

# Hoofstuk 1

## Inleiding

### 1.1 Agtergrond

Die Sentrum vir Tekstegnologie (CTexT) van Noordwes-Universiteit was vanaf 2010 tot 2013 betrokke by 'n annoteringsprojek van die *National Centre for Human Language Technology* (NCHLT). Die projek is befonds deur die Departement van Kuns en Kultuur en het ten doel gehad om, buiten Engels, ongeveer 50 000 woorde in elkeen van die oorblywende tien amptelike Suid-Afrikaanse tale op vier vlakke te annoteer. Hierdie vier vlakke sluit in (i) tekseenheid-identifikasie; (ii) lemmatisering; (iii) woordsoortetikettering (WS-etikettering); en (iv) morfologiese analise.

Daar is in totaal 60 318 Afrikaanse tekseenhede in die NCHLT-projek geannoteer. Hierdie groototaal kan verdeel word in twee subkorpuse, naamlik die NCHLT-proefkorpus van 55 484 tekseenhede (verder genoem die NCHLT-korpus) en die NCHLT-toetstek van 5 834 tekseenhede wat vir kwaliteitskontrole aangewend is. Tydens die annotering van die NCHLT-korpus het die ontwikkelaars en annoteerders telkens voor moeilike vrae te staan gekom met betrekking tot die annotering van die deelwoord in Afrikaans. Die ontwikkelaars moes onder andere besluit of die deelwoord gesien word as 'n resultaat van fleksie of afleiding. Afhangend of deelwoorde as fleksie of afleiding geklassifiseer word, lewer dit verskillende annoterings tydens die proses van lemmatisering. Die ontwikkelaars het wel uiteindelik daarvoor gekies om deelwoorde te tel onder fleksie (CTexT, 2013b).

Vir WS-etikettering en morfologiese analise was dit ook nie altyd duidelik watter kriteria om aan te lê om te besluit of 'n deelwoord funksioneer as 'n adjektief, sowel as 'n verledetydsvorm of passiefvorm van die werkwoord nie. Hoewel dit nog nie geëvalueer is nie, is daar 'n sterk vermoede dat hierdie onsekerheid tot inkonsekwente annotering sou kon lei.

Deurdadig dat daar onsekerheid bestaan oor hoe die deelwoord in Afrikaans beskou moet word, is dit nodig om op 'n verantwoordelike wyse Afrikaanse bronne te bestudeer om

vas te stel wat presies die deelwoord is en volgens watter morfologiese proses die deelwoord tot stand kom (fleksie of afleiding). Eers wanneer 'n duidelike beeld gevorm is van wat presies 'n deelwoord is, kan klinkklare riglyne vir praktiese annotering daargestel word.

## 1.2 Probleemstelling

Die deelwoord word tradisioneel (De Villiers, 1983:111; *eWAT*, 2009) gesien as 'n woord wat van 'n werkwoord gevorm is en dan as 'n ander woordsoort gebruik word (meestal as 'n adjektief), byvoorbeeld *brekende golwe* (onvoltooide deelwoord); *gebreekte glas* (swak voltooide deelwoord); en *gebroke hart* (sterk voltooide deelwoord). Uit 'n oorsig van bestaande Afrikaanse bronne (o.a. Carstens, 2011; Combrink, 1990; De Villiers, 1948; Le Roux, 1952; Ponelis, 1989), blyk dit dat hierdie tradisionele siening van die deelwoord nie so voor die hand liggend is nie en dat daar in werklikheid glad nie eenstemmigheid in die bronne bestaan oor wat 'n deelwoord is nie. Verwysings na die deelwoord (en beskrywings daarvan) is versplinterd in die sin dat verskillende aspekte rakende dié begrip oor verskillende bronne heen verskyn. Dit wil voorkom asof die deelwoord, op enkele uitsonderings na (bv. De Villiers, 1948: 95-108; Taalkommissie, 2009: 62-64, 162-167; en Van Schoor, 1983: 212-220), slegs in die verbygaan genoem word en dat hierdie skynbaar komplekse verskynsel dus ongenoegsaam bestudeer is.

