bana ba-mo-pheh-el-a nama* SObj-VO
 children AGR-OBJ-cook-APL-FV meat
 ‘The children are cooking her some meat’

(15) *Bana ba-pheh-el-a mme nama* SVOO
 children AGR-cook-APL-FV mother meat
 ‘The children are cooking meat for my mother’

This seems to be in conflict with the notion that the collectivist nature of the BSAE speakers predisposes them to profile what is being received as opposed to the receiver.

The second broad motivation for the differences may be attributed to psycholinguistic explanations. In this regard, Mukherjee and Hoffmann (2006:154-155) attribute the high frequency of the prepositional dative in Indian English in part to the high frequency with which the ditransitive verb *give* is used in light-verb constructions (e.g. *give an explanation*). They go on to explain that the use of these light-verb constructions typically results in the use of the prepositional dative allostruction. This is illustrated in (16) and (17) below:

(16) The potential impact of **giving** additional **assistance** to [~assisting] pension grant recipients in the form of food vouchers to relieve them of the increased spending on basic necessities. (BSAE)

(17) ...3 of them refer to Christians giving aid to [~aiding] needy people (Romans 15:27; Philippians 2:25,30). (BSAE)

Although a small number of examples (such as the ones above) were found in the BSAE corpus, the use of *give* as a light verb in the BSAE corpus is relatively limited, and in the instances where they do occur, there was no editorial intervention to remove the light verb. As such, it is not possible to draw strong conclusions in this regard in BSAE.

Proposing another hypothesis, Mukherjee and Hoffmann (2006:166-167) speculate that it may be possible that the prepositional dative occurred more frequently in 19th century British English than it does today, and that the Indian English speakers preserved the

usage patterns of the prepositional dative exhibited in 19th century British English. Therefore, they call for a diachronic investigation of the ditransitive constructeme in Indian English and British English to determine whether the patterns of ditransitive complementation that are observed in Indian English today is reminiscent of the ditransitive complementation patterns used in British English in the 19th century.

Should their aforementioned hypothesis be proven false, however, Mukherjee and Hoffmann (2006:166-167) speculate that the higher proportion of prepositional datives in Indian English could be ascribed to postcolonial divergence. In answer to Mukherjee and Hoffmann's hypotheses, De Cuypere and Verbeke (2013:180) quote Wolk, Bresnan, Rosenbach and Szmrecsanyi's (2013:393) observation that the proportion of prepositional datives in the ARCHER corpus has remained fairly stable since the 19th century and that there has been no significant decrease in the use of the prepositional dative in British English. They then compare the proportion of prepositional datives they found in their sample of late 20th century Indian English with the proportion of prepositional datives found in 19th century British English and "found strong evidence that the number of *to*-datives in [their] Kolhapur sample was significantly higher than would be expected if there were no difference in proportions between the two samples". Based on these results, they support Mukherjee and Hoffmann's (2006:166) alternative hypothesis by arguing that the higher proportion of prepositional datives in Indian English can likely be ascribed to postcolonial divergence.

Similar hypotheses could be posited for the higher proportion of prepositional datives found in BSAE than in WSAE. But in order to determine whether the difference is to be ascribed to the preservation of the ditransitive patterns of an earlier form of WSAE or to postcolonial divergence, a diachronic investigation into the ditransitive patterns of early WSAE should be launched. Such an investigation, however, falls outside the scope of this thesis.

Another view to consider is that BSAE users produce language in the face of certain constraints. Lanstyák and Heltai (2012:100) point out that although all communication is always constrained in some way, certain communicative events take place under conditions where certain constraints have a greater than average role. In this regard, Kruger and Van Rooy (2016a:27) point out that in situations where language contact

takes place, the communicative event is necessarily constrained. Consequently, L2 varieties demonstrate a preference for the more explicit encoding of information (Van Rooy *et al.*, 2010). In L2 research, this phenomenon is known as hyperclarity (Williams, 1987:179).

In the case of the ditransitive construction, it could be argued that the prepositional dative is a more explicit lexicogrammatical encoding of the ditransitive construction where the preposition maps the trajectory of the patient more explicitly, encoding the process of transfer more explicitly. Williams (1987:181) calls this specific form of hyperclarity **mapping-transparency**. The purpose of the explicitation of the information is two-fold. From a psycholinguistic perspective, it enables the L2 users to keep track of their own processing tasks. From a sociolinguistic perspective, the explicitness eases the task of decoding for the listener. The speaker does this to accommodate the listener in an attempt to enhance the chances of communicative success.

To gain more insight into why the L2 English users choose to use the prepositional dative construction more frequently than the L1 English users do, a distinctive collexeme analysis is performed. The results are presented in the next section.

4.5.2 *Distinctive collexeme analysis*

Earlier in Section 2.4.2.2, the notion of SEMANTIC COHERENCE was addressed. Semantic coherence entails the semantic compatibility of a lexeme and the construction in which it occurs. Such semantic coherence can be investigated by means of Stefanowitsch and Gries's (2003) collostructional analysis which can be regarded as an extension of collocational analysis. In a collostructional analysis, "the interaction of lexemes and the grammatical structures associated with them" are investigated (Stefanowitsch & Gries, 2003:209). They explain that a collostructional analysis basically determines which lexemes are attracted or repelled by a particular slot in a given construction. This is determined on the grounds of the frequency with which a lexeme occurs in that specific slot of the construction (Stefanowitsch & Gries, 2003:214-215). The distinctive collexeme analysis, in particular, is concerned with "determining the degree to which particular slots in a grammatical structure prefer, or are restricted to, a particular set or semantic class of lexical items" (Stefanowitsch & Gries, 2003:211).

Table 4.2: Collexemes distinguishing between the ditransitive and to-dative

BSAE				WSAE			
ditransitive (n=150)		prepositional dative (n=72)		ditransitive (n=370)		prepositional dative (61)	
collexeme	distinctiveness	collexeme	distinctiveness	collexeme	distinctiveness	collexeme	distinctiveness
tell (36:1)	5.6390691***	provide (18:44)	13.0600007***	tell (133:2)	8.10891117***	provide (27:19)	5.93218212***
urge (7:0)	1.2120421	blame (0:1)	0.4890205	ask (64:2)	2.78254213***	offer (13:10)	3.38886699***
promise (6:0)	1.0359508	get (1:1)	0.2640388	get (14:0)	0.94332566	give (76:25)	3.13448891***
ask (17:4)	0.8961740	offer (4:2)	0.1981279	show (28:2)	0.76288265	make (0:1)	0.84914744
prepare (5:0)	0.8608591			urge (7:0)	0.46745588		
show (8:1)	0.8219312			prepare (4:0)	0.26610481		
give (46:18)	0.6218898			warn (3:0)	0.19932711		
warn (2:0)	0.3414677			blame (1:0)	0.06627555		

- *** Collocational strength > 3 = $p < 0.001$
- ** Collocational strength > 2 = $p < 0.01$
- * Collocational strength > 1.30103 = $p < 0.05$

In this study, a distinctive collexeme analysis is performed to determine which set of lexemes are attracted to the double-object allostruction and which set of lexemes are attracted to the prepositional dative allostruction. The analysis is executed in both the BSAE corpus and the WSAE corpus to determine whether the collostructs⁹⁵ are similar to one another in the different varieties, or whether they differ across the varieties. The results of these analyses are presented in Table 4.2 above, where the verbs are ranked according to their respective collocational strengths.⁹⁶ In the BSAE corpus, 222 ditransitive constructemes were analysed, of which 150 were double-object allostructions, and 72 were prepositional dative allostructions. In the WSAE corpus, 431 ditransitive constructemes were analysed, of which 370 were double-object allostructions, and 61 were prepositional dative allostructions. The results of the analysis are presented in Table 4.2.

The results for the distinctive collexeme analysis indicate that the only verb that is distinctive of the double-object allostruction in BSAE is the verb *tell*, while the verb *provide* is the only verb that is highly distinctive of the prepositional allostruction. In WSAE, *tell* is also the most distinctive verb for the double-object allostruction, while *provide* is also the most distinctive verb for the prepositional allostruction. Another verb that is highly distinctive of the double-object allostruction in WSAE is the verb *ask*, while the verbs *offer* and *give* are also distinctive of the prepositional allostruction in WSAE.

4.5.3 Conditional inference tree

A conditional inference tree was extracted to investigate the interaction between the variables that have an influence on speakers' use of the double-object allostruction or the prepositional dative allostruction. The tree is visualised in Figure 4.4. It was drawn using the `ctree()` algorithm in the `party` package in R (Hothorn *et al.*, 2015). Conditional inference trees are used to inspect a data set recursively to determine which independent variables (categorical or

⁹⁵ Stefanowitsch and Gries (2003:215) explain that the lexemes that are attracted to a specific construction are known as the **COLLEXEMES** of that particular construction. On the other hand, the constructions that are associated with specific lexeme are known as a **COLLOSTRUCT**. The combination of a collexeme and a collostructs is known as a **COLLOSTRUCTION**.

⁹⁶ The collocational strengths also reflect the p-values. A collocational strength larger than three is significant at $p < 0.001$; a collocational strength larger than two is significant at $p < 0.01$; and a collocational strength larger than 1.30103 is significant at $p < 0.05$.

numeric) predict best the known outcome of the dependent variable. The independent variables used to draw the tree was discussed in Section 4.3 are listed in Table 4.3. All the possible splits are significant at $p < 0.05$. The conditional inference tree (depicted in Figure 4.4 on p. 187) returned a high prediction accuracy of 92%.

Table 4.3: Summary of the variables expected to drive the dative alternation

Variable	Description	Value
LEMMA	The ditransitive verb lemma	e.g. <i>give</i>
RECSYNTAX/PATSYNTAX	Indicates whether the recipient/patient is pronominalised, i.e. the syntactic structure by which the recipient is realised	'pronominal', 'nominal'
RECLENGTH/PATLENGTH	The length of the recipient/patient (measured in graphemic words)	numerical value
RECANIM/PATANIM	Indicates how alive or sentient the recipient/patient is	'animate', 'inanimate'
RECDEF/PATDEF	Indicates whether the recipient/patient is easily identifiable and distinguishable from other referents	'definite', 'indefinite'
RECLocal	Indicates the person of the recipient, i.e. 1 st person and 2 nd person (local) vs 3 rd person (non-local)	'local', 'non-local'
PATCONCRETE	Indicates whether the patient can be observed with one of the five senses	'concrete', 'abstract'
VARIETY	Indicates variety	'BSAE', 'WSAE'

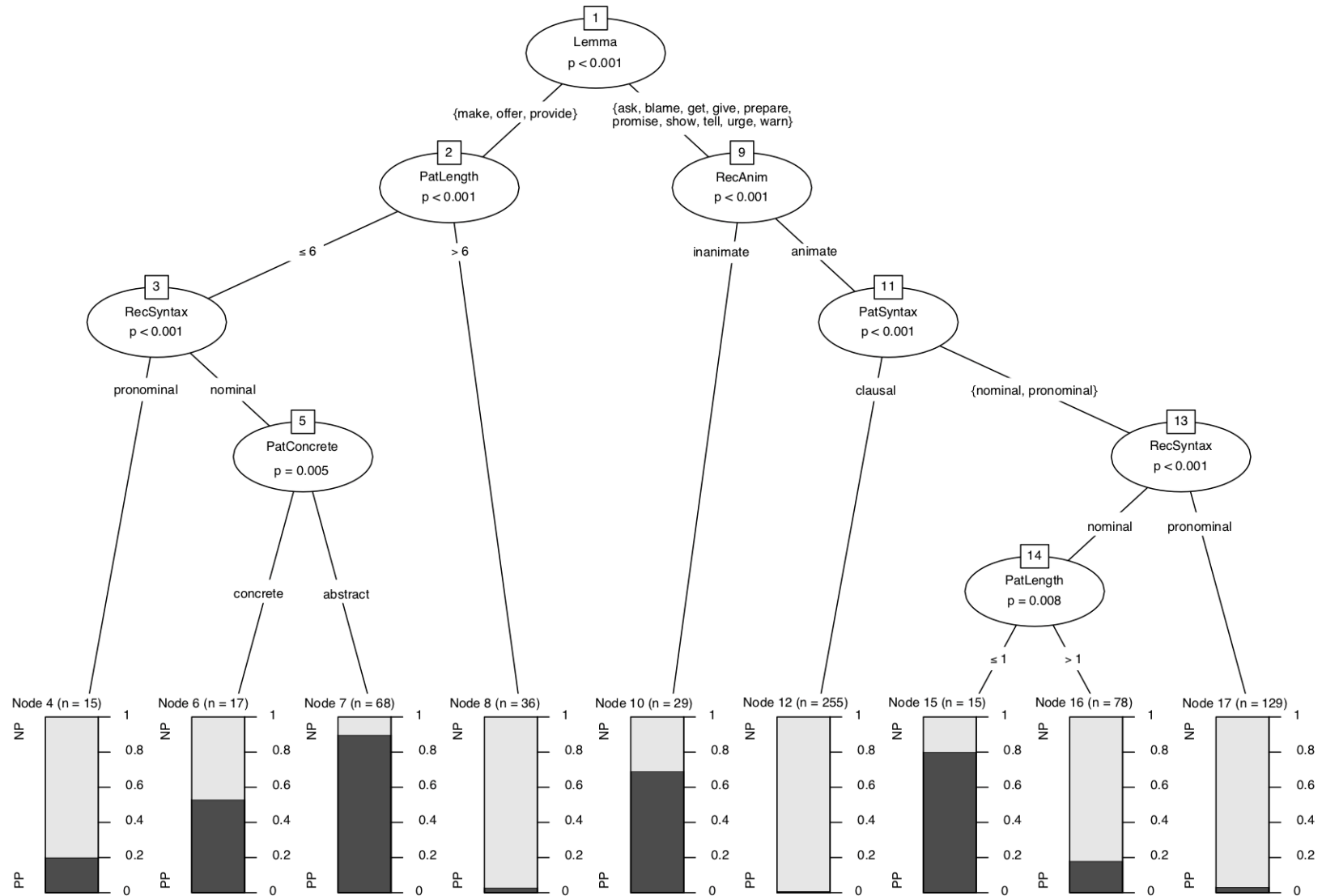


Figure 4.4: A conditional inference tree indicating the variables that influence the ditransitive constructeme in BSAE and WSAE

The first split at the top of the tree points to the variable LEMMA being the first determinant with a probability of $p < 0.001$ (cf. Node 1). The branch on the left-hand side is determined by the verbs listed in (18):

(18) make, offer, provide

On the other hand, the branch on the right-hand side is determined by the verbs listed in (19):

(19) ask, blame, get, give, prepare, promise, show, tell, urge, warn

The first variable that determines the language users' choice in allostruction when using the verbs *make*, *offer*, and *provide* is the length of the patient (cf. Node 2). In those instances where the patient consists of more than six words, the language users choose to use the double-object allostruction 98% of the time (cf. Node 8). This is exemplified in (20):

(20) This **offers** [the young artists]_{IO:NP} [an opportunity to produce their own demo tapes]_{DO:NP}. (WSAE)

However, if the patient consists of six words or fewer, the language users' choice is influenced by the syntax of the recipient (cf. Node 3). If the recipient is realised by a pronoun, the language users choose the double-object allostruction 80% of the time (cf. Node 4). This is exemplified in (21):

(21) I **am offering** [you]_{IO:NP} [my heart]_{DO:NP}. (BSAE)

In those instances where the recipient is realised by a noun phrase (cf. Node 5), the language users will consider whether the patient is concrete in nature or abstract in nature. Should the patient be concrete in nature, the users will use the double-object allostruction approximately 50% of the time (cf. Node 6). On the other hand, should the patient be abstract in nature, the language users will use the prepositional dative 90% of the time (cf. Node 7). This is exemplified in (22):

- (22) Segment-based satisfaction analysis **may offer** [a number of benefits]_{DO:NP} [to managers]_{IO:PP}. (WSAE)

The verbs on the right-hand side of the decision tree are first of all split on the basis of the animacy of the recipient. In those cases where the recipient is inanimate in nature, the prepositional dative is preferred in 70% of the instances (cf. Node 10). This is exemplified in (23):

- (23) A ruling party **gives** [shape and direction]_{DO:NP} [to government_NN1 priorities and strategies]_{IO:PP}. (BSAE)

If the recipient is animate, though, the choice is further influenced by the syntax of the patient (cf. Node 11). In those cases where the patient is realised by a clause, the double-object construction will always be chosen (cf. Node 12). It is important to note that Node 12 comprises 255 instances. (The node with the most instances after Node 12 is Node 17, comprised of 129 instances.) The instances comprising Node 12 contain communication verbs that convey verbiage. This is exemplified in (24):

- (24) He **promised** [marchers]_{IO(recipient)} [they would receive a response within a specific period]_{DO(verbiage)}. (BSAE)

On the other hand, in those cases where the patient is realised by a pronoun or a noun phrase, the language users' decisions are further influenced by the syntax of the recipient (cf. Node 13). Should the recipient be realised by a pronoun, the language users will choose to use the double-object construction 98% of the time (cf. Node 17). This is exemplified in (25):

- (25) The doctors **gave** [us]_{IO:NP} [discount]_{DO:NP}. (BSAE)

If, however, the recipient is realised by a noun phrase, the language users further take the length of the patient into account. In those instances where the patient consists of more than one graphemic word, the double-object construction is used almost 82% of the time (cf. Node 16). This is exemplified in (1):

- (1) The government ... **must give** [learners]_{IO:NP} [healthy food]_{DO:NP}... (BSAE)

However, in those instances where the patient consists of only one graphemic word, the prepositional dative is chosen 80% of the time (cf. Node 15). This is exemplified in (26) and (27):

- (26) God **gave** [instructions]_{DO:NP} [to Adam]_{IO:NP}. (BSAE)

- (27) They thought of cooking and **giving** [food]_{DO:NP} [to the sick]_{IO:PP}. (BSAE)

Given the length and complexity of the description of the conditional inference tree, the results are briefly summarised. Aside from the variable LEMMA, the variables that seem to have the strongest effect on the choice in ditransitive allostruction is the length of the recipient (RECLENGTH) and the patient respectively (PATLENGTH), and the pronominality of the recipient (RECSYNTAX). These variables are the same variables that were found to have a strong influence on the choice of ditransitive allostruction in the Asian Englishes (Bernaisch *et al.*, 2014; De Cuypere & Verbeke, 2013). The influence of these variables is largely based on the principles of harmonic alignment, which state the preference of placing shorter objects before longer objects (i.e. the end-weight principle), placing pronominals before nominals, placing concrete objects before abstract objects, and placing animate objects before inanimate objects.

From a semantic perspective, it is interesting to note that the lemmas on the right-hand branch of the conditional inference tree are primarily communicative in nature expressing metaphorical situations of transfer, e.g. *ask*, *blame*, *promise*, *tell*, *urge*, and *warn* (cf. Node 1).

Furthermore, it is important to point out that the variable VARIETY does not feature strongly in the conditional inference tree. This supports Bernaisch *et al.*'s (2014) finding that the processing-related factors that are known to drive the allostructional choices made by L1 language users also drive the allostructional choices made by L2 language users.

4.5.4 Random forest

Sometimes, conditional inference trees do not reflect highly predictive interactions (cf. Bernaisch *et al.*, 2014:14-15; Tagliamonte & Baayen, 2012:163-164), and random forests is a useful complement to confirm the findings of the inference tree. The random forest is visualised in Figure 4.5. It was drawn using the `cforest()` algorithm in the `party` package in R (Strobl, Hothorn & Zeileis, 2009a). Unbiased variable selection is the key to reliable prediction and interpretability in both individual trees and forests. However, while a single tree's interpretation is straightforward, in random forests extra effort is necessary to assess the importance of the variables.

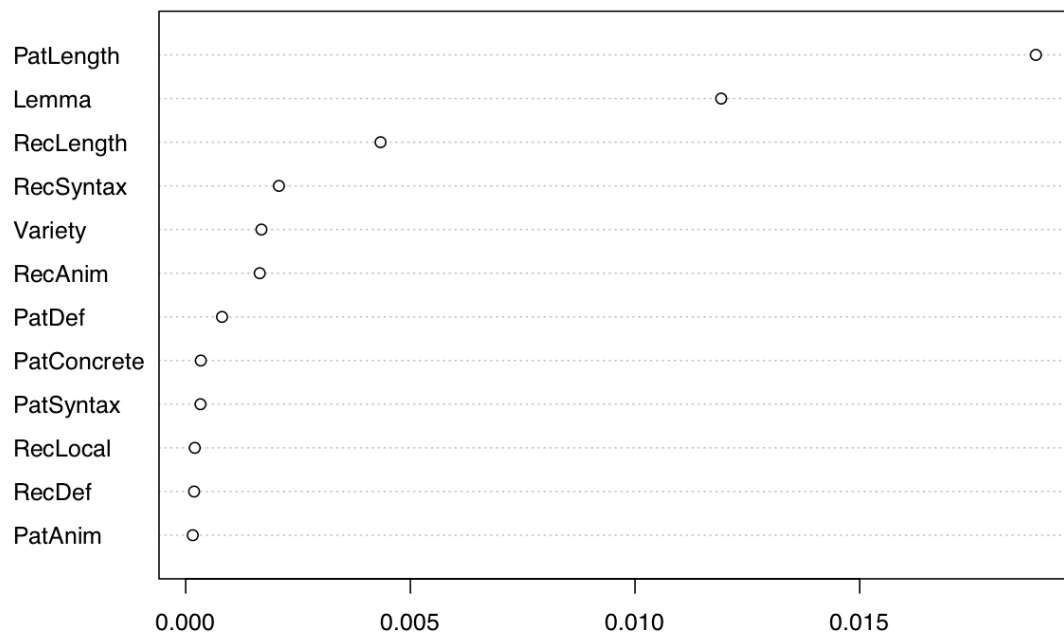


Figure 4.5: Conditional importance of variables

The random forest ranks the variables in order of importance: PATLENGTH (0.027) > LEMMA (0.005) > RECLENGTH (0.004) > RECSYNTAX (0.003) > VARIETY (0.002) > RECANIM (0.002) > PATDEF (0.001) > PATCONCRETE (0.001). The classification accuracy of the random-forest approach is slightly higher than that of the conditional inference tree at 94%.

The high predictive values of the patient length and recipient length variables once again confirm that the end-weight principle has a strong influence on the processing effort of

the language user. The high predictive value of the verb lemma, in turn, highlights the extent to which certain verb lemmas are semantically coherent with certain allostructions. Given these results, it can be said that the random forest confirms the findings of the conditional inference tree to a large extent. It is also important to note that though variety is a significant predictor, it is not as significant as factors such as the length of the patient, the verb lemma, the length of the recipient, and the syntactic realisation of the recipient. This fact indicates that the same cognitive factors that constrain the WSAE users' choices in terms of the ditransitive alternation constrain the BSAE users' choices.

4.6 Conclusion

The most important finding of the research presented in this chapter is that while variety is a significant factor amongst the factors driving dative alternation, it is not the most significant. Instead, other factors (which mostly constrain the cognitive processing of language users) drive dative alternation in both the L1 variety of English and the L2 variety of English. This is an important finding since it demonstrates that the same factors influencing the processing effort of the WSAE speakers, also influence the processing effort of the BSAE speakers.

From a psycholinguistic point of view, the factors that have the biggest influence on the processing effort of both the BSAE and WSAE language users are those that pertain to the end-weight principle, i.e. the principle of organising the constituent elements of the clause in order of increasing weight to reduce cognitive effort. The length of the recipient and the length of the patient relate to one other in that the heaviest element occupies the final position. In the corpora that were analysed, a large number of patients were realised by dependent clauses. In all of these instances, the double-object allostruction was chosen. In those instances where the patient was realised as a pronoun, the language users chose to use the double object construction in almost all the instances. In those instances where both the recipient and the patient were realised by noun phrases, the speakers based their choice of allostruction on the length of the noun phrases so that the shorter noun phrase precedes the longer noun phrase.

As pointed out in the random forest, variety is a significant predictor; the BSAE users seem to use the prepositional dative construction proportionally more often than the WSAE users (as illustrated in Section 4.5.1). It would seem as if these differences in frequency can largely be attributed to the uneven presence of the conditioning grammatical variables.

Aside from the fact that the prepositional dative construction occurs more often in BSAE and WSAE, not much difference can be observed between the BSAE users' and the WSAE users' use of the ditransitive construction. Perhaps this is indicative of the fact that the BSAE users do not view English as a symbol of social integration and feel that their identity is primarily determined by their home language and their multilingualism rather than English (Coetzee-Van Rooy, 2014:52). Since they use English to acquire social mobility and to achieve economic and educational success and not to express their identity, it could be posited that they accommodate the WSAE users' use of English rather than using English as their main identity marker. Based on these findings, it seems as if the norms as it pertains to the ditransitive construction are aligned across the STL and IDG components of the broader SAE community, and that the BSAE norm is not in the process of establishing a radically different norm from the WSAE norm. These findings are similar to the findings of Mesthrie (2010) and Mesthrie *et al.* (2015).

Chapter 5

Anti-deletion and the finite *that*-complementiser

5.1 Contextualisation

Scholars investigating L2 varieties have observed that L2 speakers produce language under certain constraints, the biggest of which is the bilingual processing restraint. One of the consequences of this bilingual processing constraint is a tendency towards structural elaboration and explicitness (Kruger & Van Rooy, 2016a:42). Williams (1987:178) ascribes this feature to a tendency towards **hyperclarity**. She explains that:

Natural language is often redundant, but from the point of view of the speaker, certain shortcuts may be taken in speech production which eliminate these redundancies for the sake of efficiency. The principle of hyperclarity, on the other hand, is more attentive to the needs of the listener. Any ambiguities which are introduced through the operation of the economy principle may be counteracted by the hyperclarity principle.

It is quite possible that speakers follow the principle of explicitness not only to ease the decoding task of the listener, but also to keep track of their own production, particularly in the production of complex sentences or semantic concepts.

(Williams, 1987:179)

This tendency of L2 English users to convey information as explicitly as possible has been observed by scholars investigating L2 varieties. Van Rooy *et al.* (2010), for example, finds that East African English language users prefer to use language that is more explicit, i.e. they use more explicit referencing, more explicit constructions, and increased repetition. Similarly, Kruger and Van Rooy (2016b) find that L2 language users prefer to use the *that*-complementiser explicitly in reported speech. In their multidimensional model of constrained language, Kruger and Van Rooy (2016a) also observe L2 English users' preferences for clausal elaboration and the overt marking of relationships between propositions.

