

'n Korpusgebaseerde ondersoek na kohesiepatrone as moontlike stilistiese kenmerk van outeurstyl

Anneen Church (Fleischmann)

2007024942

Hierdie verhandeling word voorgelê ingevolge die vereistes van die graad MA (met spesialisasie in Afrikaans en Nederlands) in die Departement Afrikaans en Nederlands, Duits en Frans in die Fakulteit Geesteswetenskappe aan die Universiteit van die Vrystaat

Januarie 2020

Studieleier: Dr. R. Wasserman

Medestudieleier: Prof. A. van Niekerk

Verklaring

Ek verklaar dat die verhandeling wat hiermee ingedien word vir die graad MA (met spesialisasie in Afrikaans en Nederlands) aan die Universiteit van die Vrystaat my eie, onafhanklike werk is en nie voorheen deur my vir 'n graad aan 'n universiteit of fakulteit ingedien is nie. Ek doen voorts afstand van die outeursreg in die verhandeling ten gunste van die Universiteit van die Vrystaat.

Naam: Anneen Church

Datum: Januarie 2020

Dankwoord

Ek gee eerstens die eer aan my Hemelse Vader vir Sy teenwoordigheid in hierdie studie. Hy is die Alfa en Omega en sonder Sy genade, is ek niks.

Hiermee bedank ek ook die volgende persone sonder wie se hulp en ondersteuning die uitvoering van hierdie studie nie moontlik sou gewees het nie:

- my man, Louis, vir jou volgehoue ondersteuning deur hierdie maraton. Al het die worsteling nie altyd vir jou sin gemaak nie, het jy my altyd geduldig en enduit ondersteun. Baie dankie.
- my twee seuntjies, Arthur en Petri. Julle is my motivering vir die voltooiing van hierdie studie. Julle lus vir die lewe dring my om aan te hou.
- my hoofstudiesleier, Ronel, vir jou deeglike en sinvolle insette en kommentaar.
- my medestudiesleier, Angelique, vir jou spoedige terugvoer en nugter benadering tot studiesleiding.
- my goeie vriendin, oudskollega en medereisiger in die akademie, Elmarie, vir jou geduldige aanpor deur die jare, onwrikbare vertrouwe in my vermoëns en al die liefde en gebede.
- my ouers, Johan en Elma, dis julle voorbeeld wat my tot hier gelei het. Dankie vir die ondersteuning.
- my skoonouers, Charles en Elsuna, vir julle volgehoue belangstelling en ondersteuning ondanks die stadige vordering. Ek waardeer dit.
- ander vriende en familie, HAML in die besonder, baie dankie vir jul belangstelling, aanmoediging en gebede.
- my OLS-kollega, Rohan Posthumus, vir sy tyd en geduld terwyl hy my gehelp het met die statistiese aspekte van hierdie studie. Sonder jou sou ek waarlik nie veel gehad het om te sê na al die analyses nie.
- Amanda Swart, vir die taalversorging van die finale dokument.
- Prof. Bertus van Rooy wat 'n dag opgeoffer het om korpuslinguistiek en WordSmith aan my bekend te stel. Dit was die katalisator vir die uiteindelijke produk.
- my lynhoofde wat ek tydens my studie gehad het, dr. Annette de Wet en Michelle Joubert, vir hulle begrip en ondersteuning wanneer ek dit nodig gehad het.
- die Universiteit van die Vrystaat vir sy finansiële bydrae.
- Sammy, wat nooit toegelaat het dat ek alleen werk nie.

Abstract

Writers leave a trail behind when they write a text because individuals have preferences for certain linguistic items (Louwse, 2004:307). In theory, this means that texts can be identified as the work of a specific author, based on his or her linguistic choices. The identification of an author as the producer of a certain text, is a part of the academic field of authorship attribution studies. Joula (2008:249-251) suggests that there are certain aspects within authorship attribution studies which remain uncertain. These aspects mainly relate to making the most appropriate decisions about methodology when conducting authorship attribution studies. Another aspect he (Joula, 2008:249-251) mentions relates to the choice of linguistic item to be studied as possible indication of authorship. Both the question about appropriate methodological considerations and that of an (linguistic) indication of authorship is explored in this study.

This study mainly aims to investigate whether the use of cohesive devices in a text (or texts) can be an indication of authorship. It also attempts to describe the way(s) in which the use and distribution of cohesion links can be used as a basis in authorship attribution studies. The five types of cohesive devices as described by, among others, Halliday and Hasan (1976) will be used as linguistic variables to show possible authorship. These categories include reference, substitution, ellipsis, conjunctions and lexical cohesion. As an additional aspect to the research, the way in which cohesion links appear in two text types (and not just the texts of two different authors) is also described.

A corpus-based methodology was followed to study the use of cohesive devices in the texts of two popular Afrikaans authors, André le Roux and Dana Snyman. The corpus for this study consisted of four sub-corpora, namely columns and short stories by Dana Snyman and columns and short stories by André le Roux. The data analysis process was made up of different stages, namely the tagging of cohesive linguistic items, sorting of the tags by means of computer software (WordSmith Tools), classifying the sorted tags to indicate the type of cohesion and statistically analysing of the use of cohesive devices.

The above-mentioned statistical analysis consisted of multiple stages (just like the process of data analysis) with the goal of systematically and reliably identifying and delineating relevant data. In each stage a dataset was analysed by means of a specific statistical test to identify parts of the dataset which could be unique to the author or text type in question. Data that were identified as a result of the statistical analysis became the dataset for the subsequent stage of statistical analysis. This process is explained in the flow diagram at the beginning of Chapter 6. Through this process *relevant* data (as identified by statistical analysis) could be used for *relevant* statistical tests in order to optimize the process of statistical analysis and improve efficiency.

After the literature review, data analysis and various stages of the analyses, the nature of cohesive devices as an element of language variation between the texts of the two authors mentioned is discussed. A reflection of the methodological and statistical approaches of this study forms part of the final chapter, together with a discussion of the implications of the findings and recommendations for future research.

Key words:

Writing style, authorship style, authorship attribution, linguistic fingerprint, textuality, cohesion, cohesion link, text type, short story, column, corpus-based methodology

Opsomming

Skrywers laat 'n spoor van hul outeurskap agter wanneer hulle 'n teks skryf, aangesien elke persoon 'n voorkeur vir sekere linguistiese items het (Louwerse, 2004:207). Dit beteken in teorie dat tekste uitmekaar geken kan word as die werk van 'n spesifieke skrywer, op grond van sy/haar toepassing of gebruik van sekere linguistiese items. Hierdie identifisering van die outeur van 'n spesifieke teks val onder die studieveld van outeurskapidentifikasie. Joula (2008:249-251) stel aspekte binne outeurskapstudies voor waaroor daar nog sekerheid verkry moet word. Hierdie aspekte hou grootliks verband met die kies van gepaste metodologiese benaderings in die veld van outeurskapstudie. 'n Ander aspek wat Joula (2008:249-251) noem, is die keuse van linguistiese item wat die beste aanduier van outeurskap sou wees. Beide die 'n metodologiese oorwegings en die (linguistiese) aanduier van outeurskap is in hierdie studie ondersoek.

In hierdie studie word hoofsaaklik ondersoek of die gebruik van kohesieskakels in 'n teks (of tekste) 'n aanduier kan wees van outeurskap. Daar is ook gepoog om te beskryf op watter manier(e) die voorkoms van kohesieskakels 'n basis kan bied vir outeurskapstudies. Die vyf tipes kohesieskakels soos, onder andere, deur Halliday en Hasan (1976) uiteengesit, is as moontlike tekens van outeurstyl ondersoek. Hierdie vyf tipes is verwysing, substitusie, ellips, konjunksie en leksikale kohesie. As addisionele aspek van die navorsing, is die voorkoms van kohesieskakels in twee tekstipes (en nie net die tekste van onderskeie skrywers nie), ook beskryf.

'n Korpusgebaseerde benadering is gevolg in 'n ondersoek na die voorkoms van kohesieskakels in die tekste van twee populêre Afrikaanse skrywers, naamlik André le Roux en Dana Snyman. Die korpus wat vir hierdie studie gebruik is, het uit vier subkorpusse bestaan, naamlik kortverhale en rubrieke van Dana Snyman en kortverhale en rubrieke van André le Roux en rubrieke van André le Roux. Die data-analiseproses het uit verskeie fases bestaan, naamlik etikettering van kohesief-verwante items, sortering van etikette met die hulp van rekenaarsagteware (WordSmith Tools), klassifikasie van etikette (vir tipes kohesieskakels) en statistiese analise van die voorkoms van die kohesieskakels.

Die bogenoemde statistiese analise het ook uit verskeie fases bestaan (net soos die data-analiseproses) wat hoofsaaklik die sistematiese en geldige afbakening van relevante data ten doel gehad het. Tydens elke fase is 'n datastel aan 'n spesifieke statistiese toets onderwerp, ten einde potensiële uniekhede in elke subkorpus te identifiseer. Data wat na afloop van die betrokke statistiese toets geïdentifiseer is, is dan in die volgende fase van die statistiese ontleding gebruik. Die resultate van elke fase het dus as die basis gedien vir die volgende fase van ondersoek soos duidelik blyk in die vloeiagram aan die begin van hoofstuk 6. Deur hierdie proses is daar gepoog om *relevante* data (soos deur statistiese analise aangedui is) aan *relevante* statistiese toetsing te onderwerp en sodoende 'n meer doeltreffende werkswyse te volg.

Op grond van die literatuurstudie, data-analise en statistiese analiseresultate is die aard van kohesieskakels as element van variasie tussen outeurs se skryfstyl bespreek. Daar is ook gereflekteer oor die nut van die metodologiese benadering en statistiese analise wat in hierdie studie gevolg is en aanbevelings vir toekomstige studies, is gemaak.

Sleutelwoorde:

Skryfstyl, outeurstyl, outeurskapidentifikasie, linguistiese vingerafdruk, tekstualiteit, kohesie, kohesieskakel, tekstipe, kortverhaal, rubriek, korpusgebaseerde metodologies

Inhoudsopgawe

| | | |
|----------|--|----------|
| 1 | Inleiding | 1 |
| 1.1 | Agtergrond en rasionaal | 1 |
| 1.2 | Navorsingsdoelstellings en -doelwitte | 4 |
| 1.2.1 | Doelstelling | 4 |
| 1.2.2 | Doelwitte..... | 4 |
| 1.3 | Navorsingsvrae..... | 4 |
| 1.4 | Navorsingsontwerp..... | 5 |
| 1.5 | Navorsingsmetodologie..... | 5 |
| 1.5.1 | Literatuurstudie | 5 |
| 1.5.2 | Data-analise..... | 6 |
| 1.5.3 | Datastel | 6 |
| 1.6 | Waarde van die navorsing | 7 |
| 1.7 | Vooruitskouing..... | 7 |
| 2 | Styl, Stilistiek en Outeurskapidentifikasie | 9 |
| 2.1 | Inleiding | 9 |
| 2.2 | Wisselwerking tussen stilistiek en outeurskapidentifikasie | 9 |
| 2.3 | Outeurskapstudies | 12 |
| 2.3.1 | Definisies, kwessies en kritiek..... | 12 |
| 2.3.2 | Oorsig van die geskiedenis van moderne outeurskapstudies..... | 13 |
| 2.3.2.1 | Klassieke benaderings | 13 |
| 2.3.2.2 | Navorsing sedert 1990 | 17 |
| 2.3.3 | Probleme, oplossings en metodologiese benaderings..... | 17 |
| 2.3.3.1 | Probleme | 17 |
| 2.3.3.2 | Oplossings | 18 |
| 2.3.4 | Navorsing sedert 2000 | 19 |
| 2.3.4.1 | Oplossings: outomatiese rekenaargesteunde analise | 20 |

| | | |
|-----------|---|-----------|
| 2.3.4.2 | Die metafoor van 'n <i>linguistiese vingerafdruk</i> | 20 |
| 2.3.5 | Teoretiese raamwerke | 22 |
| 2.3.5.1 | Stilisties-gebaseerde benadering | 23 |
| 2.3.5.2 | Kognitiewe benadering | 24 |
| 2.3.5.3 | Sistemies-Funksionele Grammatika (SFG) | 25 |
| 2.3.5.3.1 | Registervariasie | 26 |
| 2.3.5.3.2 | Dialekvariasie | 26 |
| 2.3.5.3.3 | Kodevariasie | 27 |
| 2.4 | Samevatting | 27 |
| 3 | Teks, tekstualiteit en kohesie | 28 |
| 3.1 | Inleiding | 28 |
| 3.2 | Tekslinguistiek as studieveld | 28 |
| 3.3 | Tekstuur en tekstualiteit | 29 |
| 3.4 | Tekstualiteitsbeginsels | 30 |
| 3.4.1 | Kohesie | 32 |
| 3.5 | Kohesieskakels | 33 |
| 3.5.1 | Verwysing | 33 |
| 3.5.2 | Substitusie | 37 |
| 3.5.2.1 | Naamwoordelike substitusie | 37 |
| 3.5.2.2 | Werkwoordelike substantivering | 38 |
| 3.5.2.3 | Sinstuk-substitusie | 39 |
| 3.5.2.4 | Samevatting - substitusie | 40 |
| 3.5.3 | Ellips | 40 |
| 3.5.3.1 | Tipes ellips soos aanvanklik deur Halliday en Hasan uiteengesit | 41 |
| 3.5.3.1.1 | Naamwoordelike ellips (“nominal ellipsis”) | 41 |
| 3.5.3.1.2 | Werkwoordelike ellips (“verbal ellipsis”) | 42 |
| 3.5.3.1.3 | Sinstuk-ellips (“Clausal ellipsis”) | 42 |

| | | |
|-----------|--|-----------|
| 3.5.3.1.4 | Ellips soos dit deur Carstens (1997) uiteengesit is | 42 |
| 3.5.4 | Leksikale kohesie | 43 |
| 3.5.4.1 | Leksikale herhaling | 44 |
| 3.5.4.1.1 | Herhaling van 'n woord of frase | 44 |
| 3.5.4.1.2 | Sinonieme | 44 |
| 3.5.4.1.3 | Superordinate | 45 |
| 3.5.4.1.4 | Epiteta | 45 |
| 3.5.4.2 | Kollokasie | 46 |
| 3.5.5 | Konjunksie | 48 |
| 3.6 | Samevatting | 54 |
| 4 | Genre en tekstipe | 55 |
| 4.1 | Inleiding | 55 |
| 4.1.1 | Algemeen aanvaarde vertrekpunte met betrekking tot genre | 56 |
| 4.2 | Benaderings tot genre | 56 |
| 4.2.1 | Nuwe Retoriek | 57 |
| 4.2.2 | Sistemies-Funksionele Linguistiek | 57 |
| 4.2.3 | Taal vir Akademiese Doeleindes | 58 |
| 4.3 | Genre en korpusgebaseerde navorsing | 58 |
| 4.3.1 | Genre en tekstipe | 59 |
| 4.4 | Kortverhale | 62 |
| 4.4.1 | Oorsig | 63 |
| 4.4.2 | Stylkenmerke | 63 |
| 4.5 | Rubrieke | 64 |
| 4.5.1 | Oorsig | 64 |
| 4.5.2 | Stylkenmerke | 65 |
| 4.5.2.1 | Inhoud | 65 |
| 4.5.2.2 | Skrywer se persoonlikheid | 66 |

| | | |
|-----------|---|-----------|
| 4.6 | Genre en korpusstudies | 67 |
| 4.6.1 | Biber 1988 – Multidimensionele studie | 67 |
| 4.6.2 | Louwerse <i>et al.</i> - Die implikasie van kohesieskakels as linguistiese veranderlike 69 | |
| 4.7 | Samevatting | 71 |
| 5 | Navorsingsmetodologie | 73 |
| 5.1 | Inleiding | 73 |
| 5.2 | 'n Korpusgebaseerde benadering: oorsig | 73 |
| 5.3 | Tipes navorsing wat met 'n korpusgebaseerde metode gedoen word..... | 75 |
| 5.4 | Navorsingsvrae..... | 76 |
| 5.5 | Navorsingsontwerp..... | 77 |
| 5.5.1 | Aard van redenering..... | 79 |
| 5.5.2 | Tipe data..... | 80 |
| 5.5.2.1 | Keuse van datastel | 80 |
| 5.5.2.1.1 | Beskrywing van die korpus | 81 |
| 5.5.2.1.2 | Korpusseleksie..... | 82 |
| 5.5.3 | Metode van data-analise | 83 |
| 5.6 | Analiseproses | 83 |
| 5.6.1 | Fase 1: Etikettering | 84 |
| 5.6.2 | Fase 2: WordSmith-verwerking..... | 85 |
| 5.6.3 | Fase 3: Kategorisering van kohesief verwante items..... | 87 |
| 5.6.3.1 | Kategorisering van verwysing, leksikale kohesie en substitusie..... | 87 |
| 5.6.3.2 | Kategorisering van verbandsmerkers | 90 |
| 5.6.3.3 | Kategorisering van ellips | 91 |
| 5.7 | Statistiese analysesagteware..... | 92 |
| 5.8 | Statistiese analisemetode..... | 93 |
| 5.8.1 | Chi-kwadraattoetse | 93 |

| | | |
|----------|--|------------|
| 5.8.2 | Binêre logistiese regressie-analise | 94 |
| 5.8.3 | Beskrywing van datastel vir statistiese analise | 96 |
| 5.8.4 | Doelwitte van die statistiese analise | 98 |
| 5.9 | Samevatting | 99 |
| 6 | Data-analise en interpretasie | 100 |
| 6.1 | Inleiding | 100 |
| 6.2 | Beskrywende data: Relatiewe hoeveelhede | 102 |
| 6.2.1 | Algemene voorkoms van kohesieskakels | 102 |
| 6.2.1.1 | Algemene voorkoms van subkategorieë | 106 |
| 6.2.2 | Posisionele voorkoms van kohesieskakels: Sinsvlak | 109 |
| 6.2.3 | Posisionele voorkoms van kohesieskakels: Paragraafvlak | 113 |
| 6.3 | Beskrywende data: Chi-kwadraattoets | 118 |
| 6.3.1 | Algemene voorkoms van kohesieskakels | 119 |
| 6.3.1.1 | Algemene voorkoms van subkategorieë van kohesieskakels | 121 |
| 6.3.2 | Posisionele voorkoms van kohesieskakels: Sinsverband | 123 |
| 6.3.3 | Opsomming: Beskrywende data | 125 |
| 6.4 | Binêre logistiese regressie-analise | 131 |
| 6.4.1 | Binêre logistiese regressie-analise: Voorstelling van nulmodelle | 131 |
| 6.4.2 | Binêre logistiese regressie-analise: Model 1 | 132 |
| 6.4.3 | Binêre logistiese regressie-analise: model 2 | 134 |
| 6.5 | Samevatting | 135 |
| 7 | Samevatting, beperkinge en aanbevelings | 137 |
| 7.1 | Inleiding | 137 |
| 7.2 | Doelstellings en doelwitte van die studie | 137 |
| 7.3 | Literatuurstudie | 137 |
| 7.4 | Die empiriese ondersoek | 138 |
| 7.5 | Beantwoording van navorsingsvrae | 139 |

| | | |
|-------|---|------------|
| 7.5.1 | Navorsingsvraag 1 | 139 |
| 7.5.2 | Navorsingsvraag 2a..... | 140 |
| 7.5.3 | Navorsingsvraag 2b | 140 |
| 7.6 | Beperkinge van die studie | 140 |
| 7.7 | Implikasies van die huidige ondersoek | 141 |
| 7.8 | Aanbevelings vir verdere navorsing..... | 142 |
| 7.9 | Samevatting en slot | 142 |
| | Bronnelys | 143 |
| | Bylaag A..... | 153 |
| | Bylaag B..... | 154 |
| | Bylaag C..... | 156 |

Lys van tabelle

| | |
|---|-----|
| Tabel 2-1 Vergelykende opsomming van uitgangspunte van die stilistiese en kognitiewe benadering..... | 24 |
| Tabel 3-1: Vergelykende opsomming van Halliday & Hasan (1967) en Stotsky (1983) se benaderings tot leksikale kohesie..... | 48 |
| Tabel 3-2: Voorbeelde van Carstens (1997) se indeling van verbandsmerkers..... | 49 |
| Tabel 3-3: Voorbeelde van herkategorisering van verbandsmerkers volgens Jordaan (2014) | 51 |
| Tabel 4-1: Beskrywing van die verhouding tussen genre en tekstipe wanneer genre geklassifiseer word op situasionele faktore en tekstipe geklassifiseer word op grond van teksinterne eienskappe | 60 |
| Tabel 4-2: Meintjes se illustrasie van die wisselwerking tussen genre en tekstipe | 61 |
| Tabel 4-3: Uittreksel uit Biber 1998 se bevindinge ten opsigte van taalkundige eienskappe tussen genres | 68 |
| Tabel 5-1: Woordhoeveelhede per korpus | 81 |
| Tabel 5-2: Voorbeelde van woorde wat deur outomatiese sifting as moontlike verbandsmerkers geïdentifiseer is, maar 'n ander funksie vervul..... | 90 |
| Tabel 5-3: Uiteensetting van die tipes data wat vir die statistiese ontleding beskikbaar was . | 96 |
| Tabel 5-4: Illustrasie van die aanduiding van die posisie van die kohesieskakels in die tekste | 98 |
| Tabel 6-1: Voorkoms van kohesieskakels in tekste van onderskeie skrywers | 103 |
| Tabel 6-2: Genormaliseerde hoeveelhede kohesieskakels per 100 woorde in die korpus..... | 104 |
| Tabel 6-3: Genormaliseerde data vir hoeveelhede van verskillende tipes Leksikale Kohesie, Verbandsmerkers en Verwysing per 100 woorde in die korpus | 107 |
| Tabel 6-4: Genormaliseerde data ten opsigte van die voorkoms van kohesieskakels op sinsvlak | 110 |
| Tabel 6-5: Genormaliseerde data ten opsigte van die voorkoms van kohesieskakels op paragraafvlak..... | 114 |
| Tabel 6-6: Opsomming van moontlike afwykings wanneer die hoeveelheid voorkoms van kohesieskakels in ag geneem word | 117 |
| Tabel 6-7: Chi-kwadraattoets: Algemene voorkoms van die kohesieskakels in die twee skrywers se tekste | 120 |

| | |
|--|-----|
| Tabel 6-8: Chi-kwadraattoets: Algemene voorkoms van subkategorie van Leksikale Kohesie in die tekste van die onderskeie skrywers..... | 121 |
| Tabel 6-9: Chi-kwadraattoets: Algemene voorkoms van subkategorieë van Verbandswoorde in die tekste van die onderskeie skrywers..... | 122 |
| Tabel 6-10: Chi-kwadraattoets: Voorkoms van subkategorieë van Verwysing in tekste van die onderskeie genres vir die onderskeie skrywers..... | 122 |
| Tabel 6-11: Chi-kwadraattoets: Voorkoms van Leksikale Kohesie in die sinne van tekste van die onderskeie skrywers..... | 123 |
| Tabel 6-12: Chi-kwadraattoets: Voorkoms van leksikale kohesie in die tweede en vierde sinsposisie in tekste van die onderskeie genres van die onderskeie skrywers..... | 123 |
| Tabel 6-13: Chi-kwadraatstatistiek: Voorkoms van Verbandsmerkers in die vier sinsposisies van die onderskeie skrywers..... | 124 |
| Tabel 6-14: Chi-kwadraattoets: Voorkoms van Verwysing in die sinne in tekste van die onderskeie genres van die onderskeie skrywers..... | 125 |
| Tabel 6-15: Opsomming van resultate van beskrywende data..... | 126 |
| Tabel 6-16: Voorstelling van nulmodelle van onderskeie logistiese regresse-analises..... | 132 |
| Tabel 6-17: Logistiese regressie-analise van die algemene voorkoms van leksikale kohesie, verbandsmerkers en verwysing: SPSS analise..... | 134 |
| Tabel 6-18: Logistieke regressie-analise van die algemene voorkoms van die subkategorieë binne leksikale kohesie, verbandsmerkers en verwysing..... | 135 |

Lys van Figure

| | |
|--|-----|
| Figuur 3-1: Tipes verwysing in terme van strukturele oorwegings (Carstens,1997:140)..... | 35 |
| Figuur 5-1: Brown (2004) se voorstelling van die uiteenlopende aard van taalkundige navorsing | 78 |
| Figuur 5-2: Grafiese voorstelling van korpusseleksie | 82 |
| Figuur 5-3: Voorbeeld van etikettering van data | 84 |
| Figuur 5-4: Voorbeeld van etikette binne die sintaktiese konteks WordSmith. | 86 |
| Figuur 5-5: Voorbeeld van Microsoft Excel-lêer waarin die voorkomste van kohesie aangedui is tesame met ander statistiese inligting | 88 |
| Figuur 5-6: Voorbeeld van sortering van etikette binne Microsoft Excel | 88 |
| Figuur 5-7: Voorbeeld van kategorisering van kohesieskakels in Microsoft Excel | 92 |
| Figuur 6-1:Hoofstukuiteensetting: Statistiese ontleding (Afdeling 6.2 - 6.4) | 101 |
| Figuur 6-2: Uiteensetting van gevalle van subkategorieë van verbandsmerkers in die hele datastel | 108 |

1 Inleiding

1.1 Agtergrond en rasionaal

In hierdie studie word 'n moontlike voorkeur vir sekere kohesieskakels of patrone in die gebruik van kohesieskakels van 'n bepaalde outeur ondersoek. Hierdie voorkeur word ondersoek as 'n stilistiese voorkeur in die rangskik van taalelemente, ongeag van die genre waarin geskryf word of watter moontlike stilistiese riglyne en voorskrifte daar vir die genre bestaan.

Louwerse (2004:207) noem dat skrywers (individue) hul sogenaamde merk op dokumente wat hul skryf, laat en dat so 'n dokument uitkenbaar is op grond van die voorkeur vir spesifieke linguistiese kenmerke in 'n teks. Dit beteken dat een van die faktore wat tot variasie in taalgebruik bydra persoonlike styl is. Studies wat styl in taalgebruik ondersoek, word binne die teoretiese raamwerk van stilistiek gedoen.