Ook is daar in die beskrywing van deelwoorde in bestaande Afrikaanse bronne teenstrydige sieninge van die deelwoord. Die teenstrydighede in die bronne sentreer rondom: (i) vae terminologiegebruik en formulerings; (ii) die onderskeid tussen fleksie en afleiding; en (iii) die siening van die deelwoord as 'n vorm van die werkwoord (wat insluit of die vorm van die werkwoord in verledetyd- en passiefkonstruksies ook as 'n deelwoord gesien word of nie). Vergelyk in dié verband die volgende voorbeelde:

- (i) Fleksie impliseer verskillende vorme van dieselfde woord (vormverandering), en dit lewer verbuigings by naamwoorde of adjektiewe en vervoegings by werkwoorde. Die morfeme betrokke by fleksie word uitgange genoem. Hierteenoor impliseer afleiding woordvorming van nuwe woorde en afleidings word gelewer. Die morfeme betrokke by afleiding word affikse genoem (Geerts,

- Haeseryn, De Rooij & Van der Toorn, 1984:25). Wanneer bronne hierdie terme lukraak gebruik, is dit moeilik om enige betroubare afleiding te maak of die bron die deelwoord sien as fleksie of afleiding (kyk ook Kempen, 1982:141, 160, 262).
- (ii) Combrink (1989:233, 1990:30-31) maak die onderskeid tussen fleksie en afleiding in Afrikaans as ongegrond af; Du Toit (1986:101) en De Villiers (1983:68-69, 72-73, 80, 111-112, 161) sien die vorm van die werkwoord in beide die verledetyds- as passiefkonstruksies as fleksie en die adjektiewiese gebruik van deelwoorde as afleiding; en Van Schoor (1983:209) sien deelwoordvorme as fleksiewerking.
- (iii) Byna die helfte van die bronne wat gebruik is (vergelyk Hoofstuk 2), sien die deelwoord as 'n vorm van die werkwoord (o.a. De Villiers, 1983:69,72; *eWAT*, 2009; Kempen, 1982:262; Odendal & Gouws, 2005:146; Pfeiffer, 1989:83-88; Van Schoor, 1983:353, 378, 209), maar slegs De Villiers (1983:111), Van Schoor (1983:212) en die *eWAT* (2009) verduidelik dié vorm van die werkwoord verder as 'n woord wat gelyktydig twee woordklasse dien.

Behalwe vir sporadiese nagraadse studies en enkele artikels, het die mees resente en omvattende werk oor die Afrikaanse morfologie meer as twee dekades gelede verskyn (Combrink, 1990). Gevolglik is nuwe teoretiese insigte nie toegepas ter verheldering van die beskrywing van die deelwoord in Afrikaans nie. Om daarom in pas te bly met moderne en toonaangewende teorieë, kan 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk (Langacker, 2000; 2008a) gebruik word om nuwe insigte in die beskrywing van die deelwoord te kry. 'n Kognitiewe gebruiksgebaseerde beskrywingsraamwerk word vir hierdie studie gekies om lig te werp op die deelwoord, aangesien hierdie raamwerk tans een van die leidende beskrywingsparadigmas in die linguistiek is.

Ook in praktiese analises, soos byvoorbeeld by die annotering van korpusdata, blyk dit soms moeilik te wees om te onderskei tussen die verskillende vorme van die werkwoord. Wanneer is 'n bepaalde werkwoordvorm 'n adjektief, 'n verledetydsvorm van die werkwoord, of die vorm wat gebruik word in die passiefkonstruksie? Eers wanneer daar presies vasgestel word wat 'n deelwoord is, kan duidelike riglyne vir die praktiese annotering (lemmatisering en WS-etikettering) van deelwoorde daargestel word.

Die drie probleme wat hierdie studie rig, kan soos volg saamgevat word:

- a) Uit die literatuur is dit nie duidelik wat presies 'n deelwoord in Afrikaans is nie.
- b) Nuwe teoretiese insigte, soos die kognitiewe gebruiksgebaseerde beskrywingsraamwerk) is nog nie op die deelwoord toegepas nie.
- c) Omdat dit nie duidelik is wat presies 'n deelwoord is nie, is daar ook nie duidelike riglyne vir die annotering (spesifiek lemmatisering en WS-etikettering) van deelwoorde nie.

### **1.3 Navorsingsvrae**

In ooreenstemming met die agtergrond en probleemstelling, word die volgende navorsingsvrae in hierdie studie gestel:

- a) Hoe word die deelwoord in bestaande Afrikaanse bronne omskryf of gedefinieer, en hoe kan Nederlandse en Engelse beskrywings van die deelwoord daartoe bydra om die deelwoord beter te omskryf en te definieer?
- b) Hoe kan 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk bydra tot 'n verhelderende beskrywing van die Afrikaanse deelwoord?
- c) Tot watter mate lei nuwe insigte in die deelwoord tot alternatiewe lemmatisering en WS-etikettering van deelwoorde in die NCHLT-korpus?

### **1.4 Doelstellings**

Na aanleiding van die bostaande navorsingsvrae, beoog die studie die volgende:

- a) om 'n oorsig te gee van bestaande beskrywings van die deelwoord in Afrikaans, en om met behulp van gesaghebbende Engelse en Nederlandse bronne vas te stel waarom daar so min ooreenstemming in die bestaande beskrywings van die deelwoord in Afrikaans is;
- b) om te illustreer hoe 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk verhelderende insigte kan bied in die karakterisering van die deelwoord as 'n

simboliese eenheid wat op die fonologiese pool sowel as die semantiese pool beskryf kan word; en

- c) om te bepaal of nuwe insigte in die deelwoord alternatiewe analyses van die deelwoord in die Afrikaanse gedeelte van die NCHLT-korpus sal hê.