One feature in particular that pertains to the idea of hyperclarity and conveying information as explicitly as possible is anti-deletion. Mesthrie (2006:115) coins the notion **ANTI-DELETION** to refer to the instances where deletion does not take place as commonly expected in generative analyses of English, or as reported in the dialectological literature. Anti-deletion is comprised of three types: (i) **undeletion** – restoring an underlying element that is thought to have been deleted or has an empty node in generative linguistic terms; (ii) **non-deletion** – conserving a feature of standard English is sometimes deleted, i.e. the feature is regarded as optional in standard English; and (iii) and **insertion** – inserting grammatical morphemes that are not present in the underlying structure of standard English. Mesthrie (2006:115) posits that the notion of ANTI-DELETION enables linguists to characterise the syntax of individual second language varieties of English.

Mesthrie and Bhatt (2008:92) provide several examples where BSAE speakers disfavour the deletion of elements like the infinitive *to*, resumptive pronouns in relative clauses, dummy *it* after verbs like *make clear*, dummy *it* before the verb *be*, and the

complementiser *that* after *as you know* and *as I said*.⁹⁷ A recent study by Botha (2013) uncovered further undeletion tendencies in the noun phrases of BSAE, but she is careful to point out that the constructions with deletions are more frequent than the constructions with undeletions. Her findings suggest that while undeletion is a feature of BSAE, it may not be as pervasive as scholars thought.

In this chapter, the notion of non-deletion come under consideration. Although the use of the *that*-complementiser is mandatory in extraposed clauses (cf. (1)) and almost categorically deleted in direct speech in standard English (cf. (2)), it is optional when it occurs as a complementiser in postverbal complementation (cf. (3)).

- (1) **That** we were grateful to Miss Stonedale and Miss Stott was frequently evidenced by entries in the Food Suggestion Book. (BNC)
- (2) Mahlaba said \emptyset : “Through hard work and dedicated staff this is always possible. However, with students who are willing to take education as key to their success it’s more than possible.” (BSAE)
- (3) The majority of respondents (72.7 percent) said **that/** \emptyset the pension grant is not enough to cater for all their household needs, and they hope that it will be increased by R400. (BSAE)

This chapter is concerned with the undeletion of the *that*-complementiser in postverbal complementation (as illustrated in (3)) and to a lesser extent with the insertion of the *that*-complementiser when using direct speech (as illustrated in (2)) that has been reported in studies of BSAE (cf. Makalela, 2013; Mesthrie, 2006).

⁹⁷ Not all L2 varieties exhibit the tendency to be as explicit as possible. Mesthrie (2006) posits that new varieties of English can be placed on a continuum with the notion DELETION on one end and the notion UNDELETION on the other. African varieties of English, such as BSAE will be situated more towards the *undeletion* pole, whereas Singaporean English will be situated more towards the *deletion* pole.

In order to describe the non-deletion of the *that*-complementiser, an overview is provided of the notion **OPTIONALITY** in Section 5.2. Section 5.3 provides an overview of the factors driving *that*-omission, while the results of the analysis are conveyed in Section 5.4.

5.2 Theoretical overview of optionality

In the literature, scholars are increasingly investigating the notion of **OPTIONALITY** in grammar. In this study, **OPTIONALITY** is taken to refer to those circumstances in which the presence and absence of a linguistic item are in free variation (cf. McGregor, 2013:1150-1151). Linguists operate on the assumption that the variants of optionality are exact synonyms, i.e. that the coded variant has exactly the same meaning as the non-coded variant (Biber *et al.*, 1999:680-683; McGregor, 2013:1151). For example, the presence or the absence of the *that*-complementiser in (4) and (4a) below does not have an effect on the experiential meaning^{98,99} of the utterance:

- (4) Mahlangu, who also started his amateur boxing career in the area, **believes** [∅ there is a need for such a boxing club in Boipatong]. (BSAE)
- (4a) Mahlangu, who also started his amateur boxing career in the area, **believes** [**that** there is a need for such a boxing club in Boipatong].

Instead, the absence of the linguistic sign contrasts semantically with its presence. McGregor (1997:380, 2013:1156-1157) explains this idea against the background of the notion of **JOINT ATTENTION FRAMES** used in psychology. This term is used to refer to situations in which an individual alerts another individual to the presence of an object by pointing to it, gazing at it, or speaking about it. In linguistics, scholars such as Tomasello (2003) and Tomasello and Farrar (1986) have observed that infants acquire grammatical

⁹⁸ The notion of **EXPERIENTIAL MEANING** is used in line with Halliday and Matthiessen's (2014:30) understanding of the concept. Experiential meaning is said to be representative of the language users' experiences in real life.

⁹⁹ There is a point to be made, though, that the omission or retention of the *that*-complementiser may have an influence on the formality of the text, and by extension the social and pragmatic meaning. In this regard, refer to the discussion of Gries's (2003a:139-140) argument in Section 4.1.

constructions by means of joint attention frames. McGregor (1997:380, 2013:1156-1157) argues that in the case of optionality in grammar, the meaning of the linguistic sign is determined by means of its presence or absence. These meanings are illustrated in (5):

- (5) use [± prominent]
 non-use [± backgrounded]

(McGregor, 2013:1157)

When a linguistic sign is used, the language user makes it prominent by drawing attention to it. In other words, the language user brings the linguistic sign “into relief against the other material in the joint-attentional frame” (McGregor, 2013:1157). Non-usage, on the other hand, backgrounds the linguistic sign by taking “it outside of the domain of what is being (or will be) concentrated on; the meaning is shifted outside of the joint-attentional frame” (McGregor, 2013:1157). This is illustrated in Figure 5.1.

The options of using a given linguistic sign or not raise the question of when and why speakers would choose to use the linguistic sign – consequently making it more prominent – and when and why they would choose to omit the linguistic sign – consequently backgrounding it (cf. Hilpert, 2008:395).

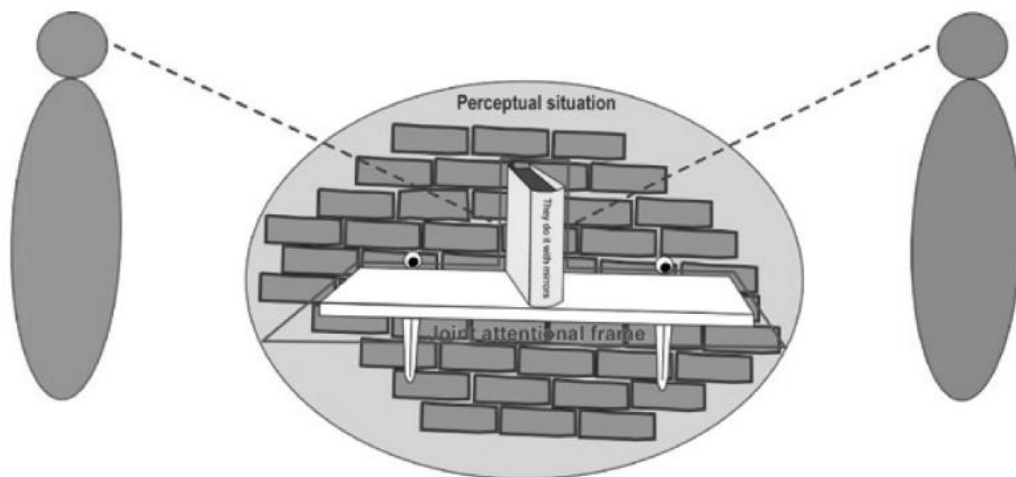


Figure 5.1: *Highlighting and greying in the joint attention frame (McGregor, 2013:1158)*

In the case of BSAE, Makalela (2013:103) and Mesthrie (2006:119) have reported the undeletion of the *that*-complementiser. They attribute these features of the *that*-complementiser in BSAE primarily to a strong substrate influence. In addition to the non-deletion of the *that*-complementiser, Makalela (2013:102-103) observes that the BSAE speakers substitute the *that*-complementiser with the *gore*-complementiser borrowed from Sepedi, one of the indigenous African languages. He argues that the non-deletion of the *that*-complementiser together with the lexical borrowing from Sepedi confirm the strong substrate influence. He also maintains that the lexical borrowing of the *gore*-complementiser may be a result of the informal nature of the language used in radio talk shows (Makalela, 2013:103).

This chapter is concerned with investigating the non-deletion of the *that*-complementiser in BSAE. Whereas Makalela (2013) and Mesthrie (2006) investigated the feature in spoken corpora, this study investigates the feature in a corpus of written texts comprised of several registers.

However, before investigating this feature, Section 5.3 provides an overview of the factors that have been found to drive *that*-omission in standard varieties of English.

5.3 Factors driving *that*-omission

In the literature, scholars identify several factors driving the retention or omission of the *that*-complementiser. The most salient factors are described in the paragraphs below.

The **frequency of the verb lemma** is one of the factors believed to play a role in the retention and omission of the *that*-complementiser. Scholars such as Tagliamonte and Smith (2005:291) and Thompson and Mulac (1991:244) have observed that highly frequent verbs such as *think*, *say*, and *tell* are also the verbs with the highest *that*-omission rates. This can be attributed to the fact that the highly frequent verbs have become semantically coherent with the constructions in which the verb controls a dependent finite complement clause. The high semantic coherence, together with the high type frequency, results in constructions with high degrees of productivity (cf. Section 2.3.1.2). The higher degrees of productivity, in turn, result in the schematisation of construction,

making it easier to process cognitively and consequently easier to omit the *that*-complementiser.

Register has also been determined to be a factor that drives *that*-omission. More formal registers (e.g. academic registers) are said to have a higher rate of *that*-retention, whereas less formal registers (e.g. spoken registers) have a higher rate of *that*-omission (Biber *et al.*, 1998:71-76; Biber *et al.*, 1999:680-683; Kearns, 2007:296-297). According to Biber (1999:144-145), the high *that*-retention ratio in academic registers can be attributed to “careful production circumstances; an expository, informational purpose; and a formal tone”. The careful production circumstances characteristic of formal registers, such as academic registers, can consequently be associated with increased processing strain. In this regard, Rohdenburg (1996:149) put forward the complexity principle which states that “more explicit grammatical alternatives tend to be preferred in cognitively more complex environments”. In a similar vein, Mondorf (2014:212) and Jaeger (2005:107) posit that analytic variants (such as the retention of the *that*-complementiser) are favoured in cognitively complex environments that exert additional processing load. This variable is encoded as REGISTER in this study and is coded for the following registers: academic (‘A’), instructional (‘I’), popular (‘P’) and reportage (‘R’).

Another factor said to influence *that*-omission is **verb semantics**. Previous studies have indicated that epistemic verbs (such as *think*, *believe*, and *know*) typically take zero complementisers. This is attributed to the fact that epistemic verbs are often used as epistemic modifiers of the embedded clause, as opposed to being an expression of the main proposition. Thompson and Mulac (1991:249) explain this phenomenon as follows:

As this happens, the distinction between ‘main’ and ‘complement’ clause is being eroded, as suggested by Emonds and Underhill, with the omission of *that* a strong concomitant. That is, the more the ‘main’ subject and verb are taken as an epistemic phrase, the less the ‘complement’ is taken as a ‘complement’, and the less likely is the complementizer that to be used.

Similarly, Boye and Harder (2007:597-598) argue that the simple independent clause has become a fixed routine that obtained an adverbial function, as illustrated in (6):

(6) I **believe** [Ø everything is better with wine]. (WSAE)

In the example, the thematic prominence has moved from the independent clause to the dependent clause, and this thematic prominence is emphasised by the lack of an overt *that*-complementiser. Since the dependent clause realises the asserted proposition, the independent clause is processed cognitively as an adverbial function and not as an independent matrix clause. (Boye & Harder, 2007:597-598; Dor, 2005:348-349). Scholars point out that this phenomenon is only possible with verbs of an epistemic nature.

Consequently, verb semantics are also considered a driving factor in the retention and omission of the *that*-complementiser. The variable is encoded as VERBSEMANTICS in this study. The following values are used in the semantic classification: (i) verbs belonging to the domain of 'activity', e.g. *show* and *indicate*, (ii) verbs belonging to the domain of 'cognitive activity', e.g. *think* and *learn*; (iii) verbs belonging to the domain of 'cognitive state', e.g. *believe* and *know*; (iv) verbs belonging to the domain of 'emotional/attitudinal state', e.g. *feel*, verbs belonging to the domain 'perception', e.g. *see*; and verbs belonging to the domain 'communication', e.g. *say* and *tell*.

On a syntactic level, syntactic complexity is also believed to be a driving factor in the retention and omission of the *that*-complementiser. Scholars have observed that the more complex the matrix clause is in terms of syntactic complexity, the more likely it is that the *that*-complementiser will be retained. Several factors are considered when determining the level of syntactic complexity of the matrix clause. The first is **the syntactic realisation of the subject of the matrix clause** (Tagliamonte & Smith, 2005:294-295; Thompson & Mulac, 1991:242-243). If the subject of the matrix clause is realised by a pronoun, it is more likely that the zero complementiser will be used (cf. (7)). If the subject of the matrix clause is realised by a noun phrase, though, it becomes more likely that the *that*-complementiser will be retained (cf. (8)). This variable is encoded as SUBJMATRIXSYNTAX in this study and are coded for the following: 'pronoun' and 'NP'.

(7) He **said** [Ø the majority of the poor have the view that they are not been given attention by the municipality]. (BSAE)

- (8) US research findings indicate [**that** increasing individual engagement increases performance by 20% and reduces intention to leave by 87%]. (BSAE)

In addition to the syntactic realisation of the subject of the matrix clause, scholars have also determined that **number** and **animacy of the subject** (cf. (9)) in the matrix clause, **voice** (cf. (10)), and **the form of the verb** (cf. (11)) all contribute to the syntactic complexity of the clause (Tagliamonte & Smith, 2005:295-296; Torres Cacoullos & Walker, 2009:26).

- (9) Furthermore, the demographic analysis of questionnaires also **showed** [**that** 71.6% (n = 53) SMTs were from primary schools while those from secondary schools were 28.4 % (n = 21)]. (BSAE)
- (10) During the June 2003, the Growth and Development Summit it was agreed [**that** the EPWP should be implemented to provide and/or improve the basic and essential infrastructure in the communities with an objective to utilise public sector budgets to reduce and alleviate unemployment]. (BSAE)
- (11) Calvin was not interested in some abstract definition of the church's essence, but he was interested in **showing** [**that** true spirituality and faithfulness as seen in the "notae" are the evidence of a true church]. (BSAE)

The number and animacy of the subject in the matrix clause are encoded as SUBJMATRIXNUMBER (with the values 'singular' and 'plural') and SUBJMATRIXANIM (with the values 'animate', 'inanimate', and 'n/a') respectively. Voice is represented by the variable VOICE, and it is coded with the values 'active' and 'passive', while the **form of the verb** is represented by the variable VERBINFLECT, and it is coded with the values 'present', 'past', 'modal', 'infinitive', and 'ing-participial'.

The aforementioned variables are said to determine the use of the zero *that*-complementiser. They are listed in Table 5.7 on p. 211.

5.4 Results

The general results of the investigation are provided in Section 5.4.1. In order to gain better insight into the variables discussed in Section 5.3 above, the results of the conditional inference tree and the random forest are provided in Section 5.4.2 and Section 5.4.3 respectively.

5.4.1 General results

As mentioned earlier in Chapter 3, given the fact that the data is not distributed normally, descriptive statistics are reported by means of a boxplot. The median values are regarded as a measure of central tendency, while the interquartile ranges are regarded as a measure of dispersion. Since the normal distribution of values is assumed for parametric tests, significance tests are done by means of non-parametric tests.

The *that*-complementiser is omitted more frequently in the WSAE corpus (with a median value of 0.58) than in the BSAE corpus (with a median value of 0.15). This difference is illustrated in Figure 5.2. A Kruskal-Wallis rank test shows that the difference between the two varieties is significant (Kruskal-Wallis chi-squared = 54.537, $df = 1$, $p = 1.525e^{-13}$). A Wilcoxon rank test confirms this finding ($W = 323\ 700$, $p = 1.526e^{-13}$).

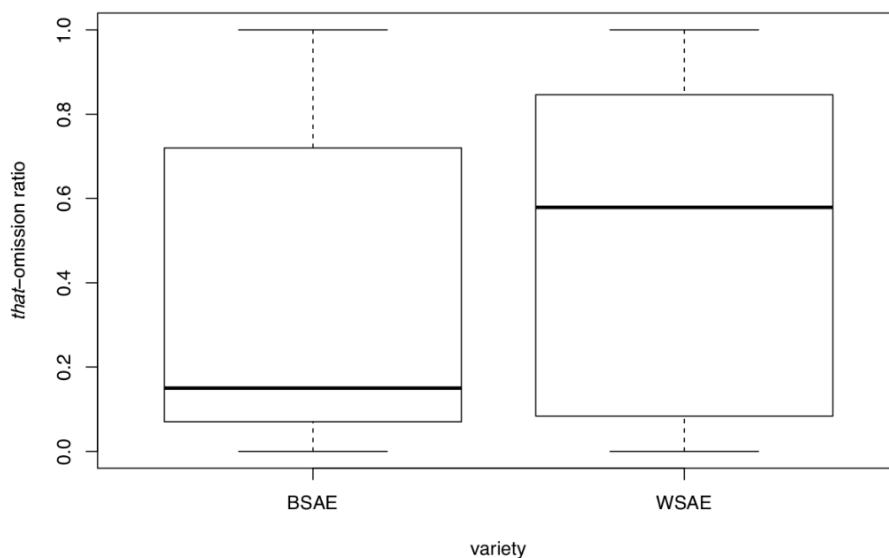


Figure 5.2: *That-omission ratio for BSAE and WSAE*

A breakdown of the proportional frequency of the omission of the *that*-complementiser across registers between BSAE and WSAE is illustrated in Figure 5.3. (p. 206). The differences in the distribution of *that*-omission were calculated by means of a chi-squared (χ^2) test to determine whether the differences in the omission of the *that*-complementiser are significant between the two corpora for each of the registers. An overview of these results is provided in Table 5.1. The differences in distribution were significant for all the registers but the academic register.

Table 5.1: An overview of the percentage of omission of the *that*-complementiser in BSAE and WSAE across registers

	BSAE	WSAE	χ^2
Academic	11	8	0.64 (ns)
Instructional	13	57	31.23 ****
Popular	28	76	32.07 ****
Reportage	86	55	44.84 ****
Total	24	50	133.05 ****

The results show that the BSAE speakers and the WSAE speakers follow the same conventions in the academic register as far as the omission of the *that*-complementiser is concerned. Both groups have a very small rate of *that*-omission which is characteristic of the academic register (Biber, 1999:144; Biber *et al.*, 1999:680; Kearns, 2007:296-297).

However, in the instructional and popular registers, the BSAE users prefer the overt forms of the *that*-complementiser in contrast to the WSAE users' preference for the zero *that*-complementiser. The difference in the instructional register is significant at $p = 2.30 \times 10^{-8}$, while the difference in the popular register is significant at $p = 1.49 \times 10^{-8}$. In this respect, though, it is important to point out that it was difficult to source texts written by BSAE authors in these registers. Since the BSAE users do not often write in these registers, their unfamiliarity with the registers may have led them to rely on overly formal register conventions. This unfamiliarity together with the accompanying cognitive strain may account for the high rate of overt *that*-complementisers in this corpus.

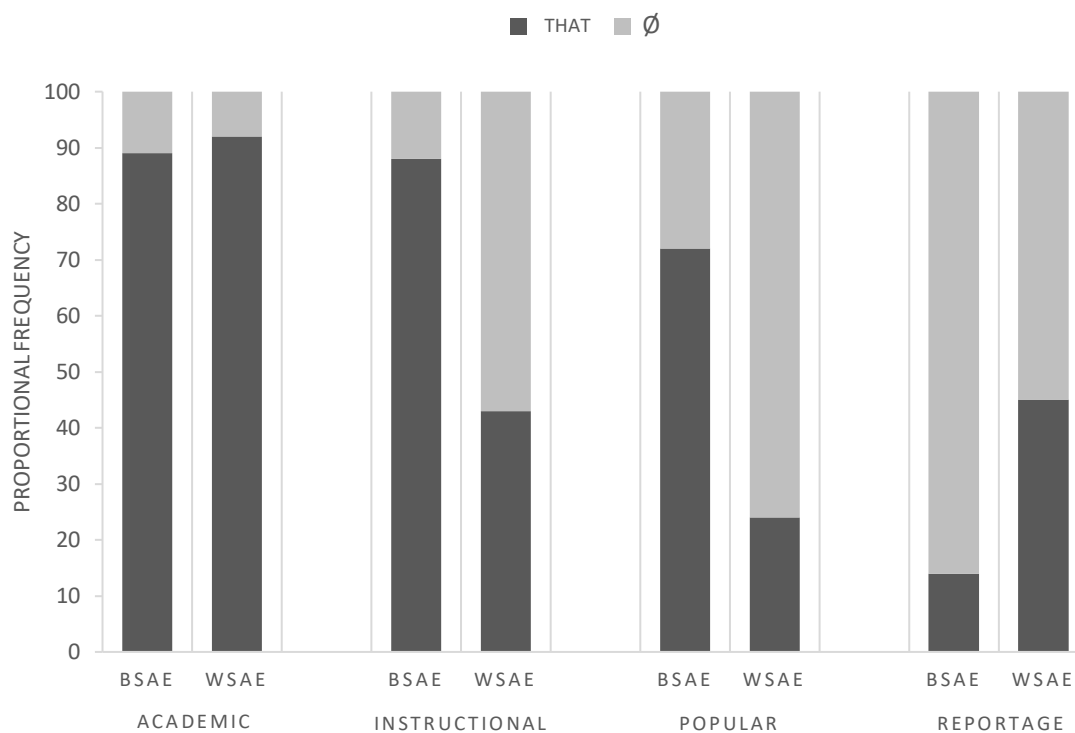


Figure 5.3: An overview of the proportional frequencies of the overt *that*-complementisers and the zero complementisers in WSAE and BSAE across four registers

Finally, in the reportage register, both the BSAE and WSAE authors have a higher rate of *that*-omission than *that*-retention. This is contrary to findings reported in the literature (Biber, 1999:144-145; Biber *et al.*, 1999:680). Kruger and Van Rooy (2016b:128) find that Afrikaans is characterised by a pragmatic preference for complementiser omission, especially in the reportage register. They attribute the higher rate of *that*-omission in the South African varieties of English to the close contact between South African Englishes and Afrikaans. The close contact between English and Afrikaans is compounded further by the use of the Afrikaans style guide, *Skryf Afrikaans van A-Z* (Müller & Pistor, 2011) in News24 (the largest English and Afrikaans news publishing company in South Africa). Since there is no comparable style guide in English, it may be possible that the Afrikaans publishing norm of omitting the *that*-complementiser after the verb *sê* (*say* in English) have been transferred through CLI to the reportage register in English (cf. Müller & Pistor, 2011:708).

In this regard, it would be prudent to point out that the verb *say* is the verb in both BSAE and WSAE that takes a finite complement clause most often. Of the 391 occurrences of the

lemma *say* with such complement clauses in the BSAE corpus, 44 take an overt complementiser, while 230 take a zero complementiser. Of the 339 occurrences of the lemma *say* in the WSAE corpus, 52 take an overt complementiser, while 111 take a zero complementiser. The difference in terms of *that*-omission is significant with $\chi^2 = 14.06$ ($p = 0.0001$). This information is conveyed in Table 5.2:

Table 5.2: *The raw frequencies of the complementation patterns of the verb say*

Type of complementation	BSAE	WSAE
<i>that</i> -complementiser	44	52
zero-complementiser	230	111
other kinds of complementation	117	176
Total	391	339

Thus far, it has become clear that BSAE is not characterised by non-deletion to the extent that the results from spoken corpora in previous studies may have led us to believe. Instead, the results from this study indicate that the BSAE users omit the *that*-complementiser more frequently than the WSAE users in the registers in which they are experienced writers (i.e. the academic and reportage registers). This brings into question the substrate influence that has been posited by Mesthrie (2006:119) and Makalela (2013:103). Instead, it may simply be that while the BSAE speakers acquire English as a second language, they expend more cognitive effort to produce English utterances, and therefore tend to use the more analytical linguistic options (cf. Mondorf, 2014:212; Rohdenburg, 1996:149). The more proficient the BSAE users become in English, the less they have to rely on the more analytical linguistic options.

The aligned edited concordances of the BSAE corpus are analysed to determine whether the editors (as normative gatekeepers) accept the BSAE speakers' rates of *that*-omission. From the analysis, it seems that in those cases where editors did make any changes, the overt *that*-complementiser was added (cf. Table 5.3). In none of the cases did the editors remove an overt *that*-complementiser. Here, it has to be noted that Kruger and Van Rooy (2016b:128) point out that the editors experience their own cognitive strain during the editing process and that this cognitive strain may account for the addition of the overt

complementisers. These changes are minimal, though, which point to the acceptance of the BSAE users' rates of *that*-omission.

Table 5.3: Overview of changes made in terms of the *that*-complementiser by the editor in the BSAE corpus

Register	Overt <i>that</i> - complementiser added	Overt <i>that</i> - complementiser removed	Other changes ¹⁰⁰
Academic	5	0	3
Instructional	0	0	0
Popular	2	0	0
Reportage	9	0	4
Total	16	0	7

Table 5.4 demonstrates that the changes made by the editors are not significant and provides an overview of the percentage of omission of the *that*-complementiser in the original BSAE texts and the edited BSAE texts across registers. In all the instances, the changes made by the editors are not significant with or $p > 0.05$ and mostly approaching the value $p = 1$.

Table 5.4: An overview of the percentage of omission of the *that*-complementiser in the original BSAE texts and the edited BSAE across registers

	BSAE (original)	BSAE (edited)	χ^2
Academic	11	10	0.00 (ns)
Instructional	13	13	0.00 (ns)
Popular	28	27	0.00 (ns)
Reportage	86	82	0.84 (ns)
Total	24	23	0.02 (ns)

¹⁰⁰ In some instances, editors made changes to the structure of the entire sentence by replacing the finite dependent *that*-clauses with non-finite dependent *to*-clauses or sometimes even noun phrases. These changes account for the slightly lower total of *that*-complementisers in the edited version of the corpus than the original version of the corpus.

Mesthrie (2006) addresses not only the non-deletion of the *that*-complementiser but also the undeletion of the *that*-complementiser in instances where direct speech is quoted.¹⁰¹ The results for the undeletion of the *that*-complementiser in those instances where direct speech is quoted are reported in Table 5.5, and examples are quoted in (12) and (13).

Table 5.5: Normalised frequencies of the non-deletion of the *that*-complementiser in the direct speech construction of BSAE and WSAE

Register	BSAE		WSAE	
	that	∅	that	∅
Academic	22	25	0	0
Instructional	0	0	0	0
Popular	4	5	0	2
Reportage	0	5	0	3
Total	26	36	0	5

- (12) Parents **said that**: “We are ill and do not know what to do. Illiteracy is killing us, so we cannot read any pamphlet about this HIV/AIDS on how to deal with the issues related to HIV and that complicates the issues.” (BSAE)
- (13) The programme record **shows that**: “The churches are paying school fees to at least 10 primary school children at \$10,00 per term for each child (i.e. \$300 per year). Stationery such as exercise books, pens and pencils have also been distributed to these needy children.” (BSAE)

The aligned concordance lines reveals that of the 36 instances of *undeletion* detected in the BSAE corpus, only ten of the overt complementisers were removed by the editors. Once again, this suggests that this feature is accepted to a certain extent by the WSAE speakers.