Volgens Nasiscione (2010:10) is toegepaste stilistiek 'n terrein wat die praktiese gebruik van die beginsels, ontdekkings en teorieë oor taal, literatuur en stilistiek (insluitende kognitiewe stilistiek) ondersoek. Hierdie sambreelbenadering is gerig op die toepassing van die taalkenner se stilistiese bedreweheid op 'n ander terrein (soos forensiese linguistiek). Binne hierdie benadering behels *stilistiese bewustheid* die waarneming van onder meer variasie in vorm en betekenis, stilisties-kohesiewe skakels in die teks en die skep van nuwe betekenis in die teks.

Onder hierdie sambreelterm is daar egter nie duidelike afbakening oor hoe verskillende ondersoekvrae wat met stilistiese ondersoeke verband hou, benader moet word nie. Een van die subterreine van stilistiek is outeurskapstudie en Craig (2004:273) verduidelik kortliks die doel van stilistiese ondersoek teenoor dié van outeurskapstudie. Hy skryf dat stilistiese analise in die algemeen “ongestruktureerd en verklarend” is, terwyl outeurskapstudie poog om antwoorde op bestaande vrae in verband met outeurskap te gee. Stamatatos (2008:538) sê dat rekenaargesteuende outeurskapidentifikasie hoofsaaklik ten doel het om tussen tekste van verskillende outeurs te onderskei deur tekstuele kenmerke van die onderskeie outeurs se tekste te ondersoek. Hy verduidelik dat daar uiteenlopende benaderings en metodologieë is om hierdie doelwit te bereik en skryf die metodologiese onreëlmatighede binne hierdie veld hoofsaaklik aan die interdisiplinêre aard van outeurskapidentifikasiestudies toe.

Soos uit die bogenoemde bekendstelling gesien kan word, bied outeurskapidentifikasiestudies aan die een kant ruimte vir eksperimentele ondersoeke en intuïtiewe afleidings oor styl en aan die ander kant, groot toepassingsmoontlikhede indien die ondersoeke, afleidings en metodologieë wetenskaplik begrond kan word. Die geskiedenis van outeurskapidentifikasiestudie sowel as huidige navorsing handel steeds oor die kwessie van wetenskaplike begronding vir studies wat poog om iets so kwalitatief soos skryfstyl te “kwantifiseer”.

Die grootste bron van voorbehoud van kritici was nog altyd die kwessie van watter linguïstiese kenmerke ’n aanduiding van outeurskap is, hoe daar in die ondersoek besluit word watter kenmerke om te gebruik en hoe dit betroubaar getoets kan word. In hierdie studie word die linguïstiese veranderlike vooraf gekies (kohesieskakels) en die outeurs van die tekste is bekend sodat die geldigheid van kohesiepatrone as moontlik tekenend van ’n individuele styl ondersoek kan word.

Kohesie hou verband met hoe die oppervlakstruktuur (die woorde wat ons kan sien of hoor) van ’n teks onderling of wedersyds (“mutually”) aan mekaar verbind is binne ’n reeks (“sequence”) (De Beaugrande & Dressler, 1981:3).

Kohesieskakels is skakels of verbindings (“ties”) wat verwys na ’n spesifieke “...instance of cohesion...” (Halliday & Hasan, 1976:3). Volgens Halliday en Hasan (1976:3) verwys dit na items wat kohesief aan mekaar verbind is. Items kan op verskillende maniere aan mekaar verbind is om ’n samehangende geheel te vorm. Hierdie verskillende maniere word gekategoriseer om die volgende kohesieskakels te vorm: verwysing, substitusie, ellipsis, konjunksie en leksikale kohesie.

Hoewel die bogenoemde kategorieë ’n breë teoretiese onderbou het, gaan ’n kort definisie uit *A dictionary of linguistics and phonetics* (Crystal, 2008) en ’n voorbeeld uit *Afrikaanse Tekslinguïstiek* (Carstens, 1997) van elke kategorie gegee word, ten einde die basiese verskil tussen die kategorieë uiteen te sit.

- a. **Verwysing** word gebruik om ’n identiteitsverhouding tussen leksikale items uit te druk. ’n Voornaamwoord kan byvoorbeeld verwys na ’n spesifieke selfstandige naamwoord of frase en só die identiteitsverhouding uitdruk (Crystal, 2008:407).

Voorbeeld: *Annari se poppe lê die hele huis vol; sy het dit al weer nie gebêre nie* (Carstens, 1997:137).

- b. *Substitusie*** is 'n term wat gebruik word wanneer 'n leksikale item deur 'n ander een vervang word binne 'n spesifieke struktuur (Crystal, 2008:463).

Voorbeeld: *My potlood* is nou te kort om te skryf. Ek moet beslis 'n ander een kry (Carstens, 1997:220).

- c. *Ellips*** is 'n term wat gebruik word wanneer daar 'n gedeelte van die struktuur van 'n sin weggelaat is. Die semantiese implikasie is egter steeds duidelik wanneer daar na die konteks gekyk word (Crystal, 2008:166).

Voorbeeld: Hansie stap versigtig nader en [^] vryf die groot hond liggies op die kop (Carstens, 1997:244).

- d. *Konjunksie*** is 'n term wat gebruik word vir leksikale items waarvan die hoofdoel is om konstruksies aan mekaar te verbind (Crystal, 2008:101).

Voorbeeld: André het haar nie geglo nie *omdat* sy reeds by 'n vorige geleentheid aan hom leuens vertel het (Carstens, 1997:268).

- e. *Leksikale kohesie*:** "...leksikale kohesie gaan om bepaalde semantiese verhoudinge wat tussen besondere leksikale elemente bestaan..." (Carstens, 1997:313)

Voorbeeld: *Kyk daardie seun* wat soos 'n bok teen die rotse uitklouter! *Die knaap* moet net versigtig wees – 'n mens val maklik teen hierdie kranse af (Carstens, 1997:319).

In hierdie studie sal die voorkeure wat 'n skrywer vir die bogenoemde kategorieë van kohesieskakels mag hê en of daar 'n patroon in die gebruik met betrekking tot posisie en frekwensie van kohesieskakels is, ondersoek word, ten einde 'n bydrae te lewer tot die veld van outeurskapidentifikasie.

1.2 Navorsingsdoelstellings en -doelwitte

1.2.1 Doelstelling

Hierdie studie het hoofsaaklik ten doel om die populêre skrywers André le Roux en Dana Snyman se stilistiese voorkeur vir sekere kohesiemiddele binne die voorgeskrewe stilistiese riglyne van 'n spesifieke genre te ondersoek en te beskryf. Hoewel daar nie noodwendig binne die stilistiese riglyne vir 'n genre (of genres) spesifiek aan kohesieskakels aandag gegee word nie, is linguistiese items interafhanklik van mekaar en kan sekere skakels ook stilisties aangewend word. Hierdie onderskeie riglyne vir betrokke genres vermeng uiteraard met die skrywer se persoonlike styl (Olsson, 2008:30). Dit lei tot die hipotese dat daar 'n sekere “stilistiese vingerafdruk” moet wees wat die skrywer agterlaat, ongeag van die stylkenmerke/-vereistes van die betrokke genre.

1.2.2 Doelwitte

Die ondersoek het ten doel om:

- te bepaal of spesifieke kohesieskakels patroonmatig met 'n bepaalde skrywer se skryfstyl verbind kan word;
- te bepaal in welke mate 'n individu se stilistiese voorkeur in terme van kohesiemerkers deur die reëls van 'n betrokke genre beïnvloed word.

1.3 Navorsingsvrae

Vanuit die bogenoemde navorsingdoelwitte, kan die volgende primêre navorsingsvraag vir hierdie studie geformuleer word:

- Kan die voorkoms van kohesieskakels in die werk van Dana Snyman en André le Roux 'n aanduiding van outeurskap gee?

Twee **sekondêre navorsingvrae** sal ook aandag geniet:

- Beïnvloed die stilistiese genrekonvensies van rubrieke en kortverhale Dana Snyman en André le Roux se gebruik van kohesieskakels?

- Is daar algemene kohesiepatrone wat as kenmerkend van onderskeidelik rubrieke en kortverhale beskou kan word?

1.4 Navorsingsontwerp

In hierdie studie is van 'n korpusgebaseerde benadering gebruik gemaak. Biber, Conrad en Reppen (1998:4) noem 'n aantal karaktereenskappe van 'n korpusgebaseerde benadering.

Die korpusgebaseerde benadering is empiries van aard en analiseer patrone in “werklike” tekste wat op 'n natuurlike wyse geskep is. Dit word verder gebruik wanneer groot versamelings tekste geanaliseer moet word. Hierdie benadering kan ook gebruik word om patrone, reëlmatighede en onreëlmatighede in tekste te identifiseer, wat analise deur die navorser moontlik maak. Laastens maak die korpusgebaseerde benadering grootliks van rekenaars en doelontwerpte sagteware gebruik om spesifieke verskynsels te soek, te groepeer, ensovoorts.

Die korpusgebaseerde benadering is 'n benadering “wat empiriese ondersoeke na taalvariasie en -gebruik (fasiliteer) met die resultaat dat navorsingsbevindinge meer veralgemeenbaar en geldig is as wat met 'n ander benadering haalbaar is” (Biber, 2012; eie vertaling) en maak staat op beide kwalitatiewe en kwantitatiewe analisemetodes om navorsingsdoelwitte te bereik. Kwantitatiewe data-analise word grootliks gebruik om verskynsels in outentieke tekste te tel en op te teken of te annoteer, waarna die navorser tydens die kwalitatiewe analise hierdie verskynsels analiseer en probeer om onderliggende patrone raak te sien (Biber *et al.*, 1998:4-5).

1.5 Navorsingsmetodologie

In hierdie studie is daar eerstens 'n oorsig oor die literatuur met betrekking tot outeurskapstudie voorgehou, waarna die metodes en prosesse van data-analise beskryf is. Uiteindelik is gevolgtrekkings en aanbevelings op grond van die data-analise gemaak.

1.5.1 Literatuurstudie

'n Toepaslike literatuurstudie wat hoofsaaklik op drie teoretiese uitgangspunte fokus, is gedoen ten einde die mees aanvaarbare en heersende beskouings oor die studieveld te bekom. Die fokus is eerstens op styl, stilistiek en outeurskapstudie wees ten einde die breë teoretiese raamwerk waarbinne die studie plaasvind, daar te stel. Volgende is die relevante tekslinguistiese aspekte

bespreek, met 'n fokus op kohesieskakels en die verdeling daarvan vir die doel van kategorisering. Laastens is 'n teoretiese oorsig van genres en tekstipes, met spesifieke fokus op rubrieke en kortverhale, gegee, ten einde die moontlike stilistiese voorskrifte (of standaardpraktyke) te probeer vasstel.

1.5.2 Data-analise

In die kwalitatiewe ondersoek van hierdie studie is die skrywer se gebruik van kohesieskakels gemerk en geanaliseer. Kohesieskakels is 'n stilistiese veranderlike (McMenamin, 1993:256), en 'n skrywer kan 'n stilistiese voorkeur vir sekere kohesieskakels of patrone of groeperings van kohesieskakels hê. Vir hierdie studie is van die rekenaarprogram WordSmith Tools¹ se weergawe 7.0 gebruik gemaak word. WordSmith Tools is analysesagteware wat linguistiese data kan help sorteer en dit word gebruik wanneer leksikonitems binne hul tekstuele konteks ondersoek word. In hierdie studie is die data (elektroniese korpora in txt-formaat²) vooraf getiketteer deur die relevante kohesieskakel tussen skerp hakies (<skakel>) aan te dui. Wanneer die data in WordSmith Tools ingevoer is, is die etikette of gevalle van kohesie gesoek en gesorteer en dit het gedien as die metadata wat verder geklassifiseer is.

Carter (1982:6) omskryf 'n benadering waar stilistiese elemente sistematies beskryf word as 'n objektiewe benadering tot stilistiese ondersoek en sê dat hierdie tipe linguistiese analise "...bewyse (bied), al dan nie, van intuïtiewe interpretasies van styl en informele en lukrake observasies van taalvariasie uitsluit. Die sistematiese en eksplisiete resultate van so 'n analise onderskei stilistiese ondersoeke van literêre kritiek" (eie vertaling). Hierdie studie laat dus letterkundige interpretasie of die aanwesigheid of afwesigheid van spesifieke letterkundige verskynsels buite rekening.

1.5.3 Datastel

Vir hierdie studie word is vanuit twee genres van twee skrywers te gebruik. Daar is besluit op tekste van André le Roux en Dana Snyman. Hierdie skrywers kom beide uit 'n joernalistiek-

¹ Scott (2008) is die ontwikkelaar van *WordSmith Tools*. Daar sal vervolgens slegs na *WordSmith* of *WordSmith Tools* verwys word.

² Hardekopie-weergawes van tekste is geskandeer en in Word-formaat gestoor. Tekste is nagegaan vir korrektheid aangesien outomatiese omskakeling na elektroniese kopie soms letterfoute tot gevolg het. Daarna is die Word-lêers in txt-formaat gestoor.

milieu en hul werk word onder populêre genre geklassifiseer. Daar is 'n groot aantal faktore wat die styl of stilistiese voorkeure van skrywers kan beïnvloed (Olsson, 2008:30-31). Hierdie studie fokus nie op die moontlike *redes* vir die stylvariasie van skrywers nie, maar op die mate van die skrywer se stylvariasie wanneer daar binne verskillende genres gewerk word. In die keuse van tekste is dit dus belangrik om eerder na vergelykbare *tekste* te kyk as na skrywers met vergelykbare opleiding, agtergrond, ouderdom, ensovoorts. Beide skrywers is manlik en werk binne die mediabedryf en skryf ook kreatiewe tekste. Die tekste wat vir hierdie studie gekies is, is kortverhaal- en rubriektekste van beide die gekose skrywers. In die keuse van spesifieke tekste is vergelykbaarheid ten opsigte van publikasiedatums en genre ook verreken.

Die kortverhaalttekste wat vir hierdie studie gebruik is, is tekste uit Dana Snyman se kortverhaalbundel *Op die agterpaai* (2008) en uit André le Roux se bundel met kortverhale en sketse *Kom ons sit sommer by die tafel* (2008). Daar is 'n soortgelyke seleksie van tekste uit hierdie twee bundels gemaak. Verder is 'n willekeurige seleksie van beide skrywers se rubrieke wat rondom 2008 verskyn het, gemaak. Hierdie rubrieke is uit vergelykbare publikasies, naamlik Afrikaanstalige koerante, gekies.

1.6 Waarde van die navorsing

Die hoofbydrae van die studie is op die terreine van die stilistiek en outeurskapidentifikasie. Soos uit die literatuuroorsig ten opsigte van hierdie terreine duidelik is (sien hoofstuk 2), beveel die meeste navorsers verdere empiriese ondersoeke aan ten einde meer legitimiteit aan die teoretiese aannames binne die velde te gee sowel as om die bestaande benaderings te toets. In hierdie studie is gepoog om juis 'n bydrae op hierdie gebied te lewer. Die keuse van linguistiese veranderlikes en oorwegings ten opsigte van navorsingsontwerp en metodologie, is getoets binne 'n gekontroleerde omgewing met tekste waarvan die skrywers bekend is, sodat die effektiwiteit van hierdie spesifieke benadering bepaal kan word.

1.7 Vooruitskouing

Die volgende hoofstukindeling word as 'n vooruitskouing van die studie gegee:

In hoofstuk 2 word die aard van outeurskapidentifikasiestudies en die oorvleueling tussen stilistiese ondersoeke en outeurskapstudie bespreek. In hoofstuk 3 word ondersoek ingestel na die aard van tekstualiteit en kohesie as een van die beginsels van tekstualiteit. Die verskillende

tipes kohesie word ook bespreek. In hoofstuk 4 val die fokus op die omskrywing van die terme ‘genre’ en ‘tekstipe’ en daar word gepoog om hierdie omskrywings in te span om as raamwerk te dien waarbinne verskillende soorte tekste³ geanaliseer kan word. Daar word ook ondersoek ingestel na die stilistiese reëls (indien enige) of aanvaarde stilistiese norme wat by die skep van verskillende teksoorte ’n rol speel. In hoofstuk 5 word die navorsingsmetodologie uiteengesit, waarna die data-analise en interpretasie van die resultate in hoofstuk 6 bespreek word. In hoofstuk 7 word ’n samevatting van die studie gegee. Daar word ook aandag gegee aan die navorsingsvrae en beperkinge van die studie. Aanbevelings vir verdere navorsing word ook gemaak.

³ ‘Soorte tekste’ word op hierdie stadium gebruik as ’n breë term vir *tekstipes* en *genres*. Die duidelike afbakening tussen hierdie twee konsepte word in hoofstuk 4 hanteer.

2 Styl, Stilistiek en Outeurskapidentifikasie

2.1 Inleiding

Dit wil voorkom asof daar nie duidelike afbakening is tussen die velde van stilistiek, rekenaargesteuende stilistiese ondersoeke, stilometrie en ondersoeke na die styl van genres en die skryfstyl van outeurs nie. Die terme wat in hierdie velde gebruik word, word dikwels afwisselend gebruik om konsepte te beskryf wat oorvleuel, baie nou verwant of inklusief is. Om hierdie rede is dit nodig om terminologie en benaderings binne die studieveld vir hierdie subvelde af te baken.

Ten einde 'n agtergrond van die bogenoemde subvelde en benaderings te gee, en dit ook te kontekstualiseer vir die doel van hierdie studie, word die wisselwerking tussen stilistiek en outeurskapidentifikasie vervolgens bespreek. Uit hierdie bespreking sal dit duidelik wees dat hierdie studie binne die veld van outeurskapstudies val, hoewel daar elemente van stilistiese ondersoekmetodes teenwoordig is. Vervolgens word 'n oorsig oor die veld van outeurskapstudies gegee, deur eerstens die relevante definisies en kwessies binne die veld te bespreek. Daarna volg 'n oorsig van die ontwikkeling en ontstaan van die studieveld, asook hoe die probleme ten opsigte van navorsing oor outeurskap in die verlede aangepak is. Voorts word die meer onlangse ontwikkelinge binne die studieveld bespreek, waarna die voorgestelde teoretiese raamwerke vir die uitvoer van outeurskapidentifikasiestudies voorgehou word.

2.2 Wisselwerking tussen stilistiek en outeurskapidentifikasie

Volgens Crystal (2008:460) kan stilistiek beskryf word as 'n "...vertakking van linguistiek wat die kenmerke van konteksgebonde gebruike (variasies) van taal ondersoek en probeer om beginsels vas te stel wat die keuses van die individu en (of) sosiale groep kan verduidelik" (eie vertaling). Na aanleiding van die definisies, kom dit dus voor asof stilistiek as 'n oorkoepelende term gebruik word vir twee tipes studies. Eerstens word dit gebruik vir studies wat verband hou met die *ondersoek van konteksgebonde variasies* in die linguistiese kenmerke van 'n taalgebruiker se taal. Tweedens het dit betrekking op studies wat probeer om die moontlike *redes vir die variasie* te gee of in berekening te bring. Die onderskeid tussen hierdie twee uitgangspunte word egter in die praktyk soms duideliker gestel en ander kere glad nie gemaak nie.

In sy boek, *Stylistics – A Resource Book for Students* (2004:2), sê Simpson dat die stilistiese ondersoekmetode hoofsaaklik op taalgebruik fokus ten einde tekste te interpreteer. Hy verduidelik verder dat die vorm van die teks wat die teksstruktuur vorm, 'n aanduiding van die funksie van die teks is. Hy voer aan dat 'n verslag of verrekening van taalkundige kenmerke kan verduidelik *hoekom* sekere tipes betekenis moontlik is (eie klem). Die taalvariasie wat vroeër beskryf is, word dus ondersoek met die fokus op die rede vir die variasie. *Hoekom* word 'n sekere taalstruktuur of stel linguistiese kenmerke bo 'n ander verkies, wanneer betekenis binne 'n teks geskep word (Simpson, 2004:22).

Outeurskapidentifikasie of outeurskapstudie is 'n studieveld wat baie nou verwant is aan die stilistiek. As bogenoemde definisies en afbakenings in ag geneem word, wil dit voorkom asof dit 'n onderafdeling van stilistiek beskou kan word, aangesien dit 'n stel taalkundige kenmerke ondersoek wat variasie tussen tekste meebring. Die doelwitte van outeurskapidentifikasie verskil egter van stilistiese analise. Craig (2004:273-288) skryf in 'n hoofstuk genaamd “Stylistic Analysis and Authorship Studies” dat hierdie twee velde verwant is, maar voor-die-hand-liggende verskille toon.

Dit is belangrik om te onderskei tussen die vakterm, *stilistiek*, wat die taalkundige ondersoekveld beskryf en terme soos *stilistiese variasie* en *stilistiese beskrywing* wat te doen het met enige ondersoek wat poog om die styl van taalgebruik te beskryf, te analiseer of as basis vir afleidings te gebruik. Outeurskapidentifikasiestudies poog byvoorbeeld om tekste as die produk van 'n spesifieke outeur *toe te ken* op grond van *spesifieke stilistiese kenmerke* binne die teks, maar deel nie noodwendig die primêre *beskrywende* doelwitte van *stilistiese* analise nie.

Craig beskryf stilistiese analise as “open-ended and explanatory” (2004:273) terwyl hy sê dat outeurskapidentifikasie probeer om antwoorde vir bestaande vrae (in verband met outeurskap) te gee. Studies in outeurskapidentifikasie volg oor die algemeen 'n kwantitatiewe metodologie en Stamatatos (2008:538) is van mening dat rekenaargesteunde outeurskapidentifikasie hoofsaaklik ten doel het om tussen tekste van verskillende outeurs te kan onderskei deur tekstuele kenmerke van hulle onderskeie tekste te ondersoek. Hoewel hierdie tipe studies 'n lang geskiedenis het, bestaan daar nog baie teenstrydige veronderstellings onder navorsers binne die veld oor die metodologiese benadering en teoretiese begroning wat die mees betroubare resultate vir so 'n studie sal oplewer. Stamatatos (2008:553) skryf die uiteenlopende benaderings en metodologiese onreëlmatighede binne hierdie veld gedeeltelik toe aan die

interdisiplinêre aard van outeurskapidentifikasiestudie. Hy sê egter dat daar sedert die 21ste eeu merkwaardige stappe in die “regte” rigting gemaak is. Stamatatos *et al.* (2014:877-878) verduidelik dat outeurskapanalise met drie aspekte van die persoonlike styl van outeurskap te make het, naamlik outeurskapidentifikasie, outeurskapprofilering en outeurskapgroepering. Die eerste het meestal ten doel om ’n teks - of tekste - as die werk van ’n sekere outeur te kan uitken. Meestal word die outeur vanuit ’n aantal moontlike kandidate gekies (Stamatatos *et al.*, 2014:877). Outeurskapprofilering en -groepering handel onderskeidelik oor die identifisering van demografiese inligting van die outeur op grond van sy/haar skryfstyl en met die verdeling van tekste met verskeie outeurs deur die gedeeltes met dieselfde veronderstelde outeur saam te voeg (Stamatatos *et al.*, 2014:878). Die huidige studie val in die kategorie van outeurskapidentifikasie of -toewysing.

Rudman (1998:357) verduidelik dat daar drie dissiplines is waarvan ’n ‘ideale navorser’ kennis behoort te hê wanneer outeurskapstudie gedoen word, naamlik linguistiek, rekenaarprogrammering en statistiek. Hy merk tereg op dat weinig navorsers volkome toegerus is vir die eise van so ’n studie en dat navorsers uit verskillende spesialisasievelde dikwels saamwerk ten einde kennis van al die nodige vakgebiede in ag te kan neem. In hierdie studie word een linguistiese veranderlike getoets en is die veld van linguistiek dus van toepassing. ’n In-diepte teoretiese oorsig van die rekenaargesteuende en statistiese metodes val dus buite die bestek van hierdie studie.

Daar word gereeld na die konsep daarvan om styl te “kwantifiseer” deur rekenaargesteuende statistiese studies (of andersins) verwys as stilometrie. Stilometrie kan beskryf word as ’n studieveld waarin gepoog word om stilistiese patrone, meestal met die hulp van rekenaarsagteware, te kwantifiseer en die “statistiese struktuur” van (meestal literêre) tekste te bestudeer (Crystal, 2008:460). Joula (2008:238) noem in sy omvattende artikel *Authorship Attribution* dat die terme *outeurskapidentifikasie* en *stilometrie* byna as sinonieme gebruik kan word.

Soos uit die bogenoemde oorsig blyk, bied die aspek van outeurskapidentifikasiestudie wat te make het met stilistiese analise aan die een kant ruimte vir eksperimentele ondersoeke en intuïtiewe afleidings oor styl. Indien die ondersoeke en afleidings geldig wetenskaplik begrond kan word, is daar vir die ondersoeke in die veld toepassingsmoontlikhede, byvoorbeeld vir wetstoepassing en binne onderrig-en-leer-kontekste. Dit kan ook van groot waarde wees vir die studie van artefakte ongeag of dit vir literêre of antropologiese doeleindes is. Regdeur die

geskiedenis van outeurskapstudie sowel as binne huidige tendense is pogings om op 'n *wetenskaplike manier* iets so kwalitatief as skryfstyl te kwantifiseer, merkbaar.