## **1.5 Navorsingsmetodes**

Die navorsingsmetodes wat in hierdie studie aangewend gaan word, betrek 'n literatuurstudie (1.5.1), 'n beskrywing (1.5.2) en 'n toepassing (1.5.3).

### **1.5.1 Literatuurstudie**

Vir doelstelling (a) gaan 'n literatuurstudie gedoen word oor bestaande beskrywings van die deelwoord in Afrikaans om vas te stel wat presies met die begrip *deelwoord* bedoel word. Resente nagraadse studies en artikels sowel as gesaghebbende Engelse en Nederlandse bronne oor die *participle* in Engels en die *deelwoord* in Nederlands sal ook betrek word om vas te stel watter onduidelikhede, gapings en teenstrydighede daar in die huidige beskrywings van die deelwoord in Afrikaans is. Uit die literatuurstudie gaan 'n gekonsolideerde beskrywing voorgestel word van hoe die deelwoord beskou kan word, terwyl aspekte wat steeds onopgelos bly met betrekking tot die Afrikaanse deelwoord, ook geïdentifiseer sal word.

Vir doelstelling (b) gaan eweneens 'n literatuurstudie gedoen word om 'n kort oorsig van 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk te gee. Die belangrikste vertrekpunte en konsepte wat nodig is vir die beskrywing van 'n deelwoord, sal in Hoofstuk 3 verduidelik word. Daar word voortgebou op die uiteensetting van Van Huyssteen (2000a; 2000b; 2005), wat aangevul sal word uit Langacker (2000; 2008a; 2008b) en Booij (2002; 2006; 2010a; 2010b).

Vir doelstelling (c) sal 'n literatuurstudie van die protokoldokumente (CText, 2013b; 2013c) van die NCHLT-korpus gemaak word. Die protokoldokument vir lemmatisering en die protokoldokument vir WS-etikettering sal in detail bestudeer word in terme van die leiding wat hierdie twee dokumente vir die annotasie van die deelwoord gee.

### 1.5.2 Beskrywing

Vir doelstelling (b) gaan die deelwoord beskryf word vanuit 'n kognitiewe gebuiksgebaseerde beskrywingsraamwerk. Die data (gebruiksvoorbeelde) vir die beskrywings sal uit die *Afrikaanse woordelys en spelreëls (AWS)* geneem word. Eers gaan die onvoltooide deelwoord op die fonologiese sowel as die semantiese pool gekarakteriseer word, en daaruit gaan 'n kategoriseringsnetwerk van die onvoltooide deelwoord gepostuleer word.

Daarna sal die voltooide deelwoord ook op beide die fonologiese en semantiese pole gekarakteriseer word, en daar sal ook 'n kategoriseringsnetwerk vir die voltooide deelwoord gegee word. Ter afsluiting van die beskrywings, sal geïllustreer word hoe die perifrastiesekonstruksievorme (PK-vorme) van die werkwoord (afdeling 2.5 en 4.4) in die verledetyd- en passiefkonstruksies, verskil van die deelwoordvorm van die werkwoord.

In alle Germaanse tale word twee tipes voltooide deelwoorde onderskei, naamlik sterk voltooide deelwoorde en swak voltooide deelwoorde (Booij, 2002:58-59). Sterk voltooide deelwoorde word gevorm deur 'n morfologiese variasie van die stam deur middel van ablaut (Matthews, 1997:2). Hierdie sterk vorme eindig altyd in Afrikaans op 'n *-e*, ongeag of dit attributief of predikatief gebruik word. Sterk voltooide deelwoorde is verder volkome onproduktief en word as 'n afgeslote groep gesien. Van Schoor (1983:216) redeneer dat indien mens sinchronies na die sterk vorm van voltooide deelwoorde sou kyk, dit bloot as byvoeglike naamwoorde gesien kon word, aangesien dit nooit in 'n werkwoordelike funksie as deel van die predikaat gebruik word nie (Van Schoor, 1983:216). Aangesien sterk voltooide deelwoorde morfologies oninteressant en volkome onproduktief is, val sterk voltooide deelwoorde *per se* buite die omvang van hierdie studie en sal daar slegs op swak voltooide deelwoorde gefokus word.