Mesthrie (2006:119) hypothesises that the undeletion may, in the first instance, most likely be attributed to the substrate influence, and in the second instance, may be the

¹⁰¹ Grammars of the standard varieties of English report direct speech without an overt *that*-complementiser (cf. Chapter 1).

result of analogy or overgeneralisation that takes place while the second language is acquired and speakers conflate direct and indirect speech. When the undeletion of the *that*-complementiser in the BSAE academic register is considered in conjunction with the high rate of *that*-omission in the reportage register, though, it seems less likely that the undeletion of the *that*-complementiser is the result of the substrate influence or overgeneralisation that takes place while the second language is acquired. These points, together with the fact that the editors did not remove many of the non-deleted *that*-complementisers, rather seem to suggest that the speakers use the analytical form due to the cognitive constraints associated with the academic register. The fact that the editors did not remove these non-deleted *that*-complementisers may provide tentative support for this hypothesis.

Table 5.6: Raw frequencies of the editorial intervention in the non-deletion of the *that*-complementiser in the direct speech construction in BSAE

Register	Undeleted <i>that</i>	Undeleted <i>that</i> -complementiser removed
Academic	30	7
Instructional	0	0
Popular	6	3
Reportage	0	0
Total	36	10

5.4.2 Conditional inference tree

A conditional inference tree was drawn (cf. Figure 5.4) to obtain an overall picture of the variables that may have an influence on language users' rates of *that*-omission. It was drawn using the `ctree()` algorithm in the `party` package in R (Hothorn *et al.*, 2015). The tree is created by means of a number of recursive decisions. These decisions are depicted hierarchically in the tree. The aim of the tree is to split the dataset into smaller groups of observation points to determine which independent variables predict the known outcome of the dependent variable best. The variables used to draw the conditional inference tree is discussed earlier in Section 5.3, and they are listed in Table 5.7. All the possible splits are significant at $p < 0.05$. The conditional inference tree returned a high prediction accuracy of 89%.

Table 5.7: *The hypothesised variables conditioning finite that-omission*

Variable	Description	Value
VARIETY	Indicates variety	'BSAE', 'WSAE'
LEMMA	Indicates the verb lemma	Verb lemma e.g. <i>say</i>
REGISTER	Indicates register	'A' = academic; 'I' = instructional 'P' = popular; 'R' = reportage
VERBSEMANTICS	Indicates the semantic domain to which the verb belongs	'activity', 'cognitive activity', 'cognitive state', 'communication', 'emotional/attitudinal state', 'perception'
SUBJMATRIXSYNTAX	Indicates the syntactic realisation of the subject of the matrix clause	'pronoun', 'NP'
SUBJMATRIXNUMBER	Indicates whether the subject of the matrix clause	'singular', 'plural'
SUBJMATRIXANIM	Indicates how alive or sentient the subject is	'animate', 'inanimate'
VERBINFLECT	Indicates whether the verb is simple or inflected	'present', 'past', 'modal', 'infinitive', 'ing-participial'
VOICE	Indicates whether the construction occurs in the active or passive form	'active', 'passive'

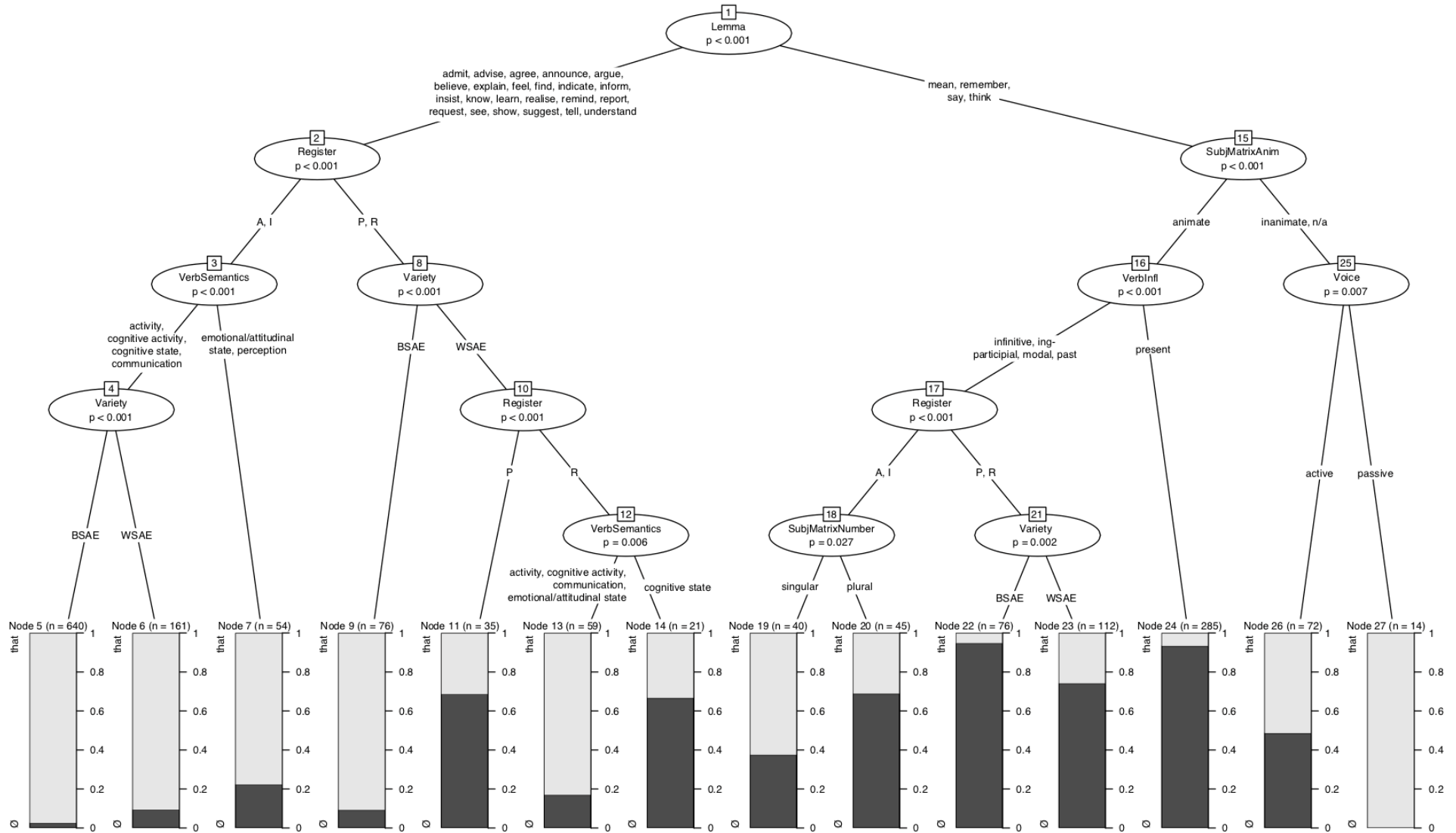


Figure 5.4: A conditional inference tree indicating the factors determining the retention and omission of the *that*-complementiser

The first split at the top of the tree points to the variable LEMMA being the first determinant with a probability of $p < 0.001$ (Node 1). The verbs that prefer overt *that*-complementation are found on the branch on the left-hand side. These verbs are listed in (14):

- (14) *admit, advise, agree, announce, argue, believe, explain, feel, find, indicate, inform, insist, know, learn, realise, remind, report, request, see, show, suggest, tell, and understand*

On the other hand, the verbs that prefer zero complementation are found on the branch on the right-hand side. These verbs are listed in (15):

- (15) *mean, remember, say, think*

On the left-hand side of the tree, the statistical model indicates that the subgroup of verbs listed in (14) are split on the basis of register (Node 2). The branch that splits to the left from Node 2 is dedicated to the academic and instructional registers. At Node 3, the academic and instructional registers are split on the basis of verb semantics. The branch that splits towards the left, take all the activity (e.g. *show*), cognitive activity (e.g. *learn*), cognitive state (e.g. *know*), and communication (e.g. *report*) verbs, whereas the right-hand branch takes all the emotional/attitudinal state (e.g. *believe*), and perception verbs (e.g. *see*). The verbs on the right-hand branch, have an almost 80% likelihood of taking the overt *that*-complementiser (cf. Node 7). This is illustrated in (16):

- (16) Employees prefer meaningful work; they want to **feel [that they are adding value and their contribution means something]**. (BSAE)

The verbs on the left-hand branch, though, is further split on the basis of variety (cf. Node 4). The BSAE users are 97.5% likely to use the overt *that*-complementiser with activity, cognitive activity, cognitive state, and communication verbs in academic and informational registers, whereas the WSAE users are approximately 6.8% more likely to use the zero complementiser than the BSAE users with verbs belonging to the same semantic category. This is illustrated in (17) and (18) below:

- (17) She **explained** [**that** by calling for an “African” leadership, these African leadership thinkers imply Africa is homogeneous rather than multicultural]... (BSAE)
- (18) CAST testing **suggested** [Ø she was not sensitised to the beta-lactam ring, but rather specific side-chains]... (WSAE)

The path from Node 1 through to Node 7 in the decision tree supports the findings in the literature that language users prefer to use the overt *that*-complementiser in the more formal registers, but also the earlier findings in Section 5.4.1 that both BSAE users and WSAE users prefer to use the overt *that*-complementiser in the academic register. It is interesting to note, though, that the BSAE speakers are almost as likely to use the zero complementiser as the WSAE users since previous studies on this feature in BSAE have found that BSAE speakers tend to not delete the *that*-complementiser.

Whereas the academic and instructional registers are split on the grounds of verb semantics, the popular and reportage registers are split on the grounds of variety (cf. Node 8). Here, the BSAE users are approximately 90.8% likely to use an overt *that*-complementiser (cf. Node 9). The WSAE users, on the other hand, make a further distinction on the grounds of register. In the popular register, there is a 68.6% likelihood that the WSAE users will use the zero complementiser (cf. Node 11). This is illustrated in (19):

- (19) Some of the community **feels** [**that** government is not doing anything to substrate the mayhem that has covered their area]. (BSAE)

In the reportage register, on the other hand, the WSAE users are 83.1% likely to use the overt *that*-complementiser with activity, cognitive activity, communication, and emotional/attitudinal state verbs (cf. Node 12 and Node 13). Cognitive state verbs though have a 66.7% likelihood of being used with the zero complementiser (cf. Node 12 and Node 14). This is illustrated in (20):

- (20) In Russia, Boris Yeltsin and a few robber barons appear to **believe** [Ø they can steal the June 16 election as well as everything else not nailed down]. (WSAE)

As mentioned earlier, the verbs on the right-hand side of the tree are more likely to take zero complementisers. It is interesting to note that two of the four verbs, i.e. *mean* and *think*, on the right-hand side of the decision tree, are epistemic verbs. This is in line with the findings of Tagliamonte and Smith (2005) and Thompson and Mulac (1991) that epistemic verbs are more likely to take zero complementisers. This is illustrated in (21) and (22) below:

- (21) Do you **think** [Ø the increase in fraud and corruption can affect your employment?] (BSAE)
- (22) Tandem flights **mean** [Ø you're strapped tightly to an experienced instructor so you can sit back and relax]. (WSAE)

The verb *say* is also situated on this branch. As mentioned earlier, the more frequently a verb occurs that can take finite complementation, the more likely it is that the verb will take a zero complementiser (Tagliamonte & Smith, 2005; Thompson & Mulac, 1991).

The data that are grouped together at the first split on the right-hand side of the decision tree (with the verbs *mean*, *remember*, *say*, and *think*) are grouped on the basis of the animacy of the subject in the matrix clause (cf. Node 15). Those verbs that take an animate subject are further split on the grounds of the inflections that the verb takes (cf. Node 16). The verbs that do not take any kind of inflection have a 93.3% likelihood of taking a zero complementiser (cf. Node 24). This is exemplified in (23):

- (23) Stranded community members **say** [Ø they have been waiting years for RDP houses]. (WSAE)

The verbs that do take inflections or are marked for modality, though, are further split on the grounds of register (cf. Node 17). Those verbs that occur in the academic and instructional registers are finally split on the grounds of the number of the subject in the

matrix clause (cf. Node 18). Those subjects that are singular have a 62.5% likelihood of taking an overt *that*-complementiser (cf. Node 19). This is exemplified in (24):

- (24) The deputy chief executive of the National Prosecuting Authority **said** [**that** it has been an embarrassment for everyone]. (BSAE)

However, should the subject be plural, the verbs have a 68.9% likelihood of taking a zero complementiser (cf. Node 20). This is illustrated in (25):

- (25) Some finance officers **said** [Ø the reporting on finances was done by them]. (BSAE)

The verbs that occur in the popular and reportage registers are finally split on the grounds of variety (cf. Node 21). If the verbs are used in BSAE, there is a 94.7% likelihood that the zero complementiser will be used (cf. Node 22). If the verbs are used in WSAE, there is an approximately 74.1% likelihood that the zero complementiser will be used (cf. Node 23). This is exemplified in (26) and (27):

- (26) She **said** [Ø residents had to pay what they consumed or use]. (BSAE)

- (27) He **said** [Ø although the universities and technikons fell within the jurisdiction of Ministry of Education, Bengu often elicited the support of the MECs in certain matters]. (WSAE)

If one returns to Node 15 and look at the verbs that take an inanimate or any other kind of subject, the next split is made on the grounds of whether the verb is used in the active or the passive voice (cf. Node 25). The branch on the left-hand side splits into Node 26. This branch is indicative of the verbs that are used in the active voice. These verbs have a 51.4% likelihood of taking an overt *that*-complementiser. The branch on the right-hand side splits into Node 27, and are indicative of the verbs that are used in the passive voice. These verbs have a 100% likelihood of taking the overt *that* complementiser. This is illustrated in (28):

- (28) It **is said** [**that** Fish Hoek with its swell and wind direction, has the best downwind paddling in the world]. (WSAE).

Overall, the decision tree highlights the importance of the verb lemma, register, and verb semantics. It confirms that the overt *that*-complementiser is most likely to occur in the academic and instructional registers, while the zero complementiser is most likely to occur in the popular and reportage registers. These findings confirm the findings of the general analysis.

What the general analysis did not point out, though, is the role of verb semantics. The use of activity, cognitive activity, and communication verbs are more likely to elicit the overt *that*-complementiser. In the academic registers, the emotional/attitudinal state and perception verbs elicit a slightly higher rate of *that*-omission. In the reportage registers, though, the verbs of emotional/attitudinal state are more likely to take the *that*-complementiser whereas the cognitive state verbs are more likely to elicit zero complementisers. This links to the idea that when epistemic verbs are used, the *that*-complementiser is sometimes removed to emphasise the move of thematic prominence from the independent clause to the dependent clause (cf. Section 5.3).

On the right-hand side of the decision tree, the factors that contribute to syntactic complexity seem to play an important role. Here, the inflections and modals that the verbs take, together with the characteristics of the subject of the matrix clause, determine the rate of *that*-omission. If the verb lemma is marked for modality or by means of inflection, the rate of *that*-retention rises slightly. Also, animate subjects in the matrix clause seem to result in higher rates of *that*-omission whereas inanimate subjects in the matrix clause seem to result in higher lower rates of *that*-omission.

However, variety did not play as big a role as expected. In those instances where variety is deemed to have an influence, the differences between the BSAE and WSAE are not in direct contrast with one another but are rather indicative of differences in the degree to which the overt *that*-complementiser or the zero complementiser is realised in particular contexts.

Sometimes, highly predictive interactions between variables are not included in conditional inference trees. Therefore, a random forest is considered in Section 5.4.3 in addition to the decision tree.

5.4.3 Random forest

In this section, a random forest is drawn to complement and confirm the findings of the inference tree in Section 5.4.2. The random forest is visualised in Figure 5.5. It was drawn using the `cforest()` algorithm in the `party` package in R (Strobl *et al.*, 2009a). The classification accuracy of the random-forest approach is slightly higher than that of the conditional inference tree at 90%.

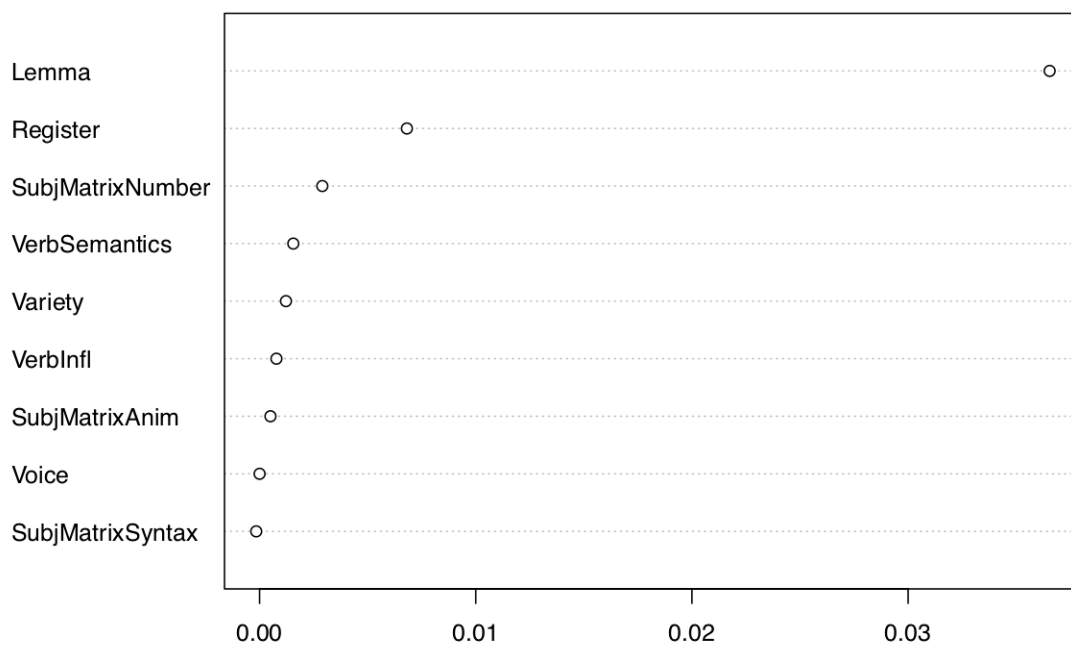


Figure 5.5: Conditional importance of the variables determining the retention and omission of the *that*-complementiser

The variable importance scores in the random forest demonstrate that the variable LEMMA (0.037) is the most important predictor, followed by REGISTER (0.007), SUBJMATRIXNUMBER (0.003), VerbSemantics (0.002), Variety (0.001), SubjMatrixAnim (0.001), Voice (0.000), and SubjMatrixSyntax (0.000).

Here it is interesting to note that the variable REGISTER, has the highest predictive value after the variable LEMMA. The implication is that the rest of the variables, i.e. the variables that primarily influence syntactic complexity, do not have as significant an effect on the cognitive processing load as one would expect.

Furthermore, the random forest confirms that variety has a low predictive value.

5.5 Conclusion

The results in this chapter indicate that in some regards, BSAE demonstrate some of the features characteristic of L2 English, but in others demonstrate some of the features that are characteristic of L1 English.

In the academic register, the BSAE users conform to the L1 conventions in terms of the use of the overt *that*-complementiser. In this register, they omit the *that*-complementiser slightly more often than the WSAE users but the difference is not statistically significant. The general picture in the academic register is similar, with formal conventions guiding the choice of BSAE and WSAE users alike. However, the BSAE users sometimes do not delete the *that*-complementiser when employing direct speech in the academic register. This could be a form of hyperclarity.

In terms of the instructional and popular registers, the BSAE users still seem to retain the *that*-complementiser. This is in line with the findings of other studies conducted in L2 varieties of English where scholars found that BSAE favours features of hyperclarity. These include: Mesthrie's (2006) observation that the BSAE speakers demonstrate a preference for the non-deletion of the *that*-complementiser in spoken forms of BSAE; Kruger and Van Rooy's (2016b) observation that L2 language users prefer using the explicit *that*-complementiser in reported speech; Van Rooy *et al.*'s (2010) finding that East African English speakers exhibit a preference for more explicit referencing, explicit constructions, and increased repetition; and Kruger and Van Rooy's (2016a) finding that the L2 varieties demonstrate a preference for clausal elaboration and the overt marking of relationships between properties.

At this point, the notion of JOINT ATTENTION FRAMES becomes important. Earlier, in Section 5.2, it was stated that when the *that*-complementiser is used, it is brought into relief against the other material in the joint-attentional frame. When the *that*-complementiser is omitted though, it is removed from the joint-attentional frame. Since the BSAE users are less experienced in writing instructional and popular registers in English, they are less familiar with the expectations that the readers of these registers have. In an attempt to accommodate the readers and to make sure that their message is conveyed, the BSAE users bring the *that*-complementiser into the joint-attentional frame.

Another explanation could be that the BSAE users' unfamiliarity with the conventions of the instructional and popular registers results in increased levels of processing effort, which in turn results in the BSAE users producing more analytical utterances that are characterised by features of hyperclarity (cf. Mondorf, 2014:212; Rohdenburg, 1996:149).

It is interesting to note, though that the BSAE users use the zero complementiser significantly more often than the WSAE users in the reportage register. Previous research in L1 varieties of English have demonstrated that the *that*-complementiser tends to be retained in reportage registers (cf. Biber, 1999; Biber *et al.*, 1999). A possible explanation for this phenomenon is language contact with Afrikaans, and more specifically the publishing conventions of Afrikaans that are followed in the dissemination of information in the media. In this regard, Kruger and Van Rooy (2016b:128) find that Afrikaans is characterised by a preference for complementiser omission, especially in the reportage register. This tendency is formalised in the Afrikaans style guide, *Skryf Afrikaans van A-Z* (cf. Müller & Pistor, 2011:708), which is used by News24 (the largest English and Afrikaans news publishing company in South Africa). Since there is no comparable style guide in English, it is believed that the Afrikaans publishing norm of omitting the *that*-complementiser after the verb *sê* (*say* in English) has found a foothold in the reportage register in English.

The results of the conditional inference tree and the random forest confirm that variety is not a contributing factor amongst the factors that drive *that*-omission. As was the case with the BSAE users' use of the ditransitive construction, it is found that the factors

driving *that*-omission in the L1 varieties of English, also drive *that*-omission in the L2 varieties of English.

The findings in the chapter make it clear that BSAE is not characterised by non-deletion to the extent that the results from spoken corpora in previous studies may have led us to believe. Instead, the BSAE speakers demonstrate sensitivity to the conventions of the registers in which they are comfortable. In fact, the BSAE users omit the *that*-complementiser even more frequently than the WSAE users in the academic and reportage registers. These findings bring into question the substrate influence that has been posited by Mesthrie (2006:119) and Makalela (2013:103). Instead, the non-deletion of the *that*-complementiser could possibly be the result of increased processing effort.

The findings of this chapter can be described against the background of the theories presented in Chapter 2. When attempting to explain the differences of the findings in this chapter with the findings of other researchers in other studies, the most important thing to consider is that earlier studies on the undeletion of the *that*-complementiser were focused on learner speakers of BSAE. In this study, the focus was on experienced BSAE speakers. This difference is an important one because the internal grammar of the learner speaker differs from the internal grammar of the experienced speaker. When the speaker is inexperienced, the more explicit form (in this case the overt *that*-complementiser) is stored in the speaker's internal grammar. Rohdenburg's (1996:151) complexity principle (otherwise known as the transparency principle) explains this phenomenon as follows: "In the case of more or less explicit grammatical options, the more explicit one(s) will tend to be favored in cognitively complex environments." As the learner gains more experience and experiences fewer cognitive constraints, the grammatical feature becomes more entrenched and it becomes easier for the learner to access the structure (Bybee, 2006:715).

While this is happening within the internal ecology, the BSAE speakers are contributing their utterances with the non-deleted *that*-complementiser to what Mufwene (2001a) calls the pool of linguistic features. There, these non-deleted instances are in competition with the zero complementisers contributed by the WSAE speakers. Gradually, as the BSAE speakers become more proficient, they start to select the zero complementiser more often.

Here it is important to note that the BSAE language users do not randomly choose to use zero complementiser. As the results clearly reveal, they are sensitive to the registers in which they produce the texts (i.e. more overt complementisers in the academic register and more zero complementisers in the reportage register). Croft (2000:92) explains that:

[A] linguistic code belongs to a (speech) community, and a community is defined by a domain. Every person is a member of multiple (speech) communities, hence every person speaks multiple codes, depending on which communities he or she belongs to.

On the one hand, the fact that the BSAE language users have mastered the codes used (in terms of *that*-deletion) in the academic and reportage domains, indicate that they regard themselves as members of those speech communities. On the other hand, the fact that the editors accept the BSAE language users' codes (in terms of *that*-deletion) signals the editors' acceptance of the BSAE language users' use of the feature in the speech community, even in cases where they adhere to slightly different norms. There appears to be sufficient alignment of norms across the STL and IDG components of the broader SAE community, but ultimately, a radically different BSAE norm is not in the process of becoming established as far as the undeletion of the *that*-complementiser is concerned.