2.3 Outeurskapstudies

2.3.1 Definisies, kwessies en kritiek

Joula (2008:233) beskryf outeurskapidentifikasie in sy omvattende artikel oor dié studieveld as “...die wetenskap (daarvan) om karaktereienskappe van 'n skrywer af te lei van karaktereienskappe van 'n dokument wat deur die betrokke skrywer geskryf is” (eie vertaling). Hy noem wel later (Joula, 2008:238) dat hierdie definisie opsetlik geformuleer is om nie te spesifiek te wees nie. Dit is duidelik dat daar met ‘karaktereienskappe’ nie slegs na persoonlikheid verwys word nie, maar na enige aantal faktore wat moontlik die skrywer en die teks (geskrewe of gesproke) met mekaar kan verbind. Stamatatos *et al.* (2014:877) se verduideliking van outeurskapanalises studies is ook relatief onspesifiek en hulle sê bloot dat hierdie tipe studies met “...die persoonlike styl van outeurs [te make het]...”. Joula (2008:233) noem ook dat vrae oor die outeurskap van dokumente sedert die ontstaan van geskrewe dokumente al bestaan en studies daarvoor 'n lang geskiedenis het (vergelyk ook Stamatatos, 2008 en Baayen *et al.*, 2002). Hy (Joula, 2008:169-170) noem egter dat kwessies in verband met outeurskap beide van die oudste en nuutste vraagstukke is. Beide Kotzé (2007) en Nel (2014:83) beaam dat kwessies in verband met die outeruskapbepaling van tekste al vir honderde jare jare ondersoek word. Sedert die begin van formele outeurskapstudie, het die metodologiese en teoretiese basis waarop hierdie studies gegrond is, groot kritiek ontlok, beide deur navorsers en kritici buite die veld (vergelyk Rudman, 1998 en Joula, 2012). Die meeste kritiek is gebaseer op die oënskynlike lukrake en subjektiewe metodiek wat in outeurskapstudies gebruik word. Die grootste bron van kommer vir kritici was, volgens beide Rudman (1998:252-253) en Joula (2012:275-276), nog altyd die kwessie van watter linguistiese kenmerke 'n aanduiding van outeurskap is, hoe die navorsers besluit watter kenmerke om te gebruik en hoe dit betroubaar getoets kan word. Die resultaat van hierdie kritiek is deurlopende voorstelle oor verbeterde tegnieke om navorsingsresultate herhaalbaar

en betroubaar te maak. Dit is dus sinvol om 'n kort oorsig te gee oor sommige van die metodes wat in die moderne era voorgestel en getoets is, asook die resultate van hierdie ondersoek.

2.3.2 Oorsig van die geskiedenis van moderne outeurskapstudies

Ten einde te illustreer watter oorwegings en metodes getoets is in die soeke na wetenskaplike betroubaarheid van bevindinge in die veld, word 'n oorsig van die belangrikste mylpale in die geskiedenis van moderne outeurskapstudies gegee. Holmes (1994:87-106) en Joula (2008:233-334) gee beide 'n volledige en ooreenstemmende opsomming van hierdie geskiedenis⁴.

2.3.2.1 Klassieke benaderings

Van die vroegste formele benaderings tot outeurskapidentifikasie het op woordvlakkenmerke gefokus. Holmes (1994:88) identifiseer die vroegste statistiese publikasie van 'n studie oor outeurskapidentifikasie as dié van Mendenhall (1887) wat ná aanvanklike analyses van tekste deur verskillende outeurs voorgestel het dat woordlengte 'n linguistiese kenmerk kan wees wat die tekste van verskillende skrywers van mekaar kan onderskei. Hy het egter later bevind dat sy metodologie tesame met die keuse van linguistiese kenmerk nie 'n akkurate aanduiding is van outeurskap nie, aangesien daar soortgelyke variasie is tussen tekste van dieselfde outeur as tekste van verskillende outeurs. Ander woordvlakkenmerke wat in Holmes (1994:88) se historiese oorsig na vore kom, is studies oor lettergrepe, woordeskatverskeidenheid en woordfrekwensies, en sg. Funksiewoorde. Dit is woorde wat hoofsaaklik 'n grammatikale funksie het, byvoorbeeld voornaamwoorde, voorsetsels, ensovoorts (Crystal, 2008:203). Uit hierdie woordvlakkenmerke het studies na die gebruik van funksiewoorde as linguistiese veranderlike die beste resultate gehad ten opsigte van 'n eienskap van 'n spesifieke skrywer se skryfstyl.

Sedertdien het verskeie navorsers met ander metodologiese benaderings probeer om deur die analise van woordlengte tog 'n positiewe resultaat te verkry, maar met relatief min sukses. Onder hierdie studies noem Holmes (1994:88) dié van Brinegar (1963), Mosteller en Wallace (1964) asook Smith (1983). Mosteller en Wallace (1964) se studie val onder die belangrikste

⁴ Sommige studies en hul bydraes tot die ontwikkeling van outeurskapstudie, sowel as hul verwysings, is uit onderskeidelik Joula (2008) en Holmes (1994) verkry. Die verwysings wat hier ter sake is, is: Bruno (1974), Brainerd (1973), Brainerd (1974), Brinegar (1963), Fucks (1952), Fucks en Bruno (1965), Mendenhall (1887), Mosteller en Wallace (1964), Yule (1944) en Zipf (1932).

mylpale in die geskiedenis van moderne outeurskapidentifikasiestudies (Joula, 2008:241-242). Smith (1983) het egter, nes Mendenhall (1887) bykans 'n eeu tevore, bevind dat die analise van woordlengte alleen nie 'n akkurate aanduiding van outeurskap is nie – veral wanneer tekste van verskillende genres of uit verskillende eras bestudeer word.

Fucks (1952), Fucks en Bruno (1965), Bruno (1974) en Brainerd (1974) het in hul studies op die taalkundige uniekheid in die gebruik van lettergrepe gefokus. Hierdie studies het meestal bevind dat daar wel variasie is tussen verskillende tekste op lettergreepvlak, maar dat die variasie nie noodwendig aan verskillende outeurstyle toegeskryf kan word nie. Dit is steeds 'n relevante veranderlike om in multidimensionele studies⁵ in ag te neem, maar eindelik maak Brainerd (1974) die gevolgtrekking dat analises van lettergrepe in 'n teks, meer betroubaar is as 'n aanduiding van genre (bv. brief teenoor koerantartikel) as dié van outeurskap.

Volgens Holmes (1994:93) is die idee dat woordeskatverskeidenheid ondersoek kan word as kenmerkend van 'n spesifieke outeur gebaseer op die aanname dat 'n skrywer 'n sekere woordeskat of leksikon het wat eie aan die individu is, en dat daar sekere voorkeure bestaan vir sekere van die leksikonitems. In hierdie verband is die gebruik van sinonieme al ondersoek deur onder meer te bepaal of 'n skrywer 'n voorkeur het vir 'n sekere leksikonitem bo 'n sinoniem daarvan met dieselfde betekenis. Probleme wat Mosteller en Wallace (1964) egter met hierdie metode ondervind het, is dat daar dikwels nie genoeg data is om genoeg voorbeelde van bogenoemde leksikonspesifieke voorkeur te kan illustreer nie.

'n Veranderlike wat verwant is aan woordeskatverskeidenheid, is woordfrekwensie. Holmes (1994:98-91) sê dat die ondersoek na woordfrekwensies nie die verspreiding van 'n woord per x hoeveelheid woorde behels nie, maar die voorkoms van 'n individuele woord in die hele teks. Zipf (1932) is van die eerste navorsers wat 'n studie gepubliseer het waarin hy woorde uit 'n teks volgens frekwensie rangskik het. Holmes verwys egter na Tallentire (1972) wat uitgewys het dat baie van hierdie rangorde-lyste sterk ooreenkomste sal toon omdat omtrent 90% van tekste in Engels uit slegs 10% van die Engelse woordeskat bestaan. Volgens inligting van die *Oxford English Corpus* (OEC) bestaan 25% van alle Engelse tekste in dié korpus, uit slegs tien verskillende woorde⁶. So ook bestaan 50% van die tekste in die OEC uit slegs 100 verskillende woorde (Oxford Dictionaries, 2018).

⁵ word volledig in afdeling 4.6.1 bespreek.

⁶ *the, be, to, of, and, a, in, that, have* en *I*

In die soeke na taalkundige eienskappe wat moontlik 'n aanduiding van outeurskap kan gee, is bogenoemde benaderings almal voorgestel en nagevors. Die bevindinge van hierdie verkennende ondersoeke dui oor die algemeen aan dat die voorkoms van elk van hierdie eienskappe tot 'n mindere of 'n meerdere mate 'n aanduiding van outeurskap is. Hierdie metodes word deesdae saam met ander metodes ingespan om 'n meer volledige verslag van 'n outeur se linguistiese keuses saam te stel soos wat byvoorbeeld in multidimensionele studies⁷ die geval is. Die woordvlak-kenmerk wat as die mees betroubare aanduiding van outeurskap geag word, is die verspreiding en frekwensie van grammatikale funksiewoorde. Volgens Holmes en Forsyth (1994:111) is die bekendste studie wat hierdie stelling bewys het, die studie deur Mosteller en Wallace (1964) wat die sogenaamde *Federalist Papers* as datastel gebruik het.

Die *Federalist Papers* is 'n stel van 85 koerant-essays wat tussen 1787 en 1788 deur 'n anonieme skrywer in 'n koerant gepubliseer is. Intussen het dit aan die lig gekom dat daar drie skrywers was wat die opstelle geskryf het. Joula (2008:242) beskryf hierdie tekste as 'n ideale toetsgrond vir nuwe metodologie in verband met outeurskapidentifikasie. Volgens Joula is dit ideaal omdat dit gedurende dieselfde tyd gepubliseer is, in dieselfde koerant en in dieselfde genre, en ook omdat dit aanvanklik anoniem gepubliseer is, maar die skrywers later bekendgemaak is. Mosteller en Wallace (1964) het hierdie tekste gebruik om hulle benadering in verband met grammatikale funksiewoorde te toets. Hulle het die verspreiding van 30 grammatikale funksiewoorde wat uit die tekste verkry is, getoets en aan sekere outeurs toegeken. Sedertdien is vele outeurskapstudies op die *Federalist Papers* uitgevoer, wat grotendeels Mosteller en Wallace se bevindinge beaam. Joula (2006:243) noem die implikasie van hierdie onderskraging deur te sê dat hierdie fokus op grammatikale funksiewoorde 'n belangrike tema in moderne outeurskapidentifikasiestudies geword het.

Behalwe vir die keuse van linguistiese veranderlikes, word verskeie modelle vir die statistiese berekening en voorstelling van die variasie voorgestel. Holmes (1994:92) en Joula (2008:240) maak beide melding van Simpson (1949) se benadering tot die meting van woordeskatverskeidenheid. Hierdie benadering is gebaseer op die waarskynlikheid (D) dat "...twee willekeurig geselekteerde voorkomste van taaltekens aan dieselfde taaltipe sal behoort" (Holmes, 1994:92, eie vertaling). 'n Ander statistiese benadering is dié van Yule (1944), naamlik, die "Characteristic K". Dit is 'n maatstaf van woordeskatverskeidenheid wat

⁷ Multidimensionele outeurskapidentifikasiestudie is studies wat die voorkoms van verskeie linguistiese veranderlikes ondersoek. Vir 'n volledige bespreking, vergelyk afdeling 4.6.1

gebaseer is op die aanname dat die voorkoms van enige gegewe woord gesien kan word as 'n Poisson-verdeling⁸. Verder is verskeidenheid in terme van woordeskatgebruik al deur 'n uitdrukking vir die sogenaamde entropie van 'n sisteem⁹ en deur die gebruik van die Waring-Herdan-model¹⁰ geïllustreer. Hierdie metodes is meestal gebruik om die frekwensieverspreiding van die aantal leksikale items wat 'n spesifieke hoeveelheid keer verskyn, voor te stel.

Behalwe vir linguistiese kenmerke op woordvlak, was daar 'n aantal studies wat kenmerke op sintaktiese en grammatikale vlak ondersoek het. Hierdie studies het onder meer woordsoortfrekwensies en die voorkoms van sekere woordpare, asook sinslengte en sintaktiese struktuur ondersoek.

Holmes (1994:89) verduidelik dat jou rasonale proses of “manier” van dink die tipe leksikonitems waarvoor jy 'n voorkeur het, kan beïnvloed. So verbind hy onder meer 'n meer rasonale tipe denkwysie met 'n hoër voorkoms van selfstandige naamwoorde en naamwoordfrases en 'n gesindheid van “dinamiese empatie en [’n] aktiewe houding” (Holmes, 1994:89, eie vertaling) met 'n voorkeur vir werkwoorde. Dit wys op die intuïtiewe afleiding dat 'n sekere geestestoestand of denkwysie sekere grammatikale besluite kan beïnvloed. Brainerd (1973) het byvoorbeeld gepoog om tussen romantiese fiksie en 'n roman te onderskei deur die gebruik van lidwoorde en voornaamwoorde te analiseer, aangesien hierdie aspekte beïnvloed word deur formele en informele registers. Hy het gevind dat daar in romantiese fiksie minder voornaamwoorde en meer lidwoorde voorkom as in 'n roman. Hy het egter ook bevind dat daar groot variasie tussen skrywers is en sy bevindinge nie noodwendig aan stylverskille tussen genres toegeskryf kan word nie.

⁸ 'n Poisson-verdeling is 'n voorbeeld van 'n diskrete waarskynlikheidsverdeling. Die term *diskreet* beteken in hierdie konteks dat die data bekom word deur 'n verskynsel te tel (en nie te meet nie). 'n Poisson-verdeling wys die waarskynlikheid dat 'n verskynsel gaan voorkom gedurende 'n sekere tyd of ruimte. Hierdie verspreiding word veral gebruik in kontekste waar die waarskynlikheid dat 'n verskynsel gaan voorkom baie laag is of in kontekste waar die voorkoms oënskynlik toevallig is, soos die hoeveelheid oproepe wat 'n skakelbord ontvang gedurende 'n gegewe tyd (Oakshott, 2006:193-200)

⁹ Entropie word in die natuurwetenskappe gebruik om die mate van ewekansigheid in 'n stelsel te meet en voor te stel (Oakes, 1998:58).

¹⁰ Die Waring-Herdan model is in die gang van taalkundige studies ontwikkel en poog om die waarskynlikheid dat 'n sekere leksikale item 'n sekere hoeveelheid kere (n) in 'n teks gaan voorkom, te voorspel (Holmes, 1994:94)

2.3.2.2 Navorsing sedert 1990

Sedert die 1990's het daar verskeie studies verskyn wat die stand van die veld van outeurskapidentifikasie, sowel as probleme en moontlike oplossings en metodologiese benaderings gee (vergelyk Stamatatos, 2008 en Joula, 2008). Die meeste studies is dit eens dat betroubaarheid van resultate en herhaalbaarheid van die studies die grootste struikelblokke in die veld is. Hierdie studies volg gewoonlik 'n bepaalde patroon en raak een of meer van drie navorsingsareas van outeurstoewysing aan. Navorsers identifiseer bepaalde probleme binne die veld. Hierdie probleme handel eerstens meestal oor metodologie en/of teoretiese begronding. Tweedens maak die navorser dikwels voorstelle in verband met hoe die probleme wat hy/sy uitgelik het moontlik opgelos kan word. Laastens is daar publikasies wat die metodologiese opsies voorstel om betroubaarheid en geldigheid in outeurskapstudie te kan verbeter. Dit word gewoonlik gedoen aan die hand van toetsing van ou en nuwe prototipes in 'n navorsingsontwerp waarvan sekere veranderlikes soos die outeurs bekend is. Dit is ook wat in hierdie studie onderneem word.

2.3.3 Probleme, oplossings en metodologiese benaderings

Rudman (1998:351-465) begin sy artikel, "The state of Authorship Attribution Studies: Some Problems and Solutions", deur te sê dat nie-tradisionele outeurskapstudies - dit is studies wat gebruik maak van stilistiek, statistiek en die rekenaar - lankal deur die eksperimentele fase moes beweeg na 'n meer wetenskaplike een. Hy gaan voort deur die veld in 'n negatiewe lig te stel. Tog stel hy oplossings vir die probleme voor wat verdere, meer omvattende navorsing aanmoedig. Die probleme wat hy identifiseer, is egter geldig en behoort oorweeg te word wanneer 'n studie binne die veld van outeurskapidentifikasie aangepak word. Hy hou die gebrek aan konsensus onder kenners oor metodiek en resultate (soos ook bo genoem) voor as bewyse van die metodologiese uitdagings in die veld (Rudman, 1998:352-353).

2.3.3.1 Probleme

Baie van die probleme wat Rudman (1998:354-357) bespreek, hou verband met aspekte van die navorsingsproses, vanaf die kennis, vaardighede en ervaring van die navorser tot by die kies van relevante metodologiese benaderings. Probleme wat verband hou met die navorser sluit onder meer 'n gebrek aan toewyding tot die veld in. Volgens hom lei dit tot 'n gebrekkige verkenning van die literatuur. Verder sê Rudman (1998:345) dat navorsers dikwels die band

wat outeurskapstudies met ander navorsingsvelde soos statistiek en linguistiek het, ontken en dus onbewus is van die insigte in hierdie velde.

Wanneer 'n outeurskapstudie onderneem word, kan daar ook probleme ontstaan wat verband hou met navorsingsontwerp en data (Rudman, 1998:354-357). Die kwaliteit van die data en oorwegings in die seleksie van die data kan probleme skep. Data wat beskikbaar en in die regte (elektroniese) formaat is, kry voorkeur bo data wat werklik relevant tot die probleem of die beste eksemplaar van 'n spesifieke verskynsel is. Verder verduidelik Rudman (1998:254-257) ook dat eksterne faktore wat die data kan beïnvloed dikwels nie in berekening gebring word nie. Die mondelinge tradisie en die rol van 'n klerk of skriba het byvoorbeeld 'n beduidende invloed op die formaat van historiese tekste. In hedendaagse tekste word die invloed van die redaksieproses ook dikwels buite rekening gelaat. Laastens is daar 'slaggate' in die dataversamelingsproses, byvoorbeeld tikfoute, skandeerfoute, ensovoorts.

Die laaste probleem wat Rudman (1998:357) noem, hou verband met die gebruik en soms misbruik van statistiese metodes en analise. Hy noem dat statistiek soms bloot gebruik word in 'n poging om wetenskaplike geldigheid aan die studie of resultate toe te dig en in die proses word die geldigheid van die bevindinge prysgegee.

In hierdie studie is gepoog om alle redelike maatreëls te volg om nie bogenoemde foute te maak nie. In terme van dataversameling, -verwerking en -analise, is oorweeg watter data en metodes vir die *navorsingsvrae* van hierdie ondersoek relevant is en daar is gepoog om nie bloot data of verwerkingsmetodes op grond van beskikbaarheid te kies nie. Daar word ook van statistiese analisemetodes in hierdie studie gebruik maak. Die relevansie van die analisemethode ten opsigte van die navorsingsvrae sal oorweeg word

2.3.3.2 Oplossings

Soos verduidelik is, is Rudman se oplossings gemik op verdere navorsing en onderwerpe waarop daar in toekomstige studies gefokus kan word. Hy lê klem op die feit dat navorsingstudies en metodologie beter ontwerp en beplan moet word en die beginsels van wetenskaplike navorsing gevolg moet word (Rudman, 1998:359). Verder stel hy (Rudman, 1998: 359-361) voor dat akademici in die veld meer navorsing moet publiseer wat die veld sal afbaken ten einde standaardwerke oor aanvaarde praktyke en metodiek binne outeurskapstudie daar te stel. Hy betoog ook vir 'n omvattende teoretiese raamwerk wat getoets en bespreek moet

word. Hy moedig samewerking aan ten einde bestendigheid ten opsigte van praktyke, terminologie en ander fasette van die navorsingsproses te bevorder. Ander oplossings wat hy in sy artikel noem, hou onder meer verband met die opleiding van persone wat 'n invloed het op 'n spesifieke veld se navorsingsproses, soos tydskrif- en konferensieresensente, agentskappe wat navorsing befonds, ensovoorts.

2.3.4 Navorsing sedert 2000

In 'n 2012-artikel deur Joula (2012) haal hy onder meer Rudman (1998) aan in sy betoog dat die stand van sake nie noodwendig tans beter is nie, al is daar sedert 2000 'n toenemende hoeveelheid navorsing gepubliseer wat bestaande teorieë en metodologieë se resultate met mekaar vergelyk.

Joula (2008: 245-337) sê in sy omvattende oorsig oor die veld van outeurskapidentifikasie dat die wetenskaplike akkuraatheid van 'n ondersoek steeds die grootste probleme veroorsaak. Hy is van mening dat daar drie verwante aspekte is wat deeglik oorweeg moet word as navorsers studies met akkurate resultate wil oplewer, naamlik tegniese (metodologiese) akkuraatheid, tekstuele oorwegings en die oordeel wat die navorser aan die dag lê.

Volgens hom hou tegniese akkuraatheid eerstens verband met die metodologiese oorwegings soos watter analisetegniek die mees gepas is vir die betrokke ondersoek. Aspekte wat analisetegnieke beïnvloed, is aspekte soos genre, korpusgrootte en of die voorbeeldteks verteenwoordigend van die skrywer se styl is. Die kombinasie van hierdie aspekte vra vir deeglike oorweging van 'n metodologiese benadering. Tweedens noem hy dat kwessies soos programmering en die kennis wat die navorser daarvoor het, oorweeg moet word in die gebruik van rekenaargebaseerde ondersoeke omdat onkunde oor of die misbruik van rekenaarsagteware onreëlmatighede kan veroorsaak en tot ongeldige of onakkurate resultate kan lei (Joula, 2008:246-247).

Navorsers is dit eens dat 'n onakkurate datastel 'n swak basis vorm vir navorsing (Joula, 2008:247). Joula (2008:247-248) wys egter die struikelblokke wat verband hou met die verkryging van "suiwer" tekste, uit. Met "suiwer" tekste word verwys na tekste waarvan die grootste gedeelte sonder twyfel die werk van die outeur is. Hy noem dat daar 'n aantal eksterne faktore is wat kan veroorsaak dat 'n teks nie as 'n "suiwer" teks van 'n bepaalde outeur gesien kan word nie. Hierdie faktore sluit onder meer die invloed van proeflesers, redakteurs,

medeskrywers, ensovoorts in. Joula (2008:248) beklemtoon weer Rudman (1998) se standpunt dat daar navorsers binne die veld is wat nie “...verstaan wat ’n geldige studie konstitueer” nie (eie vertaling). Hy noem verder dat navorsers, veral in die lig van bogenoemde aspekte, uiters bekend moet wees in beide hulle navorsingsveld en navorsingskonvensies binne die veld.

2.3.4.1 Oplossings: outomatiese rekenaargesteuende analise

Joula (2008:249-251) stel oplossings voor wat probeer om optimisties na die veld van outomatiese, rekenaarmatige analises te kyk, alhoewel hy noem dat daar ’n aantal navorsers (insluitend Rudman) is, wat betoog dat die metodologiese struikelblokke en die risiko vir onakkurate resultate groter is as die wins van ’n outomatiese analisesisteam. Hoewel Rudman se voorstelle minder konseptueel is, is Joula se voorstelle grootliks gerig op aansporing tot verdere navorsing.

Joula (2008:249-251) stel voor dat daar ’n aantal areas is waarvoor sekerheid verkry moet word ten einde betroubare akkurate resultate te verseker. Hy noem eerstens weer dat die veld interdisiplinêr is en dat navorsers wat in hierdie dissiplines staan, byvoorbeeld linguïste, statistici of letterkundiges, hulle navorsing so wyd as moontlik beskikbaar sal maak vir persone op die periferie, soos joernaliste, onderwysers, prokureurs, speurders, ensovoorts. Tweedens moet verseker word dat analise-instrumente soos rekenaarprogrammatuur wat gereeld betroubare en akkurate resultate gee, ontwikkel en gebruik word sodat sistemiese foute uitgeskakel kan word. Hy noem ook dat ’n goeie analisesisteam bekragtig moet word met empiriese bewyse van geldige resultate wat oor ’n wye verskeidenheid van navorsingsprobleme verkry is. Verder sê Joula (2008:250) dat resultate van veral rekenaargebaseerde analises duideliker aangebied moet word, selfs d.m.v. die programmering van die betrokke sagteware, sodat die resultate meer toeganklik kan wees. Laastens betoog Joula dat ’n ideale sisteem stewig in ’n teoretiese benadering gegrond moet wees. Hierdie benadering moet die metodologie kan motiveer in terme van die kenmerke van teksproduksie. Die teoretiese en metodologiese benadering in hierdie studie word volledig in afdeling 2.3.5 en in afdeling 5.2 bespreek.

2.3.4.2 Die metafoer van ’n *linguïstiese vingerafdruk*

Saam met die moontlikhede wat rekenaargesteuende outeurskapidentifikasie studie bring, kom daar ook ’n verandering in die verwagte uitkoms van hierdie studies. Dit word soms genoem

dat daar 'n verskynsel soos 'n tipe linguistiese vingerafdruk is, waarvolgens skrywers maklik en vinnig uitgeken kan word (vergelyk Vandermay *et al.*, 2008 en Junghaenel, Smyth & Santner, 2008). Hierdie metafoer is gewortel in die teoretiese konsep van idiolek. Die term 'idiolek' word deur Crystal (2008:235) beskryf as een wat in linguistiek gebruik word om na "die linguistiese sisteem van 'n individuele spreker [te verwys]" (eie vertaling). Coulthard (2005:23) verwys na die feit dat elke taalgebruiker oor 'n aktiewe woordeskat beskik, wat nie net uit 'n "lys" van items bestaan nie, maar ook 'n *voorkeur* vir sekere items het. Hy verwys ook na Halliday (1964) wat sê dat die aanname dat 'n skrywer se idiolek deur "kenmerkende en idiosinkratiese keuses in tekste" dus 'n moontlike teoretiese basis vir outeurskapidentifikasie-ondersoeke bied. Nel (2014:83-83) brei ook uit op die waarde van die konsep van idiolek vir outeurskapidentifikasiestudies. Louwerse (2004:207-221) noem in sy studie oor semantiese variasie in taalgebruik en kohesie in geskrewe en gesproke genres dat taalgebruikers (individue) hul sogenaamde 'merk' laat op dokumente wat hul skryf en dat so 'n dokument uitkenbaar is op grond van die voorkeur vir spesifieke linguistiese kenmerke in 'n teks; amper soos 'n mens wat herkenbaar is aan sy/haar vingerafdruk. Hoewel bogenoemde aan die begin van die artikel as feite gestel word, erken die skrywer dat hipoteses wat van die bestaan van die "linguistiese vingerafdruk"- teorie afgelei is meestal nie empiries bewys is nie (sien ook Olsson, 2008:25). In die studie van Louwerse (2004:209-220) gebruik hy tekste deur beide mans en vroue, vanuit twee literêre tydperke, om die beginsel van idiolek en sosiolek te ondersoek; die tekste wat hy gebruik het, is deur George Elliot, Charles Dickens, Virginia Woolf en James Joyce geskryf. Hy (Louwerse, 2004:209-220) lê in sy studie bewyse voor van semantiese variasie tussen tekste van verskillende individue asook van groepe skrywers. Olsson (2008:24-26) noem in hierdie verband dat die probleem dalk lê by die feit dat die linguistiese vingerafdruk gewoonlik in die konteks van forensiese linguistiek genoem word (Olsson, 2008:26) en by implikasie as van soortgelyke betroubaarheid gesien kan word as 'n werklike vingerafdruk of afdruk van iets anders soos 'n skoen- of bandafdruk.