Buiten sterk verlede deelwoorde, word twee ander groepe sogenaamde deelwoorde by die studie uitgesluit. Die een groep is deelwoorde wat as voorsetsels optree, omdat (i) daar slegs enkele voorbeelde hiervan is, byvoorbeeld *gedurende*; en (ii) omdat dié enkele voorbeelde nie meer 'n natuurlike verbale parafrase toelaat nie, aangesien hulle geleksikaliseer het. Die ander groep is deelwoorde met 'n naamwoordelike basis, byvoorbeeld *gestreepte trui* of *hy is bebaard*. Kempen (1982:141-143) verwys na

hierdie deelwoorde as preverbale en Booij (2002:77) noem hulle pseudodeelwoorde, omdat die geïmpliseerde verbale basis nooit as 'n werkwoord realiseer nie.

### **1.5.3 Toepassing**

Vir doelstelling (c) gaan insigte wat oor die deelwoord ingewin is, toegepas word op die geannoteerde Afrikaanse korpus van die NCHLT. Nadat die protokolle bestudeer is, sal voorstelle gemaak word ter verbetering van die protokoldokumente, sowel as voorstelle vir alternatiewe lemmatiserings en WS-etiketterings.

Daar sal hoofsaaklik van die NCHLT-toetstek van 5 834 tekseenhede (CTexT, 2013a) gebruik gemaak word om gebruiksgebaseerde voorbeelde te vind. Indien nodig, sal die voorbeelde aangevul word uit die NCHLT-korpus van 55 484 tekseenhede (CTexT, 2013a), die *AWS* (Taalkommissie, 2009) of die *eWAT* (2009). Daar sal deurgaans bepaal word of nuwe insigte in die deelwoord tot alternatiewe analyses en etiketterings lei.

## **1.6 Ontplooiing**

As vertrekpunt vir hierdie studie word daar in Hoofstuk 2 'n literatuuroorsig van vyftien verskillende Afrikaanse bronne gemaak om vas te stel hoe die deelwoord in die literatuur beskryf word. Die doel van die literatuurstudie is om te bepaal wat presies met 'deelwoord' bedoel word en ook om te bepaal hoe ooreenstemmig of teenstrydig verskillende bronne die deelwoord beskryf. In die sintese van die literatuuroorsig gaan daar spesifiek gekyk word na (i) verwarrende terminologiegebruik en formulerings; (ii) of deelwoorde gesien word as fleksie of afleiding; en (iii) die deelwoord as 'n ander vorm van die werkwoord.

Hoofstuk 3 verduidelik die belangrikste uitgangspunte en konstrukte vanuit 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk. Anders as in tradisionele taalteorieë, word die fleksie-afleiding-onderskeid eerder op 'n kontinuum gesien as 'n absolute binêre onderskeid. Binne so 'n beskrywingsraamwerk word woorde ook in woordklasse verdeel op grond van hulle semantiese konseptualisering. Alvorens 'n beskrywing van die deelwoord in Hoofstuk 4 gemaak kan word, word die uitgangspunte en fundamentele konstrukte van so 'n beskrywingsraamwerk verduidelik. Die konstrukte wat spesifiek in die hoofstuk bespreek gaan word, sluit onder andere in

domein, profiel, trajektor, landmerk, gesigspunt, skandering, verwysingspunt, skopus, en temporele dimensie.

In Hoofstuk 4 word die deelwoord gekarakteriseer vanuit 'n kognitiewe gebruiksgebaseerde beskrywingsraamwerk. Beide die onvoltooide en voltooide deelwoord word op die fonologiese sowel as die semantiese pool beskryf, voordat 'n kategoriseringsnetwerk van elke tipe deelwoord gegee word. Aandag word ook in hierdie hoofstuk geskenk aan die karakterisering van die PK-vorme van die werkwoord (d.i. die vorm in die verledetyds- en passiefkonstruksies). Die verskil tussen die deelwoordvorm en die PK-vorme word ook bespreek en geïllustreer.

Die insigte wat verkry is oor die beskrywing van die deelwoord word in Hoofstuk 5 prakties van toepassing gemaak. Tydens die NCHLT-projek het die deelwoord op twee vlakke van annotering onsekerheid laat ontstaan: op die vlak van lemmatisering sowel as op die vlak van WS-etikettering. Die protokoldokumente van die NCHLT-projek vir hierdie twee vlakke van annotering word bestudeer en voorstelle word gemaak ter verbetering van die lemma-identifisering en WS-etikettering van deelwoorde.

Hoofstuk 6 gee 'n samevatting van die studie. In hierdie hoofstuk sal die navorsingsvrae en die doelstellings van die studie weer in oënskou geneem word. Verder sal daar ook 'n opsomming en gevolgtrekking van die studie gemaak word. Die hoofstuk sal eindig met 'n verwysing na verdere navorsing wat moontlik uit hierdie studie kan voortspruit.