Chapter 6

Anti-deletion and the infinitive *to*-marker

6.1 Contextualisation

The notion of OPTIONALITY is investigated in relation to the finite *that*-complementiser in Chapter 5. In this chapter, the same notion is investigated in relation to the infinitive *to*-complementiser. While *that*-omission occurs quite freely, the contexts in which *to*-omission occur are more constrained. McEnery and Xiao (2005:161) point out that the lemma “*help* is the only verb that can both control either a full infinitive or a bare infinitive and occur[s] either with or without an intervening noun phrase”. This construction is illustrated in (1) and (2):

- (1) The NMRS represents the highly-sophisticated technology that will **help** researchers **to** identify new topics or areas of research... (BSAE)

- (2) Notwithstanding the challenges, participant SMTs seem to have had strategies for **helping** them **∅** succeed. (BSAE)

In addition, Biber *et al.* (1999:708) point out that in verb + NP + infinitive constructions, verbs like *let*, *make*, and *have* only control bare infinitive clauses (cf. (3)), while verbs like *cause*, *enable*, and *force* only govern overt infinitive clauses (cf. (4)):

- (3) Hertzberg investigated what **makes** people \emptyset feel good or bad about their work... (BSAE)
- (4) The survey recommended that appropriate mechanisms should be implemented to **allow** external parties **to** report allegations of fraud. (BSAE)

Finite clauses profile processes, and tense marking grounds these processes by relating them to time (Langacker, 2008:124). Non-finite clauses, on the other hand, are atemporal. According to Langacker (2008:108-127), temporal finite clauses are processed by means of sequential scanning and are viewed by the producers and the receivers of the clausal-encoded message as a process unfolding in time. This means that language users can trace the process as it unfolds through conceptual time. Atemporal non-finite clauses, though, are processed by means of summary scanning and are viewed by the producers and the receivers of the clausal-encoded message as an atemporal relation that is frozen in time. This means that language users take in the entire process at a glance. Langacker's (2008:117) depiction of these temporal and atemporal processes can be seen in Figure 6.1 and Figure 6.2 respectively:

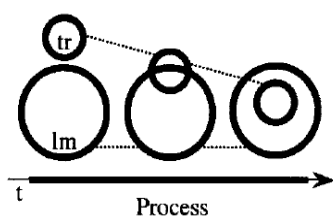


Figure 6.1: A diagrammatic representation of a temporal, finite clause (Langacker, 2008:117)

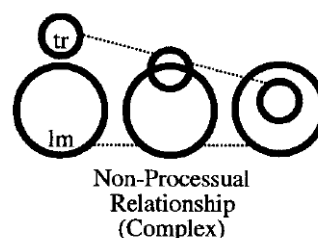


Figure 6.2: A diagrammatic representation of an atemporal, infinitive clause (Langacker, 2008:117)

The focus in this chapter is on the infinitive clauses as verb complements, and more specifically, on the use of the infinitive versus the use of the bare infinitive. Egan (2008:91) observes that the bare infinitive is used with six kinds of matrix words. In same-subject constructions, it occurs with verbs of effort (cf. (5)) and verbs of applied attitude (cf. (6)). In different-subject constructions, it occurs with verbs of perception (cf. (7)), communication (cf. (8)), enablement (cf. (9)), and causation (cf. (10)):

- (5) Reporting in the context of HIV, the report indicates that communities **help** [Ø mobilise resources, significantly improve knowledge of people, increase the use of services and affect outcomes of social processes]. (BSAE)

- (6) But we **dare not** [Ø forget that the achievement of true peace in the province also depends on political parties...] (WSAE)

- (7) As we braaied, we **heard** [a hyena Ø cry from the camp while the lions kept roaring].

- (8) Then, once inside the sitting-room, she **bade** [him Ø take a seat] (BNC)

- (9) Ethics are not difficult to define but it is difficult to **let** [different people Ø define them in their own terms as morals or values mean different things to different people]. (BSAE)

- (10) Therefore, while community-based responses are hailed as a critical sustainable development approach, their weakness **make** [potential supporters Ø cringe]. (BSAE)

Egan (2008:92) goes on to say that aside from the fact that all bare infinitive complement constructions have immediacy and certainty of realisation in common, the event or state in the infinitive complement is always profiled as “an undifferentiated whole”. In (7), for example, the sound of the hyena’s cry takes place in the moment, and cannot be construed as an ongoing event (unlike the sound of the lions’ roaring). Duffley (1992:141) makes a similar point when he states that “unlike the *-ing* form, the infinitive always produces a

representation of an event seen as a whole". Consequently, Egan (2008:91-92) argues that "[t]he bare infinitive complement form differs from its *to*-infinitive counterpart precisely in that it points to an almost certain, rather than to a merely possible, or perhaps probable eventuality".

The explanation above, however, does not account for verbs such as *help* that can govern either an overt infinitive *to*-marker or a bare infinitive marker. Scholars, such as Dixon (2005:268-269), Duffley (1992:26-29), and Lind (1983:270) argue that in those cases where the matrix verb subject plays an active role in realising the situation depicted in the complement clause, the bare infinitive is used (11). However, should the verb *help* encode no more than the transference of skill, the overt infinitive *to* complementiser will be used (cf. (11)):

- (11) John **helped** Mary [\emptyset eat the pudding] (he ate half). (Quoted by Dixon, 2005:201)
- (12) John **helped** Mary [**to** eat the pudding] (by guiding the spoon to her mouth, since she was still an invalid). (Quoted by Dixon, 2005:201)

Although Mair (1995:262-263) agrees in principle with this argument, he cautions that Dixon's examples are not representative of real-life instances of text. In this regard, Mair (1995:263) says that "[Dixon] is careful to couch his analysis in probabilistic terms" and that "such fine minimal pairs will hardly ever occur in contexts". Therefore, Mair (1995:263) states that these and other similar semantic constraints should not be considered narrowly:

These and similar semantic constraints should probably not be considered [...] with the intention only of accounting for the distribution of the various patterns of complementation found with the verb *help*. Rather, they are specific instantiations of a more general principle of iconicity in syntactic coding which could roughly be formulated as follows: direct acts of causation and assistance seem to follow a more reduced type of complement clause than indirect ones, in which assistance is rather like

advice and in which enabling condition/cause and resultant state are easy to keep apart.

Following Mair's advice, an overview is provided of some of the variables that have been found in the literature to account for the variability in the use of the verb *help* with either the overt or the bare infinitive marker.

The alternation in the use of the *help + to* and the *help + Ø* allostructions in the *help + infinitive* constructeme has been studied predominantly in standard varieties of English (e.g. Kjellmer, 1985; Lind, 1983; McEnery & Xiao, 2005) but not extensively in New Englishes. Consequently, this chapter will investigate the alternation in the use of the *help + to* and the *help + Ø* allostructions in BSAE.

Another feature that has enjoyed some attention in New Englishes is the undeletion of the infinitive *to*-marker. The notion **ANTI-DELETION**¹⁰² in terms of the infinitive marker has not been addressed extensively in the literature on BSAE. Initially, this feature has been recognised in passing as a BSAE feature by scholars such as De Klerk and Gough (2002) and Gough (1996) (also cf. Section 1.2). Later, though, Mesthrie (2006) discusses the anti-deletion of the infinitive marker in more detail.

Not all L2 varieties of English undelete the infinitive *to*-marker. Instead, they delete the infinitive marker in constructions where the infinitive *to*-marker is prototypically not omitted (e.g. in the *enable + NP + to* construction). This, however, is not only a feature of L2 varieties such as East African English but is also a feature that has been recorded in L1 varieties of English.

In the sections that follow, the alternation in the use of the infinitive marker in the *help + infinitive* constructeme (Section 6.2), the undeletion of the infinitive marker (Section 6.3), and the deletion of the infinitive marker (Section 6.4) are investigated. The findings are summarised in Section 6.5.

¹⁰² A definition of the notion is provided in Chapter 5.

6.2 The *help* + infinitive constructeme in BSAE

6.2.1 Introduction

Section 6.2.2 provides an overview of the factors driving the choice of the *help* + *to* allostruction or the *help* + \emptyset allostruction in the *help* + infinitive constructeme. In Section 6.2.3.1, the general results of the investigation into the feature is provided. The results of the inferential decision tree in Section 6.2.3.2 and the results of the random forest in Section 6.2.3.3 shed light on the factors that drive language users' decisions to omit or retain the infinitive marker in the *help* + infinitive constructeme.

6.2.2 Factors driving infinitive *to*-omission: the case of *help* + infinitive

In the literature, several factors driving the alternation in the *help* + infinitive constructeme are identified. The most prevalent factors are described in the paragraphs that follow.

Before the factors are discussed, Table 6.1 below provides an overview of the constructional variants of the *help* + infinitive constructeme and serves to elucidate the discussion of the factors driving the omission of the infinitive marker.

Table 6.1: The various structural realisations of the *help* + infinitive constructeme

Subject	Verb	Object of <i>help</i>	Infinitive verb	Object of infinitive verb
Teachers [helper]	help	\emptyset [helpee]	to build	the education system (BSAE)
Provincial road departments [helper]	help	them [helpee]	\emptyset implement	projects (BSAE)

In the *help* + infinitive constructeme, the verb *help* may be followed by an object (reflected in the third column of the table). In the discussion below, this object will be referred to as the “object of *help*”. In the same clause, the infinitive verb may also take an object. This object will be referred to as the “object of the infinitive verb”.

One of the first significant factors believed to contribute to the alternation in the *help* + infinitive constructeme is **register**. Scholars such as Lind (1983:156) and McEnery and Xiao (2005:122) have observed that language users tend to use the *help* + *to* allostruction in formal situations and written texts, whereas they tend to use the *help* + \emptyset allostruction in informal situations and spoken texts. Therefore, the variable **REGISTER** is included in the analysis, and coded with the values 'A' (academic), 'I' (instructional), 'P' (popular), and 'R' (reportage).

In the literature, scholars have also observed that the language users of different English **varieties** prefer different allostructions (Kjellmer, 1985; Mair, 2002, 2006; McEnery & Xiao, 2005). Research has revealed that although American English and British English follow the same development trend regarding the *help* + infinitive constructeme (i.e. an increasing preference for the *help* + \emptyset allostruction), the *help* + \emptyset allostruction is used more frequently in American English than in British English. Consequently, language variety is included as the variable **VARIETY** and coded with the values 'TLE', 'BSAE', and 'WSAE' in the analysis. The learner variety corpus (i.e. the TLE) is included in the analyses of this chapter since it is suspected that the differences observed in terms of the anti-deletion of the infinitive marker in BSAE may not be "errors", but might be ascribed to issues pertaining to complexity.

Another prominent factor is whether there is any intervening material between the lemma *help* and the infinitive (Kjellmer, 1985:158; Lind, 1983:272; Lohmann, 2011:515; Mair, 2002:122, 2006:136; McEnery & Xiao, 2005:176). In this regard, Lind (1983:272) finds that those forms with **an intervening noun phrase** (the object of *help*) have a proportionally higher degree of omission than the forms without an intervening nominal. This factor is illustrated in (13) below:

- (13) According to Creswell (2009:178), the idea behind qualitative research is to purposefully select participants or sites (or documents or visual material) that will best **help the researcher** [\emptyset understand the problem and the research question]. (BSAE)

Lind (1983:269) ascribes this phenomenon to the fact that the infinitive is dropped in similar constructions with the lemmas *see*, *hear*, and *make* (illustrated in (14) below). In this analysis, the variable INTERVENINGNP is used and coded with the values ‘yes’ and ‘no’.

- (14) The role of the leader is to **make** people [\emptyset succeed].

Rohdenburg (1995:375, 1996:155-158) finds that the intervening noun phrase is an important variable since it influences the processing complexity of the language user (cf. the discussions on processing complexity in Chapter 2 and Chapter 4.) It is also important to keep in mind that noun phrases are also realised by syntactic structures of various complexity (i.e. pronouns, which are relatively easy to process, and noun phrases, which are typically more difficult to process). Against this background, scholars have found that when the nominal phrase is realised by a noun phrase, the language user is more likely to use the overt infinitive marker. Consequently, the variable INTERVENINGNPSYNTAX is included in the analysis and coded with the values ‘pronoun’ and ‘NP’.

Since noun phrases can vary in length, and other elements – aside from an intervening nominal – can occur in this position (cf. (15)) where and adjunct occurs between the lemma *help* and the infinitive marker), one could argue that the **distance between the verb *help* and the infinitive marker** should also be considered (Lohmann, 2011:504; McEney & Xiao, 2005:176-177). In this study, the distance is measured by counting the number of graphemic words between the lexical verb and the infinitive marker. In (15), for example, the distance between the lexical verb and the infinitive marker is three (3). This factor is accounted for with the variable DISTANCEVERBTO and coded with a numeral value.

- (15) The teacher **can help** in this regard [**to** remind them]. (BSAE)
- 1 2 3

It should be noted, though, that Lohmann (2011:504) points out that the distance variable is epiphenomenal to the syntactic realisation of the intervening nominal and that this should be considered when the results are interpreted.

Another variable to consider is the **animacy of the subject** in the matrix clause. Animacy is a semantic feature and indicates how alive or sentient an entity is. Therefore, humans and animals are conceived as being animate, while everything else is conceived as being inanimate. Scholars have found that animate subjects tend to result in the *help + Ø* allostruction (Lind, 1983:270; Lohmann, 2011). Lind (1983:270) postulates that it seems more natural to associate animate subjects with a closer, more direct involvement in the helping process. This close relationship results in the omission of the infinitive marker. On the other hand, it seems more natural to associate inanimate subjects with a more distant, indirect involvement in the helping process (also refer to the discussion in Section 6.1). As such, it results in the retention of the infinitive marker. This feature is accounted for with the SUBJECTANIMACY variable, and it takes the values “animate”, “inanimate”. In those instances where the subject is not explicit (e.g. in the passive voice), the value “n/a” is used.

Mair (1995:262-263), however, argues that linguists should not consider semantic constraints (such as the animacy of the subject in the matrix clause) too narrowly as explanations for the distribution of complementation patterns, because these semantic explanations are not always as clear-cut in examples quoted from corpora as it would seem in examples generated by linguists themselves. In corpora, for example, verb lemmas that take animate subjects often do not govern bare infinitives clauses (cf. (16)), and verb lemmas that take inanimate subjects often do not govern overt infinitive clauses (cf. (17)).

(16) I, however, ensured that the data reflect the actual views of participants and discussed this with my supervisor, who, on the basis of his expertise and knowledge of the subject, help [to verify the accuracy thereof]. (BSAE)

(17) What suggestions can **help** [Ø improve the program]? (BSAE)

Based on this observation, Mair (2002:124-126, 2006:137-140) and Leech *et al.* (2009:189-190) argue that *help + to* and *help + Ø* are undergoing a process of semantic bleaching and are gradually taking on grammatical properties, pointing to an incipient process of grammaticalisation. In this view, *help* takes on a more general meaning of

[contribute to] or [provide a favourable environment for] (Mair, 2002:124).¹⁰³ This view is illustrated with an example from the BSAE corpus in (18) below:

- (18) Scaling-up has positive effects on the public policy climate – scaled-up programmes' visibility and effectiveness **help** [Ø shift social norms and foster greater acceptability and support]. (BSAE)

This process of grammaticalisation, as put forward by Mair (1995, 2006), is probabilistic (rather than a discrete variable). Since Mair's hypothesis is not congruent with Lind's (1983) view that the animacy of the subject is a determining factor in the retention/omission of the infinitive marker, the absence of the variable SUBJECTANIMACY as a determining factor in the retention/omission of the infinitive marker is considered as possible support for Mair's hypothesis.

Lind (1983:268-269) finds that the bare infinitive occurs more frequently with the **uninflected form of help**. This corresponds with findings by Pullum and Zwicky (1999) that the inflectional form of the verb may play a determining role in whether users use the *help* + *to* allostruction or the *help* + Ø allostruction. Following this finding in previous studies (e.g. Lind, 1983; Pullum & Zwicky, 1999), the different forms of the verb lemma in the matrix clause were also coded in this analysis. The variable VERBINFLECT is coded with the values 'present', 'past', 'present progressive', 'past progressive', 'present perfect', 'past perfect', 'modal', 'infinitive', 'ing-participle' and 'ed-participle'. An issue related to the verb inflection is the issue of **negated infinitives**, which tend to collocate with the overt form of the infinitive.

- (19) The teacher's job is not to correct mistakes the pupil has already made, but to **help** him **not to make** that mistake next time. (Quoted by Leech *et al.*, 2009:188)

Rohdenburg (1995:378-380) ascribes this phenomenon to the fact that negation in less explicit infinitive clauses leads to increased processing capacity. However, the negated

¹⁰³ McEnergy and Xiao (2005:175-176) contest Mair's (2002) hypothesis, but do not provide corpus-evidence for their argument.

infinitive is not attested in either the BSAE corpus or the WSAE corpus. Therefore, this factor was not considered.

Finally, scholars such as Lind (1983:269) and Rohdenburg (1995:381-382) find that the bare infinitive is more likely to be chosen when it follows **an infinitive form of help**:

- (20) The basic purpose of Code of Ethics is practically to **help** employees [\emptyset do what is expected of them in a particular role or function]. (BSAE)

While Lind (1983:269) attributes this phenomenon (known as the *horror aequi* principle) to euphony, Rohdenburg (1995:381-382) argues that it is a special manifestation of structural complexity. On the basis of Bever, Carroll and Hurtig (1976:172), he explains that “sequences of clauses standing in a relationship of subordination to each other are more difficult to process than coordinated clauses”. He attributes the complexity of an infinitive clause embedded in another infinitive clause to the fact that the language user needs to recover two understood subjects in rapid succession. The second infinitive marker is omitted to facilitate the recovery of these subjects. The *horror aequi* principle is accounted for with the variable INFHELP and is coded with the values ‘yes’, ‘no’, and ‘bare’.

The results of the analyses are presented in Section 6.2.3 below.

6.2.3 Results

The general results of the investigation are provided in Section 6.2.3.1. In order to gain better insight into the interaction between the variables discussed in Section 6.2.2 above, the results of the conditional inference tree and the random forest are provided in Section 6.2.3.2 and Section 6.2.3.3 respectively.

6.2.3.1 General results

Table 6.2 provides an overview of the frequency of the *help + infinitive* constructeme in the TLE, BSAE, and WSAE corpora. A chi-squared test indicates that there is no significant difference between the results of the three corpora ($\chi^2 = 4.38$; $p = 0.11$). There is also no significant difference between the results of the TLE corpus and the BSAE corpus ($\chi^2 = 1.07$; $p = 0.30$) and between the BSAE corpus and the WSAE corpus ($\chi^2 = 1.06$; $p = 0.30$). The difference between the TLE corpus and the WSAE corpus, though not significant, approaches significance ($\chi^2 = 3.55$; $p = 0.06$).

Table 6.2: Summary of the raw alternation frequency of the *help + infinitive* constructeme in the TLE, BSAE and WSAE corpora

	TLE	BSAE	WSAE
<i>help + to</i>	43	71	27
<i>help + ∅</i>	13	34	20
Total	56	105	47

At first glance, it seems as if the *to*: \emptyset -ratio of 1:0.30 in the TLE corpus is smaller than the ratio of 1:0.48 in the BSAE corpus. In turn, the *to*: \emptyset -ratio of 1:0.48 in the BSAE corpus is smaller than the ratio of 1:0.74 in the WSAE corpus.¹⁰⁴ These ratios can be seen in perspective by considering the feature in British English and American English. The *to*: \emptyset -ratio that Mair (2002:112) observes for British English and American English are also considered (also cf. Quirk *et al.*, 1985:1205). In 1961, British English had a *to*: \emptyset -ratio of 1:0.29, while American English had a *to*: \emptyset -ratio of 1:1.27. By 1991/1992 the picture changed with British English having a *to*: \emptyset -ratio of 1:1.58 and American English having a

¹⁰⁴ The ratios are calculated as follows:

$$\frac{\text{frequency of help + to}}{\text{frequency of help + to}} : \frac{\text{frequency of help + } \emptyset}{\text{frequency of help + to}}$$

$$\text{TLE: } \frac{43}{43} : \frac{13}{43}$$

$$1:0.30$$

$$\text{BSAE: } \frac{71}{71} : \frac{34}{71}$$

$$1:0.48$$

$$\text{WSAE: } \frac{27}{27} : \frac{20}{27}$$

$$1:0.74$$

to: \emptyset -ratio of 1:4.61. In this context, BSAE has a lower *to*: \emptyset -ratio than WSAE, but in comparison to the 1991/1992 *to*: \emptyset -ratios for British English and American English, the ratios for TLE, BSAE, and WSAE are quite low. Future work in which the *to*-omission rate is investigated in a diachronic corpus of WSAE may provide more information regarding usage trends over time and whether BSAE and WSAE are following the same trends as British English and American English by moving towards a higher rate of *to*-omission over time.

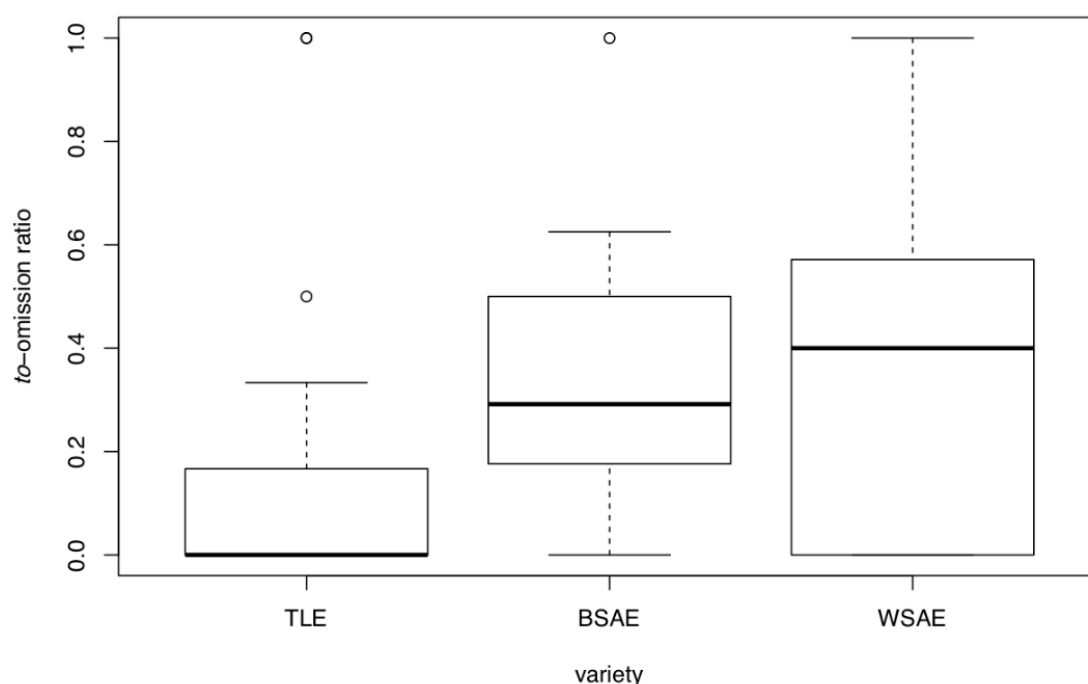


Figure 6.3: Boxplot of the *to*-omission ratio in the *help + infinitive constructeme* for BSAE and WSAE

In terms of dispersion, the TLE corpus has a lower median (0.00) than the BSAE corpus (0.29). And the BSAE corpus, in turn, has a lower median than the WSAE corpus (0.40). These dispersion statistics are illustrated in Figure 6.3. A Kruskal-Wallis rank sum test shows that these differences in dispersion are significant (Kruskal-Wallis chi-squared = 19.694, $df = 2$, $p = 5.291e^{-05}$). A post-hoc pairwise Wilcoxon rank sum test (cf. Table 6.3) shows that the difference between BSAE and WSAE is not significant at $p = 0.415$, while the difference between the TLE and BSAE is significant at $p = 0.0001$ and the difference between the TLE and WSAE is significant at ($p = 0.0003$).

Table 6.3: Results of the post-hoc pairwise Wilcoxon rank sum test for variable *to*-omission

	TLE	BSAE
BSAE	0.00011	-
WSAE	0.00027	0.41452

The results of the frequency analysis in this section do not provide any insight into the interaction between the variables that drive speakers' choices in allostruction. Consequently, a conditional inference tree was drawn. The results of this tree are provided in Section 6.2.3.2.

6.2.3.2 Conditional inference tree

A conditional inference tree was drawn to gain more insight into the interaction between the variables that drive speakers' use of either the *help + to* allostruction or the *help + ∅* allostruction. The tree was drawn using the `ctree()` algorithm in the `party` package in R (Hothorn *et al.*, 2015) and is visualised in Figure 6.4 on p. 238. The conditional inference tree highlights which independent variables predict language users' choice. The independent variables used to draw the tree are discussed in Section 6.2.2 and listed in Table 6.4 below. All the possible splits are significant at $p < 0.05$. The conditional inference tree returned a relatively high prediction accuracy of 75%.

Table 6.4: *The hypothesised variables conditioning infinitive to-omission*

Variable	Description	Value
VARIETY	Indicates variety	'TLE', 'BSAE', 'WSAE'
REGISTER	Indicates register	'A' = academic; 'I' = instructional 'P' = popular; 'R' = reportage; "SW" = student writing
INTERVENINGNP	Indicates the presence of an intervening noun phrase	'yes', 'no'
INTERVENINGNPSYNTAX	Indicates the syntactic structure by which the intervening NP is realised	'pronoun', 'NP'
DISTANCEVERBTO	Indicates the distance between the verb in the matrix clause and the infinitive marker, measured in graphemic words	numeral value
SUBJECTANIMACY	Indicates how alive or sentient the subject in the matrix clause is	'animate', 'inanimate'
VERBINFLECT	Indicates whether the verb is simple or inflected	'present', 'past', 'modal', 'infinitive', 'ing-participial'
HELPINFIN	Indicates whether the verb <i>help</i> in the matrix clause is an infinitive verb or not	'bare', 'yes', 'no'

The decision tree shows that only two variables play a role in the decision-making process. The first split in the tree is determined by whether the matrix lemma *help* is an infinitive verb or not (Node 1). If the lemma *help* is an infinitive verb and is preceded by an overt infinitive marker, there is a 67% likelihood that the language users will use the *help* + \emptyset allostruction (Node 5). This is exemplified in (21):

- (21) Joint Aid Management (JAM), a Christian-based, humanitarian organisation, implemented school feeding projects in South Africa **to help** [\emptyset improve the nutritional status of the children]. (BSAE)

It confirms Rohdenburg's (1995:381-382) assertion that embedded dependent clauses are difficult to process and that the omission of the second infinitive leaves the language user with more processing resources to retrieve the subjects of the two dependent clauses.