Binne die Sistemies-Funksionele Grammatika word idiolek beskryf as die "...pad deur die netwerk van taalsisteme wat 'n individu tipies kies binne 'n sekere konteks as gevolg van sy/haar unieke sosiolinguistiese agtergrond" (Nini & Grant, 2013:180, eie vertaling). Nini en Grant verduidelik verder dat idiolek uit die kodevariasie en dialektvariasie van 'n individu se taalgebruik bestaan en dat idiolek gemeet kan word deur hierdie tipes variasie te meet, maar die registervariasie buite rekening laat deur die konteks dieselfde te hou.

Kognitiewe studies oor outeurskapidentifikasie steun sterk op die idee dat daar 'n stel of stelle linguistiese merkers bestaan wat 'n linguistiese vingerafdruk van 'n outeur kan vorm. Grieve (2007:250-270) het 'n verskeidenheid kwalitatiewe outeurskapidentifikasiestegnieke geëvalueer deur dit op dieselfde stel tekste van toepassing te maak. Om dus die metodologie te toets, het hy uit tekste van bekende outeurs probeer vasstel hoe akkuraat die outeur geïdentifiseer kan word. In hierdie studie is die veranderlike ook nie die outeur nie, maar die linguistiese kenmerk wat moontlik voorkeur geniet by sekere, reeds bekende, outeurs.

Hoewel dit voorkom asof daar nog nie 'n antwoord is op die vraag in verband met 'n linguistiese vingerafdruk nie, word Coulthard (2005:14) se standpunt in hierdie studie as uitgangspunt geneem. Hy is van mening dat die probleem meer by die metafoor lê as by die teorie. Die metafoor van 'n vingerafdruk skep die idee dat daar 'n databank van uitkenningstemplate vir individue se skryfstyl geskep kan word en dat tekste daarteen getoets kan word om skrywers uit te ken. So iets is volgens hom tans onprakties en amper onmoontlik. Hy voer egter in sy artikel, "Author Identification, Idiolect and Linguistic Uniqueness" (Coulthard, 2004:1-2) aan dat daar nie onder navorsers twyfel is oor die basiese uitgangspunt dat daar idiolekvariasie en uniekheid onder individuele taalgebruikers se tekste en spraak bestaan nie. Die onsekerheid hou meer verband met die metodologiese aspekte wat die mees betroubare resultate kan oplewer.

2.3.5 Teoretiese raamwerke

Soos dit uit die bogenoemde oorsig duidelik is, is navorsers in die veld van outeurskapidentifikasie deurlopend besig om geldigheid aan hul metodologie en teoretiese begronding te probeer gee. Sedert 1990 was die fokus daarop om geldigheid en betroubaarheid te verkry deur te poog om 'n outomatiese sisteem te skep wat volkome rekenaargebaseerd is. Aanvanklike navorsing ten opsigte van outeurskapidentifikasie soos dié van Mendenhall (1887) het lukraak voorgekom en daar was 'n groot hoeveelheid metodologieë en teorieë wat ontwikkel en getoets is. Die massa dataverwerking wat deur die ontwikkeling in rekenaartegnologie moontlik gemaak is en die aanspraak op wetenskaplikheid het navorsers krities na die veld laat kyk in 'n poging om teorieë en metodologiese benaderings te verfyn. Daar het sedert 2010 'n hele aantal studies verskyn wat poog om van hierdie verfynde benaderings te integreer. Vervolgens word die teoretiese benaderings tot outeurskapidentifikasiestudie bespreek.

Solan (2012:551-576) wys in sy artikel *Intuition versus “Algorithm: The Case of Forensic Authorship Attribution”* op die uiteenlopende aard van teoretiese benaderings wanneer kenners outeurskap toeken. Johnson en Wright (2014:39) verduidelik dat daar polariteite bestaan binne die benaderings wat gevolg word in outeurskapidentifikasiestudies. Aan die een kant is daar die kwalitatiewe en stilistiese benaderings en aan die ander kant die rekenkundige en statistiese benaderings. Hierdie uiteenlopende benaderings bied voor-die-hand-liggende uitdagings in die lig van wetenskaplike en teoretiese begronding.

Soos dit uit die bogenoemde oorsig van struikelblokke en moontlike oplossings duidelik is (vergelyk afdeling 2.3.3), is daar nie ’n onbetwiste teoretiese raamwerk waarop navorsers hulle aannames en metodiek grond nie. Nini en Grant (2013:174-202) beaam Johnson en Wright (2014) se observasie dat daar polariteite bestaan in die benaderings wat kenners volg en identifiseer die twee belangrikste teoretiese uitgangspunte waarop metodologieë gebou is as stilisties- en kognitief-geïnspireerde teorieë. Hulle stel dan ’n teorie voor wat die relevante sterkpunte van die afsonderlike teorieë ondervang en ’n moontlikheid van ’n omvattende teorie, naamlik *Sistemies-Funksionele Grammatika*¹¹ (SFG) deur Halliday en Hasan (1976).

2.3.5.1 Stilisties-gebaseerde benadering

Volgens Nini en Grant (2013:174) is daar ’n aantal metodologiese aannames wat deur die navorsers wat ’n stilisties-gebaseerde metodologie volg, gemaak word. Eerstens sê hulle dat ’n teorie wat idiolek in ag neem belangrik is, maar nie noodsaaklik vir die ontwikkeling van ’n geldige metodologie nie. ’n Tweede aanname is dat ’n individu uniek en uitkenbaar is as gevolg van sy/haar styl, wat al die aspekte van mense se lewens waarin vrye keuse ’n rol speel, beïnvloed. Derdens kom ’n individu se styl in sy/haar taalgebruik na vore deur ’n aantal stylmerkers wat uit gewoonte uit die individu se linguistiese repertoire aanwesig is. Die linguistiese repertoire van die individu, sowel as die keuses wat hy/sy maak, varieer van ander individue as gevolg van historiese en sosiologiese verskille tussen individue. Navorsers wat hierdie benadering volg, redeneer dat hierdie linguistiese keuses as stylmerkers in ’n individu se skryfwerk sigbaar is, en gebruik kan word om ’n teks te identifiseer as ’n produk van ’n spesifieke individu (Nini & Grant, 2013: 175). Hierdie benadering ontvang meestal kritiek op grond van die subjektiewe, lukrake aard van die ondersoek. Daar is nie ’n voorafbepaalde stel stylmerkers wat ondersoek gaan word as bewyse van variasie nie; die teks word bestudeer deur na onderskeidende stylkenmerke te soek (Nini & Grant, 2013: 175-176). In die lig van die

¹¹ Vergelyk afdeling 2.3.5.3 vir ’n volledige bespreking van hierdie teorie.

uitdagings ten opsigte outeurskapstudies wat in afdeling 2.3.3.1 beskryf is, kan die bogenoemde benadering uiteraard tot sekere vooraf-geïdentifiseerde probleme lei, soos dié wat Joula (2008) en Rudman (1998) beskryf het.

2.3.5.2 Kognitiewe benadering

Soortgelyk aan hul werkswyse vir die beskrywing van die stilistiese benadering tot outeurskapstudies, beskryf Nini en Grant (2013:175) ook die aannames waarop die kognitiewe benadering gegrond is. Eerstens is 'n teorie van idiolek onontbeerlik. Tweedens sê hulle dat 'n individu uniek is as gevolg van die struktuur waarop hulle sin maak van die wêreld, dus die struktuur van kennis. Hierdie struktuur is waarneembaar in die taalkundige strukture wat die individu produseer. Deur verskillende sintaktiese patrone te bestudeer, kan byvoorbeeld tussen individue onderskei word. Laastens is navorsers wat van hierdie benadering gebruik maak volgens hulle van mening dat die metodologie vir outeurskapidentifikasiestudies in kognitiewe taalkundige teorie gegrond moet wees en kwantitatief getoets moet wees. Nini en Grant (2013:175) identifiseer verder die grootste bron van kritiek vir die kognitiewe benadering as die feit dat daar steeds nie duidelikheid is oor die *rede* vir sintaktiese variasie tussen individue nie.

Indien hierdie twee stelling aannames langs mekaar geplaas word soos in die onderstaande tabel (tabel 2-1) is dit duidelik dat die begroning van die stilistiese benaderings tot outeurskapstudie klem lê op taalgebruikers se keuses en optrede en dat die die begroning van die kognitiewe benaderings in die taalgebruiker se verstaan van die wêreld gesetel is.

Tabel 2-1 Vergelykende opsomming van uitgangspunte van die stilistiese en kognitiewe benadering

| Stilistiese benadering | Kognitiewe benadering |
|--|--|
| 'n Teorie van idiolek nie noodsaaklik nie | 'n Teorie van idiolek onontbeerlik |
| Individue is uniek in alle aspekte van die lewe waar vrye keuse 'n rol speel. | 'n Individu se unieke struktuur van kennis lei tot unieke taalstrukture |
| 'n Individu kies 'n aantal stylmerkers uit gewoonte uit sy/haar linguistiese repertoire en dit lei tot 'n waarneembare styl. | Benadering moet in kognitiewe taalteorie gegrond wees en kwantitatief getoets kan word |

Nini en Grant (2013:177) stel vervolgens Sistemies-Funksionele Grammatika (SFG) voor as 'n teorie wat moontlik die tekortkominge van die teorieë kan oorbrug. Hierdie teorie is oorspronklik deur M.A.K. Halliday in 1961 voorgestel en dit word as 'n holistiese teorie beskryf wat alle vlakke van taalgebruik, van fonologie tot sosiolinguistiese variasie probeer verklaar.

2.3.5.3 Sistemies-Funksionele Grammatika (SFG)

Sistemies-Funksionele Grammatika (SFG) word beskryf as "... 'n beskrywende en interpretatiewe raamwerk wat gebruik kan word om taal as 'n strategiese, betekenisvormende hulpbron te ondersoek (Eggins, 2004:2, eie vertaling). Eggins (2004:3) verduidelik verder dat voorstanders van hierdie teoretiese raamwerk vier teoretiese uitsprake oor taal maak, naamlik dat (1) taal funksioneel is, (2) taal se funksie betekenisgenererend (semanties) is, (3) hierdie betekenis deur die soiale en kulturele konteks waarbinne die taalhandeling plaasvind, beïnvloed word en (4) taalgebruik 'n semiotiese proses is, m.a.w. 'n proses waartydens betekenis geskep word deur keuses uit te oefen. Deur slegs die bogenoemde aansprake in ag te neem, kan reeds gesien word hoekom hierdie teoretiese raamwerk gebruik kan word wanneer ondersoek na outeurskap gedoen word en hoekom dit gebruik kan word om die kognitiewe en stilistiese benadering, soos deur Nini en Grant (2013) beskryf is, te versoen. Veral die laaste twee teoretiese vertrekpunte wat in hierdie teorie gemaak word ((3) en (4) hierbo), spreek tot die aannames wat in die stilistiese benadering tot outeurskapstudies gemaak word (vergelyk afdeling 2.3.5.1). Met die veranderlikes wat deur die SFG as invloede vir variasie in taal beskryf word (vergelyk afdeling 2.3.5.3.1-2.3.5.3.3), behoort daar egter ook beskrywingsmoontlikhede van 'n individu se taalkeuses en kognitiewe proses binne hierdie teorie te wees. Hierdie aspek van die teorie spreek veral tot die aannames t.o.v. idiolek wat deur die kognitiewe benadering tot outeurskapstudies voorgestel word.

Volgens Matthiessen (2007:538-539) word daar in die SFG veronderstel dat daar drie tipes variasie is wat inherent tot taal is, naamlik *register*, *kode* en *dialek*. Registervariasie verwys binne hierdie benadering hoofsaaklik na variasie wat voorkom as gevolg van kontekstuele faktore. Faktore wat aanleiding gee tot taalvariasie, maar wat verband hou met die *taalgebruiker*, bv. sosiale klas, ouderdom, geografiese herkoms ensovoorts, word as kodevariasie en/of dialektiese variasie beskryf (Crystal, 2008:409 en Nini & Grant, 2013:180).

2.3.5.3.1 *Registervariasie*

Volgens die SFG behels registervariasie om op 'n gestruktureerde manier die waarskynlikheid van sekere linguistiese keuses aan te pas op grond van kontekstuele faktore (Nini & Grant, 2013:178). 'n Voorbeeld hiervan is om van formele taal gebruik te maak wanneer 'n akademiese teks geskep word of om jou taal aan te pas wanneer jy met 'n jong kind praat. Die SFG identifiseer drie verwante (Halliday en Matthiesson, 2014:34) en abstrakte kategorieë van kontekstuele faktore wat aanleiding gee tot registervariasie, naamlik diskoersmodus (*mode*), diskoersstrekking (*tenor*) en diskoerssfeer (*field*) (Schleppegrell, 2013:22).

Diskoersmodus verwys na die rol wat die teks (of taal) in die gegewe konteks speel, bv. gesproke teenoor geskrewe taalgebruik (Eggins, 2004:9). Volgens Crystal (2008:308) kan *mode* meer volledig as “medium van diskoers” (eie vertaling) beskryf word en sluit dit die keuse tussen gesproke en geskrewe taal in, maar ook die formaat van die teks, bv. koerant, gedig, kommentaar, ensovoorts.

Diskoersstrekking verwys na die sosiale rol van 'n taalgebruiker binne 'n gegewe taalhandelingsituasie (Eggins, 2004:9). Volgens Schleppegrell (2013:22) hou diskoersstrekking verband met die interpersoonlike verhoudings binne die situasie en dit dra houding of gesindheid oor. Dit fokus spesifiek op die graad van normaliteit wat die taalgebruikers aanneem (Crystal, 2008:479).

Diskoerssfeer verwys na die onderwerp waaroor die taalhandeling handel. Sy verduidelik dat taalgebruikers verskillende linguistiese keuses sal uitoefen wanneer hulle byvoorbeeld oor draf praat as wanneer hulle oor linguistiek praat (Eggins, 2004:9). Diskoerssfeer hou dus met die inhoud van die diskoers verband (Crystal, 2008:188).

2.3.5.3.2 *Dialekvariasie*

Volgens die SFG hou dialektiese variasie verband met die wyse waarop semantiese opsies verwesenlik word. Die manier waarop hierdie semantiese opsies verwesenlik word, word 'n norm onder spesifieke groepe mense wat met mekaar interaksie het. Dit veronderstel dat hierdie norm 'n sosiogeografiese verband tussen taalgebruikers veronderstel (Matthiessen, 2007:539-540). Die variasie is dus in die *manier* waarop dieselfde semantiese eenheid uitgedruk word. In Engels word die voorbeeld van die keuse tussen “burnt” en “burned” vir die verlede tyd van ‘burn’ gebruik. Beide druk die verledetyd van ‘burn’ uit, maar die feit dat sommige individue

telkens ‘burnt’ sal gebruik in plaas van “burned” (en andersom) kan op dialektvariasie wys. Individue wat dus van dieselfde sosiogeografiese agtergrond is, behoort telkens dieselfde vorm te kies om ’n sekere semantiese eenheid (byvoorbeeld die verlede tyd van “burn”) te realiseer (Nini & Grant, 2013:179).

2.3.5.3.3 Kodevariasie

Hasan en Cloran (1990:95-97) verduidelik kode as die taalkundige styl wat mense aanneem om by hulle interpretasie van ’n sekere konteks in te pas. Omdat hierdie tipe variasie ook te doen het met die konteks, kan dit maklik met diskoersstrekking (sien 2.3.5.3.1) verwar word. Die verskil is egter dat die individu se interpretasie van die konteks ter sprake is en nie slegs die konteks, soos wat by registervariasie die geval is nie. Die variasie vind dus plaas as gevolg van ’n individu se unieke interpretasie van die konteks.

In hierdie studie word daar ook van ’n korpusgebaseerde benadering gebruik gemaak. Hierdie benadering is egter nie ’n teorie nie, maar ’n metodologiese benadering en dit word in hoofstuk 3 volledig uiteengesit.

2.4 Samevatting

In hierdie hoofstuk is die studieveld van outeurskapidentifikasie bespreek. Uit hierdie oorsig is dit duidelik dat daar verskeie uitdagings binne die veld is, ten spyte van die lang geskiedenis van studies wat die soeke na outeurskap ten doel het. In terme van die uitdagings ten opsigte van teoretiese begronding, stel Nini en Grant (2013:176-178) die Sistemies-Funksionele Grammatika voor as ’n teoretiese basis waarvolgens outeurskapstudie gerig kan word.

In die volgende hoofstuk word die linguistiese verskynsel, *kohesie*, die linguistiese veranderlike wat in hierdie studie ondersoek word, bespreek.

3 Teks, tekstualiteit en kohesie

3.1 Inleiding

Kohesie is 'n verskynsel wat te make het met die eenheidstruktuur van 'n teks. Dit word bestudeer in die studieveld van tekslinguistiek wat op sy beurt 'n deel vorm van die teoretiese benaderings van Funksionele Grammatika. 'n Omvattende beskrywing van die teoretiese uitgangspunte van die benaderings en studieveldde waarbinne die verskynsel van kohesie geplaas word, val buite die bestek van hierdie studie en daar gaan telkens slegs 'n agtergrond en 'n werksdefinisie gegee word vir begrippe wat nie primêr met hierdie studie verband hou nie. In hierdie hoofstuk word daar wel eerstens 'n oorsig van tekslinguistiek as studieveld gegee, waarna die begrippe *tekstuur* en *tekstualiteit* beskryf word. Daarna kom die beginsels vir *tekstualiteit* en spesifiek *kohesie* aan bod. Laastens word 'n volledige omskrywing van die vyf kohesieskakels, *verwysing*, *substitusie*, *ellips*, *leksikale kohesie* en *konjunksie* gegee.

3.2 Tekslinguistiek as studieveld

In hulle boek met sleutelbegrippe binne die linguistiek sê Trask en Stockwell (2007) dat die veld van tekslinguistiek gesien kan word as 'n spesifieke benadering tot teksanalise wat klem lê op tekstualiteit. Hierdie verduideliking impliseer dat tekslinguistiek nie die enigste benadering tot teksanalise is nie. Tekslinguistiek en diskoersanalise is studieveldde met 'n groot aantal oorvleuelende definisies en eienskappe (sien Trask & Stockwell (2007); Carstens (1993) en Malmkær (1991)). Ten spyte van oorvleueling van begrippe en definisies tussen studieveldde, noem Carstens (1993:3) in sy artikel oor Tekslinguistiek as studieterrrein ook dat daar baie “vaag-omlynde begrippe” in dié veld is en gaan dan voort om selfs die term *tekslinguistiek* self as 'n vaag-omlynde begrip te identifiseer.

In Carstens se omvattende boek ook die Afrikaanse tekslinguistiek, *Afrikaanse tekslinguistiek* (1997:71-81), word daar verduidelik hoekom die begrip *teks* binne die linguistiek so moeilik is om neer te pen. In sy hoofstuk oor tekslinguistiek in die boek, *Discursive Pragmatics*, stel De Beaugrande (2011:286-295) weer die vraagstuk oor die duidelike afbakening van die term teks in die studieveld tekslinguistiek voorop. Carstens (1997:74-78) bied tientalle moontlike definisies van teks om aan te toon dat daar 'n verskeidenheid van definisies is wat al gebruik is en wat gebruik kan word. Hy sê egter dat die begrip as 'n pre-teoretiese term gesien moet

word wat beteken dat wanneer die term gebruik word, daar vooraf 'n werksdefinisie vir die betekenis van *teks* in die bepaalde studie/vakgebied daargestel moet word.

Uit die definisies wat Carstens bied, is daar 'n duidelike fokusverskuiwing van die strukturelemente van 'n teks na 'n wyer definisie van teks as 'n mensgemaakte konstruksie. Hierdie studie handel egter nie oor die onderskeid tussen tekste en nie-tekste nie en daarom is die definisie van Halliday en Hasan (1976:1) steeds relevant. Hulle sê dat wanneer daar in taalkunde na die woord “teks” verwys word, dit na “... enige passasie (taal), gesproke of geskrewe, van enige lengte, wat 'n eenheid vorm [verwys]” (eie vertaling). Hierdie definisie, hoewel vereenvoudig, bring implikasies of vereistes vir 'n teks mee. Vir die doel van hierdie studie is die belangrikste implikasie dat 'n teks 'n eenheid moet wees. Oor hierdie implikasie van 'n eenheid skryf De Beaugrande en Dressler (1981:3) in hulle standaardwerk *Introduction to Text Linguistics* dat 'n teks gedefinieer kan word as 'n “kommunikatiewe gebeurtenis wat aan die sewe beginsels vir tekstualiteit voldoen” (eie vertaling). Hierdie standarde of beginsels van tekstualiteit word in afdeling 3.4 uiteengesit. In die volgende paragrawe word verduidelik wat die verband tussen die teks as eenheid en tekstualiteit en/of tekstuur is.

3.3 Tekstuur en tekstualiteit

Die feit dat Halliday en Hasan (1976:1) se definisie so duidelik daarop fokus dat die teks as 'n eenheid ervaar moet word, impliseer dat 'n stuk taalgebruik wat nie oor eenheid beskik nie, nie 'n teks is nie. Dit word bevestig in 'n definisie van teks wat klem lê op faktore wat eenheid bewerkstellig. Hierdie definisie kom voor binne die konteks van die studieveld tekslinguistiek wat sê dat 'n teks taalgebruik is wat “... gekarakteriseer [word] deur beginsels soos kohesie, koherensie en informatiwiteit...” (Crystal, 2008:482, eie vertaling). Kohesie, koherensie en informatiwiteit dra volgens Crystal (2008:482) by tot 'n teks se tekstuur of tekstualiteit. Dit lyk of hierdie verduideliking van die term tekstuur 'n ander patroon volg as dié van Halliday en Hasan (1976:2) wat sê dat *tekstuur* in 'n teks verkry word deurdat die teks as 'n eenheid funksioneer en dat ‘tekstuur’ dus 'n teks van 'n nie-teks’ (“nontext”), onderskei. Die verwantskap tussen die teks as 'n eenheid, die teks se tekstuur en elemente soos kohesie en koherensie is egter duidelik. Wanneer 'n teks oor kohesie, koherensie en informatiwiteit beskik, funksioneer die teks as 'n eenheid en is daar tekstuur in die teks. Dit is dan hierdie tekstuur wat 'n teks van 'n nie-teks onderskei.

Die vraag wat op grond van bogenoemde definisies ontstaan, is of daar ’n verskil tussen tekstuur en tekstualiteit is. In Crystal se *A Dictionary of Linguistics and Phonetics* (2008:482-483) definieer hy tekstuur en tekstualiteit gelyktydig:

“Texts are seen as language units which have a definable communicative function, characterized by such principles as *cohesion*, *coherence*, and informativeness, which can be used to provide a *formal* definition of what constitutes their identifying textuality or texture.”

Vir die doel van hierdie studie word daar dus aanvaar dat beide die begrippe *tekstuur* en *tekstualiteit* te make het met die mate waarin die teks as ’n eenheid funksioneer en dus die mate waarin ’n teks as sodanig geklassifiseer kan word. Verdere onderskeid is nie in hierdie studie nodig nie.

3.4 Tekstualiteitsbeginsels

Halliday en Hasan lê in hulle boek *Cohesion in English* (1976:2) klem daarop dat ’n teks nie net ’n strukturele eenheid vorm nie, maar merendeels ’n semantiese eenheid. Hiermee word bedoel dat die strukturele elemente slegs die semantiese elemente enkodeer. In terme van wat voorheen gesê is oor tekstualiteit as kriterium vir die bepaling van ‘teks-heid’ in gedagte, kan ’n mens sê dat tekstualiteit grootliks afhang van die semantiese eenheid binne ’n teks.

Om die tekstualiteit van ’n teks te “meet”, het De Beaugrande sewe beginsels van tekstualiteit geïdentifiseer wat moontlik kan bydra tot eenheid/tekstuur/tekstualiteit in ’n teks, naamlik: kohesie, koherensie, intensionaliteit, aanvaarbaarheid, informatiwiteit, relevansie en intertekstualiteit.

Volgens De Beaugrande en Dressler (1981:113, eie vertaling) is “(d)ie kohesie van oppervlakelemente en die onderliggende koherensie van die tekstuele woorde die aangewese vereistes vir tekstualiteit”. Kohesie en Koherensie hou dus verband met die tekstuele aspekte. Die ander tekstualiteitsbeginsels, intensionaliteit, aanvaarbaarheid, relevansie en intertekstualiteit, hou meer verband met die gebruiker en die verbale en/of pragmatiese konteks. Carstens (1997) merk ook op dat hierdie beginsels tekstualiteit van *taal in gebruik* bevorder en dus ’n sterk basis in Pragmatiek het. Die pragmatiese aspekte van die kohesieskakels val egter buite die bestek van hierdie studie.

Volgens De Beaugrande en Dressler (1981) hou koherensie verband met die wyse waarop die dieptestruktuur (die semantiese konsepte en verhoudings) wat deur die oppervlakstruktuur geskep word, wedersyds toeganklik (“mutually accessible”) en relevant is. Volgens Carstens (1997:116) hou koherensie verband met konseptuele skakeling wat tekseenheid bevorder. Hy (Carstens, 1997:472) sê ook dat die konseptuele skakeling grootliks op drie faktore berus, naamlik, (1) die deelnemers aan die teks/diskoers se gedeelde agtergrondkennis, (2) hulle vermoë om relevante afleidings te maak binne ’n sekere konteks en (3) die verwagtinge van die betrokke partye.

De Beaugrande & Dressler sê in hul standaardwerk *Introduction to Text Linguistics* (1981:113) dat hoewel kohesie en koherensie as talige elemente grootliks tot die tekstualiteit van ’n teks bydra, daar gereeld tekste voorkom wat nie volkome kohesief en koherent voorkom nie. Om hierdie rede is dit volgens hulle belangrik om die houdings van die teksgebruiker ook in ag te neem. *Intensionaliteit* en *aanvaarbaarheid* hou verband met die doel/intensie van die skrywer/spreker met die skep van die teks en hoe die leser/hoorder die teks aanvaar. Die aard van hierdie tekstualiteitsbeginsel leen homself tot gebruik in Pragmatiek, met baie voorbeelde van intensionaliteit en aanvaarbaarheid wat in gesproke taal voorkom. Dit is egter belangrik om op te merk dat hierdie verskynsels nie uitsluitlik by gesproke taal voorkom nie.

’n Verdere beginsel wat tot die semantiese eenheid in ’n teks bydra, is *relevansie*. Carstens (1997:387) wys op die duidelike verband wat daar is tussen ’n teks en die kontekstuele faktore wat ’n invloed op die teks het. Volgens De Beaugrande en Dressler (1981:161) verleen die konteks relevansie aan die teks. Hierdie relevansie speel ’n belangrike rol in die interpretasie en semantiese eenheid van die teks.

Die beginsel van *intertekstualiteit* berus op die nosie dat een teks geïnterpreteer kan word (gedeeltelik) as gevolg van kennis van voorafgaande verwante tekste. So verwys De Beaugrande en Dressler (1981) na die feit dat ’n sekere teks inligting waarvan die taalgebruiker in vorige tekste kennis geneem het in ’n mate aktiveer en dat ’n redelike taalgebruiker ’n teks sal interpreteer deur op die inligting staat te maak wat hy/sy in vorige verwante tekste kan vind.

Carstens (1997:427-429) verduidelik die tekstualiteitsbeginsel van *informatiwiteit* as een wat verband hou met die wyse waarin inligting oorgedra word in tekste. Hy verwys na Neubert en Schreve (1992:89) wat van mening is dat tekste inligting bevat wat nie bestaan het voordat die teks geskep is nie. Dit impliseer dat alle tekste in ’n mate informatief van aard is. Hoe hierdie

inligting oorgedra word en hoe die sender nuwe en bekende inligting in die teks deel, bepaal die aard van die *informatiwiteit* van die teks.

Kohesie is die sewende tekstualiteitsbeginsel. Dit is die linguistiese veranderlike wat in hierdie studie ondersoek word. Dus volg 'n breedvoerige bespreking van dié beginsel in afdeling 3.4.1 en 3.5.

3.4.1 Kohesie

Kohesie hou verband met die wyse waarop die oppervlakstruktuur (die woorde wat ons kan sien of hoor) van 'n teks onderling of wedersyds (*mutually*) aan mekaar verbind is binne 'n reeks (*sequence*) (De Beaugrande & Dressler, 1981:3; vergelyk ook Carstens & Van de Poel, 2010:106 en Fouché, 2016:17-18). In sy ondersoek oor die verband tussen kohesie en semantiese interpretasie in letterkundige tekste wys Janjua (2012:150) op die verbande tussen 'n aantal opvattinge oor kohesie. Hy noem eerstens dat die *Concise Oxford Dictionary of Linguistics* (Matthews 1997) kohesie as 'n *sintaktiese eenheid* sien. Crystal (2008:85) plaas die klem elders en beskou kohesie eerstens as 'n *grammatikale eenheid*. Hy noem laastens Halliday en Hasan se fokus op kohesie as 'n *semantiese eenheid*. Hierdie skynbare teenstrydigheid kan gedeeltelik verklaar word deur na twee basisdefinisies van kohesie deur De Beaugrande en Dressler (1981) en Halliday en Hasan (1976) te kyk. De Beaugrande en Dressler (1981:3) fokus meer op die oppervlakstruktuur terwyl Halliday en Hasan (1976:4) sê dat *kohesie* 'n konsep is wat oor semantiese skakeling handel. Hierdie twee klaarblyklik teenstrydige sienings kan egter versoen word deur die veronderstelling dat konsepte binne die semantiese sisteem deur die leksikogrammatiese sisteem tot uiting kom. Die feit dat die semantiese skakeling op 'n waarneembare wyse sigbaar word deur die oppervlakstruktuur van die teks maak dat kohesie as ondersoekinstrument gebruik kan word. In die volgende definisie deur Malkmjar (1991:463) smelt die twee bogenoemde sienings saam: “Kohesie het te make met die manier waarop die linguistiese items waaruit 'n teks saamgestel is, *betekenisvol binne 'n reeks aan mekaar verbind is* binne die raamwerk van die *grammatikale reëls* van die taal” (eie vertaling, eie klem). Dit impliseer dat kohesie verband hou met die manier waarop semantiese eenheid verkry word, met behulp van die grammatikareëls binne taal.

Die manier waarop die bogenoemde linguistiese items aan mekaar verbind word, is die hoofokus van Halliday en Hasan se boek, *Cohesion in English* (1976). In hierdie basiswerk

oor kohesie bespreek hulle die kategorieë van die verskillende maniere waarop linguistiese items aan mekaar verwant kan wees.

3.5 Kohesieskakels

Kohesieskakels is skakels of *ties* wat verwys na 'n spesifieke "...[i]nstance of cohesion..." van linguistiese items wat kohesief aan mekaar verwant is (Halliday & Hasan, 1976:3). Daar is verskillende strategieë wat gebruik word om items wat aan mekaar verwant is, linguisties aan mekaar te skakel om 'n samehangende geheel te vorm. Hierdie verskillende strategieë is gekategoriseer om die volgende kohesieskakels te vorm: Verwysing, substitusie, ellips, konjunksie en leksikale kohesie.

Bogenoemde kategorieë het 'n breë teoretiese onderbou wat in die afdelings wat volg, bespreek sal word. Kort definisies en voorbeelde van elke tipe kohesieskakel kan in afdeling 1.1 gekry word.

Deur kennis te neem van vereenvoudigde voorbeelde (Afdeling 1.1) kan gesien word dat die struktuur van sommige van hierdie kohesieskakels meer kompleks is as ander. Verwysing is byvoorbeeld relatief maklik waarneembaar deur na voornaamwoorde te soek, maar leksikale kohesie behels 'n mate van interpretasie en die waarneming daarvan is meer subjektief. Tydens hierdie studie se loodsondersoek (2012) na die gebruik van kohesieskakels in kortverhale en rubrieke van Nataniël, skrywer, sanger en vermaaklikheidspersoonlikheid, het dit duidelik geword dat, indien kohesieskakels as veranderlike in die ondersoek gebruik gaan word, daar 'n duidelike afbakening moet wees in die klassifikasie van hierdie skakels. Duidelike definisies en omskrywings van wat elke kohesieskakel behels, maak die teksanalise meer akkuraat en betroubaar. Die teoretiese agtergrond van die kohesieskakels gaan in hierdie hoofstuk bespreek word terwyl die toepassing daarvan as analiseveranderlike in hierdie studie in afdeling 5.6.3 bespreek sal word.

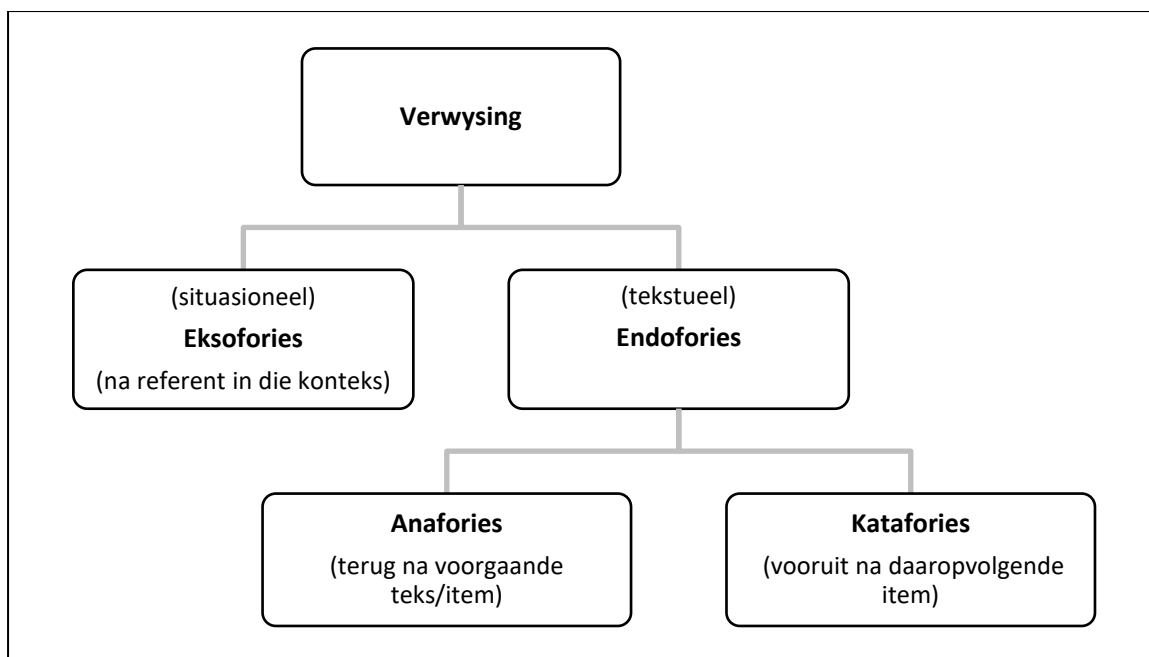
3.5.1 Verwysing

Soos vroeër vermeld, word *verwysing* deur Crystal (2008:407) gedefinieer as 'n kohesieskakel wat gebruik word om 'n *identiteitsverhouding* tussen leksikale items uit te druk (eie klem). 'n Voornaamwoord kan byvoorbeeld verwys na 'n spesifieke selfstandige naamwoord of frase en só word die identiteitsverhouding uitgedruk. Carstens (1997) sit egter verskillende

klassifikasiestrukture vir verwysing uiteen en verwys veral na De Stadler (1989) en Halliday en Hasan (1976).

Halliday en Hasan (1976:31) lê aanvanklik klem daarop dat *verwysing* verband hou daarmee om na spesifieke inligting te gaan soek. Hulle definisie van verwysing hang dus saam met die feit dat daar 'n semantiese verhouding tussen die verwysende item en die referent is, waarin die verwysende item aan die leser of hoorder kommunikeer om dit waarna verwys word, die referent, elders te gaan vind.

Carstens (1997:140) gebruik 'n diagram wat aanvanklik deur Halliday en Hasan voorgestel is, om die tipes strukturele oorwegings wat met verwysing verband hou, te illustreer (sien Figuur 3-1). Die eerste verdeling wat voorgestel word, is tussen sg. eksoforiese en endoforiese verwysing. In wese beteken dit dat die antesedent in die geval van eksoforiese verwysing buite die teks staan (in die konteks) en in die geval van endoforiese verwysing, binne die teks gevind kan word. 'n Verdere verdeling wat van endoforiese verwysing gedoen kan word, is dié van 'n kataforiese en 'n anaforiese verwysingsverhouding. Kortliks dui hierdie onderskeid die verhouding van die referent en die verwysde item aan, in terme van *waar* (in die teks) die verwysde item ten opsigte van die referent gevind kan word. Indien die referent terug verwys na 'n leksikale item wat reeds genoem is, is dit 'n anaforiese verwysing. Aan die ander kant dui 'n kataforiese verwysing aan dat die referent vooruit verwys na 'n leksikale item wat nog in die teks genoem moet word (Carstens, 1997:139-140).



Figuur 3-1: Tipes verwysing in terme van strukturele oorwegings (Carstens,1997:140)

Behalwe vir die bogenoemde verdeling wat hoofsaaklik op die struktuur van die voorkoms van verwysing betrekking het, is daar verdere indelings wat deur verskeie navorsers voorgestel word. Fouché (2016:30-32) verduidelik Halliday en Hasan (1976:37-87) se indeling en noem die kategorieë van persoonlike verwysing (“personal reference”), aantoonende verwysing (“demonstrative reference”) en vergelykende verwysing. Sy (Fouché, 2016:31-32) gee verder voorbeelde van die realisering van hierdie kategorieë in Engels. Hierdie voorbeelde sluit onder meer die volgende in:

Persoonlike verwysing; Persoonlike voornaamwoorde wat selfstandig optree; persoonlike voornaamwoorde wat attributief optree; en selfstandige besitlike voornaamwoorde.

Aantoonende verwysing: Die neutrale “the”; die selektiewe “this”, “that”, “these” en “those”; en bywoorde van tyd en plek en plek, soos “here”, “there”, “now” en “then”.

Vergelykende verwysing: Woorde wat ooreenstemming, gelyksoortigheid of verskil uitdruk (bv. onderskeidelik “same”, “similarly” of “likewise”); en woorde wat aandui dat ’n eienskap van twee linguïstiese items

in terme van kwaliteit of kwantiteit verskil (bv. onderskeidelik “easier” of “fewer”)

Nog ’n verdeling van tipes verwysing word deur Carstens (1997:136-137) gegee wanneer hy De Stadler (1989:18) se verdeling van hoof tipes verwysing uiteensit, nl. bepaalde verwysing, niebepaalde verwysing, generiese verwysing en unieke verwysing. Bepaalde en niebepaalde verwysing hou beide daarmee verband of die referent binne die referensiële uitdrukking herkenbaar is, al dan nie. Die toepassing of reikwydte van voorbeelde wat onder bepaalde verwysing ingesluit word, is uitgebreid, maar word óf as “Entiteite met ’n bepaaldheidstatus” óf as “Besondere tipes referensiële uitdrukkings” deur (Carstens, 1997:149-150) gekategoriseer. Sommige van hierdie voorbeelde sluit in:

Entiteite met ’n bepaaldheidstatus: sg. unieke entiteite waarvan die algemene persoon veronderstel is om bewus te wees, bv. *die son, die maan, die sterre*; sg. prototipiese entiteite wat ’n sekere klas verteenwoordig, bv. *Jan Publiek, die Afrikaanssprekende Suid-Afrikaner*; en sg. relationele entiteite wat slegs binne die konteks van die teks identifiseerbaar is, bv. *hy, sy, dit* (Carstens, 1997:149-150).

Besondere tipes referensiële uitdrukkings: Eiename, wat ook insluit die name van entiteite wat in die buite-talige konteks bestaan, bv. *Die Burger, Nelson Mandela*, ensovoorts; sg. identifiseerbare beskrywings wat meestal bestaan uit die bepaalde lidwoord die wat met ’n selfstandige naamwoord verbind word; sowel as bepaalde voornaamwoorde, in die besonder persoonlike, betreklike en aanwysende voornaamwoorde (Carstens, 1997:149-150).

Aangesien die identifikasie en klassifikasies van die verskillende tipes verwysing staatmaak op ’n uitgebreide stel kategorieë en vereistes (soos in die voorgaande paragrawe uiteengesit is), is daar besluit om vir die doel van hierdie studie slegs op twee van die onderafdelings te fokus. Die doel van hierdie afbakening is grootliks om betroubaarheid van die analise te verhoog deur die analiseproses te vereenvoudig. Die voorkoms van verwysing is nie die enigste linguistiese element wat ondersoek word nie en ’n gedetailleerde verdeling van subkategorieë is nie haalbaar binne die huidige studie nie.

Dus, vir die doel van hierdie ondersoek sal twee tipes verwysing gemerk word. Die eerste is eiename wat na ’n algemene entiteit buite die teks verwys, bv. *Kaapstad, Pieter-Dirk Uys, die Toring van Babel*, ensovoorts. Die hoofrede dat hierdie kategorie gekies word, is die relatiewe

gereelde voorkoms van hierdie tipe verwysing in die tekste¹². Die tweede kategorie wat ingesluit gaan word, kom veral deur die voorkoms van persoonlike, betreklike en aanwysende voornaamwoorde tot stand kom (bv. onderskeidelik *hy, sy, wat, wie, hierdie, daardie, ensovoorts*). Beide hierdie gekose subkategorieë val onder die oorhoofse kategorie van bepaalde verwysing.

3.5.2 Substitusie

Substitusie is 'n kategorie van kohesie wat verband hou daarmee om een element met 'n ander te vervang. Halliday en Hasan (1967:89) beskryf die kohesiewe aard van substitusie as 'n "leksikogrammatiese" een. Dit beteken dat substitusie nie met die betekenis van die items wat kohesief aan mekaar verbind is verband hou nie, maar met die bewoording. Om hierdie rede behoort die element wat die ander element vervang dieselfde grammatikale funksie te hê as die een wat vervang is. Dus moet 'n werkwoord met 'n werkwoord vervang word, 'n naamwoord met 'n naamwoord, ensovoorts.

As gevolg van die feit dat elemente vervang moet word met ander elemente met dieselfde grammatikale funksie, word die soorte substitusie ook in grammatikale kategorieë ingedeel. Volgens Halliday en Hasan (1976:90) bepaal die doel van die item wat vervang word, watter grammatikale kategorie van substitusie gebruik moet word. Hulle gaan voort deur te sê dat 'n naamwoord, werkwoord of 'n sinsnede dienooreenkomstig vervang kan word met 'n *naamwoordelike, werkwoordelike* of *klousale* substitusie (vergelyk ook Carstens, 1997:225-236). Ponelis (1979:566-573) sit in sy boek, *Afrikaanse Sintaksis*, die kategorieë van substitusie uiteen.

3.5.2.1 Naamwoordelike substitusie

Naamwoordelike substitusie vind plaas wanneer 'n naamwoord vervang word. Volgens Ponelis (1979:566-573) kan 'n naamwoordelike tipe-anafoor weggelaat word en vervang word met 'n element wat as 'n naamwoord funksioneer; hy noem dit PRO-substantivering en bespreek vier tipes PRO-substantiewe wat in Afrikaans gebruik kan word: (i) *een/ene*, (ii) *myne/joune*, ensovoorts, (iii) *s'n* en (iv) gesubstantiveerde adjektiewe. In Carstens (1997) word hierdie kategorieë uitgebrei en verduidelik aan die hand van voorbeelde en variasies. Vir hierdie studie

¹² Soos tydens die informele loodsanalises opgemerk

gaan daar egter enkele voorbeelde (uit Carstens, 1997:226-236) per kategorie ter illustrasie gebruik word.

- (i) Substitusie deur *een/ene*
 - a. My kinders ken die storie van Dawid en Goliat, maar nie die **een/ene** oor David en Jonathan nie.
 - b. Hierdie skrywer se boeke verkoop wel goed, maar daardie **een** se werk is van hoë literêre kwaliteit.
 - c. Hulle het met heelwat voorstelle vorendag gekom, maar dit het nie gehelp nie, want slegs die **een** is op die ou end aanvaar.

- (ii) Substitusie deur *myne/joune*, ensovoorts
 - a. Ek weet jou tas is in die kamer, maar waar is **syne/myne/hare**?
 - b. My motor staan hier om die hoek. Waar is **joune/syne**?

- (iii) Substitusie deur *s'n*
 - a. Ons atlete is hier, maar waar is julle **s'n**?
 - b. Sarel se voorstel is van die hand gewys, terwyl hulle **s'n/julle s'n/ons s'n** aanvaar is.

- (iv) Substitusie met behulp van gesubstantiveerde adjektiewe.
 - a. Sorteer asseblief hierdie klompie appels vir my uit: die **vrottes** kan jy maar in die drom gooi.
 - b. Die een hand gebruik hy om hom op die tafel te steun, en met die **ander** beduie hy woes.
 - c. A: “Kan ek asseblief appeltert met room kry?”
B: “Ek wil **dieselfde** hê, dankie.”

3.5.2.2 Werkwoordelike substantivering

Volgens Carstens (1997:233-234) kan werkwoorde op drie maniere vervang word. Hierdie maniere is (i) deur doen *dit/so* (die Engelse “do”), (ii) deur ’n werkwoord + *dieselfde* en (iii)

deur 'n vorm van klousale substitusie. Die volgende voorbeelde ter illustreer kom uit Carstens (1997:234)

- (i) Substitusie deur doen *dit/so*
 - a) Ek sal nog skottelgoed was as dit nodig is, maar hy sal **dit** beslis nie doen nie.
 - b) Johan het gisteraand met sy voete op die tafel gesit en nou dat ek ook **so maak/dit doen**, is Ma kwaad vir my.

- (ii) Substitusie deur *dieselfde*
 - a. Hulle het almal begin gil. Toe doen ons **dieselfde**.
 - b. My bankbestuurder het sy goudaandele verkoop. Waarom doen jy nie **dieselfde** nie?

3.5.2.3 Sinstuk-substitusie

Volgens Carstens (1997:234) is klousale substitusie 'n belangrike deel van substitusie in Afrikaans. 'n Sinstuk of klous word beskryf as 'n "...strukturele eenheid wat kleiner as 'n sin (is), maar langer as frases of woorde" (Crystal, 2008:78, eie vertaling). Dit impliseer dat die klous wat vervang word, 'n langer strukturele eenheid as in die geval van 'n naamwoordelike of werkwoordelike frase kan wees.

In Afrikaans word 'n sinstuk normaalweg deur die substituuat *so* of *nie/so nie* vervang. In al die gevalle van klousale substitusie kan die substituuat slegs verstaan word indien die vorige sinstuk begryp is. Carstens (1997:236) gebruik die volgende voorbeelde om bogenoemde te illustreer:

“Almal dink hy is onskuldig. Indien dit *so* is, waarom het hy dan weggehardloop?” (**so** – hy is onskuldig)

“Is almal dan huis toe? Ek hoop tog *nie so nie!*” (**nie so** – almal is nie huis toe nie)

“Kan ek vir jou 'n koppie koffie en 'n stukkie koek aanbied?”

“O, nee, beslis *nie* – ek probeer tans minder inneem, want ek wil gewig verloor.” (**nie** – jy kan nie vir my 'n koppie koffie en 'n stuk koek aanbied nie)

3.5.2.4 Samevatting - substitusie

By substitusie is daar dus twee elemente wat oorweeg moet word om die tipe substitusie te identifiseer. Aan die een kant moet daar opgelet word na die woordsoortlike funksie van die sinsnede wat vervang word en aan die ander kant na die element wat dit vervang. Carstens (1997:237) som die Afrikaanse woorde wat gebruik word om woorde/sinsnedes te vervang soos volg op:

Naamwoordelike substitusie:

- lidwoord = een/ene/enetjie (bv. die een)
- demonstratief/vraagwoord + een (bv. hierdie/watter een)
- rangtelwoord/vraagwoord + een (bv. hierdie/ watter een)
- adjektief + een (bv. die sesde een)
- myne, ensovoorts
- s'n
- gesubstantiveerde adjektiewe (bv. die rooie)
- die + ander
- dieselfde

Werkwoordelike substitusie:

- doen/maak, ensovoorts (+ dit/so/dieselfde)

Klousale substitusie:

- so
- nie
- + nie

3.5.3 Ellips

Carstens en Van de Poel (2010:109) merk op dat ellips tesame met substitusie relatief min as kohesieskakel in tekste ingespan word in vergelyking met van die ander kohesiemiddele (sien ook Carstens, 1997:239 en Van de Poel, Carstens & Linnegar, 2012:76). Hierdie resultate word ook in hierdie studie opgemerk (sien hoofstuk 6). Omdat die gebruik van ellips die weglating van 'n element of elemente behels, soos vervolgens verduidelik word, is dit ook nie so duidelik identifiseerbaar soos sommige van die ander kohesiemiddele nie.

Carstens (1997:239-240) verduidelik dat ellips in 'n groot mate ooreenkom met substitusie en verwysing, maar uiteraard ook 'n aantal verskille toon. In hierdie verband gaan Halliday en Hasan (1976:142) so ver as om te sê dat hul bespreking oor substitusie en ellips eintlik in een hoofstuk geplaas sou kon word. Die rede hiervoor is dat substitusie en ellips beide 'n

verhouding tussen woorde of sinsnedes veronderstel teenoor byvoorbeeld verwysing, wat 'n semantiese verhouding tussen kohesiewe elemente veronderstel. Hulle gaan voort deur te sê dat daar op strukturele vlak 'n onderskeiding tussen die twee kohesieskakels is, m.a.w. ten opsigte van die oppervlakstruktuur van die teks.

Volgens Halliday en Hasan (1976:143) kom ellips voor in sinne of sinsnedes waar dit uit die struktuur duidelik is dat daar inligting ontbreek wat elders verkry moet word om die sin/sinsnede te verstaan (sien ook Carstens & Van de Poel, 2010:110). Hulle stel dit egter duidelik dat dit slegs gaan om inligting wat binne die teks verkry kan word, en nie inligting wat uit die pragmatiese konteks verkry word nie. Dit impliseer dat die weggelate gedeelte elders in die teks konkreet identifiseerbaar is.

Dit wil voorkom asof voorbeelde wat in die literatuur gebruik word om tipes ellips te illustreer op een van twee tipes binding fokus. Dit fokus óf hoofsaaklik op binding deur middel van ellips tussen twee (of meer) sinne, soos Halliday en Hasan (1976) illustreer óf dit fokus op binding binne 'n neweskikkende konstruksie (Ponelis 1979, in Carstens, 1997:251-256) fokus.