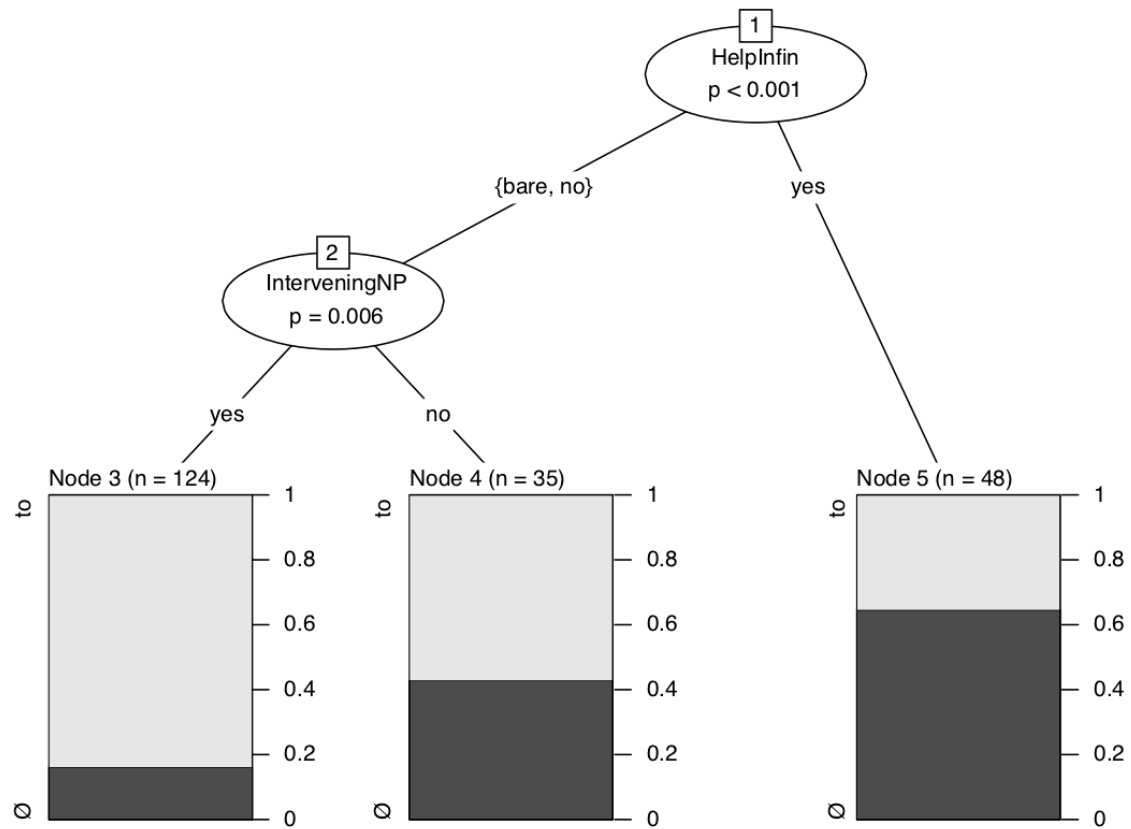


Figure 6.4: A conditional inference tree of the variables predicting infinitive to-omission in the help + to constructeme in BSAE and WSAE

If the lemma *help* is a bare infinitive verb or not an infinitive verb at all, the language users' choice is determined by whether there is an intervening noun phrase in the construction or not (Node 2). The left-hand split indicates that if there is an intervening noun phrase between the verb *help* and the infinitive marker, there is a more than 80% likelihood that the *help + to* allostruction will be used (Node 3). This split confirms that the intervening noun phrase makes the construction more difficult to process. To make the process easier to process, language users use the more explicit allostruction – in this case, the allostruction *help + NP + to* instead of *help + NP + ∅*. If, however, there is no intervening NP in the construction, the likelihood of the *help + to* allostruction being chosen decreases with approximately 15% (Node 4). This is exemplified in (22) and (23):

- (22) It **helps** [∅ track the progress that government is making in achieving results and help collect evidence about what work and what did not, to help improve planning and implementation on an annual basis]. (BSAE)
- (23) This factor **helped** some SMTs [to achieve the intended training results]. (BSAE)

Table 6.5 shows that the *help + NP + to* allostruction occurs more frequently in all the corpora. However, a χ^2 -test indicates that there is no significant difference between the results of the three corpora ($\chi^2 = 2.76$; $p = 0.25$). There is also no significant difference between the results of the TLE corpus and the BSAE corpus ($\chi^2 = 0.88$; $p = 0.35$) and between the BSAE corpus and the WSAE corpus ($\chi^2 = 0.33$; $p = 0.57$). The difference between the TLE corpus and the WSAE corpus, is also not significant ($\chi^2 = 1.95$; $p = 0.16$).

Table 6.5: Summary of the *help + NP + to* allostruction and the *help + NP + ∅* allostruction

	Raw frequency			Normalised frequency		
	TLE	BSAE	WSAE	TLE	BSAE	WSAE
<i>help + NP + to</i>	35	62	22	17	45	9
<i>help + NP + ∅</i>	8	24	12	4	17	5
Total	43	86	34	21	63	14

Furthermore, the animacy of the subject of the matrix clause does not feature in the decision tree above. In Section 6.2.3.3, this variable also does not feature strongly. This means that the hypothesis Mair (2002:124-126, 2006:137-140) and Leech *et al.* (2009:189-190) put forward could be undergoing a process of semantic bleaching which in turn could be indicative of an incipient process of grammaticalisation, at the same time disconfirming McEnery and Xiao's (2005:179) arguments. In this regard, Lohmann (2011:515) also found confirmation for Mair (2002:124-126, 2006:137-140) and Leech *et al.*'s (2009:189-190) hypothesis.

Finally, when considering the conditional inference tree, it becomes clear that variety does not play as big a role as expected in the choices made between the *help + to* and the *help + ∅* allostructions.

6.2.3.3 Random forest

As mentioned earlier in this thesis, conditional inference trees sometimes do not reflect highly predictive interactions (cf. Bernaisch *et al.*, 2014:14-15; Tagliamonte & Baayen, 2012:163-164). Therefore, a random forest was used to confirm the findings of the inference tree. The random forest is visualised in Figure 6.5 (p. 241). It was drawn using the `cforest()` algorithm in the `party` package in R (Strobl *et al.*, 2009a). The classification accuracy of the random-forest approach is slightly higher than the classification accuracy of the conditional inference tree at 76%.

The variable importance scores confirm that whether the lemma *help* is inflected or not is the most important predictor (VERBINFLECT = 0.029). Whether the lemma *help* is an infinitive verb follows (HELPINF = 0.026). Interestingly, the form of the verb does not feature as a significant variable in the decision tree. This finding is in line with the findings of Lind (1983:168) and Lohmann (2011:512).

Another variable that is not found in the conditional inference tree, but is pointed out as influencing the decisions that language users make is register (REGISTER = 0.013). Although this variable importance score is lower than the variable importance score of the VERBSTRUCTURE variable, its score is still higher than the variable importance score of

the INTERVENINGNP (INTERVENINGNP = 0.001) variable, which does feature in the conditional inference tree.

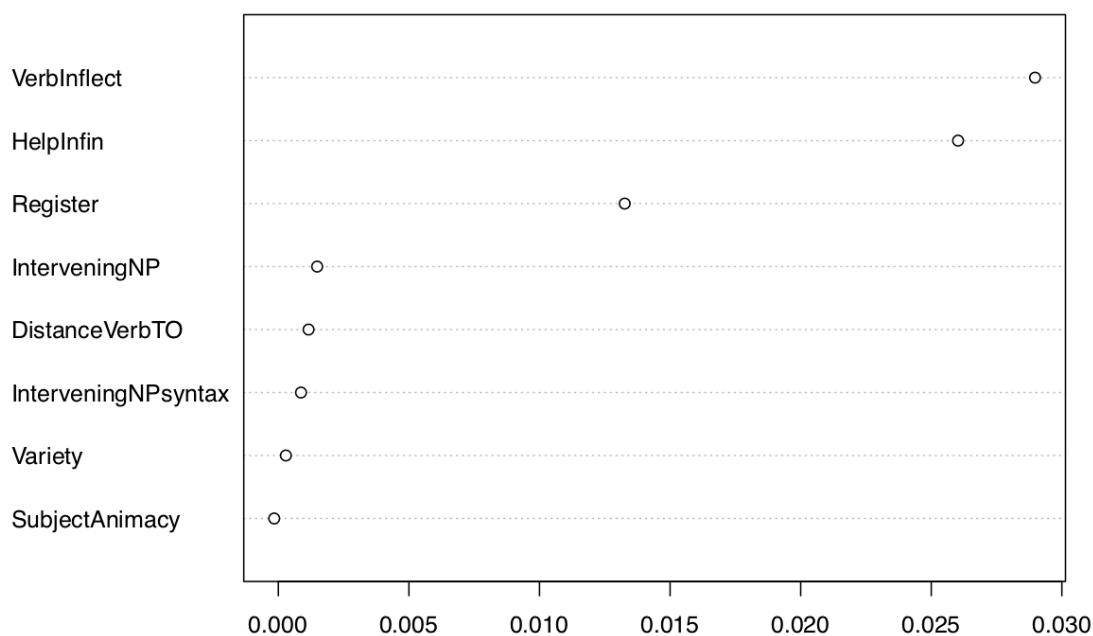


Figure 6.5: Conditional importance of the variables believed to drive the omission of the infinitive marker in the *help + infinitive* constructeme

Considering the results in Table 6.6, the BSAE speakers tend to use the the *help + to* allostruction more often in the academic register than the *help + ∅* allostruction. The WSAE speakers, on the other hand, use the *help + infinitive* marker constructeme less frequently than the BSAE speakers, with the frequency of the *help + to* allostruction occurring equally often. A Fisher exact test indicates that the difference between BSAE and WSAE is significant ($p = 0.068$). Since the academic register is regarded as a formal register, it would seem as if the BSAE users and the WSAE users follow the conventions of the British English users and the American English users (cf. Lind, 1983:156; McEnery & Xiao, 2005:122). In the instructional register, which may be regarded as less formal than the academic register, the *help + ∅* allostruction is used slightly more often than the *help + to* allostruction. The *help + infinitive* constructeme did not occur frequently enough in the popular and reportage register to draw any conclusions about the use of the allostructions in those registers.

Finally, the variable VARIETY scored a variable importance score of 0.001. This confirms the finding of the conditional inference tree that the variety does not have a high predictive value in terms of the choices that language users make between the *help* + infinitive allostruction and the *help* + \emptyset allostruction.

Table 6.6: A summary of the normalised frequencies of the *help* + *to* allostruction and the *help* + \emptyset allostruction across registers in BSAE and WSAE¹⁰⁵

	BSAE		WSAE		Fisher exact test (p-value)
	<i>help</i> + <i>to</i>	<i>help</i> + \emptyset	<i>help</i> + <i>to</i>	<i>help</i> + \emptyset	
Academic	43	16	7	8	0.068
Instructional	8	11	5	7	1.000
Popular	2	1	1	1	1.000
Reportage	3	1	2	2	1.000
Total	56	29	15	18	0.059

The general frequency and statistical results indicate that the differences in the omission of the infinitive *to*-marker in the *help* + infinitive constructeme cannot be attributed to variety. To further test the veracity of these results, the editors' acceptance of the BSAE users' choices regarding the *help* + infinitive marker is considered in Section 6.2.3.4.

6.2.3.4 Acceptability

In Chapter 2 (Section 2.2.5), it was argued that the editors' acceptance of a feature (i.e. the authoritative measure) might be a sign of the endonormative stabilisation of a feature. The implication is that if the editors accept the BSAE authors' use of the *help* + infinitive constructeme, variety is truly not a variable that contributes to the differences in the use of the constructeme. Consequently, an aligned concordance was drawn of the

¹⁰⁵ The TLE corpus does not share the same registers as the BSAE editing corpus and WSAE editing corpus. Therefore, all the TLE instances in the analysis were labelled as 'learner' in terms of the variable REGISTER. So, for the TLE corpus, in the 'learner' register, the normalised frequency of the *help* + *to* allostruction is 21, and the normalised frequency of the *help* + \emptyset allostruction is 6.

help + infinitive constructeme in the BSAE editing corpus to determine what changes the editors made. The overall results of this analysis are shown in Figure 6.6.

In the academic register, the editors made changes to four of the *help* + infinitive constructions. In the instructional register, the editor made changes to one of the constructions, and in the reportage register, the editors made two changes.

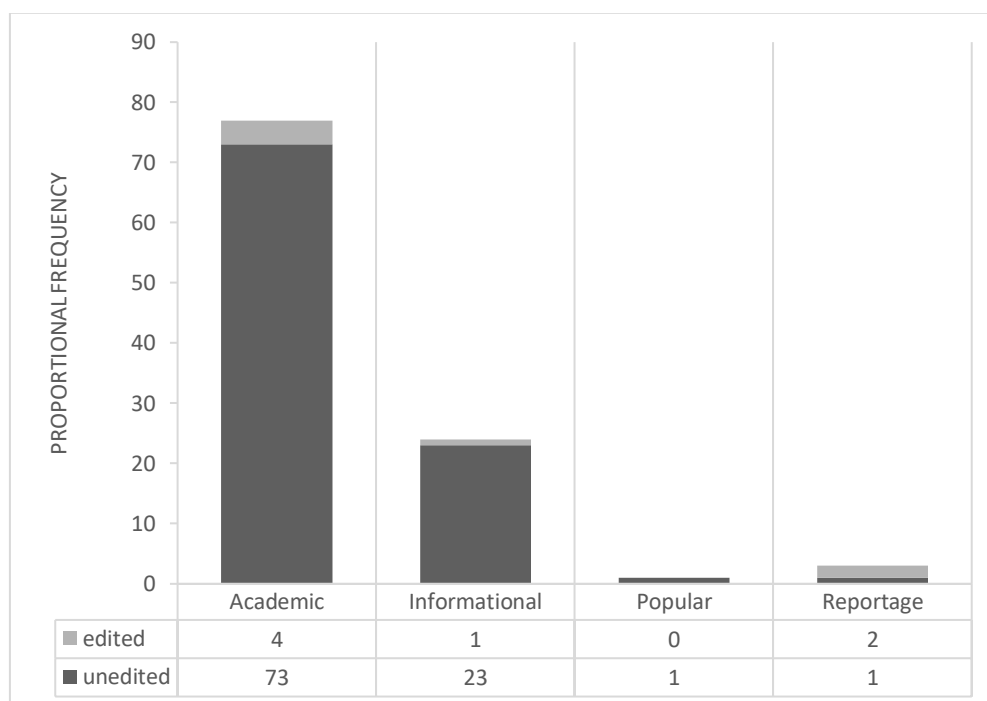


Figure 6.6: Graph illustrating the number of changes the WSAE editors made to the *help* + infinitive constructeme

The *help* + infinitive constructeme is most prevalent in the academic register. Yet, the editors only changed four of the 77 instances. In all four of the instances, the *help* + \emptyset allostruction is replaced with the *help* + *to* allostruction (cf. Table 6.7). These changes are in line with the increased explicitness that characterises the academic register.

Table 6.7: Raw frequencies of editors' changes made to the *help + infinitive* constructions in the BSAE

Register	Raw frequencies	
	<i>to</i> added	<i>to</i> removed
Academic	4	0
Instructional	1	0
Popular	0	0
Reportage	0	2

In the reportage register, though, the editors replaced the *help + to* allostruction with the *help + Ø* allostruction. These changes seem to counter Biber *et al.* (1999:736) frequency counts for the use of the bare infinitive in the reportage register. This can be explained by considering the nature of the texts found in the reportage register and the word count for each register.

Table 6.8: A summary of the raw frequencies of the *help + to* allostruction and the *help + Ø* allostruction in the original BSAE corpus and the edited BSAE corpus

Register	BSAE original	BSAE edited
<i>help + to</i>	71	76
<i>help + Ø</i>	34	29

A χ^2 -test (cf. Table 6.8) confirms that the differences in the frequencies of the *help + to* allostruction and the *help + Ø* allostruction in the original BSAE texts and the edited BSAE texts are not significant ($\chi^2 = 0.36$; p-value = 0.55).

6.2.4 Summary

In Section 6.2.3.1 it is found that the infinitive *to*-omission ratio in the *help + infinitive* marker constructeme between the TLE and the BSAE corpora is not significant. Similarly, the infinitive *to*-omission ratio in the same constructeme between the BSAE and WSAE corpora is also not significant. However, a significant difference is observed between the

TLE and WSAE corpora. This finding is the first finding suggesting that the non-deletion of the infinitive marker is a feature of the learner variety of BSAE.

The results of the conditional inference tree and the random forest provide evidence that the frequency differences can largely be attributed to the uneven pressure of the conditioning grammatical variables and that variety does not play as big a role as expected in the choices made by the language users between the *help + to* and *help + ∅* allostruction.

Furthermore, considering the few changes made by the editors to the *help + infinitive* marker constructeme, it would seem as if the BSAE users' use of the constructeme is legitimised by the WSAE speakers.

6.3 Undeletion of the infinitive *to*-marker

6.3.1 Contextualisation

Scholars such as Mesthrie (2006) and Parkinson and Singh (2007:58, 60) have reported that the undeletion of the infinitive marker in constructions such as *know + NP + to* and *make + NP + to* is a feature of BSAE. This feature is illustrated in (24) and (25) below:

- (24) The extension of the training could **make** all staff members and school authorities **to** have first-hand information. (BSAE)
- (25) He further said He will then give this race of people His Laws and **have** them **to** build a tabernacle for Him. (BSAE)

In these constructions, Mesthrie (2006) observes that the undeletion of the infinitive marker occurs 34.5% of the time in his spoken corpus. Instead of categorising the feature as an 'error', he argues that it forms part of the coherent system that is BSAE. Parkinson and Singh (2007:58, 60), on the other hand, find that 59% of their BSAE students did not remove the overt infinitive marker in the *make + NP + to* construction when given examples of the construction in grammatical judgement tests after they had been made

aware of the standard variety convention of omitting the infinitive marker. They find that the difference in the test scores before the intervention and after the intervention is statistically significant, and consequently infer that there is a shift towards standard English.

Given Mesthrie (2006) and Parkinson and Singh's (2007) findings, together with the findings above pertaining to the BSAE language users' use of the bare infinitive marker in the *help* + NP + *to* construction, it may be insightful to investigate constructions in which the infinitive marker forms part of the underlying structure in standard varieties of English. The constructions investigated here include *feel* + NP + \emptyset , *hear* + NP + \emptyset , *know* + NP + \emptyset , *have* + NP + \emptyset , *let* + NP + \emptyset , *make* + NP + \emptyset , *notice* + NP + \emptyset , *observe* + NP + \emptyset , *see* + NP + \emptyset , and *watch* + NP + \emptyset .

6.3.2 Results

No realisations of the *feel* + NP + infinitive, *know* + NP + infinitive, *notice* + NP + infinitive, and *observe* + NP + infinitive constructions were found in the TLE corpus, BSAE corpus, or the WSAE corpus. In Figure 6.7, one can see that the proportion of undeletion in the verb + NP + \emptyset constructeme in the TLE is similar to the proportions of undeletion reported by Mesthrie (2006). However, the proportion of undeletion in the BSAE corpus is lower, signalling that the feature occurs less often the more experienced the speakers becomes.

A Fisher exact test indicates that the differences in the undeletion of the verb + NP + \emptyset constructeme across all the corpora are significant ($p = 3.95 \times 10^{-10}$). Another Fisher exact test indicates that the difference in the undeletion of the verb + NP + \emptyset constructeme between the TLE and the BSAE is significant ($p = 0.014$). Yet another Fisher exact test indicates that the difference in the undeletion of the verb + NP + \emptyset constructeme between the BSAE and the WSAE is significant ($p = 0.03$). A final Fisher exact test indicates that the difference in the distribution of the undeletion in the verb + NP + \emptyset constructeme between the TLE and the WSAE is significant ($p = 5.44 \times 10^{-10}$).

One factor that may account for the difference in the results of this study and Mesthrie's (2006) study is register. Biber (1999:133) emphasises the importance of register in

studies of discourse and grammar, saying that most descriptions of linguistic features are not valid for language as a whole. Mesthrie's (2006) corpus was a corpus of spoken texts, which may account for the differences in the findings.

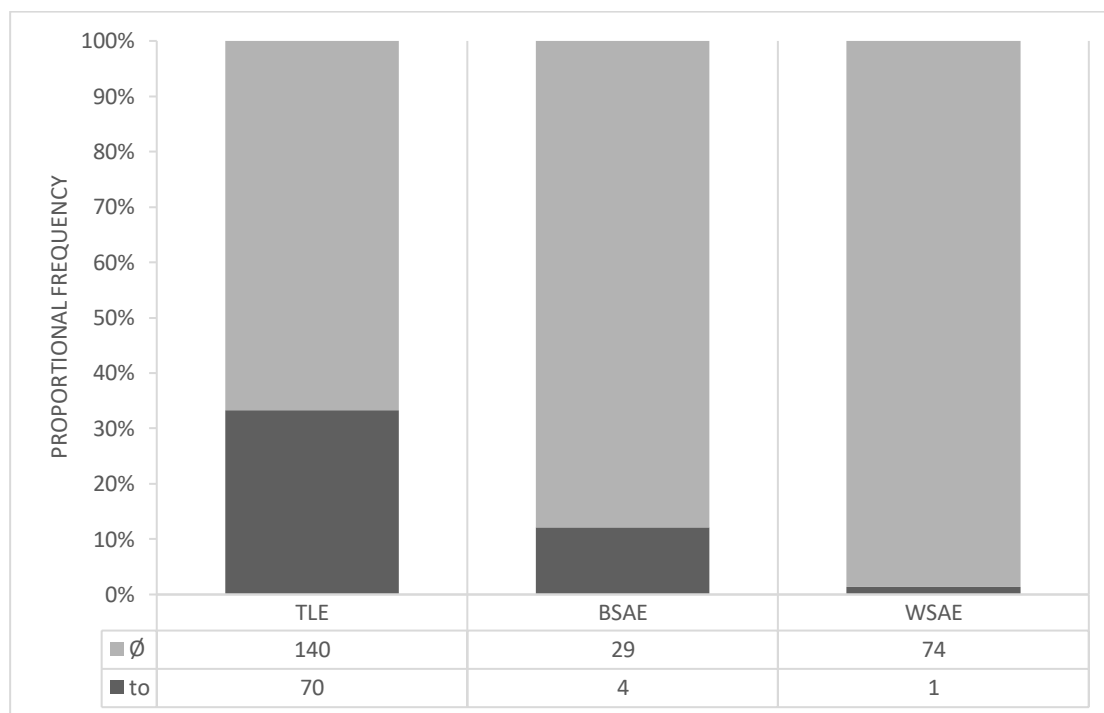


Figure 6.7: *The proportional frequencies of the overt infinitive to-complementiser and the bare infinitive to-complementiser with the lemmas feel, hear, know, have, let, make, notice, observe, see, and watch*

When the corpora were prepared, the quoted texts (from other written texts and interviews) were marked up. This prevented the inclusion of features of other varieties of English. But to determine whether this may, in fact, be a feature that may be attributed to register, some of the quotes from interviews in the academic register were perused. Below are some examples from the quoted spoken language in which the construction with the undeleted infinitive marker were found:

- (26) We do give the Life Orientation educators to **let their kids to** exhibit their work. (BSAE)
- (27) The support given by their communities **made them to** achieve compelling results. (BSAE)

- (28) I think it is the support of the parents that **made us to** achieve these good results. (BSAE)
- (29) **Let no situation to** demoralise you. Stand up and keep you head held high. (BSAE)

Consequently, Parkinson and Singh's (2007) findings (and the findings of this thesis) should be read with the understanding that only written registers were investigated and that the same findings may not necessarily be true for more informal registers such as spoken language.

Another interesting observation is that the verb + NP + *to* allostruction is also attested in WSAE:

- (30) ...some of whom are suffering from a severe case of technophobia, as well as those who have decided to throw up their hands in surrender and **let their children to** do what they like in the digital world. (WSAE)

There is a possibility that this feature may be transferred from BSAE to WSAE by means of language contact, but the claim cannot be made on the grounds of a single observation. So, to determine whether the feature is acceptable to the WSAE, a parallel concordance was drawn to determine whether the WSAE editor removed the infinitive marker or not.

Table 6.9 shows that of the four instances of undeletion that have been identified, the overt infinitive marker was only accepted in one instance. In all the other instances the infinitive marker was removed. A Fisher exact test shows that the difference between the original BSAE texts and edited BSAE texts is not significant ($p = 0.36$).

Table 6.9: A summary of the raw frequencies of the verb + NP + *to* allostruction and the verb + NP + *to* allostruction in the original BSAE corpus and the edited BSAE corpus

Register	BSAE original	BSAE edited
<i>help + to</i>	29	32
<i>help + ∅</i>	4	1

Considering the fact that the use of the infinitive marker occurs less often in the BSAE corpus than in the TLE corpus, together with the fact that the editors tend to remove this feature from the experienced BSAE users' writing, it can be assumed that the undeletion of the infinitive *to*-marker is a transient feature that occurs less frequently the more experienced the language users become. Consequently, it can be said that the undeletion of the infinitive marker is a feature of the learner variety.

In light of these unexpected findings, it was decided that it would be prudent to investigate the deletion of the infinitive *to*-marker in constructions where the infinitive marker is prototypically present as well.

6.4 Deletion of the infinitive *to*-marker

6.4.1 Contextualisation

In his study of non-finite complementation in L1 English, Egan (2008) notices that some L1 speakers omit the infinitive *to*-marker with verbs such as *assist*, *allow*, and *enable*. He states there are very few instances of this kind and he describes them as “exceptions [...] so rare as to render it unnecessary to revise the classification proposed” (Egan, 2008:196-197). Callies (2013), however, does not agree that these instances are exceptions. Consequently, he specifically searches for structures of this kind, and find that these instances are not “mere exceptions”. In fact, he claims that these structures occur so often in his data that it “challenge[s] the long-standing claim that *help* “is the only verb that can both control either a full infinitive or a bare infinitive and occur either with or without an intervening noun phrase” as claimed by McEnery and Xiao (2005:161). Examples (31) and (32) below are examples quoted from Callies (2013:7) to illustrate such instances:

- (31) At worst, food will be **allowed** \emptyset develop only to the extent that it serves the needs of agriculture. (BNC)
- (32) I can't see him **allowing** me \emptyset take the children so far away from London. (BNC)

Buregeya (2006:205) also observes this feature in East African English. In his analysis of the feature in East African English, Van Rooy (2015) observes the feature (especially with the verb *enable* which takes the bare infinitive approximately 20% of the time) and states that the feature is expanding and becoming more systematic.

Given the presence of the feature in L1 varieties and L2 varieties, the fact that the bare infinitive marker is used more often with the verb *help* than expected and less frequently with verbs such as *make* and *have*, this feature is investigated in the sections that follow. The feature is investigated with the verbs *force*, *allow*, *assist*, *cause*, and *enable*. The results of the investigation are presented in Section 6.4.2.

6.4.2 Results

The results of the use of the bare infinitive with the verbs *force*, *allow*, *assist*, *cause*, and *enable* are illustrated in Figure 6.8.