3.5.3.1 Tipes ellips soos aanvanklik deur Halliday en Hasan uiteengesit

Halliday en Hasan (1976) klassifiseer tipes ellips op dieselfde manier as wat hulle tipes substitusie klassifiseer, naamlik naamwoordelike, werkwoordelike en sinstuk-ellips.

3.5.3.1.1 Naamwoordelike ellips (“nominal ellipsis”)

Halliday en Hasan (1976:147-148) verduidelik dat daar by naamwoordelike ellips altyd 'n naamwoord (of naamwoordstuk) weggelaat word en dat die kwalifiseerders van die betrokke naamwoord in die elliptiese frase agterbly. Onder die voorbeelde wat hulle ter toeligting gee, is die volgende:

- *Four other Oysters followed them, and yet another four.*
- *Which last longer, the curved rods or the straight rods? – The straight are less likely to break.*

In die eerste voorbeeld bly die hoeveelheidsaanduiding oor as aanduiding van waar die naamwoord wat weggelaat is, gevind kan word.

In die tweede voorbeeld is die tipe-aanduiding “straight” steeds in die sin, maar die naamwoord “rod” is weggelaat.

3.5.3.1.2 Werkwoordelike ellips (“verbal ellipsis”)

Net soos naamwoordelike ellips die weglating van die naamwoord of naamwoordstuk inhou, behels werkwoordelike ellips dat die werkwoord weggelaat word. Net soos met die naamwoordelike ellips is daar sekere strukturelemente wat “oorgedra” word, ten einde steeds die boodskap te verstaan. By werkwoordelike ellips word daar veronderstel dat “...een of meer woorde uit ’n vorige werkwoordelike uitdrukking..” struktureel weggelaat is, maar dat die betekenis daarvan, steeds duidelik in die sin blyk (Halliday & Hasan, 1967:167), terwyl slegs die hoofwerkwoord weggelaat word.

Halliday en Hasan tref etlike fyner onderskeidings in subkategorieë van verbale ellips. Vir die doel van hierdie studie gaan daar egter volstaan word met bogenoemde beskrywing.

3.5.3.1.3 Sinstuk-ellips (“Clausal ellipsis”)

Onder die subkategorieë van verbale ellips word “operator ellipsis” en “lexical ellipsis” aangedui. Beide hierdie twee tipes steun egter baie sterk op Engelse grammatikareëls en dit word nie in hierdie studie ondersoek nie (vergelyk afdeling 3.5.3.1.4).

3.5.3.1.4 Ellips soos dit deur Carstens (1997) uiteengesit is

Buiten die bogenoemde klassifikasie deur Halliday en Hasan, merk Carstens (1997:251) op dat ellips baie keer tot stand kom wanneer neweskikkende sinne/sinsnedes saamgevoeg word. Hy identifiseer drie tipes ellips, naamlik neweskikkingsreduksie, werkwoordellips en bondeling. Hy merk op dat neweskikkingsreduksie en bondeling baie ooreenkomste toon en tot ’n mate saam gegroeper kan word met werkwoordellips wat los van hierdie groepering staan. Hy (Carstens, 1997:254) wys uit dat neweskikkingsreduksie en werkwoordellips groot strukturele ooreenkomste toon. Die grootste verskil lê in die feit dat daar hoofsaaklik werkwoordelike elemente by werkwoordellips weggelaat word, terwyl nominale elemente hoofsaaklik by

neweskikkingsreduksie weggelaat word. By die neweskikkingsreduksie wys hy deur middel van voorbeelde hoe die subjek en/of die objek en adjunk(te) weggelaat kan word om neweskikkingsreduksie te bewerkstellig. 'n Voorbeeld van neweskikkingsreduksie is die volgende (Carstens, 1997:253):

In die vakansie swem Waldo daaglik in ons swembad en [^] ry Stefan met sy fiets in die straat rond.

In die bogenoemde voorbeeld is die adjunk “In die vakansie” weggelaat.

Die ander tipe ellips wat hy (Carstens, 1997:254-256) bespreek, is werkwoordellips. Hy is van mening dat hierdie tipe gereeld in Afrikaans voorkom en wys uit dat die hoof- of hulpwerkwoord of bykomende elemente tesame met die werkwoordelement weggelaat kan word. 'n Voorbeeld wat hy gee van 'n hoofwerkwoord wat saam met nominale elemente weggelaat word, is die volgende (Carstens, 1997:255):

My ma het nog nie die perskes ingelê nie, maar sy sal seker later [^^].

In die bogenoemde voorbeeld is die nominale element ‘die perskes’ en die werkwoord ‘inlê’ na die voegwoord weggelaat.

In hierdie studie gaan slegs ellips wat in 'n neweskikkende konstruksie voorkom, ondersoek word. Die rede hiervoor is om die elemente wat ondersoek word duideliker af te baken. As gevolg van die neweskikkingskonstruksie is dit haalbaar om sulke elliptiese elemente op te spoor, aangesien dit duidelik aangedui word deur die voegwoord(e) wat die neweskikkende sinne/sinsnedes verbind.

3.5.4 Leksikale kohesie

Die kohesieskakels wat in die voorafgaande afdelings bespreek is, hou verband met binding op grammatikale vlak en daar is telkens 'n beperkte of selfs 'n vaste versameling merkers wat hierdie kohesieskakels aandui. Leksikale kohesie is egter nie 'n grammatikale bindmiddel nie, maar 'n semantiese een. Leksikale kohesie word deur Halliday en Hasan (1976:274) beskryf as “...die kohesiewe effek wat deur woorkeuse bewerkstellig word” (eie vertaling). Hulle gaan voort om te sê dat leksikale kohesie in 'n groot mate verskil van die ander vorme van kohesie. Die grootste verskil is dat leksikale kohesie deel is van 'n sogenaamde “open set” wat beteken

dat daar nie 'n eindigende versameling leksikale items is wat leksikale kohesie bewerkstellig nie.

Halliday en Hasan identifiseer aanvanklik vier kategorieë van leksikale kohesie, wat hulle later in twee afdelings indeel, naamlik leksikale herhaling (“reiteration”) en kollokasie (“collocation”). Onder leksikale herhaling klassifiseer hulle die herhaling van dieselfde leksikale item om te bind met die voorafgaande een, sinonieme, superordinate en 'n epiteton of ‘algemene woord’ (Carstens,1997:318-320; Carstens & Van de Poel, 2010:111).

3.5.4.1 Leksikale herhaling

3.5.4.1.1 Herhaling van 'n woord of frase

Die fisiese herhaling van die woord is die duidelikste waarneembaar en Carstens (1997:318) gebruik 'n uittreksel uit Jan Rabie se kortverhaal, *Drie kaalkoppe eet tesame* om die verskynsel te illustreer:

Die skraal kelner bedien hulle swygend. Hy toon geen emosie voor sy kliënte nie; maar ander mense in die restaurant is verbaas oor die etery van die *drie kaalkoppe*, en verlaat hul eie tafels om in 'n kring om die drie eters te kom staan.

Die *drie kaalkoppe* swaai op en af in die woeste vaart van hulle eet.

In bogenoemde uittreksel word “drie kaalkoppe” herhaal en dit is duidelik dat die tweede voorkoms van dié item na die eerste verwys

3.5.4.1.2 Sinonieme

Die gebruik van 'n sinoniem om terug te verwys na 'n spesifieke leksikale item val ook onder die kategorie leksikale kohesie. Volgens Carstens (1997:319) kan die volgende as 'n voorbeeld van die gebruik van 'n sinoniem as bindingselement gesien word:

Kyk net daardie *seun* wat soos 'n bok teen die rotse uitklouter! Die *knaap* moet net versigtig wees – 'n mens val maklik teen hierdie kranse af.

3.5.4.1.3 Superordinate

Superordinate verwys volgens Crystal (2008:465) na sogenaamde “hoër orde” leksikale elemente wat semanties ander “laer orde” elemente insluit. ’n Voorbeeld is dat *taal* ’n superordinaat kan wees van *Afrikaans*, *Engels*, *Nederlands*, ensovoorts. Volgens Carstens (1997:319) kan superordinate as ’n skakelende element optree. Die voorbeeld wat Carstens gee, is as volg:

Kyk *daardie seun* wat soos ’n bok teen die rotse uitklouter! Die **kind** moet net versigtig wees – ’n mens val maklik teen hierdie kranse af.

Die begrip “kind” sluit onder meer woorde soos “seun” en “dogter”, maar selfs begrippe soos “boetie” en “sussie” in en in bogenoemde voorbeeld tree “kind” duidelik as oorkoepelende term op en is verbind aan “seun”.

3.5.4.1.4 Epiteta

Volgens Carstens (1997:175) word epiteta “...algemeen as anafore gebruik en word hoofsaaklik gekenmerk deur hulle affektiewe waarde; daarom dien hulle veral as etikette, dit wil sê as waarnemings of oordele oor die referent gesien vanuit ’n bepaalde oogpunt (positief of negatief)”. Dit wil sê deur gebruik te maak van ’n algemene woord om terug te verwys na ’n bepaalde referent word die betrokke elemente nie slegs referensieel aan mekaar verbind nie, maar word verdere semantiese inligting oor die referent gegee. Vergelyk die volgende voorbeeld:

Ons het toe uiteindelik gehoor van *Julius*. Die *eensame ou siel* het dit ook maar nie te maklik daar in Johannesburg nie. (Carstens 1997:175)

Dit is duidelik dat “die arme ou siel” verwys na “Julius” wat in die eerste sin genoem is. Die epiteta dra egter ook ’n affektiewe betekenis. Volgens Carstens (1997:175) kom epiteta veral in spreektaal voor, as gevolg van die metaforiese gebruikswaarde daarvan.

Bogenoemde is die vier kategorieë van Halliday en Hasan se *leksikale herhaling* (as een van twee onderdele van leksikale kohesie) wat hoofsaaklik in die literatuur bespreek word.

3.5.4.2 Kollokasie

Die ander kategorie van leksikale kohesie wat deur Halliday en Hasan (1976) bespreek word, is kollokasie. Hulle beskryf kollokasie as kohesie wat verkry word deur die “...assosiasie van leksikale items wat gereeld saam voorkom”. Crystal (2008) gebruik steeds hierdie definisie in *A dictionary of Linguistics and Phonetics*. Volgens Carstens (1997:322) is daar egter ’n soort vaagheid wanneer dit kom by die begrip “kollokasie” soos dit in tekslinguistiese terme gebruik word. Hy gaan dan voort om ’n aantal navorsers aan te haal ter illustrasie van sy standpunt.

Halliday en Hasan (1976:285) noem selfs dat kollokasie ’n problematiese deel van leksikale kohesie is. Hulle meen dat daar drie faktore is wat die mate van kohesie tussen kollokasies beïnvloed. Eerstens word die ‘kohesiewe krag’ beïnvloed deur die voorkoms van die kollokasies in algemene taalgebruik. Tweedens beïnvloed die voorkoms van die kollokasies in tekste in die algemeen die sg. kollokatiewe krag. Laastens beïnvloed die fisiese afstand tussen die woorde binne die betrokke teks ook die ‘kohesiewe krag’.

Stotsky (1983) lewer in haar artikel oor leksikale kohesie in verklarende/beskrywende skryfwerk, kritiek op Halliday en Hasan se riglyne oor kollokasie. Sy sê dat daar veral twee tekortkominge is wanneer dit by hulle skema vir kollokasie kom. Eerstens praat sy van die gebrek aan uitgebreidheid wanneer die kollokatiewe kohesiepatrone in verklarende/beskrywende skryfwerk ondersoek word. Tweedens is sy van mening dat daar ’n gebrek aan ondersteunende argumente is vir hulle twee subkategorieë binne leksikale kohesie (leksikale herhaling en kollokasie) en dat redes vir die indeling onder hierdie subkategorieë ontbreek.

In verband met die gebrek aan uitgebreidheid by kollokatiewe patrone in verklarende/beskrywende skryfwerk gebruik sy voorbeelde uit opstelle van haar studie se tekste. Die vrae wat sy vra, is egter relevant vir hierdie studie. Sy bespreek hoofsaaklik twee probleme wat met hierdie ‘uitbreidheid’ te doen het. Eerstens sê sy dat Halliday en Hasan se skema nie plek het vir woorde wat met mekaar verband hou as gevolg van ’n soortgelyke stam met morfeme wat hulle semanties van mekaar verwyder nie. Sy gebruik die voorbeeld van *krag/kragtig* (eie vertaling). Volgens haar is dit onduidelik waar hierdie woordpaar sou inpas binne Halliday en Hasan se skema. Dit pas nie werklik in by hulle beskrywing van ’n kollokasie nie, maar kan ook nie onder leksikale herhaling geklassifiseer word nie. Haar tweede probleem in verband met die ‘uitbreidheid’ van hul skema, hou verband met die

superordinaat/hiponiem-verhouding. Volgens haar beskryf Halliday en Hasan nie die skakel wat ontstaan wanneer die superordinaat voor die subordinate genoem word nie.

In verband met die gebrek aan ondersteunende argumente vir die klassifikasie van die subkategorieë is Stotsky (1983:433) van mening dat dit veral problematies is dat daar nie konstante beginsels vir die klassifikasie van onderafdelings in hierdie subkategorieë is nie.

Vanuit Halliday en Hasan se skema met die bygewerkte kritiek, herorganiseer Stotsky (1983:438-441) die teoretiese raamwerk van Halliday en Hasan (1976). Sy stel ook 'n sisteem voor met twee kategorieë. Die eerste is 'n kategorie wat leksikale kohesiepatrone insluit wat 'n sistematiese semantiese verband tussen woorde het, ongeag daarvan of die woorde gereeld saam verskyn. Hieronder val pare soos sinonieme, antonieme en superordinate – soortgelyk aan Halliday en Hasan se skema. Verder word afgeleide woorde soos *intens/intensiveer* en dele van 'n georganiseerde of ongeorganiseerde stel (een, twee, drie of *tafel/stoel*) ook onder hierdie kategorie geklassifiseer. Dus val woorde wat 'n herkenbare semantiese relasie tot mekaar het, ongeag van die taalkundige rede, onder hierdie kategorie. Sy noem hierdie skakels semanties verwante woorde.

Aan die ander kant beskryf Stotsky (1983:432-433) woorde wat kollokatief aan mekaar verbind is as woorde wat nie 'n sistematiese semantiese relasie het nie, maar wat aan mekaar verbind is deur middel van die onderwerp van die betrokke teks. Voorbeelde van woorde wat sy as kollokatief verwant sien, is “bevoegdheidstoetse, hoërskooldiploma, leerder, moeilikheidsgraad, lees, skryf, wiskunde” en “werkgewer, werker, ambag, betrekking” (Stotsky 1983:439). Dit lyk asof sy tog aanvaar dat die leser se agtergrondkennis van die onderwerp van die teks, 'n rol sal speel in die ervaring van hierdie tipe kollokasie, as 'n meganisme wat kohesie bewerkstellig.

Stotsky (1983) en Halliday en Hasan (1976) erken beide dat die kohesieskakel *kollokasie* nie 'n duidelike een is om af te omskryf nie en dat die interpretasie van die skakel tot 'n mate afhanklik is van die leser. In hierdie studie word die uitgangspunt van Stotsky geneem in die klassifikasie van kollokasie – dus met die bykomende aspek van terme wat dalk eerder lyk of dit deel van dieselfde semantiese veld is. Daar sal in afdeling 5.6.3.1 van 'n aantal voorbeelde uit hierdie studie se datastel gebruik gemaak word om te verduidelik hoe kollokasie in die analise “toegeken” is.

'n Opsomming van Halliday en Hasan asook van Stotsky se benadering, lyk dus soos volg:

Tabel 3-1: Vergelykende opsomming van Halliday & Hasan (1967) en Stotsky (1983) se benaderings tot leksikale kohesie

| Halliday & Hasan (1967) | Stotsky (1983) |
|---|---|
| 1. Leksikale herhaling | 1. Semanties verwante woorde |
| <ul style="list-style-type: none"> • Herhaling • Sinoniem • Superordinaat • Epiteta/Algemene Woord | <ul style="list-style-type: none"> • Herhaling • Sinoniem • Kontras/Antoniem • Hiponimie • Lid van 'n geordende stel/reeks • Lid van 'n ongeordende stel/reeks • Afleiding |
| 2. Kollokasies | 2. Kollokatief verwante woorde |
| <ul style="list-style-type: none"> • Assosiasie met 'n spesifieke onderwerp • Kontras of antoniem • Lid van geordende stel/reeks • Lid van ongeordende stel/reeks | <ul style="list-style-type: none"> • Skakel waarin leksikale items aan mekaar verwant is deurdat hulle gereeld saam verskyn in soortgelyke kontekste. |

3.5.5 Konjunksie

Konjunksie verskil van die kohesiemiddele wat tot dusver bespreek is in die sin dat dit nie 'n anaforiese verhouding met 'n antesedent het nie. Volgens Halliday en Hasan (1976:226) is verbandsmerkers 'n ander tipe semantiese verhouding in die sin dat die skakel self nie semanties, of andersins, verbind is aan 'n referent nie, maar eerder 'n verhouding van 'n gedeelte wat volg teenoor 'n gedeelte wat afgehandel is, aandui. Halliday en Hasan (1976:238-239) stel moontlike kategorieë voor waarin verbandsmerkers geplaas kan word. Die kategorieë is gebaseer op tipes verbande wat daar tussen teksgedeeltes (kan) ontstaan en dit is aaneenskakelend (“additive”), teenstellend (“adversative”), redegewend (“causal”) en tydsaanduidend (“temporal”). Carstens (1997:292) gebruik hierdie kategorieë om 'n lys van verbandsmerkers in Afrikaans wat binne hierdie kategorieë pas, saam te stel (sien tabel 3-2).

Tabel 3-2: Voorbeelde van Carstens (1997) se indeling van verbandsmerkers

| Subkategorie | Voorbeelde |
|------------------------|---|
| AANEENSKAKELEND | |
| analiserend | naamlik |
| beoordelend | egter |
| herbewoordend | eerder, met ander woorde |
| oorsaak en gevolg | en, gevolglik |
| ordenend | ten laaste, ook |
| rangskikkend | voorts, vervolgens, verder |
| ruimtelik | tussenin |
| toevoegend | bygesê, addisioneel |
| verklarend | eerder, soos gesê, ten minste |
| versterkend | tog, ook, buitendien |
| ervattend | anders gestel, dit wil sê |
| voortgang | en, plus |
| voorwaardelik | naamlik, te wete, op die voorwaarde dat |
| TEENSTELLEND | |
| alternerend | of, óf...óf, hetsy, mits |
| kontrasterend | intendeel, egter, dus, maar |
| korrektief | in plaas van, in teenstelling met |
| ooreenstemmend | dieselfde, dienooreenkomstig |
| toegewend | al, ten spyte van |
| vergelykend | as, soos, vergeleke met |
| vervangend | meer, alternatiewelik |
| voorwaardelik | as, met dien verstande |
| wyse | dus, asof |
| REDEGEWEND | |
| afleidend | met ander woorde |
| doel | sodat |
| illustreerend | byvoorbeeld, te wete |
| kontrasterend | anders, andersins, afgesien hiervan |
| oorsaaklik | want, omdat |

| | |
|-----------------------|---|
| opsommend | op die ou end |
| plek | waar |
| resulterend | as gevolg van, gevolglik |
| toegewend | in elk geval, nogtans |
| verontagsamend | hoe dit ook al sy, wat ook al |
| voorwaardelik | as, mits, tensy |
| TYDSAANDUIDEND | |
| herhalend | die vorige keer, 'n volgende keer |
| hervattend | om te hervat, vervolgens, hieropvolgend |
| kontrasterend | terwyl |
| onderbroke | tans, later, 'n kort rukke tevore |
| oorsaaklik | dan, toe |
| opeenvolgend | dan, vandat, volgende |
| opsommend | ten slotte, laastens |
| ordenend | eerstens, verder, vervolgens |
| ruimtelik | daar |
| spesifiserend | die volgende dag, vyf minute vroeër |
| stiptelik | op daardie oomblik, toe |
| voorafgaande | voordat, alvorens |
| voortgang | intussen, onderwyl, die hele tyd |
| voorwaardelik | behalwe, indien, wanneer |

Jordaan (2014:39-79) dui in haar korpusgebaseerde studie oor die gebruik van verbandsmerkers (sy gebruik die term verbandsmerkers) in akademiese skryfwerk deur studente aan dat die bogenoemde indeling van Carstens (1997) nie “...klakkeloos nagevolg...” (Jordaan, 2014:39) kan word nie, aangesien haar data toon dat die verbandsmerkers nie altyd in die verband gebruik word wat Carstens (1997:292) aangetoon het nie. Sy haal 'n aantal voorbeelde uit haar data aan om haar argument (sien tabel 3-3) te staaf (Jordaan, 2014:39-40).

Tabel 3-3: Voorbeelde van herkategorisering van verbandsmerkers volgens Jordaan (2014)

| Merker | Carstens (1997:292) se indeling | Herkategorisering volgens data van Jordaan (2014) |
|---|--|---|
| egter | Onder andere aaneenskakelende verbandsmerker – beoordelend | Teenstellende verbandsmerker |
| Voorbeeld: ...wat beide die herinneringe laat uitwis het wat hulle van mekaar gehad het. Herinnering is egter nie iets wat maklik uitgewis kan word nie... | | |
| gevolglik | Onder andere aaneenskakelende verbandsmerker – oorsaak en gevolg | Redegewende verbandsmerker |
| Voorbeeld: Die studente voltooi nie hulle studie kursus nie, gevolglik styg die werkloosheidsyfer weer. | | |
| dieselfde | Teenstellende verbandsmerker – ooreenstemmend | Aaneenskakelende verbandsmerker |
| Voorbeeld: Die ek maak wilde aannames oor die Ander wat nie gebaseer is op feite nie, maar eerder op 'n veralgemening. Dieselfde gebeur vandag nog wat betref baie blankes se idees oor die intellek van swartes. | | |

Op grond van die gebruik van verbandsmerkers herverdeel sy (Jordaan, 2014:43-80) die Afrikaanse voorbeelde wat Carstens (1997:292) gegee het soos volg:

Aaneenskakelende verbandsmerkers

- bowendien
- daarby
- ook nog
- buitendien
- daarenbowe
- in elk geval
- bygenoem
- bygesê
- addisioneel
- dit wil sê
- dit wil voorkom
- dieselfde
- en
- asook
- om te hervat
- soos genoem
- soos voorheen
- soos gesê
- ook
- selfs
- net so
- dienooreenkomstig
- dieselfde
- daarmee ooreenkomende
- stem ooreen
- opsommend
- ter opsomming
- ten minste
- minstens
- ter illustrasie
- om te illustreer
- verder

- voorts
- vervolgens
- wat die geval ook mag wees
- hoe dit ook al sy
- die feit bly staan

Teenstellende verbandsmerkers

- al
- hetsy
- alternatiewelik
- andersins
- asof
- behalwe
- behalwe vir
- buiten
- benewens
- eerder
- liewer
- desondanks
- ten spyte hiervan
- ten spyte van
- afgesien van
- afgesien hiervan
- in weerwil van
- in weerwil hiervan
- niteenstaande
- dienooreenkomstig
- dan
- egter
- ewenwel
- in kontras met
- in teenstelling
- teenstellend
- daarteenoor
- daarenteen
- in stede van
- in plaas van
- pleks van
- in vergelyking met
- vergeleke met
- intendeel
- maar
- nietemin
- tog
- of
- ofskoon
- hoewel
- alhoewel
- ondanks
- sonder dat
- terwyl

Redegewende verbandsmerkers

- afhangende van
- hang af van
- berustend op
- aangesien
- siende dat
- met die dat
- omdat
- omrede
- deurdat
- anders
- as
- indien
- waar
- ingeval
- mits
- as gevolg van
- daarom
- om die rede
- om dié rede
- weens die rede
- vir dieselfde rede
- nog een rede
- 'n ander rede
- kan 'n rede wees
- is 'n rede
- die rede wat
- die rede vir
- 'n rede vir
- die rede hiervoor
- 'n rede hiervoor

- een rede hiervoor
- die rede daarvoor
- 'n rede daarvoor
- een rede daarvoor
- die rede waarom
- derhalwe
- dienooreenkomstg
- dus
- gevolglik
- hierom
- om hierdie rede
- namate
- na aanleiding van
- na aanleiding hiervan
- nogtans
- om
- op die ou end
- eindelik
- uiteindelik
- op die voorwaarde dat
- op die kondisie dat
- met dien verstande dat
- op grond van
- sodat
- tensy
- wan

Tydsaanduidende verbandsmerkers

- alvorens
- voordat
- voor
- daarna
- dan
- voorts
- vorige
- hieropvolgend
- hierna
- hiervandaan
- intussen
- onderwyl
- solank
- later
- latere
- 'n kort rukkie tevore
- nadat
- noudat
- op hierdie oomblik
- op daardie oomblik
- die volgende oomblik
- op die ou end
- eindelik
- uiteindelik
- sedert
- vandat
- sedertdien
- sodra
- telkens
- terwyl
- toe
- tot
- totdat
- vantevore
- voorheen
- vroeër
- wanneer

Hoewel Jordaan (2014) nie aanvoer dat haar lys al die verbandsmerkers wat bestaan, insluit nie, gaan die woorde in hierdie lys gebruik word om die voorkoms van verbandsmerkers in die tekste van André le Roux en Dana Snyman aan te dui. Hoewel nie gewaarborg volledig nie, is Jordaan (2014:43-80) se lys van verbandsmerkers 'n omvattende een wat onlangs hersien en op grond van empiriese data saamgestel is. Die doel van hierdie studie is nie om verbandsmerkers (of ander kohesiemerkers) op te spoor nie, maar slegs die gebruik daarvan te

ondersoek. Dit is ook 'n afgebakende korpus van woorde/frases wat betroubaar opgespoor kan word.

3.6 Samevatting

In hierdie hoofstuk is die teoretiese veronderstellings rakende kohesie as 'n tekstualiteitsbeginsel bespreek. Daar is verduidelik watter kohesieskakels bestaan en wat elke kohesieskakel behels. Met die uitsondering van kollokasie binne leksikale kohesie, is kohesieskakels 'n waarneembare verskynsel wat dit 'n geskikte veranderlike maak om as moontlik kenmerkend in 'n skrywer se styl te ondersoek.