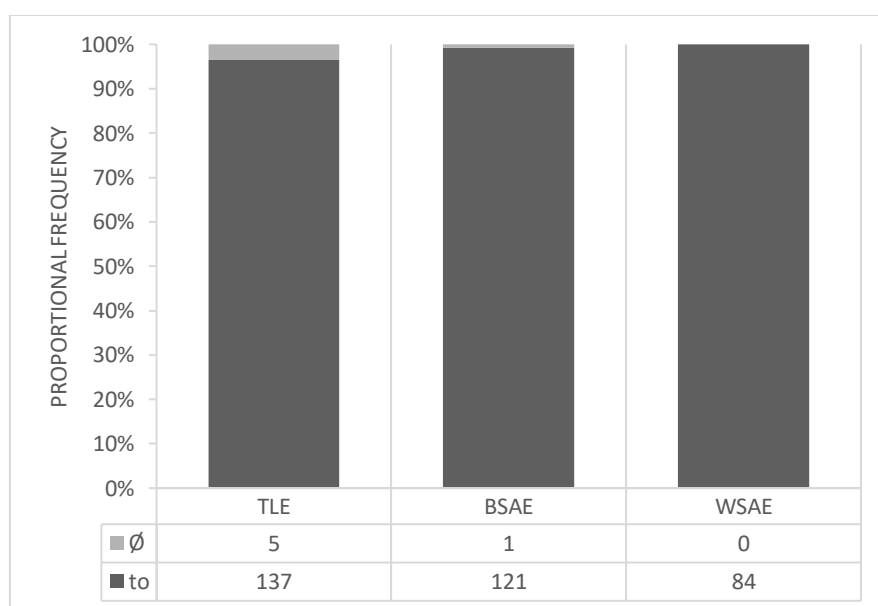


Figure 6.8: The proportional frequencies of the overt infinitive *to*-complementiser and the bare infinitive complementiser with the lemmas *force*, *allow*, *assist*, *cause*, and *enable*

The results demonstrate that the bare infinitive occurs, less than 3.5% of the time in the TLE corpus, once in the BSAE corpus, and never in the WSAE corpus. This suggests that while this feature seldom occurs in the learner variety, it is not used often enough to be

conventionalised in the experienced BSAE variety. Furthermore, the one instance in which the infinitive marker was omitted in the BSAE corpus was modified when the editor inserted the infinitive marker.

A Fisher exact test indicates that the differences between all the corpora are not statistically significant ($p = 0.1406418$). There are also no statistically significant differences between the TLE corpus and the BSAE corpus ($p = 0.2215303$), the BSAE corpus and the WSAE corpus ($p = 1$), and the TLE corpus and the WSAE corpus ($p = 0.160229$).

Consequently, it can be said that the feature does occur in the learner variety but it is never conventionalised in the experienced variety.

6.5 Conclusion

The findings in this chapter show that the BSAE users tend to follow the norms of the WSAE users. In terms of the *help* + infinitive marker constructeme, the bare infinitive allostruction has a slightly lower occurrence in the BSAE corpus than the WSAE corpus, but the difference is not statistically significant. When the results of the TLE corpus were included, it was found that the bare infinitive allostruction occurs less frequently in the TLE corpus than in the BSAE corpus, but that the difference is not significant. The differences in distribution between the TLE and WSAE, however, is significant. This is interesting because it shows that as the BSAE speakers become more experienced, they tend to omit the infinitive *to*-marker more often. This finding suggests that the overt use of the infinitive marker may sometimes be required by the inexperienced language users when they experience high levels of cognitive strain.

In addition, it was found that not all the factors that are said to drive the omission of the infinitive marker in the *help* + infinitive marker constructeme predict infinitive *to*-omission. In this regard, the factors that exhibited the strongest influence were the infinitive use of the lemma *help*, the complexity of the verb phrase, the presence of an intervening noun phrase, and register. The fact that variety was not a predictor indicates

that the BSAE users and the WSAE users experience a similar cognitive strain when using the *help* + infinitive construction.

The undeletion of the infinitive marker did not occur as often as expected. In this study, the proportional undeletion of the infinitive *to*-marker in the TLE seemed to align with the proportional undeletion that Mesthrie (2006) observes in his spoken corpus. In addition, the fact that the undeletion of the infinitive marker occurs less often in the BSAE than in the TLE suggests that this is another transient feature of BSAE, and as the L2 users become more experienced, they tend to use the feature less frequently.

In terms of the deletion of the infinitive *to*-marker, it was found that the BSAE users do not delete the infinitive *to*-marker in constructions where the WSAE users deem the presence of the marker compulsory.

The findings from this chapter can be described against the background of the theories presented in Chapter 2. In terms of the theory on the ecology of language, it can be said that both psycholinguistic and sociolinguistic factors influence the BSAE users' use of the verb + NP infinitive marker construction.

From a psycholinguistic perspective, the learner speakers of BSAE tend to use the infinitive *to*-marker explicitly when they first acquire English. This form is the more analytical form, and it assists the inexperienced speakers in keeping track of how the information contained in the dependent clause relates to the information in the rest of the clause. By using the more transparent form, the pressures of the cognitive constraints are alleviated (Rohdenburg, 1996:151). As the construction becomes more entrenched in the speaker's internal grammar, though, it becomes easier for the speaker to access the structure (Bybee & Thompson, 1997:715).

At the same time, from a sociolinguistic point of view, the learner BSAE speakers come in contact with the WSAE speakers. While the BSAE speakers contribute their explicit forms of the infinitive marker to the pool of linguistic features, they are also exposed to bare infinitive forms that the WSAE speakers and other proficient BSAE speakers contribute to the pool. There, these two features are in competition with one another. Since the BSAE speakers wish to achieve social mobility, they gradually start selecting the bare infinitive

feature more often to accommodate the WSAE speakers. This finding is similar to the findings on phonetic selections reported by Mesthrie (2010) and Mesthrie *et al.* (2015).

This finding emphasises the importance of the distinction that Kachru draws between performance varieties of English and institutionalised varieties of English. While there is an overlap between the performance varieties (represented by the TLE corpus here) and the institutionalised varieties (represented by the BSAE editing corpus), it is important to recognise that not all features of the performance varieties will necessarily be represented as strongly in the institutionalised varieties (as is the case with the non-deletion and the undeletion of the infinitive *to*-marker in BSAE).

From the aforementioned, it becomes clear that the language users from the IDG strand are not in the process of establishing a new norm in terms of the non-deletion and the undeletion of the infinitive *to*-marker, but are instead moving towards the STL strand norms established by the WSAE language users. To the extent that they contribute new forms to the feature pool and eventual selection, the innovations related to infinitive clauses do not seem to become conventionalised in the shared norms of the STL and IDG strands.

Chapter 7

Conclusion

7.1 Introduction

Early studies on BSAE provided lists of characteristic features of the variety (Buthelezi, 1995; De Klerk & Gough, 2002; Gough, 1996). These studies were inclined to not only spot unusual linguistic features but also to overestimate the frequency of these features, all the while overlooking other potentially characteristic features of BSAE. This is typical of what Kahneman (2011:81) refers to as confirmation bias. Consequently, there was doubt whether BSAE is “a coherently defined variety, characteristic of a particular group of speakers, and subject to a sufficient degree of generational transmission as a second language” (Mesthrie, 2006:112).

In 2003, though, De Klerk published an article in which she investigated some of the features regarded as features of BSAE using a spoken corpus of Xhosa English. This study was followed by Mesthrie’s (2006) work on anti-deletions in a corpus of spoken language

and Van Rooy's (2006) description of the progressive aspect in the writing of BSAE learners. Some of these studies include the work of Botha (2013), De Klerk (2006), Minow (2010), Van Rooy (2008a, 2008b, 2014b), and Van Rooy and Terblanche (2010).

However, many of these studies focused on studying BSAE as a learner variety and not as an institutionalised variety. Kachru (1992a:54-55) points out that although learner varieties overlap with institutionalised varieties, they are in fact two distinct varieties, each with its own characteristics. When scholars say that they are investigating a New English variety, they are in fact investigating an institutionalised variety. In light of this distinction, BSAE is investigated as an institutionalised variety in this thesis.

By investigating BSAE as an institutionalised variety, it is possible to determine whether the linguistic features that have been identified in the learner variety of BSAE have in fact conventionalised and found its way into the institutionalised variety of BSAE. This is an important next step in studies on BSAE since theories pertaining to evolutionary linguistics suggest that the features contributed to the linguistic feature pool by learner speakers are in competition with the features contributed by the L1 speakers and by proficient L2 users of SAE. This means that not all features of the learner variety will be fit enough to survive and find its way into the institutionalised varieties. Those features that have survived within the linguistic feature pool and have found their way into the institutionalised variety can no longer be regarded as "errors" since they have been conventionalised by the speakers of the variety.

Since BSAE was investigated as an institutionalised variety, the acceptance of the BSAE features could be measured by observing the changes that the editors made with regard to these features. Kruger and Van Rooy (2017:21) point out that if the linguistic gatekeepers of the publishing industry accept a feature and allow it to be disseminated in published writing, one can assume that not only did the feature conventionalise in the variety, it is no longer regarded as an error, judged in relation to some exonormative standard.

Considering the fact that scholars agree that BSAE is situated in Phase 3 of Schneider's (2003) Dynamic Model, and Schneider (2003:249) identifies innovative verb-complementational patterns as one of the typical ways in which nativisation is reflected

in the grammar of a New English variety, I chose to investigate verb-complementational features of BSAE in this thesis.

7.2 The research questions and aims revisited

7.2.1 *The nature of the verb-complementational profiles of ditransitive verbs in BSAE*

Research question 1: What do the verb-complementational profiles of ditransitive verbs look like in BSAE, and what are the cross-varietal differences and similarities of these profiles between BSAE and WSAE?

It was found that the BSAE language users use the prepositional dative allostruction more often than the WSAE language users do. This corresponds with the findings of Mukherjee and Hoffmann (2006) and Olavarría de Ersson and Shaw (2003). It should be pointed out, though, that the prepositional dative occurs a lot less frequently in WSAE than is reported in British English, and that the proportional frequencies of the BSAE language users' use of the prepositional dative are closer to the proportional frequencies reported for British English than WSAE.

Mukherjee and Hoffmann (2006) and Olavarría de Ersson and Shaw (2003) put forward several hypotheses of both a sociolinguistic nature and psycholinguistic nature to explain this phenomenon. Olavarría de Ersson and Shaw (2003), for example, argue that culture may be a driving force in terms of the allostructional choices that are made. However, no proof could be found for the hypothesis in this study. In turn, Mukherjee and Hoffmann (2006) put forward a psycholinguistic hypothesis, stating that the high frequency of the prepositional dative in Indian English corresponds to the high frequency in which the ditransitive verb *give* is used in light-verb constructions. In this study though, *give* is not used often as a light verb in the BSAE corpus and when the BSAE language users do use it as a light verb, no editorial intervention is observed.

Since none of these hypotheses seemed to be applicable to BSAE, it was proposed that the BSAE users may be using the prepositional dative as a more explicit lexicogrammatical

encoding of the ditransitive construction where the preposition maps the trajectory of the patient more explicitly, encoding the process of transfer more explicitly.

To test this hypothesis, the cognitive constraints known to drive dative alternation were analysed in both BSAE and WSAE. However, the results of this analysis indicated that the same cognitive constraints that drive dative alternation in BSAE drive dative alternation in WSAE. Here, patient length and recipient length (together with the verb lemma) turned out to be the most important constraints. This finding is in line with statements made by Hawkins (1990, 1994) that processing complexity is mitigated when constituents are placed in order of increasing length. In fact, it is such a strong influence that it supersedes the semantic constraints proposed in the meaning-to-structure hypothesis. Variety, on the other hand, did not have a strong predictive value. Consequently, it was posited that the frequency differences can largely be attributed to the uneven pressure of the conditioning grammatical variables, for reasons independent of the grammatical variables as such.

7.2.2 *The relative range and frequency of the anti-deletion profiles in BSAE*

Research question 2: What is the relative range and frequency of the anti-deletion profiles of the finite *that*-complementiser and the non-finite infinitive *to*-marker in BSAE, and what are the cross-varietal differences and similarities of these profiles between BSAE and WSAE?

In Chapter 5, it was observed that previous studies in BSAE point to the non-deletion of the *that*-complementiser and the undeletion of the infinitive *to*-complementiser as potential features of BSAE (Makalela, 2013; Mesthrie, 2006). Since these features have only been investigated in spoken corpora of BSAE and the writing of BSAE learners, it was investigated in a corpus of written texts of BSAE in this study.

The analysis of the anti-deletion of the *that*-complementiser indicated that the omission of the *that*-complementiser occurs less frequently in BSAE than in WSAE and that the difference in this regard is significant. When the results were broken down according to register, it turned out that the zero complementiser occurred more frequently in BSAE

than in WSAE in two of the four registers, i.e. the academic register and the reportage register. The difference between BSAE and WSAE in the academic register with regard to the *that*-complementisers was not significant, and it was found that the BSAE users adhere to the same conventions as the WSAE users.

The only observable difference in the academic register was that the BSAE users tended to undelete the *that*-complementiser in instances where direct speech was used. This only happened in the academic register, though, and the editors left the majority of the undeleted forms unchanged, thus signalling the top-down conventionalisation of the feature.

The difference in the use of the *that*-complementiser in the reportage register, though, was significant. Here the BSAE language users selected the zero complementiser significantly more often than the WSAE users. However, the editors only added a few overt *that*-complementisers. Instead, the editors seemed more inclined to remove a few instances of the overt *that*-complementiser, signalling their acceptance of the BSAE users' use of the feature. Kruger and Van Rooy (2016b:122) point out that the South African publishing industry is dominated by Afrikaans and English. Newspapers, magazines, and books are not only often published in both English and Afrikaans from within the same publishing houses and media companies. This creates a unique contact situation in the external ecology of contact, where the same publishing norms govern the publication of both English and Afrikaans texts in an environment. Consequently, these publishing norms are "characterised by bilingualism [and] some convergence in discourse conventions".

One of these norms that are believed to be the result of the aforementioned contact is the preference for using the zero complementiser with the verb *say* in the reportage register. This norm is analogous to the norm expressed in the influential Afrikaans style guide *Skryf Afrikaans van A tot Z* (Müller & Pistor, 2011:708) which states that the *dat*-complementiser (the equivalent of the *that*-complementiser in Afrikaans) is best omitted after the verb *sê* (the equivalent of the verb *say*) in the reportage register (Kruger & Van Rooy, 2016a:708). The results are compounded by the fact that the verb *say* is one the verbs that take finite complementisers most often (overt or zero). Consequently, in the

majority of the instances where the verb lemma *say* is used, the *that*-complementiser is omitted.

In the instructional and popular registers, though, the overt *that*-complementiser was used significantly more often by the BSAE language users than the WSAE users. From a psycholinguistic perspective, this phenomenon is ascribed to the fact that the BSAE language users are not experienced authors in these registers. Since they are not familiar with these registers, they tend to fall back on overly formal register conventions. This argument is congruent with Rhodenburg's (1995) complexity thesis which posits that "any factors complicating the processing of the structural relationships involved will tend to favour a more explicit rendering of the relevant complement clause". Another possible explanation is offered by Van Rooy *et al.* (2010:346). They posit that in most post-colonial Englishes, the more formal forms of language are usually chosen because the language users' are exposed more often to formal registers in English because their home languages serve their private functions (also cf. Coetzee-Van Rooy, 2014). Similarly, Kruger and Van Rooy (2016a:125) state that in cases where the L2 language users are primarily exposed to English through formal texts, they are more likely to choose the entrenched overt complementiser as opposed to the zero form.

An analysis of the factors driving *that*-omission revealed that the factors driving *that*-omission in BSAE are the same factors driving *that*-omission in WSAE. The two factors with the highest predictive value were lemma and register. In this analysis, variety was not a highly predictive factor.

In Chapter 6, anti-deletion was investigated in the verb + NP + infinitive marker constructeme. First, this feature was investigated with the verb *help*. The initial findings did not show a significant difference in the use of the *help* + infinitive marker constructeme between BSAE and WSAE. Since the overt infinitive seemed to occur slightly more often in the BSAE corpus than in the WSAE corpus, it was decided to include the TLE corpus, which is representative of the learner variety of BSAE. The inclusion of the learner variety in the analysis revealed that there is a statistically significant difference in the use of the infinitive marker in the *help* + infinitive marker constructeme in the TLE and WSAE, indicating that the non-deletion of the infinitive marker in the *help* + infinitive marker constructeme is indeed a feature of the learner variety. This means that although the form

with the undeleted infinitive marker is added to the linguistic feature pool, it is not fit enough to be conventionalised in the institutionalised variety of BSAE.

An analysis of the factors driving the omission of the infinitive marker confirmed that variety is not a predictive variable. Instead, the same variables that drive the omission of the infinitive marker in BSAE drive the omission of the infinitive marker in WSAE.

When previous studies on BSAE listed “idiosyncratic” patterns of verb complementation as one of the features of BSAE, the undeletion of the infinitive *to*-marker in the verb + NP + \emptyset allostruction were quoted as examples. In his study of this feature in a spoken corpus of BSAE, Mesthrie (2006) has observed the undeletion of the infinitive marker in 34.5% of the occurrences of the verb + NP + \emptyset allostruction. In this study, though, the BSAE speakers undeleted the infinitive marker in only 12.5% of the occurrences of the verb + NP + \emptyset allostruction, and the difference in the use of the allostruction between BSAE and WSAE was not statistically significant. Consequently, the TLE corpus was once again included in the analysis. The learner speakers of BSAE undeleted the infinitive marker in 33.3% of the occurrences of the verb + NP + \emptyset allostruction. Although the difference between the learner variety and the institutionalised variety was not statistically significant, the difference between the learner variety and WSAE was significant. It would seem as if the undeletion of the infinitive marker in the verb + NP + \emptyset allostruction is a feature of the learner variety of BSAE that is not conventionalised in the institutionalised variety of BSAE.

Since the deletion of the infinitive marker in the verb + NP + *to* allostruction is reported in L1 varieties of English and East African English, the feature was also investigated here. Although there are a few instances of deletion in the verb + NP + *to* allostruction in the TLE (i.e. less than 4%), and one instance thereof in the BSAE, there are not enough instances of deletion to qualify it as a feature of BSAE. The results revealed that this is not a feature of BSAE or WSAE.

7.2.3 *The extent to which the BSAE features have stabilised*

Research question 3: To what extent have these constructions stabilised in BSAE?

Variety was not the most significant predictive variable in the analysis of the use of the ditransitive construction by the BSAE language users and the WSAE language users. Furthermore, the only other observable difference in the use of the BSAE and WSAE language users' use of the ditransitive construction, is that the BSAE language users seem to use the prepositional dative construction more often than the WSAE language users do.

In terms of the anti-deletion of the *that*-complementiser and the infinitive *to*-complementiser, variety also was not the strongest predictive variable. In the registers in which the BSAE language users often work (i.e. the academic and reportage registers), the BSAE users use the zero complementiser more often than the WSAE language users. In terms of the infinitive *to*-complementiser, the BSAE users seem to follow the norms of the WSAE users. With the inclusion of the TLE data, it becomes clear that the more experienced the BSAE speakers become, the more they adhere to the norms of the of the WSAE speakers.

Considering the findings as a whole, it is clear that as the BSAE speakers become more experienced and the features become more entrenched in their internal grammar, they experience fewer cognitive constraints and no longer rely as much on the more analytical grammatical forms. Consequently, when they select the features for use from the linguistic feature pool (to which other BSAE speakers and the WSAE speakers contribute features), they tend to choose the zero complementiser and the bare infinitive marker more often to accommodate the WSAE speakers. It is posited that the BSAE users accommodate the WSAE users readily, because English serves the purpose of social mobility, and does not form part of the BSAE language user's identity.

Given these findings, it seems as if the norms as they pertain to the linguistic features investigated in this thesis are aligned across the STL and IDG components of the broader SAE community, and that the BSAE learner usage is not in the process of establishing a radically different norm from the WSAE norm. These findings support the findings of

Mesthrie *et al.* (2015) and Mesthrie (2010). In light of this alignment of norms, there is no new feature that are on its way to being stabilised.

7.2.4 *The social acceptance of the BSAE features*

Research question 4: Have these constructions obtained social acceptance in the speech community (i.e. are these stabilised constructions accepted as conventionalised innovations)?

Considering the input from the editors and the minor changes they made overall, it is clear that the linguistic gatekeepers of the publishing industry in South Africa have accepted the BSAE language users' use of the linguistic features investigated in this study. Given the fact that the BSAE speakers are accommodating the WSAE language users to a large extent, and that their norms in terms of the linguistic features studied in this thesis are gradually stabilising towards the norms set by the WSAE speakers, it is not surprising that the editors accept the norms that are so similar to their own.

7.2.5 *The theoretical implications of the findings for the current scholarly discourse in New Englishes*

Research question 5: What are the theoretical implications of the findings with regard to the process of stabilisation for the current scholarly discourse in New Englishes?

The findings of this study have shown that the BSAE norm is gradually conforming to the norms established for WSAE. In terms of the use of the ditransitive constructions, the use of the *that*-complementiser, and the use of the infinitive *to*-marker, it is clear that the innovations contributed to the linguistic feature pool by BSAE learners are not becoming conventionalised in the shared norms of the IDG and STL strands. Instead, the BSAE language users seem to be accommodating the norms of the WSAE language users.

In light of some of the issues that cropped up in Chapter 2, these findings should perhaps not be entirely surprising. In Section 2.3.3, for example, it was mentioned that scholars

have found that English is not an “identity carrier” in BSAE (Coetzee-Van Rooy, 2012; Mesthrie, 2015:83; Morreira, 2012:197-198) and that BSAE language users use English to obtain social mobility instead. In the literature, scholars have pointed to this fact to explain why BSAE users confine their use of English to specific domains and are not replacing their home languages with English. In addition to identity, the contact between the IDG speakers and the STL speakers, whose norms have already been established to a larger degree, is more sustained than in other postcolonial settings. This has some important implications for Schneider’s (2003, 2007) Dynamic Model.

It is important to apply Schneider’s Dynamic Model to the contact between individual IDG and STL strands within their individual local settings before attempting to apply the Dynamic Model to an entire country as suggested by Van Rooy (2014a:22). This approach will provide for the “successive ‘waves’ of nativisation” found in countries with complex colonial histories account (e.g. South Africa) and explain why a variety such as WSAE finds itself in the phase of endonormative stabilisation, while a variety such as BSAE finds itself in the phase of nativisation (also cf. Bekker, 2009:432-433).

This approach also emphasises the importance of guarding against viewing concentric circles of Kachru’s model as hermetically sealed units that overlook the linguistic variation that exist in each of the circles. In the South African context, for example, English is an L1 for some language users, but an L2 (and sometimes an L3 or even an L4) for others. Furthermore, these language users are in constant contact with one another, influencing the way in which both the L1 and L2 language users use English.

At the same time, it is also important to disentangle the notions of ENDONORMATIVITY and HOMOGENEITY. Although these two processes often run concurrently, it is important not to conflate them into a single process to account for the fact that some varieties of English have achieved a semblance of endonormativity but not homogeneity (Van Rooy, 2014a:35-36).

Another important fact to consider is the role of identity in Schneider’s (2003, 2007) Dynamic Model. In the case of BSAE, English has not become an “identity carrier” for BSAE speakers. This explains why a radically new norm has not been established in BSAE. Since the BSAE speakers acquire English to obtain social and economic mobility and not to

express identity, they tend to accommodate the WSAE norm to a large extent. On the one hand, the similarities in the use of the *that*-complementiser in the academic register and the overall similarities in the use of the infinitive *to*-complementiser between the BSAE and WSAE language users suggest that the BSAE norms seem to be converging with the WSAE norms (e.g. the similar ways in which the infinitive *to*-marker is used in BSAE and WSAE). On the other hand, the WSAE language users seem to accept some of the innovative uses of BSAE (e.g. the high rate of *that*-omission in the reportage register and the low rate of *that*-omission in the instructional and popular registers). These findings suggest that the BSAE norm may also be influencing WSAE, but to a smaller extent. This confirms that a semblance of endonormativity is achieved, while homogeneity is not established but inclined to move closer to the STL norm.

These findings are congruent with the findings of Mesthrie (2010) and Mesthrie *et al.* (2015) and provide support for Van Rooy's (2014a:35) view that, unlike CSAE and ISAE, BSAE may be in danger of being absorbed into the STL strand.

7.3 Limitations and future research possibilities

Although this study has demonstrated the importance of studying BSAE not only in learner corpora but also in corpora containing the language of experienced BSAE language users, there are several limitations to this study which provide possibilities for future research.

In the first place, the features investigated in this study were only investigated in written registers. Since previous studies investigated the spoken language of BSAE speakers that could be considered learners of English, it would be interesting to see whether the experienced BSAE speakers conform to the norms of standard English to the same extent in the spoken registers as they do in the written registers.

In the second place, more features of BSAE need to be analysed in corpora of experienced BSAE since not enough features are studied here to come to a final conclusion about the extent to which BSAE will be absorbed into WSAE or retain some unique traces. However, if the findings of this study are read alongside the evidence for an innovative progressive

construction in BSAE (cf. Kruger & Van Rooy, 2017), then there is evidence for a limited degree of uniqueness in BSAE, but at the same time, indications are that such a difference will not be as extensive as suggested by claims made on the basis of learner data.

In the case where a higher proportion of prepositional datives was found in BSAE than in WSAE, it would be useful to conduct a diachronic study to determine whether the difference can be ascribed to the preservation of the ditransitive patterns of an earlier form WSAE or to postcolonial divergence.

Finally, Gut (2014:118-121) argues that the acceptance of linguistic features is dependent on the attitudes of the speakers themselves. Therefore, future studies should include the distribution of acceptability questionnaires to BSAE speakers to determine the extent to which they themselves find the features described in this study acceptable.

Bibliography

A Dictionary of South African English on Historical Principles. 2010. Edited by Penny Silva, Wendy Dore, Dorothea Mantzel, Colin Muller, Madeleine Wright. London and New York: Oxford University Press.

Achebe, C. 2003. The African writer and the English language. (In Okpewho, I., ed. Chinua Achebe's *Things fall apart*: a casebook. Oxford and New York: Oxford University Press. pp. 55-65.)

Aissen, J. 1999. Markedness and subject choice in optimality theory. *Natural Language and Linguistic Theory*, 17(4):673-711.

Aitchison, J. 1991. *Language change: progress or decay?* 2nd ed. Cambridge and New York: Cambridge University Press.

Altenberg, B. & Granger, S. 2001. The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics*, 22(2):173-194.

Arnold, J.E., Losongco, A., Wasow, T. & Ginstrom, R. 2000. Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1):28-55.