In hoofstuk 4 word die nosie van genre en tekstipe ondersoek. Die twee gekose genres vir hierdie studie, naamlik die rubriek en kortverhaal, word ook bespreek ten einde af te baken watter stilistiese vereistes of norme daar moontlik vir hierdie genres bestaan.

4 Genre en tekstipe

4.1 Inleiding

Soos daar in die voorafgaande hoofstukke uiteengesit is, fokus hierdie studie hoofsaaklik op die gebruik van kohesiemiddele as moontlike linguistiese veranderlike in die skryfstyl van 'n bepaalde outeur. Hierdie linguistiese element word primêr ondersoek as moontlik tekenend van die outeur se persoonlike skryfstyl. Daar is egter verskeie faktore wat 'n invloed het op 'n outeur se persoonlike skryfstyl waarvan 'n situasionele faktor soos genre-vereistes van die spesifieke skryfstuk een is (vergelyk afdeling 2.3). In hierdie afdeling word 'n aantal prominente benaderings en teorieë ten opsigte van genre kortliks uiteengesit ten einde die terme en benaderings vir hierdie studie af te baken. Daarna word die implikasies van hierdie teorie binne die konteks van hierdie studie geplaas.

Biber en Conrad (2019:4) verduidelik dat taalkundige variasie hoogs sistematies van aard plaasvind. Hulle verduidelik dat taalgebruikers 'n groot aantal linguistiese karaktereenskappe (bv. uitspraak, woordkeuse, grammatieka, ensovoorts) na gelang van nielinguistiese faktore aanpas. Hul noem dat hierdie faktore onder meer die doelwit vir kommunikasie, verhouding tussen gespreksgenote, en sosiale karaktereenskappe van die spreker insluit. Die aanname kan gemaak word dat hierdie aanpassings gedeeltelik plaasvind om by 'n sekere konteks aan te pas. Biber en Conrad (2019:4) se studies oor taalvariasie dek grootliks variasie tussen genres, registers, dialekte en style; dit is duidelik dat hul publikasies oor hierdie onderwerp variasie in beide gesproke en geskrewe taalgebruik uiteensit (Biber, 1989 en Biber, Conrad & Reppen, 1998). Taavitsainen (2001:140) noem dat genre op konvensies gebaseer is. Hyland (2002:114) verduidelik dat hierdie konvensies bestaan omdat daar 'n sosiaal aanvaarbare manier is waarop taal binne sekere kontekste gebruik word. Genreteoretici is van mening dat kennis van hierdie konvensies deelnemers se taalhandeling rig. Dit impliseer dat elke suksesvolle teks tekens sal toon van die teksproduseerder se bewustheid van die sosiale konteks waarbinne die teks geproduseer word (Hyland, 2002:114). Dit is dus relevant om die betrokke skrywer se styl in meer as een genre te ondersoek ten einde die moontlike stylvariasie in die gebruik van kohesiemiddele as gevolg van die sosiale konvensies, te beskryf.

4.1.1 Algemeen aanvaarde vertrekpunte met betrekking tot genre

Voor die onderskeie teoretiese vertrekpunte bespreek word, is dit sinvol om die aannames oor die aard van genre-analise waarvoor die meeste teoretici saamstem, te noem. Hyland (2002:114) sê dat genre-analise hoofsaaklik op twee aannames gebaseer is. Eerstens hang die eienskappe van 'n teks saam met die sosiale konteks waarbinne die teks geproduseer is en waarbinne die teks gebruik word en of tekste van 'n soortgelyke groep *moontlik* binne soortgelyke kontekste geproduseer en/of gebruik word (eie klem). Tweedens kan daardie eienskappe van 'n teks beskryf word met betrekking tot soortgelyke tekste sowel as die gevolglike keuses en beperkinge wat dit aan die teksproduseerder bied. Hy verduidelik dat taal 'n manier is om uitdrukking aan sosiale realiteite/konteks te gee. Kress (1989:49) sê dat genre dus die optrede van 'n taalproduseerder op twee maniere beïnvloed. Eerstens moet hy/sy die wesensaard en beperkinge van die spesifieke konteks in ag neem. Tweedens moet hy/sy hierdie beperkinge in ag neem terwyl kennis van generiese tekstipes ook verreken word.

Hoewel daar algemene ooreenstemming is met betrekking tot bogenoemde aannames oor die aard van genre as studieveld en genre-analise, fokus teoretici op verskillende aspekte daarvan en benodig hulle genre-indeling vir verskillende doeleindes. Dit is as gevolg van die feit dat teoretici op verskillende aspekte klem lê, wat die drie teoretiese standpunte wat in Hyon (1996) uiteengesit is, ontwikkel het.

4.2 Benaderings tot genre

Daar is 'n aantal verskillende benaderings tot die indelings van tekste in genres en/of tekstipes wat binne die literatuur beskryf word. Op 'n konseptuele vlak sê Taavitsainen (2001:139) dat genres inherent kulturele skemata (skemas) is wat gebruik word om kennis en ervaring deur taal te organiseer. Hierdie definisie wys op die interaksie tussen teks en konteks. Dit is 'n aanduiding van twee van die hoofaspekte wat betrokke is wanneer tekste geklassifiseer word en bied verskeie benaderingsopsies wanneer daar in die literatuur oor genre-analisetorie geskryf word. Hyon (1996) bespreek die drie teoretiese standpunte wat binne toegepaste linguistiek gebruik word om *genre* in sy geheel te benader, naamlik die Nuwe Retoriek, Australiese Sistemies-Funksionele Linguistiek en Taal vir Spesifieke Doeleindes (*English for Specific Purposes – ESP*).

4.2.1 Nuwe Retoriek

Een van die benaderings tot genre staan bekend as die Nuwe Retoriek. Navorsers wat hierdie benadering aanhang, plaas die doel wat 'n genre binne 'n spesifieke sosiale konteks vervul, sentraal (Flowerdew, 2005:323). Flowerdew (2005:323.) verwys na Bazerman (1998) wat die verband tussen konteks en die aanvaarbare genre-reaksie daarop uitlig. Bazerman (1998) verduidelik dat kennis van die sosiale konteks waarbinne die taalhandeling plaasvind die teksproduseerder se keuse van gepaste diskoersvorm rig. Meintjes (2014:65) maak in hierdie verband die opmerking dat die klem vir die Nuwe Retoriek val op die sosiale doel wat deur genres vervul word, aangesien die *retoriese omstandighede* van die genre ondersoek word, eerder as op die *vorm van die taalhandeling/genre* (eie klem). Hyland (2002:114) bespreek die implikasies van hierdie situasionele klem. Hy sê dat studies binne die Nuwe Retoriek in der waarheid wegbeweeg van die teks as gevolg van die fokus op kontekstuele faktore. Dit het tot gevolg dat studies eerder etnografiese metodologieë volg as dié van teksanalise (sien ook Flowerdew, 2005:324).

4.2.2 Sistemies-Funksionele Linguistiek

Sistemies-Funksionele linguistiek is 'n benadering tot taalstudie wat aanvanklik deur M.A.K. Halliday voorgestel is en wordvolgens Eggins (2004:1-2) 'n gesien as 'n analiseraamwerk waarbinne taal gesien word as 'n hulpbron wat strategies aangewend word en gemik is daarop om betekenis te skep. Sistemies-Funksionele Linguistiek benader genres as 'n breë konsep en gebruik die term tekstipe oorwegend as 'n sinoniem vir genre (Eggins, 2004:55; vergelyk afdeling 4.3.1 en 4.6 vir besprekings oor tekstipe en genre). Martin (1984 en 1985) se definisie van genre word gereeld aangehaal (sien Eggins, 2004:55 en Hyland, 2002:115) as 'n "stapsgewyse, doelgeoriënteerde proses" waarin sprekers (teksproduseerders) as lede van 'n spesifieke kultuur optree. Eggins (2004:55) sê dat Martin se definisie vereenvoudig kan word deur te sê dat genre "'n manier is om dinge gedoen te kry wanneer taal gebruik word om dit te doen" (eie vertaling). Uit die eersgenoemde definisie kom die stapsgewyse aard van genre na vore. Hyland (2002:115) wys uit dat navorsing binne hierdie genreteorie tot dusver gepoog het om hierdie stapsgewyse aard te definieer en te beskryf. Hierdie stappe, tesame met die gestruktureerde gebruik van grammatikale, leksikale en kohesiewe patrone, vorm volgens

hierdie teorie die herkenbare struktuur van die betrokke genre. Lee (2001:43) merk op dat hierdie klem op genre as proses of aktiwiteit, eerder as 'n groep tekste of produk, een van die grootste verskille tussen hierdie en ander benaderings is.

4.2.3 Taal vir Akademiese Doeleindes

Hyland (2002:115) wys uit dat die benadering van Taal vir Akademiese Doeleindes tussen bogenoemde benaderings se uitgangspunte val in die sin dat dit elemente van beide het. Dit benader genre as 'n hoogs gekontekstualiseerde aktiwiteit wat sosiale konvensies in ag neem om die gepaste retoriek te kies (Nuwe Retoriek), maar dit erken ook die strukturele aard van tekste (Sistemies-Funksionele Linguistiek). Hyland (2002:115) verduidelik dat genre in hierdie sin uit 'n *klas of groep gestruktureerde kommunikatiewe gebeurtenisse* bestaan wat deur *spesifieke diskoersgemeenskappe* gebruik word deur lede wat *soortgelyke doelwitte* in terme van kommunikasie het (eie vertaling, eie klem). Flowerdew (2005:322) wat ook hierdie definisie gebruik, noem verder dat Swales se navorsing 'n sleutelement was in die vorming van hierdie benadering tot genreteorie. Paltridge (1996:236) wys daarop dat hierdie genreteorie gereeld gebruik word wanneer taalvaardighede binne 'n spesifieke konteks onderrig word, veral binne akademiese en professionele kontekste.

4.3 Genre en korpusgebaseerde navorsing

Korpusgebaseerde studies en studies wat groot korpusse soos die BNC (British National Corpus) gebruik, benodig 'n meer toegepaste benadering tot genre. 'n Benadering wat hulle volg moet eerstens die verskillende elemente wat met genreteorie verband hou in meer besonderhede beskryf en tweedens die praktiese implikasies van 'n genreteorie duidelik stel.

Wanneer daar in die literatuur oor genre met betrekking tot korpusgebaseerde navorsing geskryf word, is daar 'n aantal verwante terme en begrippe wat altyd aandag geniet. Begrippe soos tekstipe, register, domeine, en styl kom baie keer in besprekings voor en navorsers is dit eens dat daar nie 'n duidelike afbakening tussen en definisies van hierdie begrippe bestaan nie (Lee, 2001:37; Taavitsainen, 2001:139). Hierdie beskrywingsmoontlikhede is veral nodig wanneer sogenaamde natuurlike taalgebruik bestudeer word.

As gevolg van die meer toegepaste benadering wat korpusgebaseerde studies nodig het, is duidelike afbakening van terme belangrik. Soos dit uit die volgende bespreking sal duidelik wees, blyk dit of daar nie 'n algemeen aanvaarde onderskeid tussen terme gemaak kan word nie en dat die besonderhede van die afbakening van die terme telkens binne die konteks van die betrokke studie beskryf word. Die onderskeid tussen die terme *genre* en *tekstipe* word vervolgens bespreek.

4.3.1 Genre en tekstipe

Paltridge (1996:237) is van mening dat daar dikwels nie 'n onderskeid gemaak word tussen die terme *genre* en *tekstipe* nie en dat dit dikwels óf afgewissel word, óf een gekies word en vir beide konsepte gebruik word. Lee (2001:37) sê dat daar selfs verwarring bestaan tussen die *betekenis* van die twee terme en dat dit die rede is hoekom daar nie deurlopend na dieselfde konsepte op dieselfde manier verwys word nie (eie klem).

Biber (1988:4) poog om 'n duideliker onderskeid tussen genre en tekstipe te maak deur te onderskei tussen teksinterne en -eksterne faktore. Volgens hom is genre-indeling gebaseer op eksterne faktore soos konteks terwyl tekstipes onderskei kan word na aanleiding van hulle interne linguistiese eienskappe. Tekste wat dus binne dieselfde konteks of met dieselfde doel of binne dieselfde konvensies geskep is, behoort tot dieselfde genre. Hierteenoor sal tekste wat soortgelyke linguistiese eienskappe besit tot dieselfde tekstipe behoort. Biber kan waarskynlik hierdie definisie voorstel omdat hy die teksinterne eienskappe van 67 genres in diepte ondersoek het en dus op grond van hierdie eienskappe 'n onderskeid sal kan maak. Biber se omvattende multidimensionele korpusgebaseerde studie van 67 genres (vergelyk afdeling 4.6.1) het die weg gebaan vir kritiese bespreking van en navorsing oor genre en tekstipe.

Ongeag van die voorsprong wat Biber se studie aan toekomstige studies verleen het, wys Lee (2001:39) op die feit dat daar steeds nie duidelike voorstelle of reëls is as dit by die bepaling van tekstipe kom nie. Dit wil sê dat daar nog nie 'n duidelike, geldige manier is om deur middel van teksinterne eienskappe, tekste in spesifieke tekstipes in te deel nie.

Terwyl die genre van 'n teks betreklik maklik op grond van kontekstuele faktore bepaal kan word, kan dieselfde egter nie van tekstipes gesê word nie (Lee, 2001: 39). Giltrow en Stein (2009:3) gaan so ver as om te sê dat tekstipe slegs die linguistiese sy of aspek van genre is, wat

'n baie nou verhouding tussen die twee aspekte veronderstel. Hulle lê ook klem op die verband tussen genre en tekstipe en funksie en vorm. Dit impliseer Biber (1988) se veronderstelling dat die funksie van die teks die genre bepaal en die vorm, die tekstipe.

Uit bostaande bespreking, blyk dit dat die kombinasie van linguistiese keuses in die teks, die tekstipe bepaal. Ten spyte van hierdie duidelike verband tussen genre en tekstipe is dit egter so dat spesifieke tekstipes nie uitsluitlik tot spesifieke genres behoort nie. Verskillende genres kan soortgelyke tekstipes hê en net so kan tekstipes binne dieselfde genre baie varieer. Hierdie beginsel word deur Paltridge (1996:238-239) geïllustreer deur die tabel wat op Hammond *et al.* (1992) se artikel gebaseer is, te gebruik. In die tabel word 'n lys van genres gegee wat tekste binne sekere pragmatiese kontekste lys, soos byvoorbeeld 'n resep, persoonlike brief, advertensie, polisieverslag, formele brief, ensovoorts. tabel 4-1 is 'n voorstelling van Paltridge se veronderstelling in verband met die verhouding tussen genre en tekstipe.

Tabel 4-1: Beskrywing van die verhouding tussen genre en tekstipe wanneer genre geklassifiseer word op situasionele faktore en tekstipe geklassifiseer word op grond van teksinterne eienskappe

| Genre | Tekstipe |
|--------------------------|--------------------|
| Resep | Prosedure |
| Persoonlike brief | Anekdote |
| Advertensie | Beskrywing |
| Polisieverslag | Beskrywing |
| Student se opstel | Verklaring |
| Formele brief | Verklaring |
| Formele brief | Probleem/oplossing |
| Nuusitem | Oorvertel |

Soos in tabel 4-1 gesien kan word, kan beide die genres “advertensie” en “polisieverslag” dieselfde tekstipe te wete “beskrywing” hê. So wil dit ook voorkom dat 'n genre nie verbind is aan 'n enkele tekstipe nie, maar meer as een tekstipe-moontlikheid het, bv. in die geval van die “formele brief” word “verklaring” en “probleem/oplossing” as moontlike tekstipes gegee.

Behalwe dat Lee (2001:40) kritiek lewer op bogenoemde tekstipes en sê dat dit liever diskoers/retoriese struktuurtypes is as tekstipes, verskil Meintjes (2014:65) se vergelykbare tabel van genre en tekstipe veral op die vlak van genre met Paltridge s'n.

Meintjes (2014:93-95) verduidelik, nes Paltridge, dat genre met die funksie van die teks verband hou en tekstipe met die vorm. Sy verduidelik dus dat tekste ingedeel kan word "op grond van tipiese eienskappe wat die skrywer van hulp kan wees om die korrekte formaat teks in die regte konteks te produseer" (Meintjes, 2014:64). Sy gaan egter voort om die verskil tussen genre en tekstipe soos volg uiteen te sit:

Tabel 4-2: Meintjes se illustrasie van die wisselwerking tussen genre en tekstipe

| Genre (funksie) | Tekstipe (vorm) |
|------------------------|--|
| Narratiewe | Kortverhale, novelle, romans, kinderverhale. |
| Argumente | Akademiese artikels, verslae, verhandelings, proefskrifte. |

Bogenoemde twee pogings om funksionele afdelings uit die teorie te skep, wys dat daar in die algemeen konsensus is oor die *teoretiese basis* waarop genre en tekstipe geskei kan word, maar dat die *praktiese implikasies en definiëring* van terme en konstrunkte steeds nie duidelik is nie.

Carstens (1997:78-80) klassifiseer verskeie tekstipes deur die indeling van Van Luxemburg (1987:62) te gebruik. Hy verduidelik dat daar in die algemeen referensiële tekste, ekspressiewe tekste, oorredende tekste, retoriese tekste, konvensionele tekste en literêre tekste is, sommige met subtipes wat hy soos volg aandui:

- Referensiële tekste
 - Informatiewe tekste
 - Argumentatiewe tekste
 - Instruktiwe tekste
 - Narratiewe tekste
- Ekspressiewe tekste
- Oorredende tekste
 - Evaluerende tekste
 - Direktiewe tekste
- Retoriese tekste
- Konvensionele tekste

- Literêre tekste

Dit wil voorkom asof die afbakening van genre, tekstipe en die verwante begrippe en konstrunkte binne studies gedefinieer word vir die doel van die betrokke studie. Dit lyk ook asof daar meer konstrunkte is wat gedefinieer moet word as wat daar terme is waarmee dit gedefinieer kan word. Dit is ook nie noodwendig so dat genre die funksie van die teks impliseer indien genre-indeling op grond van kontekstuele faktore gedoen word nie. Die funksie van die teks kan slegs 'n aspek wees van die kontekstuele faktore wat 'n genre-indeling beïnvloed.

Om die teorie in al hierdie wisselende benaderings tot klassifisering van tekste vir hierdie studie bruikbaar te maak, moet daar egter gedeeltes van tekste geselekteer word, waarop besluite oor tekste en die klassifikasie daarvan gebaseer kan word. Hoewel baie van die literatuur oor genre-indeling relevant is, val dit buite die bestek van hierdie studie om 'n oplossing vir die bogenoemde probleme te bied. Daar moet dus bepaal word watter van die benaderings of teorieë vir die doel van hierdie studie duidelik gedefinieer moet word.

Hierdie studie fokus primêr op die individu se skryfstyl. Individuele skryfstyl word deur 'n aantal faktore bepaal (vergelyk afdeling 2.3 en 2.3.5.3) waarvan kontekstuele vereistes van die betrokke situasie een is, dus genrevereistes, indien 'n mens Biber (1988) se model neem. Die verhouding tussen kontekstuele faktore en teksinterne eienskappe word deur Ventola in Carstens (1997:87) verduidelik. Hy sê dat dit belangrik is om bewus te wees van die genre van 'n teks (eksterne invloede) omdat dit die register beïnvloed. Soos daar egter in die voorafgaande oorsig gesien is, is register nie die enigste tekstuele aspek wat deur die eksterne faktore beïnvloed word nie.

4.4 Kortverhale

In hierdie afdeling word die stylvoorstelle wat algemeen deur literatuur of onderrigbronne aanvaar word as bydraend tot goeie kortverhaalskryf, narratiewe teks of prosateks bespreek. Die doel van hierdie bespreking is om agtergrond te gee tot watter tipe eksterne faktore of genre-vereistes moontlik op 'n skrywer inwerk wanneer hy 'n teks produseer. Daar word ook gepoog om moontlike taalkundige implikasies van hierdie genre-kenmerke te voorspel of te identifiseer, ten einde 'n maatstaf vir vergelyking te skep wanneer data eindelik geanaliseer word. Eerstens volg 'n kort oorsig oor hoe 'n kortverhaal omskryf word.

4.4.1 Oorsig

Aangesien die kortverhaalt tekste wat in hierdie studie gebruik word, almal onder 1 300 woorde is, word die verhale wat in hierdie studie gebruik word as kort-kortverhale geklassifiseer (Gouws, 2013). Gouws (2013) verwys in sy artikel oor die kortkortverhaal in *Literêre Terme en Teorieë* na Lombard (1979:124) wat sê dat “(d)ie kort-kortverhaal oor ’n duidelike aantoonbare verhaalmatigheid (beskik)”. Gouws gaan voort deur te sê dat die kort-kortverhaal ’n baie unieke aard het vanweë die feit dat die teks so kort is tesame met die verhaalmatige aard daarvan. Hy verduidelik dat die taalpatroon van die kort-kortverhaal baie soortgelyk is aan die liriek omdat “metaforiese kriptiek en onderspeling ’n belangrike kenmerk van die kort-kortverhaal” is (Gouws, 2013).

4.4.2 Stylkenmerke

Gouws (2013) verduidelik verder wat die styl van die kort-kortverhaal ten opsigte van beide letterkundige en taalkundige eienskappe is. In terme van prosa-elemente verduidelik hy dat karakters nie in hulle volledige kompleksiteit uitgebeeld word nie, maar dat slegs enkele karaktertrekke gewys word. Dikwels speel daar slegs een of enkele handeling in die verhaal af. Hy (Gouws, 2013) sê dat daar dikwels met tyd en vertellersperspektief geëksperimenteer word en dat die “vertellersperspektiefwisseling nou met die tydshantering (saamhang)” (Gouws, 2013). Bogenoemde letterkundige elemente behoort ’n aantal taalkundige verskynsels wat te make het met die kohesiemiddele wat in hierdie studie ondersoek word, tot gevolg te hê. So sou die feit dat daar met tyd en vertellersperspektief geëksperimenteer word, ’n invloed kon hê op onderskeidelik tydsaanduiding konjunksie en referensiële verwysing.

In terme van taalkundige elemente verduidelik Gouws (2013) dat die woordgebruik in die kort-kortverhaal dikwels ’n poëtiese aard het. Verder sê hy dat verskynsels soos herhaling, parallelismes en anaforiek dikwels in kort-kortverhale voorkom.

Waldell (2014) gee in sy inligtingstuk oor “Basic prose style and mechanics” ’n oorsig oor stylkenmerke van goeie prosaskryfwerk. Anders as Gouws (2013) se oorsig oor die kort-kortverhaal, is Waldell se beskrywing grootliks op die praktiese aspekte van taalkundige styl gemik en lewer hy geensins kommentaar op die literêre aard van prosatekste nie. Dit is egter

belangrik om van sy stylwenke kennis te neem, aangesien dit deel is van die ooreenkoms tussen die outeur en die leser (vergelyk afdeling 4.5.2). Hy gee 13 voorstelle vir taalkundige aspekte wat nagestreef of vermy moet word wanneer prosatekste geskep word, naamlik:

- Skryf in die lydende vorm
- Vermy die gebruik van te veel bysinne
- Vermy nominalisering
- Maak die logiese verhouding tussen jou idees duidelik
- Druk parallelle idees in parallelle grammatikale vorm uit
- Laat dele uit wat nie 'n bydrae maak nie.
- Druk stelsinne in die positiewe vorm uit (dus nie negativering nie)
- Vermy oortolligheid
- Varieer sinspatrone
- Vermy werkwoordfrases
- Kies die woorde wat jy gebruik noukeurig
- Gebruik metafore om te illustreer

Indien 'n skrywer wou poog om die genrekonvensies, soos hierbo uiteengesit, te volg, sou dit 'n invloed op sy/haar skryfstyl kon hê.

4.5 Rubrieke

Soos in die afdeling wat oor kortverhale handel (afdeling 4.4), word daar in hierdie afdeling 'n oorsig gegee oor die moontlike vereistes (talig en buite-talig) wat 'n skrywer in ag moet neem wanneer hy/sy 'n rubriek skryf. Die doel van hierdie bespreking is om agtergrond te gee tot watter tipe genrevereistes op 'n skrywer se skryfwerk mag inwerk wanneer hy/sy 'n rubriek skryf.

4.5.1 Oorsig

In sy inleiding tot die bundel *Saterdagand Tuis* met Kas van den Berg se rubrieke sê Engelbrecht (1981:9) dat 'n rubriek "...algemeen bestempel [kan] word as gewilde leesstof wat gereeld onder 'n spesifieke titel, in 'n besondere vorm en gewoonlik op dieselfde plek in 'n

medium voorkom – en meestal deur dieselfde skrywer behartig word”. In haar artikel in verband met “rubriekmaak” haal Rabe (2015:176) ’n definisie uit die *Webster’s New World Dictionary of Media and Communication* (Weiner, 1996) aan wat treffende ooreenkomste toon met dié van Engelbrecht. Hierdie definisie sê dat rubrieke as “...’n reeks sketse of artikels deur ’n spesifieke skrywer oor ’n spesifieke onderwerp [is, wat] gereeld in ’n koerant of tydskrif verskyn (eie vertaling)”. Beide bogenoemde beskrywings plaas die manier waarop rubrieke aan die leser gelewer word, voorop. Dit is dus moontlik om te sê dat bogenoemde kenmerke van ’n rubriek, sterk invloed het op die vorm wat die rubriek aanneem. Hierdie karaktereenskappe, veral die eienskap dat dit gereeld en in dieselfde publikasie verskyn, skep ’n sterk verwagting by lesers en impliseer (of veroorsaak) ’n duidelike verhouding tussen rubriekskrywer en rubriekleser. Dit is dan ook nie vreemd dat die dinamiek tussen die partye in hierdie verhouding (rubriekskrywer en leser) deel uitmaak van die stylkenmerke wat deur populêre sowel as akademiese bronne voorgestel word nie.