- Atkins, S., Clear, J. & Ostler, N. 1992. Corpus design criteria. *Literary and Linguistic Computing*, 7(1):1-16.
- Bamgboṣe, A. 1992. Standard Nigerian English: issues of identification. (In Kachru, B., ed. *The other tongue: Englishes across cultures*. Urbana and Chicago: University of Illinois Press. pp. 148-161.)
- Bamgboṣe, A. 1998. Torn between the norms: innovations in world Englishes. *World Englishes*, 17(1):1-14.
- Bardhal, J. 2008. Productivity: evidence from case and argument structure in Icelandic. Amsterdam and Philadelphia: John Benjamins.
- Barlow, M. 2000. Usage, blends, and Grammar. (In Barlow, M. & Kemmer, S., eds. *Usage-based models of language*. Cambridge: Cambridge University Press. pp. 315-345.)
- Bekker, I. 2009. The vowels of South African English. Potchefstroom: North-West University. (Thesis – PhD).
- Bekker, I. 2012a. South African English as a late 19th-century extraterritorial variety. *English World-Wide*, 33(2):127-146.
- Bekker, I. 2012b. The story of South African English: a brief linguistic overview. *International Journal of Language, Translation and Intercultural Communication*, 1:139-150.
- Bernaisch, T., Gries, S.T. & Mukherjee, J. 2014. The dative alternation in South Asian English(es): modelling predictors and predicting prototypes. *English World-Wide*, 35(1):7-31.
- Berry-Rogghe, G.L.M. 1974. Automatic identification of phrasal verbs. (In Mitchell, J.L., ed. *Computers in humanities*. Edinburgh: Edinburgh University Press. pp. 16-26.)
- Bever, T.G., Carroll, J.M. & Hurtig, R.R. 1976. Analogy or ungrammatical sequences that are utterable and comprehensible are the origins of new grammars in language acquisition and linguistic evolution. (In Bever, T.G., Katz, J.J. & Langendoen, D.T., eds. *An integrated theory of linguistic ability*. New York: Crowell Press. pp. 149-182.)
- Biber, D. 1993. Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4):243-257.
- Biber, D. 1999. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies*, 1(2):131-150.
- Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. Longman grammar of spoken and written English. Essex: Longman.
- Boas, H.C. 2013. Cognitive Construction Grammar. (In Hoffmann, T. & Trousdale, G., eds. The Oxford handbook of Construction Grammar. Oxford: Oxford University Press. pp. 233-252.)
- Bolton, K. & Botha, W. 2015. Researching English in Contemporary China. *World Englishes*, 34(2):169-174.
- Botha, Y. 2013. Corpus evidence of anti-deletion in Black South African English noun phrases. *English Today*, 29(1):16-21.
- Botha, Y.V. 2012. Specification in the English nominal group with reference to student writing. Potchefstroom: North-West University. (Thesis – PhD).
- Bowker, L. & Pearson, J. 2002. Working with specialized language: a practical guide to using corpora. London and New York: Routledge.
- Boye, K. & Harder, P. 2007. Complement-taking predicates: usage and linguistic structure. *Studies in Language*, 31(3):596-606.
- Boylan, P. 2009. Accommodation theory revisited again. (In Fatigante, M., Mariottini, L. & Sciubba, M.E., eds. *Lingua e società: scritti in onore di franca orletti*. Milano: FrancoAngeli. pp. 287-305.)
- Brandford, W. 1994. English in South Africa. (In Burchfield, R., ed. *The Cambridge history of the English language. Volume V: English in Britain and overseas – origins and development*. Cambridge: Cambridge University Press. pp. 430-496.)
- Branford, W. 1996. English in South Africa. (In De Klerk, V., ed. *Focus on South Africa*. Amsterdam: John Benjamins. pp. 35-52.)
- Bresnan, J., Cueni, A., Nikitina, T. & Baayen, R.H. 2007. Predicting the dative alternation. (In Bouma, G., Krämer, I. & Zwarts, J., eds. *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Arts and sciences. pp. 69-94.)
- Bresnan, J. & Ford, M. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language*, 86(1):168-213.
- Bresnan, J. & Hay, J. 2008. Gradient grammar: an effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua*, 118(2):245-259.
- Bresnan, J. & Nikitina, T. 2007. The gradience of dative alternation. (In Uyechi, L.A. & Wee, L., eds. *Reality exploration and discovery: pattern interaction in language and life*. Stanford: CSLI Publications. 161-184.)

- Bruthiaux, P. 2003. Squaring the circles: issues in modeling English worldwide. *International Journal of Applied Linguistics*, 13(2):159-178.
- Buregeya, A. 2006. Grammatical features of Kenyan English and the extent of their acceptability. *English World-Wide*, 27(2):199-216.
- Buthelezi, Q. 1995. South African Black English. (In Mesthrie, R., ed. *Language and social history: studies in South African sociolinguistics*. Claremont: David Philip. pp. 242-250.)
- Bybee, J. 2006. From usage to grammar: the mind's response to repetition. *Language*, 82(4):711-733.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J. 2013. Usage-based theory and exemplar representations of constructions. (In Hoffmann, T. & Trousdale, G., eds. *The Oxford handbook of Construction Grammar*. Oxford and New York: Oxford University Press. pp. 49-69.)
- Bybee, J. 2015. *Language change*. Cambridge: Cambridge University Press.
- Bybee, J. & Scheibman, J. 1999. The effect of usage on degrees of constituency: the reduction of *don't* in English. *Linguistics: an Interdisciplinary Journal of the Language Sciences*, 37(4):575-596.
- Bybee, J. & Thompson, S. 1997. Three frequency effects in syntax. *Berkeley Linguistic Society*, 23:378-388.
- Callies, M. 2013. Bare infinitival complements in present-day English. (In Aarts, B., Close, J., Leech, G. & Wallis, S., eds. *The verb phrase in English: investigating recent language change with corpora*. Cambridge: Cambridge University Press. 239-255.)
- Cappelle, B. 2006. Particle placement and the case for "allostructions". *Constructions*, Special Volume 1(1):1-28.
- Chambers, J.K. 2003. *Sociolinguistic theory*. 2nd ed. Oxford: Blackwell.
- Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. 1959. Review of verbal behavior by B.F. Skinner. *Language*, 35(1):26-58.
- Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. 1986. *Knowledge of language: its nature, origin, and use*. New York and London: Praeger.

- Christiansen, M.H. & Kirby, S. 2003. Language evolution: consensus and controversies. *Trends in Cognitive Sciences*, 7(7):300-307.
- Church, K.W., Gale, W., Hanks, P. & Hindle, D. 1991. Using statistics in lexical analysis. (In Zernik, U., ed. *Lexical acquisition: exploiting on-line resources to build up a lexicon*. Hillsdale: Lawrence Erlbaum. pp. 115-164.)
- Coetzee-Van Rooy, S. 2010. The importance of being multilingual. Inaugural lecture. Unpublished manuscript: North-West University.
- Coetzee-Van Rooy, S. 2012. Flourishing functional multilingualism: evidence from language repertoires in the Vaal Triangle region. *International Journal of the Sociology of Language*, 2012(218):87-119.
- Coetzee-Van Rooy, S. 2014. The identity issue in bi- and multilingual repertoires in South Africa: implications for Schneider's Dynamic Model. (In Buschfeld, S., Hoffmann, T., Huber, M. & Kautzsch, A., eds. *The evolution of Englishes: the Dynamic Model and beyond*. Amsterdam: John Benjamins. pp. 39-59.)
- Coetzee-Van Rooy, S. & Van Rooy, B. 2005. South-African English: labels, comprehensibility and status. *World Englishes*, 24(1):1-19.
- Coetzee-Van Rooy, S. & Verhoef, M.M. 2000. Perceptions of English proficiency: views from Southern Sotho speakers. *Southern African Journal of Linguistics*, 18(Supplement 38):164-185.
- Collins, P. 1995. The indirect object construction in English: an informational approach. *Linguistics*, 33(1):35-49.
- Conover, W.J. 1999. *Practical nonparametric statistics*. 3rd ed. New York: John Wiley.
- Cook, V. 2003. Introduction: the changing L1 in the L2 user's mind. (In Cook, V., ed. *Effects of the second language on the first*. Clevedon and Buffalo: Multilingual Matters. pp. 1-18.)
- Cook, V., Iarossi, E., Stellakis, N. & Tokumaru, Y. 2003. Effects of the L2 on the syntactic processing of the L1. (In Cook, V., ed. *Effects of the second language on the first*. Clevedon and Buffalo: Multilingual Matters. pp. 193-213.)
- Corder, S.P. 1967. The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(14):161-170.
- Croft, W. 1995. Autonomy and functional linguistics. *Language*, 71(3):490-532.
- Croft, W. 1998. Linguistic evidence and mental representations. *Cognitive Linguistics*, 9(2):151-173.

Croft, W. 2000. Explaining language change: an evolutionary approach. Essex: Pearson Education.

Croft, W. 2001. Radical Construction Grammar: syntactic theory in typological perspective. Oxford: Oxford University Press.

Crystal, D. 1997. English as a global language. Cambridge: Cambridge University Press.

Culicover, P.W. & Jackendoff, R. 1999. The view from the periphery: the English comparative correlative. *Linguistic Inquiry*, 30(4):543-571.

Darwin, C. 1871/1981. The descent of man, and selection in relation to sex. (With an introduction by John Tyler Bonner and Robert, M. May). New Jersey: Princeton University Press.

Davydova, J. 2012. Englishes in the Outer and Expanding Circles: a comparative study. *World Englishes*, 31(3):366-385.

Dawkins, R. 1989. The selfish gene. 2nd (30th anniversary) ed. Oxford: Oxford University Press.

De Cuypere, L. & Verbeke, S. 2013. Dative alternation in Indian English: a corpus-based analysis. *World Englishes*, 32(2):169-184.

De Groot, A.M.B. & Christoffels, I.K. 2006. Language control in bilinguals. *Bilingualism: Language and Cognition*, 9(2):189-201.

De Klerk, V. 1997. Encounters with English over three generations in a Xhosa family: for better or for worse? (In Schneider, E.W., ed. Englishes around the world. (Volume 2: Caribbean, Africa, Asia, Australasia). Amsterdam and Philadelphia: John Benjamins. pp. 97-118.)

De Klerk, V. 1999. Black South African English: Where to from here? *World Englishes*, 18(3):311-324.

De Klerk, V. 2003. Towards a norm in South African Englishes: the case for Xhosa English. *World Englishes*, 22(4):463-481.

De Klerk, V. 2006. Corpus linguistics and World Englishes: an analysis of Xhosa English. London and New York: Continuum.

De Klerk, V. & Gough, D. 2002. Black South African English. (In Mesthrie, R., ed. Language in South Africa. Cambridge: Cambridge University Press. pp. 356-378.)

Dediu, D., Cysouw, M., Levinson, S.C., Baronchelli, A., Christiansen, M.H., Croft, W., Evans, N., Garrod, S., Gray, R.D., Kandler, A. & Lieven, E. 2013. Cultural evolution of language.

(In Richerson, P.J. & Christiansen, M.H., eds. Cultural evolution: society, technology, language, and religion. Cambridge and Massachusetts: MIT Press. pp. 303-332.)

Demuth, K., Odatu, C., Machobane, M. & Moloji, F. 2014. The acquisition of Sesotho double object constructions. *IULC Working Papers*, 1(1):13-24.

Deshors, S.C. 2014. A case for a unified treatment of EFL and ESL: a multifactorial approach. *English World-Wide*, 35(3):277-305.

Deumert, A. 2010a. Imbodela zamakhumsha: reflections on standardization and destandardization. *Multilingua*, 29(34):243-264.

Deumert, A. 2010b. Tracking the demographics of (urban) language shift: an analysis of South African census data. *Journal of Multilingual and Multicultural Development*, 31(1):13-35.

Deumert, A. 2011. Multilingualism. (In Mesthrie, R., ed. The Cambridge handbook of sociolinguistics. Cambridge and New York: Cambridge University Press. pp. 259-282.)

Deutscher, G. 2005. The unfolding of language: the evolution of mankind's greatest invention. London: Arrow Books.

Dixon, R.M.W. 2005. A semantic approach to English grammar. 2nd ed. Oxford: Oxford University Press.

Dor, D. 2005. Toward a semantic account of *that*-deletion in English. *Linguistics*, 43(2):345-382.

Du Plessis, G. 2015. Regionality in White South African English: an acoustic dialectometric investigation. Potchefstroom: NWU. (Dissertation – MA).

Duffley, P.J. 1992. The English infinitive. London and New York: Routledge.

Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61-74.

Eaton, L. & Louw, J. 2000. Culture and self in South Africa: individualism-collectivism predictions. *The Journal of Social Psychology*, 140(2):210-217.

Eckert, P. 2000. Linguistic variation as social practice: the linguistic construction of identity in Belten High. Massachusetts and Oxford: Blackwell.

Egan, T. 2008. Non-finite complementation: a usage-based study of infinitive and *-ing* clauses in English. Amsterdam and New York: Rodopi.

Erteschik-Shir, N. 1979. Discourse constraints on dative movement. (In Givón, T., ed. *Syntax and Semantics: discourse and syntax*. New York: Academic Press. pp. 441-467.)

Fillmore, C.J. 1992. "Corpus linguistics" or "computer-aided armchair linguistics". (In Svartvik, J., ed. *Directions in corpus linguistics: proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin and New York: Mouton de Gruyter. pp. 35-60.)

Fillmore, C.J., Kay, P. & O'Connor, M.C. 1988. Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language*, 64(3):501-538.

Giles, H. 1973. Accent mobility: a model and some data. *Anthropological Linguistics*, 15(2):87-105.

Giliomee, H. & Mbenga, B. 2007. *New History of South Africa*. Cape Town: Tafelberg.

Gill, S.K. 1999. Standards and emerging linguistic realities in the Malaysian workplace. *World Englishes*, 18(2):215-231.

Gilquin, G. 2015. At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared. *English World-Wide*, 36(1):91-124.

Gilquin, G. & Granger, S. 2011. From EFL to ESL: evidence from the *International Corpus of Learner English*. (In Mukherjee, J. & Hundt, M., eds. *Exploring second-language varieties of English and learner Englishes: bridging a paradigm gap*. Amsterdam and Philadelphia: John Benjamins. pp. 55-78.)

Gilquin, G. & Gries, S.T. 2009. Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5(1):1-26.

Givón, T. 2005. *Context as other minds: the pragmatics of sociality, cognition and communication*. Amsterdam and Philadelphia: John Benjamins.

Goldberg, A.E. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A.E. 2002. Surface generalizations: an alternative to alternations. *Cognitive Linguistics*, 13(4):327-356.

Goldberg, A.E. 2003. *Constructions: a new theoretical approach to language*. *Trends in Cognitive Sciences*, 7(5):219-224.

Goldberg, A.E. 2006. *Constructions at work: the nature of generalization in language*. Oxford and New York: Oxford University Press.

- Götz, S. & Schilk, M. 2011. Formulaic sequences in spoken ENL, ESL and EFL: focus on British English, Indian English and learner English of advanced German learners. (*In* Mukherjee, J. & Hundt, M., eds. *Exploring second-language varieties of English and learner Englishes: bridging a paradigm gap*. Amsterdam and Philadelphia: John Benjamins. pp. 79-100.)
- Göttsche, H. 2013. *Deviational syntactic structures*. London & New York: Bloomsbury.
- Gough, D. 1996. Black English in South Africa. (*In* De Klerk, V., ed. *Focus on South Africa*. Amsterdam: John Benjamins. pp. 53-77.)
- Graham, A. 2013. *Statistics: a complete introduction* London: Hodder & Stoughton.
- Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. eds. 2009. *International corpus of learner English*. Louvain: Presses universitaires de Louvain.
- Green, G.M. 1974. *Semantics and syntactic regularity*. Illinois: University of Illinois.
- Gries, S.Th. 2003a. Multifactorial analysis in corpus linguistics: a study of particle placement. London and New York: Continuum.
- Gries, S.Th. 2003b. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, 1(1):1-27.
- Gries, S.Th. 2006a. Corpus-based methods and cognitive semantics: the many senses of *to run*. (*In* Gries, S.T. & Stefanowitsch, A., eds. *Corpora in Cognitive Linguistics: corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter. pp. 57-99.)
- Gries, S.Th. 2006b. Introduction. (*In* Gries, S.T. & Stefanowitsch, A., eds. *Corpora in Cognitive Linguistics: corpus-based approaches to syntax and Lexis*. Berlin: Mouton de Gruyter. pp. 1-17.)
- Gries, S.Th. 2006c. Some proposals towards more rigorous corpus linguistics. *Zeitschrift für Anglistik und Amerikanistik*, 54(2):191-202.
- Gries, S.Th. 2007. Coll.analysis 3.2a. A program for R for Windows 2.x.
<http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/coll.analysis.r>
Date of access: 15 June 2018.
- Gries, S.Th. 2010. Dispersions and adjusted frequencies in corpora: further explorations. (*In* Gries, S.Th., Wulff, S. & Davies, M., eds. *Corpus-linguistics applications: current studies, new directions*. Amsterdam and New York: Routledge. pp. 197-212.)
- Gries, S.Th. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: some necessary clarifications. *Studies in Language*, 36(3):477-510.

Gries, S.Th. 2013a. Data in Construction Grammar. (*In Hoffmann, T. & Trousdale, G., eds. The Oxford handbook of Construction Grammar. Oxford: Oxford University Press. pp. 93-108.*)

Gries, S.Th. 2013b. Statistics for linguistics with R: a practical introduction. 2nd ed. Berlin and Boston: De Gruyter Mouton.

Gries, S.Th. & Stefanowitsch, A. 2004. Extending collocation analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1):97-129.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R. & Wilson, R. 1989. The learnability and acquisition of the dative alternation in English. *Language*, 65(2):203-257.

Grosjean, F. 2011. An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16(1):11-21.

Grosjean, F. 2013. Bilingualism: a short introduction. (*In Grosjean, F. & Li, P., eds. The psycholinguistics of bilingualism. Massachusetts and Oxford: Wiley-Blackwell. pp. 5-25.*)

Gut, U. 2011. Studying structural innovations in New English varieties. (*In Mukherjee, J. & Hundt, M., eds. Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap. Amsterdam: John Benjamins. pp. 101-124.*)

Halliday, M. & Matthiessen, C. 2014. Halliday's introduction to functional grammar. 3rd ed. London and New York: Routledge.

Halliday, M.A. & Matthiessen, C.M. 2004. An introduction to functional grammar. 3rd ed. London: Arnold.

Haspelmath, M. 2015. Ditransitive constructions. *Annual Review of Linguistics*, 1(1):19-41.

Haugen, E. 1966. Dialect, language, nation. *American Anthropologist*, 68(4):922-935.

Haugen, E. 1972/2001. The ecology of language. (*In Fill, A. & Mühlhäusler, P., eds. The ecolinguistics reader: language, ecology and environment. London and New York: Continuum. pp. 57-66.*)

Hawkins, J.A. 1990. A parsing theory of word order universals. *Linguistic Inquiry*, 21(2):223-261.

Hawkins, J.A. 1994. A performance theory of order and constituency. Cambridge and New York: Cambridge University Press.

- Hawkins, J.A. 2003. Why are zero-marked phrases close to their heads? (*In* Kortmann, B. & Traugott, E.C., eds. *Determinants of grammatical variation in English*. Berlin and New York: Mouton de Gruyter. pp. 175-204.)
- Hawkins, J.A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
- Hickey, R. 2012. Internally and externally motivated language change. (*In* Hernández-Campoy, J.M. & Conde-Silvestre, J.C., eds. *The handbook of historical sociolinguistics*. West Sussex and Malden: Wiley Blackwell. pp. 388-405.)
- Hilpert, M. 2008. The English comparative – language structure and language use. *English Language and Linguistics*, 12(3):395-417.
- Hintzman, D.L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4):411-428.
- Hirson, B. 1981. Language in control and resistance in South Africa. *African Affairs*, 80(319):219-237.
- Hoffmann, S. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition. (*In* Lindquist, H. & Mair, C., eds. *Corpus approaches to grammaticalization in English*. Amsterdam and Philadelphia: John Benjamins. pp. 171-210.)
- Hoffmann, T. 2011. *Preposition placement in English: a usage-based approach*. Cambridge: Cambridge University Press.
- Hoffmann, T. & Trousdale, G. 2011. Variation, change and constructions in English. *Cognitive Linguistics*, 22(1):1-23.
- Hoffmann, T. & Trousdale, G. 2013. Construction Grammar: introduction. (*In* Hoffmann, T. & Trousdale, G., eds. *The Oxford handbook of Construction Grammar*. Oxford: Oxford University Press. pp. 1-12.)
- Hofstede, G. 2011. Dimensionalizing cultures: the Hofstede model in context. *Online Readings in Psychology and Culture*, 12(1):1-26.
- Hofstede Insights. 2018. Compare countries. <https://www.hofstede-insights.com/product/compare-countries/> Date of access: 30 September 2018.
- Hopper, P. 1987. Emergent grammar. *Berkeley Linguistic Society*, 13:139-157.
- Hopper, P. & Traugott, E.C. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

- Hothorn, T., Hornik, K. & Zeileis, A. 2015. *ctree*: Conditional inference trees. CRAN.R_project. <https://rdrr.io/rforge/partykit/f/inst/doc/ctree.pdf> Date of access: 3 October 2018.
- Huddleston, R. 2002a. Content clauses and reported speech. (*In* Huddleston, R. & Pullum, G.K., *eds.* The Cambridge grammar of the English language. Cambridge: Cambridge University Press. pp. 947-1030.)
- Huddleston, R. 2002b. Non-finite and verbless clauses. (*In* Huddleston, R. & Pullum, G.K., *eds.* The Cambridge grammar of the English language. Cambridge: Cambridge University Press. pp. 1171-1271.)
- Hudson, R.A. 1996. Sociolinguistics. 2nd ed. Cambridge: Cambridge University Press.
- Hull, D.L. 1988. Science as a process: an evolutionary account of the social and conceptual development of science. Chicago and London: University of Chicago Press.
- Hundt, M. 1998. New Zealand English grammar: fact or fiction. Amsterdam and Philadelphia: John Benjamins.
- Hundt, M. & Mukherjee, J. 2011a. Introduction: bridging a paradigm gap. (*In* Mukherjee, J. & Hundt, M., *eds.* Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap. Amsterdam: John Benjamins. pp. 1-6.)
- Hundt, M. & Mukherjee, J. 2011b. Discussion forum: New Englishes and learner Englishes – *quo vadis?* (*In* Mukherjee, J. & Hundt, M., *eds.* Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap. Amsterdam: John Benjamins. pp. 209-217.)
- Hurford, J.R. 2012. The origins of grammar: language in the light of evolution. Oxford and New York: Oxford University Press.
- Infogistics. 2001. Tag-Set. <http://www.infonetware.com/tagset.html> Date of access: 15 November 2018.
- Jaeger, T.F. Optional *that* indicates production difficulty: evidence from disfluencies. (*In* ISCA, *ed.* Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop, 2005 Aix-en-Provence, France. ISCA. pp. 103-108.)
- Jarvis, S. 2003. Probing the effects of the L2 on the L1: a case study. (*In* Cook, V., *ed.* Effects of the second language on the first. Clevedon and Buffalo: Multilingual Matters. pp. 81-102.)
- Jenkins, R. 2008. Social identity. 3rd ed. Oxon and New York: Routledge.

Kachru, B. 1982. The bilingual's linguistic repertoire. (*In* Hartford, B., Valdman, A. & Foster, C.R., eds. *Issues in international bilingual education: the role of the vernacular*. New York and London: Plenum Press. pp. 25-52.)

Kachru, B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. (*In* Quirk, R. & Widdowson, H., eds. *English in the world: teaching and learning the language and literatures*. (Papers of an international conference entitled 'Progress in English Studies' held in London, 17-21 September 1984 to celebrate the fiftieth anniversary of the British Council and its contribution to the field of English Studies over fifty years). Cambridge and New York: Cambridge University Press (for the British Council). pp. 11-30.)

Kachru, B. 1991. Liberation linguistics and the Quirk concern. (*In* Tickoo, M.L., ed. *Languages and standards: issues, attitudes, case studies*. Singapore: SEAMEO Regional Language Centre. pp. 206-226.)

Kachru, B. 1992a. Models for non-native Englishes. (*In* Kachru, B., ed. *The other tongue: Englishes across cultures*. 2nd ed ed. Urbana and Chicago: University of Illinois Press. pp. 48-74.)

Kachru, B. ed. 1992b. *The other tongue: English across cultures*. Urbana and Chicago: University of Illinois Press.

Kachru, B. 1992c. Teaching World Englishes. (*In* Kachru, B., ed. *The other tongue: English across cultures*. 2nd ed ed. Urbana and Chicago: University of Illinois Press. pp. 355-365.)

Kachru, B. 1996. World Englishes: agony and ecstasy. *Journal of Aesthetic Education (Special Issue: Distinguished Humanities Lectures II)*, 30(2):135-155.

Kahneman, D. 2011. *Thinking, fast and slow*. London and New York: Penguin.

Kamwangamalu, N.M. 2002. English in South Africa at the millennium: challenges and prospects. *World Englishes*, 21(1):161-163.

Kamwangamalu, N.M. 2003. Globalization of English, and language maintenance and shift in South Africa. *International Journal of the Sociology of Language*, 164:65-81.

Kay, P. & Fillmore, C.J. 1999. Grammatical constructions and linguistic generalizations: the *What's X doing in Y?* construction. *Language*, 75(1):1-33.

Kearns, K. 2007. Regional variation in the syntactic distribution of null finite complementiser. *Language Variation and Change*, 19(3):295-336.

Keller, R. 1994. *On language change: the invisible hand in language*. London and New York: Routledge.

Kemmer, S. 2003. Schemas and lexical blends. (*In* Cuyckens, H., Berg, T., Dirven, R. & Panther, K., eds. *Motivation in language: studies in honor of Günter Radden*. Amsterdam and Philadelphia: John Benjamins. pp. 69-97.)

Kirsner, R.S. 2014. *Qualitative-quantitative analysis of Dutch and Afrikaans grammar and lexicon*. Amsterdam and Philadelphia: John Benjamins.

Kjellmer, G. 1985. *Help to/help* Ø revisited. *English Studies*, 66(2):156-161.

Koch, C. & Bernaisch, T. 2013. Verb complementation in South Asian English(es): the range and frequency of 'new' ditransitives. (*In* Andersen, G. & Bech, K., eds. *English corpus linguistics: variation in time, space and genre*. Amsterdam: Rodopi. pp. 69-89.)

Kortmann, B. & Szmrecsanyi, B. 2009. World Englishes between simplification and complexification. (*In* Hoffmann, T. & Siebers, L., eds. *World Englishes: problems, properties and prospects*. Selected papers from the 13th IAWWE conference. Amsterdam and Philadelphia: John Benjamins. pp. 265-285.)

Krifka, M. 1999. Manner in dative alternation. *West Coast Conference on Formal Linguistics*, 18(1):260-271.

Kruger, H. 2012. A corpus-based study of the mediation effect in translated and edited language. *Target*, 24(2):355-388.

Kruger, H. 2017. The effects of editorial intervention: implications for studies of the features of translated language. (*In* De Sutter, G., Lefer, M.A. & Delaere, I., eds. *Empirical translation studies: new methodological and theoretical traditions*. Berlin: De Gruyter. pp. 113-155.)

Kruger, H. & Van Rooy, B. 2016a. Constrained language: a multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, 37(1):26-57.

Kruger, H. & Van Rooy, B. 2016b. Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences*, 56(1):118-131.

Kruger, H. & Van Rooy, B. 2017. Editorial practice and the progressive in Black South African English. *World Englishes*, 35(2):20-41.

Labov, W. 1969. Contraction, deletion, and inherent variability of the English copula. *Language*, 45(4):715-762.

Labov, W. 1994. *Principles of linguistic change: internal factors*. (Volume 1 of *Principles of linguistic change*). Oxford and Cambridge: Wiley-Blackwell.

Labov, W. 2001. Principles of linguistic change: social change. (Volume 2 of *Principles of linguistic change*). Massachusetts and Oxford: Wiley-Blackwell.

Labov, W. 2010. Principles of linguistic change: cognitive and cultural factors. (Volume 3 of *Principles of linguistic change*). West Sussex: Wiley-Blackwell.

Lakoff, G. & Johnson, M. 1999. Philosophy in the flesh: the embodied mind and its challenge to western thought. New York: Basic Books.

Langacker, R.W. 1987a. Foundations of Cognitive Grammar. (Volume 1: *Theoretical prerequisites*). Stanford: Stanford University Press.

Langacker, R.W. 1987b. Foundations of Cognitive Grammar. (Volume 2: *Descriptive application*). Stanford: Stanford University Press.

Langacker, R.W. 2000. A dynamic usage-based model. (In Barlow, M. & Kemmer, S., eds. Usage-based models of language. Stanford: CSLI Publications. pp. 1-63.)

Langacker, R.W. 2008. Cognitive grammar: a basic introduction. Oxford: Oxford University Press.

Lanham, L.W. 1996. A history of English in South Africa. (In De Klerk, V., ed. Focus on South Africa. Amsterdam: John Benjamins. pp. 19-34.)

Lanstyák, I. & Heltai, P. 2012. Universals in language contact and translation. *Across Languages and Cultures*, 13(1):99-121.

ResearchGate. 2015. Leclercq, B.'s Q&A thread: Can a construction (as in Construction Grammar) have different forms?
https://www.researchgate.net/post/Can_a_construction_as_in_Construction_Grammar_have_different_forms Date of access: 16 October 2018.

Leech, G. 1992. Corpora and theories of linguistic performance. (In Svartvik, J., ed. Directions in corpus linguistics: proceedings of the Nobel Symposium 82 held in Stockholm, 48 August 1991. Berlin and New York: Mouton de Gruyter. pp. 105-125.)

Leech, G., Hundt, M., Mair, C. & Smith, N. 2009. Change in contemporary English: a grammatical study. Cambridge: Cambridge University Press.

Lenake, J.M. 2014. *Gore* as complementiser in Sesotho [personal interview]. Vanderbijlpark.

Levin, B. 1993. English verb classes and alternations: a preliminary investigation. Chicago and London: The University of Chicago Press.

- Levshina, N. 2015. How to do linguistics with R: data exploration and statistical analysis. Amsterdam and Philadelphia: John Benjamins.
- Lewis, M.P., Simons, G.F. & Fennig, C.D. 2016. Ethnologue: languages of the world. www.ethnologue.com Date of access: 11 July 2016.
- Lexical Computing. 2018. Sketch Engine. <https://www.sketchengine.eu/> Date of access: 18 November 2018.
- Lieven, E., Behrens, H., Speares, J. & Tomasello, M. 2003. Early syntactic creativity: a usage-based approach. *Journal of Child Language*, 30(2):333-370.
- Lind, Å. 1983. The variant forms of *help to/help* Ø. *English Studies*, 64(3):263-275.
- Lohmann, A. 2011. *Help vs help to*: a multifactorial, mixed-effects account of infinitive marker omission. *English Language and Linguistics*, 15(3):499-521.
- Mahlberg, M. 2005. English general nouns: a corpus theoretical approach. Amsterdam and Philadelphia: John Benjamins.
- Mair, C. 1990. Infinitival complement clauses in English: a study of syntax in discourse. Cambridge: Cambridge University Press.
- Mair, C. 1995. Changing patterns of complementation, and concomitant grammaticalisation, of the verb *help* in present-day English. (In Aarts, B. & Meyer, C.F., eds. *The verb in contemporary English: theory and description*. Cambridge: Cambridge University Press. pp. 258-272.)
- Mair, C. 2002. Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora. *English Language and Linguistics*, 6(1):105-131.
- Mair, C. 2006. Twentieth-century English: history, variation and standardization. Cambridge and New York: Cambridge University Press.
- Mair, C. 2013. Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide*, 34(3):253-278.
- Makalela, L. 2013. Black South African English on the radio. *World Englishes*, 32(1):93-107.
- Mathibela, L.Z. 2014. *Gore* as complementiser in Sesotho [personal interview]. Vanderbijlpark.
- Matras, Y. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36(2):281-331.

- Matras, Y. 2007. The borrowability of structural categories. (In Matras, Y. & Sakel, J., eds. *Grammatical borrowing in cross-linguistic perspective*. Berlin and New York: Mouton de Gruyter. pp. 31-73.)
- Matras, Y. 2009. *Language contact*. Cambridge and New York: Cambridge University Press.
- McEney, A. & Xiao, Z. 2005. *Help or help to: what do corpora have to say? English Studies*, 86(2):161-187.
- McEney, T. & Wilson, A. 2001. *Corpus linguistics: an introduction*. 2nd ed. Edinburgh: Edinburgh University Press.
- McGregor, W.B. 1997. *Semiotic Grammar*. Oxford: Clarendon Press.
- McGregor, W.B. 2013. Optionality in grammar and language use. *Linguistics*, 51(6):1147-1204.
- McMahon, A.M.S. 1994. *Understanding language change*. Cambridge: Cambridge University Press.
- McWhorter, J.H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology*, 5(2/3):125-166.
- McWhorter, J.H. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Boston and Berlin: De Gruyter Mouton.
- Melchers, G. & Shaw, P. 2003. *World Englishes: an introduction*. London: Arnold.
- Mesthrie, R. 2003. The world Englishes paradigm and contact linguistics: refurbishing the foundations. *World Englishes*, 22(4):449-461.
- Mesthrie, R. 2004. Synopsis: the phonology of English in Africa and South and Southeast Asia. (In Schneider, E.W., Burridge, K., Kortmann, B., Mesthrie, R. & Upton, C., eds. *A handbook of varieties of English*. (Volume 1: Phonology). Berlin and New York: Mouton de Gruyter. pp. 1 099-1 109.)
- Mesthrie, R. 2006. Anti-deletions in an L2 grammar: a study of Black South African English mesolect. *English World-Wide*, 27(2):111-145.
- Mesthrie, R. 2010. Socio-phonetics and social change: deracialisation of the GOOSE vowel in South African English. *Journal of Sociolinguistics*, 14(1):3-33.
- Mesthrie, R. 2015. Towards a distributed sociolinguistics of postcolonial multilingual societies. (In Smakman, D. & Heinrich, P., eds. *Globalising sociolinguistics: challenging and expanding theory*. London and New York: Routledge. pp. 80-92.)

- Mesthrie, R. & Bhatt, R.M. 2008. *World Englishes: the study of new linguistic varieties*. Cambridge: Cambridge University Press.
- Mesthrie, R., Chevalier, A. & Dunne, T. 2015. A regional and social dialectology of the BATH vowel in South African English. *Language Variation and Change*, 27(1):1-30.
- Mesthrie, R., Swann, J., Deumert, A. & Leap, W.L. eds. 2000. *Introducing Sociolinguistics*. Edinburgh: Edinburgh University Press.
- Milroy, J. 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, 5(4):530-555.
- Milroy, J. & Milroy, L. 2012. *Authority in language: investigating standard English*. 4th ed. London and New York: Routledge.
- Minow, V. 2010. *Variation in the grammar of Black South African English*. Frankfurt am Main: Peter Lang.
- Moag, R.F. 1992. The life cycle of non-native Englishes: a case study. (In Kachru, B., ed. *The other tongue: English across cultures*. Urbana and Chicago: University of Illinois Press. pp. 233-252.)
- Mondorf, B. 2014. Apparently competing motivations in morphosyntactic variation. (In MacWhinney, B., Malchukov, A. & Moravcsik, E., eds. *Competing motivations in grammar and usage*. Oxford: Oxford University Press. pp. 209-228.)
- Morreira, K. 2012. *Social change, class formation and English: a study of young black South Africans with "Model C" school backgrounds*. Cape Town: University of Cape Town. – PhD).
- Mufwene, S.S. 2001a. Competition and selection in language evolution. *Selection*, 3(1):45-56.
- Mufwene, S.S. 2001b. *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S.S. 2008a. Colonization, population contacts, and the emergence of new language varieties: a response to Peter Trudgill. *Language in Society*, 37(2):254-258.
- Mufwene, S.S. 2008b. *Language evolution: contact, competition and change*. London and New York: Continuum.
- Mukherjee, J. 2005. *English ditransitive verbs: aspects of theory, description and a usage-based model*. Amsterdam and New York: Rodopi.

Mukherjee, J. 2007. Steady states in the evolution of New Englishes: present-day Indian English as an Equilibrium. *Journal of English Linguistics*, 35(2):157-187.

Mukherjee, J. 2010. Corpus-based insights into verb-complementational innovations in Indian English: cases of nativised semantico-structural analogy. (In Lenz, A. & Plewnia, A., eds. Grammar between norm and variation. Frankfurt: Peter Lang. pp. 219-241.)

Mukherjee, J. & Gries, S.T. 2009. Collustruational nativisation in New Englishes: verb-construction associations in the International Corpus of English. *English World-Wide*, 30(1):27-51.

Mukherjee, J. & Hoffmann, S. 2006. Describing verb-complementation profiles of New Englishes: a pilot study of Indian English. *English World-Wide*, 27(2):147-173.

Müller, D. & Pistor, S. 2011. Skryf Afrikaans van A tot Z. 2nd ed. Cape Town: Pharos.

Murphy, V.A. & Pine, K.J. 2003. L2 influence on L1 linguistic representations. (In Cook, V., ed. Effects of the second language on the first. Clevedon and Buffalo: Multilingual Matters. pp. 142-167.)

Nelson, C.L. 1984. Intelligibility: the case of non-native varieties of English. Urbana, Illinois: Graduate College of the University of Illinois at Urbana-Champaign. (Thesis – PhD).

Nosofsky, R.M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1):54-65.

Oehrle, R.T. 1976. The grammatical status of the English dative alternation. Massachusetts: Massachusetts Institute of Technology. (Thesis – Doctor of Philosophy).

Olavarría de Ersson, E. & Shaw, P. 2003. Verb complementation patterns in Indian standard English. *English World-Wide*, 24(1):137-161.

Omoto, C.K. & Lurquin, P.F. 2012. Genes and DNA: a beginner's guide to genetics and its applications. New York: Columbia University Press.

Ortega, L. 2009. Understanding second language acquisition. London and New York: Routledge.

Oxford English Dictionary. 2016a. How dictionaries are created...
oup.m.useremarkable.com/production/images/uploads/230/original/ODO_HowDictionariesareCreated_220x900_Infographic_June16_Final.png?141711385 Date of access: 21 September 2016.

Oxford English Dictionary. 2016b. How does a word qualify for inclusion in the OED?
public.oed.com/about/frequently-asked-questions/#qualify Date of access: 21 September 2016.

Pan South African Language Board (PanSALB). 1999. Board Notice 54 of 1999: PanSALB's position on the promotion of multilingualism in South Africa. http://www.gov.za/sites/www.gov.za/files/20098_bn54_0.pdf Date of access: 16 July 2016.

Pan South African Language Board (PanSALB). 2015. PanSALB history. <http://www.pansalb.org/history1.html> Date of access: 25 July 2016.

Paradis, M. 2004. A neurolinguistic theory of bilingualism. Amsterdam and Philadelphia: John Benjamins.

Parkinson, J. & Singh, V. 2007. Stability of grammatical features of Black South African English. *Per Linguam*, 23(2):54-67.

Pavlenko, A. 2003. 'I feel clumsy speaking Russian': L2 influence on L1 in narratives of Russian L2 users of English. (In Cook, V., ed. *Effects of the second language on the first*. Clevedon and Buffalo: Multilingual Matters. pp. 32-61.)

Pedersen, T. 1996. Fishing for exactness. *Proceedings of the SCSUG 96 in Austin, TX, 188-200*.

Pedersen, T., Pakhomov, S.V.S., Patwardhan, S. & Chute, C.G. 2007. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3):288-299.

Perek, F. 2015. Argument structure in usage-based construction grammar: experimental and corpus-based perspectives. Amsterdam and Philadelphia: John Benjamins.

Peters, P. 2014. Usage guides and usage trends in Australian and British English. *Australian Journal of Linguistics*, 34(4):581-598.

Peterson, B. 2012. Black writers and the historical novel: 1907-1948. (In Attwell, D. & Attridge, D., eds. *The Cambridge history of South African history*. Cambridge: Cambridge University Press. pp. 291-308.)

Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge: MIT press.

Pinker, S. 1994. *The language instinct*. London and New York: Harper Perennial Modern Classics.

Pinker, S. 1999. *Words and rules: the ingredients of language*. London: Phoenix.

Popper, K.R. 1979. *Objective knowledge: an evolutionary approach*. Revised edition ed. Oxford: Clarendon Press.

- Pullum, G. & Zwicky, A. 1999. Gerund particles and head-complement inflection conditions. (*In* Collins, P. & Lee, D., *eds.* The clause in English: in honour of Rodney Huddleston. Amsterdam and Philadelphia: John Benjamins. pp. 251-271.)
- Quirk, R. 1990/2003. Language varieties and standard language. (*In* Harris, R. & Rampton, B., *eds.* The Language, ethnicity and race reader. London: Routledge. pp. 97-106.)
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. A comprehensive grammar of the English language. 2nd ed. London and New York: Longman.
- R Foundation. 2018. The R project for statistical computing. <https://www.r-project.org/> Date of access: 18 November 2018.
- Radford, A. 1997. Syntax: a minimalist introduction. Cambridge and New York: Cambridge University Press.
- Rappaport Hovan, M. & Levin, B. 2008. The English dative alternation: the case for verb sensitivity. *Journal of Linguistics*, 44(1):129-167.
- Rayson, P. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4):519-549.
- Rohdenburg, G. 1995. On the replacement of finite complement clauses by infinitives in English. *English Studies*, 76(4):367-388.
- Rohdenburg, G. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, 7(2):149-182.
- Rosenbach, A. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language*, 81(3):613-644.
- Sankoff, D. 1988. Sociolinguistics and syntactic variation. (*In* Newmeyer, F.J., *ed.* Linguistics: the Cambridge survey. Volume IV. Language: the socio-cultural context. Cambridge: Cambridge University Press. pp. 140-161.)
- Saussure, F.d. 1916/1972. Course in general linguistics. Translated from the French by Roy Harris. Illinois: Open Court.
- Schmied, J.J. 1991. English in Africa: an introduction. London: Longman.
- Schneider, E.W. 2003. The dynamics of New Englishes: from identity construction to dialect birth. *Language*, 79(2):233-281.
- Schneider, E.W. 2004. How to trace structural nativization: particle verbs in world Englishes. *World Englishes*, 23(2):227-249.

- Schneider, E.W. 2007. *Postcolonial English: varieties around the world*. Cambridge: Cambridge University Press.
- Schneider, E.W. 2008. Accommodation versus identity? A response to Trudgill. *Language in Society*, 37(2):262-267.
- Schneider, E.W. 2013. English as a contact language: the "New Englishes". (In Schreier, D. & Hundt, M., eds. *English as a contact language*. Cambridge and New York: Cambridge University Press. pp. 131-148.)
- Schneider, E.W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes*, 33(1):9-32.
- Scott, M. 2018. WordSmith Tools. <https://lexically.net/wordsmith/downloads/> Date of access: 18 November 2018.
- Seargeant, P. & Tagg, C. 2011. English on the internet and a 'post-varieties' approach to language. *World Englishes*, 30(4):496-514.
- Sebba, M. 1997. *Contact languages: pidgins and creoles*. London: Macmillan Press.
- Seidlhofer, B. & Jenkins, J. 2003. English as a lingua franca and the politics of property. (In Mair, C., ed. *The politics of English as a world language: new horizons in postcolonial cultural studies* (ASNEL papers 7). Amsterdam and New York: Rodopi. pp. 139-154.)
- Siegel, J. 2003. Substrate influence in creoles and the role of transfer in second language acquisition. *Studies in Second Language Acquisition*, 25(2):185-209.
- Silva, P. 1997. South African English: oppressor or liberator? Paper presented at MAVEN 97, Major varieties of English. Vaxjo. 20-22 November 1997. https://www.ru.ac.za/media/rhodesuniversity/content/dsae/documents/articles/Silva_article.pdf Date of access: 17 July 2016.
- Skutnabb-Kangas, T. 2003. Linguistic diversity and biodiversity: the threat from killer languages. (In Mair, C., ed. *The politics of English as a world language: new horizons in postcolonial cultural studies – ASNEL papers 7*. Amsterdam and New York: Rodopi. pp. 31-52.)
- South African Concise Oxford Dictionary. 2002. Edited by The Dictionary Unit for South African English. Cape Town: Oxford University Press.
- Spencer, B. 2011a. International sporting events in South Africa, identity re-alignment and Schneider's EVENT X. *African Identities*, 9(3):267-278.
- Spencer, B. 2011b. Linguistic penetration at Schneider's Phase 4: acceptability ratings of entrenched features of Black South African English by South Africans outside the originating culture of the variety. *Journal for Language Teaching*, 45(2):133-149.

Sridhar, K.K. & Sridhar, S.N. 1992. Bridging the paradigm gap: second-language acquisition theory and indigenized varieties of English. (In Kachru, B., ed. *The other tongue: English across cultures*. Urbana and Chicago: University of Illinois Press. pp. 91-107.)

Statistics South Africa (StatsSA). 2012. Census 2011: census in brief. http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf Date of access: 24 July 2016.

Stefanowitsch, A. 2013. Collostructional analysis. (In Hoffmann, T. & Trousdale, G., eds. *The Oxford handbook of Construction Grammar*. Oxford and New York: Oxford University Press. pp. 290-306.)

Stefanowitsch, A. & Gries, S.T. 2003. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2):209-243.

Stefanowitsch, A. & Gries, S.T. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1):1-43.

Strang, B.M.H. 1970. *A history of English*. London: Methuen.

Strobl, C., Hothorn, T. & Zeileis, A. 2009a. Party on! A new, conditional variable importance measure for random forests available in the **party** package. <https://epub.uni-muenchen.de/9387/1/techreport.pdf> Date of access: 25 October 2018.

Strobl, C., Malley, J. & Tutz, G. 2009b. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, 14(4):232-348.

Svartvik, J. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle by Braj Kachru. (In Quirk, R. & Widdowson, H., eds. *English in the world: teaching and learning the language and literatures*. (Papers of an international conference entitled 'Progress in English Studies' held in London, 17-21 September 1984 to celebrate the fiftieth anniversary of the British Council and its contribution to the field of English Studies over fifty years). Cambridge and New York: Cambridge University Press (for the British Council). pp. 31-34.)

Svartvik, J. 1992. Corpus linguistics comes of age. (In Svartvik, J., ed. *Directions in corpus linguistics: proceedings of the Nobel Symposium 82 held in Stockholm, 4-8 August 1991*. Berlin and New York: Mouton de Gruyter. pp. 7-13.)

Swales, J.M. 1997. English as *Tyrannosaurus rex*. *World Englishes*, 16(3):373-382.

Szmrecsányi, B. 2004. On operationalising syntactic complexity. (In Purnelle, G., Fairon, C. & Dister, A., eds. *Le poids des mots: proceedings of the 7th International Conference on*

Textual Data Statistical Analysis (Volume II). Louvain-la-Neuve: UCL Presses Universitaires De Louvain. pp. 1 032-1 039.)

Tagliamonte, S.A. & Baayen, R.H. 2012. Models, forests and trees of York English: *was/were* variation as a case study for statistical practice. *Language Variation and Change*, 24(2):135-178.

Tagliamonte, S.A. & Smith, J. 2005. *No momentary fancy!* The zero 'complementizer' in English dialects. *English Language and Linguistics*, 9(2):289-309.

Taylor, C. 2008. What is *corpus linguistics*? What the data says. *ICAME Journal*, 32(1):179-200.

Taylor, J.R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Teubert, W. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10(1):1-13.

Thomason, S.G. 2001. *Language contact: an introduction*. Washington: Georgetown University Press.

Thomason, S.G. 2006. Language change and language contact. (In Brown, K., ed. *Book Language change and language contact*. (Volume 6) 2nd ed. New York: Elsevier. 339-346.)

Thompson, G. & Hunston, S. 2006. System and corpus: two traditions with a common ground. (In Thompson, G. & Hunston, S., eds. *System and corpus: exploring connections*. London: Equinox. pp. 1-4.)

Thompson, S.A. 1990. Information flow and dative shift in English discourse. (In Edmondson, J.A., Feagin, C. & Mühlhäusler, P., eds. *Development and diversity: language variation across time and space*. Dallas: SIL & University of Arlington. pp. 239-253.)

Thompson, S.A. & Mulac, A. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, 15(3):237-251.

Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam and Philadelphia: John Benjamins.

Tomasello, M. 2003. *Constructing a language: a usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Tomasello, M. 2008. *Origins of human communication*. Cambridge: The MIT Press.

Tomasello, M. & Farrar, M.J. 1986. Joint attention and early language. *Child development*, 57(6):1 454-1 463.

- Torres Cacoullos, R. & Walker, J.A. 2009. On the persistence of grammar in discourse formulas: a variationist study of *that*. *Linguistics*, 47(1):1-43.
- Traugott, E.C. & Trousdale, G. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Treffers-Daller, J. & Sakel, J. 2012. Why transfer is a key aspect of language use and processing in bilinguals and L2-users. *International Journal of Bilingualism*, 16(1):3-10.
- Triandis, H.C., McCusker, C. & Hui, C.H. 1990. Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology*, 59(5):1006-1020.
- Trudgill, P. 2001. Contact and simplification: historical baggage and directionality in linguistic change. *Linguistic typology*, 5:371-374.
- Trudgill, P. 2004. *New-dialect formation: the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Trudgill, P. 2008. Colonial dialect contact in the history of European languages: on the irrelevance of identity to new-dialect formation. *Language in Society*, 37(2):241-254.
- Tuten, D.N. 2003. *Koineization in medieval Spanish*. Berlin and New York: Mouton de Gruyter.
- Tuten, D.N. 2008. Identity formation and accommodation: sequential and simultaneous relations. *Language in Society*, 37(2):259-262.
- UCREL. 2018. CLAWS part-of-speech tagger for English. <http://ucrel.lancs.ac.uk/claws/> Date of access: 11 November 2018.
- Ulbrich, C. & Ordin, M. 2014. Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, 45(1):26-42.
- Van der Walt, J. & Van Rooy, B. 2002. Towards a norm in South African Englishes. *World Englishes*, 21(1):113-128.
- Van Rooy, B. 2004. Black South African English: phonology. (In Schneider, E.W., Burridge, K., Kortmann, B., Mesthrie, R. & Upton, C., eds. *A handbook of varieties of English*. (Volume 1: Phonology). Berlin and New York: Mouton de Gruyter. pp. 943-952.)
- Van Rooy, B. 2006. The extension of the progressive aspect in Black South African English. *World Englishes*, 25(1):37-64.
- Van Rooy, B. 2008a. An alternative interpretation of tense and aspect in Black South African English. *World Englishes*, 27(3/4):335-358.

- Van Rooy, B. 2008b. A multidimensional analysis of student writing in Black South African English. *English World-Wide*, 29(3):268-305.
- Van Rooy, B. 2008c. What counts as explanation in linguistics? Inaugural lecture. Unpublished manuscript: North-West University.
- Van Rooy, B. 2009. The status of English in South Africa. (In Granger, S., Dagneaux, E., Meunier, F. & Paquot, M., eds. International Corpus of Learner English. Louvain: Presses Universitaires de Louvain. pp. 198-204.)
- Van Rooy, B. 2010. Social and linguistic perspectives on variability in world Englishes. *World Englishes*, 29(1):3-20.
- Van Rooy, B. 2011. A principled distinction between error and conventionalized innovations in African Englishes. (In Mukherjee, J. & Hundt, M., eds. Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap. Amsterdam: John Benjamins. pp. 189-207.)
- Van Rooy, B. 2013. Corpus linguistic work on Black South African English. *English Today*, 29(1):10-15.
- Van Rooy, B. 2014a. Convergence and endonormativity at Phase Four of the Dynamic Model. (In Buschfeld, S., Hoffmann, T., Huber, M. & Kautzsch, A., eds. The evolution of Englishes: the Dynamic Model and beyond. Amsterdam: John Benjamins. pp. 21-38.)
- Van Rooy, B. 2014b. Progressive and stative verbs in Outer Circle varieties. *World Englishes*, 33(2):157-172.
- Van Rooy, B. 2015. Changing patterns of complementation in East and Black South African English. Lecture held in Chemnitz on 6 February.
- Van Rooy, B. & Terblanche, L. 2006. A corpus-based analysis of involved aspects of student writing. *Language Matters Studies in the Languages of Africa*, 37(2):160-182.
- Van Rooy, B. & Terblanche, L. 2009. A multidimensional analysis of a learner corpus. (In Renouf, A. & Kehoe, A., eds. A multidimensional analysis of a learner corpus. Amsterdam: Rodopi. pp. 239-254.)
- Van Rooy, B. & Terblanche, L. 2010. Complexity in word-formation processes in New Varieties of South African English. *Southern African Linguistics and Applied Language Studies*, 28(4):357-374.
- Van Rooy, B., Terblanche, L., Haase, C. & Schmied, J.J. 2010. Register differentiation in East African English: a multidimensional study. *English World-Wide*, 31(3):311-349.

- Van Rooy, B., Van Rooyen, S. & Van Wyk, H. 2000. An assessment of high school pupils' attitudes towards the pronunciation of Black South African English. *Southern African Journal of Linguistics*, 18(Supplement 38):187-213.
- Wardhaugh, R. 1992. An introduction to sociolinguistics. 2nd ed ed. Oxford and Cambridge: Blackwell.
- Wasow, T. 1997. Remarks on grammatical weight. *Language Variation and Change*, 9(1):81-105.
- Weinreich, U. 1953. Languages in contact: findings and problems. Paris and New York: Mouton.
- Williams, J. 1987. Non-native varieties of English: a special case of language acquisition. *English World-Wide*, 8(2):161-199.
- Wilmot, K. 2011. Socio-cultural change in two prestigious secondary schools in South Africa: a sociophonetic study of black and white females. Cape Town: University of Cape Town. (Dissertation – MA).
- Wolk, C., Bresnan, J., Rosenbach, A. & Szmrecsanyi, B. 2013. Dative and genitive variability in late modern English: exploring cross-constructural variation and change. *Diachronica*, 30(3):382-419.
- Yano, Y. 2001. World Englishes in 2000 and beyond. *World Englishes*, 20(2):119-131.
- Zanettin, F. 2012. Translation-driven corpora. Manchester and New York: St. Jerome Publishing.
- Zipf, G.K. 1935. The psycho-biology of language. Boston: Houghton Mifflin.