4.5.2 Stylkenmerke

Rabe (2015:183-184) bespreek verskeie eienskappe wat tipies aan ’n rubriek is. Hoewel haar beskrywing binne ’n akademiese konteks geplaas is, toon dit sterk ooreenkomste met populêre bronne oor rubriekskryf (Kirsten, 2012:174-177 en Engelbrecht, 1981:9-12). Dit wil voorkom asof die stylkenmerke of eienskappe van goeie rubrieke in twee kategorieë verdeel kan word. Een van die kategorieë hou verband met die inhoud van ’n rubriek, terwyl die ander met die skrywer se persoonlikheid verband hou.

4.5.2.1 Inhoud

Rabe (2015:183-184) noem dat ’n goeie rubriek ’n verskeidenheid van onderwerpe moet dek. Kirsten (2012:275) sluit hierby aan deur te sê dat ’n rubriek afwisseling moet hê. Dit blyk dus dat ’n rubriek nie week na week dieselfde tipe inhoud moet hê nie, ten spyte van die feit dat spesifieke rubrieke deur ’n spesifieke skrywer ’n baie spesifieke identiteit het. Kirsten (2012:275) noem dat ’n rubriek komies, polities/aktueel, nostalgies of vakkundig kan wees. In Engelbrecht (1981:9) se lys van tipes rubrieke word rubrieke vir verskillende belangstellingsvelde genoem, onder meer taalrubrieke, kunsrubrieke, motorrubrieke,

godsdienrubrieke en politieke rubrieke, maar noem dat geselsrubrieke moontlik die gewildste tipe is.

'n Tweede vereiste wat die inhoud betref, is dat die inhoud die skrywer se eie opinie moet wees. Kirsten (2012:275) en Rabe (2015:183-184) maak beide van hierdie feit melding. Kirsten se afdeling oor rubrieke in *Die Afrikaanse Skryfgids* (Scheepers en Kleyn, 2012) se titel is: "Rubriek: Onthou dis jou opinie". Dit impliseer dat die rubriekskrywer nie noodwendig eers dieselfde posisie as die publikasie waarin sy/haar rubriek verskyn, hoef in te neem nie. Engelbrecht (1981:9) sê dat jou mening en die vryheid om dit in jou rubriek uit te spreek 'n belangrike element is waarmee die skrywerspersoonlikheid in 'n rubriek opgebou word. Hierdie skrywerspersoonlikheid lei dan tot die volgende stel stylkenmerke.

4.5.2.2 Skrywer se persoonlikheid

Omdat 'n rubriek gereeld deur dieselfde skrywer in dieselfde publikasie verskyn (sien definisie in afdeling 4.5.1), bestaan daar by lesers 'n verwagting om 'n sekere produk op 'n sekere tyd te kry. Dit beteken ook dat die skrywer mettertyd 'n bepaalde verhouding met sy/haar lesers bou en dat die skrywerspersoonlikheid wat geskep word, baie belangrik is. Die leser sien dus amper die rubriekskrywer as 'n vriend (Rabe, 2015:183-184).

Daar is verskeie eienskappe wat van 'n rubriekskrywer 'n goeie "vriend vir sy leserspubliek" maak. Engelbrecht (1981) sê dat dit belangrik is vir 'n rubriekskrywer om voortdurend aan sy/haar skrywerspersoonlikheid te bou. Rabe (2015:12-13) verwys na die feit dat die rubriekskrywer 'n definitiewe "stem" om persoonlikheid moet hê. Die skrywer se skrywerspersoonlikheid moet dus duidelik wees en die skrywer moet sy opinie duidelik kan artikuleer (Kirsten, 2012:275). Verder noem Rabe (2015:12-13) dat die rubriekskrywer 'n hele aantal belangstellings en opinies moet hê. Dit impliseer 'n belese rubriekskrywer met 'n wye verwysingsraamwerk wat 'n sy/haar opinie oor 'n wye verskeidenheid onderwerpe kan gee.

Hoewel 'n mens uit die bogenoemde stylkenmerke 'n idee kan kry van die styl wat 'n rubriek tipies sal aanneem, is die konkrete talige/grammatikale impak van hierdie kenmerke onduidelik. Die enigste van hierdie kenmerke wat op die oog af 'n spesifieke grammatikale vergestaltung mag hê, is die feit dat rubrieke in die eerstepersoon geskryf kan word, aangesien die skrywer sy eie mening uitdruk. Wat wel duidelik is uit bogenoemde stylkenmerke van 'n

rubriek is dat die moontlik duideliker uitkenbaar as 'n produk van 'n spesifieke skrywer mag wees as gevolg van die feit dat daar relatief min reëls of voorskrifte is wat die genre as't ware voorsê.

4.6 Genre en korpusstudies

4.6.1 Biber 1988 – Multidimensionele studie

Biber (1988) se studie *Variation across speech and language* het die weg gebaan vir studies wat ten doel het om veral variasie in taalgebruik tussen genres te bestudeer. Hoewel hy later genre-variastudies uit ander perspektiewe ook onderneem het (McEnery, Xiao & Tono., 2006:160), het die eerste sinkroniese studie oor linguistiese variasie tussen gesproke en geskrewe genres die grondslag gelê vir verdere studie oor die eienskappe en aard van linguistiese variasie.

Biber (1988) het 'n multidimensionele studie gedoen van gesproke en geskrewe genres¹³ ten einde te ondersoek tot watter mate die registers stilisties van mekaar verskil. In hierdie studie het Biber die Lanchaster-Oslo-Bergen-korpus (LOB-korpus) en die London-Lund-korpus gebruik om 23 verskillende geskrewe en gesproke registers te ondersoek. Hy het na die studie verwys as multidimensioneel omdat hy vyf dimensies geïdentifiseer het wat elk bestaan uit 'n stel linguistiese kenmerke (features) wat volgens hom tot variasie in taal bydra (McEnery *et al.*, 2006:165). Daar is in totaal 67 linguistiese eienskappe ondersoek. Elke genre is dan op 'n skaal binne elk van die dimensies geplaas ten einde die aard van die styl en variasie te beskryf. Hierdie dimensies is soos volg benoem:

- Betrokke teenoor informatiewe produksie (*Involved versus Informational Production*)
- Vertelstyl teenoor nie-vertelstyl (*Narrative versus Non-narrative Concerns*)
- Uitgebreide teenoor situasie-afhanklike verwysing (*Elaborated versus Situation-Dependent Reference*)
- Overt uitdrukking van oortuiging (*Overt Expression of Persuasion*)

¹³ Die term 'genres' word in Biber (1988) gebruik vir die verskeie tipes tekste wat in die studie gebruik is, hetsy gesproke of geskrewe.

- Abstrakte teenoor nie-abstrakte style (*Abstract versus Non-abstract style*)

Hy het verder aan elk van hierdie dimensies 'n aantal linguistiese eienskappe toegeken, wat geanaliseer sou word ten einde 'n spesifieke genre binne hierdie dimensie te kan plaas. Biber se argument was dat daar nie 'n enkele stilistiese eienskap is wat al die variasie tussen genres sou kon verduidelik nie en dat elke genre deur middel van sy linguistiese eienskappe op 'n sekere plek binne 'n dimensie geplaas kan word. 'n Uittreksel uit sy bevindinge word in die volgende tabel voorgestel. In die tabel 4.3 word dimensie 1, nl. Betrokke teenoor informatief voorgestel. 'n Skaal met eienskappe wat stilisties tussen betrokke (*Involved*) en informatief (*Informational*) lê, word onder dit voorgestel. In die volgende kolom word die linguistiese eienskappe wat hy aan die onderskeie stilistiese eienskappe toegeken het, voorgestel (McEnery *et al.*, 2006: 164-165). In die derde kolom word genres wat tipies 'n sekere dimensie goed verteenwoordig, aangedui.

Tabel 4-3: Uittreksel uit Biber 1998 se bevindinge ten opsigte van taalkundige eienskappe tussen genres

| Funksie | Linguistiese eienskappe | Verteenwoordigende genres |
|--|---|--|
| Dimensie 1: Betrokke teenoor informatief | | |
| 1. Betrokke (inter)persoonlike teksproduksie | 1. Eerste- en tweedepersoonsvoornaamwoorde | Gesprekke, persoonlike briewe, publieke gesprekke |
| 2. Interaktief | 2. Vrae | |
| 3. Persoonlike oogpunt | 3. Verskansers | Informatiewe skryfwerk bv. amptelike dokumente en akademiese skryfwerk |
| 4. Aanlyn produksie | 4. Bywoordelike bysinne | |
| 5. Informatief | 5. Byvoeglike naamwoorde | |
| 6. Noukeurige produksie | 6. Setselstukkies | |
| 7. Onpersoonlik | 7. Lang woorde | |

McEnery *et al.* (2006:165) noem die afleidings wat Biber uit hierdie studie gemaak het. Eerstens sê hulle dat daar nie net een dimensie ondersoek kan word om die volle variasie in ooreenkomste en verskille in tekste te beskryf nie en dat 'n multidimensionele analise daarom

meer doeltreffend is. Die benadering van Biber se studie en die multidimensionele faktoranalise het sedertdien die norm geraak in baie korpusgebaseerde variasiestudies.

Tweedens het Biber se data geen verskil aangedui tussen gesproke en geskrewe taalgebruik ten opsigte van die gekose veranderlikes in sy studie nie. Hy het bevind dat die gesproke en geskrewe genres wat hy ondersoek het, eenders is wanneer dit volgens die genoemde dimensies getoets is (McEnery *et al.*, 2006:165). Louwersonse *et al.* (2004a) het egter in 'n soortgelyke studie as Biber s'n later bevind dat daar bykomende linguistiese veranderlikes (buiten die 67 wat getoets is) wat variasie tussen gesproke en geskrewe genres kan aandui.

4.6.2 Louwersonse *et al.* - Die implikasie van kohesieskakels as linguistiese veranderlike

Soos dit uit die bogenoemde omvattende studie deur Biber (1988) duidelik is, is variasie van teksinterne eienskappe tussen genres op verskeie vlakke. Hoewel Biber se studie baie omvattend is – veral in terme van die verskeidenheid genres – is dit amper onmoontlik om al die moontlike linguistiese veranderlikes wat tot variasie kan bydra, te ondersoek. Soos Louwersonse *et al.*(2004:843) tereg opmerk, fokus die linguistiese eienskappe wat Biber gekies het om te ondersoek, hoofsaaklik op woordvlak, tydsaanduiding, werkwoordgebruik, leksikale kategorieë, ensovoorts.

Een van die linguistiese veranderlikes wat Biber nie ondersoek nie, is patrone in kohesieskakels. Daar is ook nie baie navorsing gedoen oor kohesieskakels as linguistiese veranderlike wat tot variasie tussen genres, tekstipes of teksproduseerders bydra nie. Teksanalitiese studies wat op kohesieskakels fokus, het dikwels te make met die kwaliteit (en tekstuur¹⁴) van 'n teks met betrekking tot ESL- en EFL-studies¹⁵. Een van die studies wat hierdie linguistiese veranderlike bestudeer, is *Variation in language and Cohesion across Written and Spoken Registers* (Louwersonse *et al.*, 2004).

Louwersonse *et al.* (2004) herhaal die studie van Biber (1988), maar voeg bykomende linguistiese veranderlikes, onder meer kohesie, by. Die dimensies en die linguistiese veranderlikes wat daarmee verband hou, is ook ondersoek, maar 'n opname in kohesiepatrone is bygevoeg.

¹⁴ Vergelyk afdeling 3.3

¹⁵ Studies wat die onderrig en leer van Engels as 'n tweedetaal (English as a Second Language) en Engels as 'n vreemdetaal (English as a Foreign Language) bestudeer

Hulle verduidelik in hulle gevolgtrekking dat, hoewel hulle Biber se studie grootliks herhaal het, hulle addisionele linguistiese veranderlikes ondersoek het wat nie, soos die meeste van die veranderlikes wat Biber gebruik het, op woordvlak voorkom nie, maar wat op bysin- en sinsvlak sowel as tussen paragrawe voorkom, soos kohesiepatrone. As gevolg van hierdie bykomende aspek van die analise, het hulle veral tussen gesproke en geskrewe genres groter variasie ten opsigte van die bykomende linguistiese veranderlikes gesien as wat Biber na aanleiding van sy studie gerapporteer het.

Hierdie resultate hou groot implikasies vir variasiestudies, wat voorheen grootliks op die bevindinge van Biber staatgemaak het, in. Dit beteken dat die keuse van die linguistiese veranderlike(s) wat ondersoek word, 'n groot invloed op die beskrywing van die variasie kan hê.

Tweedens is die metodologiese oorwegings in bogenoemde studie ook van belang vir die huidige studie, omdat daar so min studies is wat as voorbeeld kan dien vir die wetenskaplike ondersoek na kohesieskakels.

Louwerse *et al.* (2004:844) verduidelik dat kohesie (saam met koherensie) nie 'n linguistiese eienskap is wat op woordvlak lê nie, maar dat dit *tussen* bysinne, sinne en paragrawe lê (eie klem). Hulle wys uit dat linguistiese elemente wat op woordvlak lê, maklik deur rekenaarmatige prosesse geanaliseer kan word, maar dat die uitgebreide aard van kohesieskakels nie so eenvoudig is om deur 'n outomatiese proses raakgesien te word nie. Vir die doel om kohesiepatrone binne tekste te analiseer, het hulle webgebaseerde sagteware ontwikkel wat spesifiek gemik is op die identifisering en analise van kohesie in tekste. Hierdie sagteware staan bekend as *Coh-Matrix*. In 'n artikel wat verduidelik hoe die sagteware werk, verduidelik Graesser *et al.* (2004) die basiese konstrakte wat deur die *Coh-Matrix* ondersoek kan word. Hulle verduidelik dat die *Coh-Matrix* meer as 50 tipes kohesieverhoudinge en meer as 200 ander linguistiese eienskappe toets en analiseer.

Dit is van belang om te verstaan dat outomatiese analisemetodes op voorafverwerkte data staatmaak om klassifikasies te motiveer. Vir 'n linguistiese eienskap soos kohesie om suksesvol deur sagteware geanaliseer te word, moet daar 'n groot hoeveelheid metadata ingevoer word sodat die teks geanaliseer kan word. Woorde wat byvoorbeeld sekere verhoudings aandui of 'n grammatikale verskynsel soos 'n tipe bysin of sintaktiese patroon aandui, moet vooraf as sulks

gemerk word, sodat die sagteware die verskynsel kan verstaan. Verder moet woordverbande soos sinonieme, hiponieme, antonieme, ensovoorts, vooraf geprogrammeer word. Sekere leksikale kohesiemiddele, soos kollokatiewe verhoudings, is bykans onmoontlik om vooraf te programmeer omdat die verhoudings baie konteksgebonde is en in die sosiale bewustheid van die leser bestaan. “Mes” en “vurk” kan byvoorbeeld semanties aan mekaar verbind wees, so ook “mes” en “wapen”. As gevolg van hierdie amper oneindige moontlikhede vir verbindings, word kollokatiewe kohesie baie keer weggelaat in analysesagteware. Verder kan sagteware wat vir hierdie doel geprogrammeer is, nie in ’n ander taal analyses doen nie, omdat die inligting wat gebruik word om die analyses te maak, taal-spesifiek is.

Behalwe dat die data vir hierdie studie in Afrikaans is en die *Coh-Matrix* nie die teks sal kan ontleed nie, stel die literatuur oor outeurskapstudies soos dit in hierdie studie relevant is (vergelyk afdeling 2.3) ook ’n integrasie tussen ’n kwantitatiewe en kwalitatiewe studie voor. Die outeurs van die *Coh-Matrix* het die sagteware tot dusver grootliks gebruik vir studies in ESL en EFL. Die doel van die sagteware is dus nie om beskrywingsmoontlikhede vir variasie in *persoonlike styl* in taalgebruik te gee nie, maar om ’n hulpbron te wees om variasie in die taalgebruik van tweede- of vreemdetaalsprekers van Engels te beskryf met die duidelike doel om die data te gebruik vir onderrig of om variasie in genre te beskryf. In hierdie studie word gepoog om die gebruik van kohesiemiddele op so ’n manier te ontleed dat dit beskrywingsmoontlikhede vir ’n outeur se styl gee. Die patroon waarin die middele gebruik word, is dus van belang en nie net die hoeveelheid of ratio van die voorkoms daarvan nie. Waar die fokus van die *Coh-Matrix* dus val op die voorkoms van hierdie linguistiese element, val die fokus in hierdie studie op die manier waarop kohesieskakels deur ’n individu gebruik word.

4.7 Samevatting

In hierdie hoofstuk is benaderings ten opsigte van genre en die relevansie daarvan vir hierdie studie uitgelig. Daar is eerstens op die oorvleueling in gebruik van die terme genre en tekstipe gewys, waarna die wesensaard van kortverhale en rubrieke ondersoek is. Daar is genoem dat beide hierdie genres stylaanpassings by ’n teksproduseerder sou kon meebring. Om mee af te sluit is kortliks na Biber (1988) se korpusgebaseerde studie ten opsigte van variasie tussen genres verwys, aangesien dit relevant is vir die benadering tot genre wat in hierdie studie geneem word.

In hoofstuk 5 word die metodologiese benadering aangebied, wat gebruik is om die navorsingsvrae in hierdie studie te beantwoord. Temas wat onder meer in hierdie hoofstuk na vore kom, is die metodologiese benadering van 'n korpusgebaseerde ondersoek, die navorsingsontwerp en analiseproses van die huidige studie asook die statistiese analisemetodes wat gevolg is om data te analiseer.

5 Navorsingsmetodologie

5.1 Inleiding

In die voorafgaande hoofstukke is 'n literatuuoroorsig verskaf wat poog om die breë teorieë en denkraamwerke van die relevante studievervelde uiteen te sit. Daar is bevind dat teorieë in die veld van outeurskapidentifikasie wyd aanvaar word, maar dat die beste metodologiese benaderings binne die veld nog onder dispuut is. Daar word dus wyd aanvaar dat meer empiriese navorsing nodig is om teorieë en metodologie te toets en te verfyn. Die doel van hierdie studie is om vas te stel of patrone in die gebruik van kohesiemerkers eie is aan 'n spesifieke outeur en indien wel, tot watter mate. Deur sommige van die veranderlikes wat in outeurskapstudies voorkom konstant te hou en met die identiteit van die skrywers bekend, kan 'n sekere aspek van skryfwerk (kohesiemerkers) geïsoleer word om sodoende die gebruik daarvan te ondersoek.

'n Korpusgebaseerde benadering word in hierdie studie gevolg ten einde die navorsingsdoelwitte te bereik. Om mee te begin word 'n kort oorsig oor korpuslinguistiek as metodologiese benadering gegee, waarna die korpusseleksie en die analiseproses sowel as die statistiese verwerkingsmetodes bespreek word.

5.2 'n Korpusgebaseerde benadering: oorsig

'n Korpusgebaseerde benadering is 'n benadering wat gebruik kan word om spesifieke navorsingsvrae te aan te pak. Hoewel daar 'n wye verskeidenheid linguistiese studies is waar 'n korpusgebaseerde benadering die navorsingsdoelwitte sou kon behaal, is dit belangrik om te noem dat dit nie die enigste benadering tot hierdie doelwitte is nie. Die eienskappe en toepassingsmoontlikhede van hierdie benadering (wat in afdeling 5.2 en 5.3 bespreek word), tesame met die vereistes van die navorsingsvrae, beïnvloed die gepastheid van hierdie metodologiese raamwerk vir hierdie studie.

Biber, Conrad en Reppen (1998:4) noem 'n aantal karaktereenskappe van 'n korpusgebaseerde benadering. Die korpusgebaseerde benadering is eerstens empiries van aard en analiseer patrone in natuurlike tekste. Empiriese navorsing is gebaseer op die veronderstelling dat kennis

verwerf kan word deur ervaring en waarneming van 'n verskynsel (Babbie & Mouton, 1998:27).

In hierdie studie word waarnemings ten opsigte van tekste gemaak en patrone ten opsigte van skryfstyl in ooreenstemming met die navorsingsdoelwitte gesoek. Tweedens word die korpusgebaseerde benadering gebruik wanneer tekste wat op 'n natuurlike wyse geskep is, geanaliseer word (Biber *et al.*, 1998:4). Tekste wat op 'n natuurlike wyse geskep word, is tekste wat deur 'n taalgebruiker geskep is, en nie deur 'n navorser ter illustrasie van teorie nie. Behalwe vir die feit dat hierdie tipe studies van natuurlike taalgebruik as voorbeelde gebruik maak, word daar ook dikwels 'n korpusgebaseerde benadering gevolg wanneer daar 'n groot hoeveelheid van hierdie tekste is wat ondersoek moet word. Deur van 'n korpusgebaseerde benadering gebruik te maak, kan patrone, reëlmatighede en onreëlmatighede in tekste geïdentifiseer word, wat analise deur die navorser moontlik maak. Derdens maak die korpusgebaseerde benadering grootliks van rekenars en doelontwerpte sagteware gebruik om spesifieke verskynsels te soek, te groepeer, ensovoorts (Biber *et al.*, 1998:4-5). Die sagteware bied outomatiese en interaktiewe moontlikhede vir die onttrekking en analise van linguistiese data uit die datastel. Hierdie outomatiese onttrekking van data uit die datastel, lei tot die laaste karaktereienskap wat Biber bespreek, naamlik dat hierdie benadering op beide kwalitatiewe en kwantitatiewe metodes staatmaak om data te analiseer en gevolgtrekkings of afleidings te maak. (Biber *et al.*, 1998:4). Kwalitatiewe data-analise word gebruik om verskynsels te merk of te annoteer en te tel, waarna die navorser na hierdie verskynsels kyk en probeer om onderliggende patrone raak te sien. Biber *et al.* (1998:15) stel dit duidelik wanneer hulle beklemtoon dat korpusgebaseerde analise nie slegs die tel van linguistiese uniekhede behels nie, maar dat “kwalitatiewe, funksionele interpretasies” onontbeerlik is (eie vertaling).

McEnery, Xiao en Tono (2006:120-121) noem vier uitkomstes wat *nie* met die hulp van 'n korpusgebaseerde benadering bereik kan word nie. Hulle sê eerstens dat die *korpusgebaseerde* analise van data nie 'n aanduiding kan gee van wat moontlik en onmoontlik in terme van taalkundige konstruksies is nie. Aangesien hierdie tipe studie probeer aanteken wat wel in tekste voorkom wat natuurlik tot stand gekom het, kan dit slegs “rapporteer” watter verskynsel(s) daar wel is en nie oor die normatiewe geldigheid daarvan nie. Die tweede uitkomstes wat nie deur die gebruik van 'n korpusgebaseerde analise bereik kan word nie, is die verklarende aspek van die ondersoek. Korpora kan dus nie gebruik word om vas te stel hoekom 'n teks 'n spesifieke

struktuur aanneem nie. Dit kan dus resultate lewer, maar nie redes vir moontlike linguistiese patrone of variasie gee nie. 'n Derde beperking wat McEnery *et al.* (2006:121) noem, is dat die keuse om 'n korpusgebaseerde studie te onderneem sekere tipes metodologiese oplossings meebring en dat dit dus by die navorsingsvrae moet pas. Hoewel dit nie 'n beperking is wat eie aan 'n korpusgebaseerde benadering is nie – alle metodologiese benaderings het tekortkominge – is hulle van mening dat dit belangrik is om navorsingsvrae so te formuleer dat dit moontlik is om die vrae te beantwoord deur 'n korpusgebaseerde metode te volg. Laastens noem hulle dat navorsers in gedagte moet hou dat afleidings wat na aanleiding van 'n korpusgebaseerde studie gemaak word, net op die spesifieke korpus van toepassing is. Hulle verduidelik dat dit moontlik is om veralgemenings na aanleiding van 'n verteenwoordigende korpus te maak, maar dat navorsers moet waak teen ongegronde veralgemenings (McEnery *et al.*, 2006:121).

Die teorie oor korpusgebaseerde navorsing bied 'n raamwerk wat dataverwerkingsmoontlikhede skep vir taalkundige studies wat ten doel het om linguistiese patrone en hulle ekstralinguistiese oorsake/gevolge te ondersoek (McEnery *et al.*, 2006). Hierdie raamwerk kan op verskeie maniere in die praktyk gebruik word. Die toepassingsmoontlikhede van hierdie benadering kan gesien word in die wyse waarop navorsers oor die jare hierdie raamwerk gebruik het met 'n verskeidenheid van veranderlikes vir 'n verskeidenheid van navorsingsvrae. 'n Korpusgebaseerde benadering is al in 'n verskeidenheid studies gebruik, onder meer studies oor grammatikapatrone, registerwisseling, taalverskeidenheid, vertaling, diskoersanalise, stilistiek en forensiese linguistiek (McEnery *et al.*, 2006:123-203). Omdat hierdie studie op spesifieke taalgebruikers (in hierdie geval skrywers) se skryfstyl binne twee gekose genres fokus, het dit sterk raakpunte met forensiese linguistiek, stilistiek, registerwisseling en genre-analise.

5.3 Tipes navorsing wat met 'n korpusgebaseerde metode gedoen word

In hulle standaardwerk, *Corpus linguistics: Investigating Language Structure and Use*, verduidelik Biber *et al.* (1998:1-3) dat linguistiese studies tradisioneel op een van twee areas fokus. Dit kan op taalkundige struktuur fokus (morfeme, woorde, sinskonstruksie, ensovoorts) of op die wyse waarop taal gebruik word. Die fokus van sy boek, en ook hierdie studie, val op hoe taal gebruik word, met ander woorde, hoe taalgebruikers die taalkundige hulpbronne tot hulle beskikking gebruik (en soms misbruik).

Volgens Biber *et al.* (1998:5) is daar twee hoofdoelwitte wat 'n studie met die klem op taalgebruik gewoonlik kan hê. Eerstens kan so 'n studie ten doel hê om 'n patroon in taalgebruik te vind en die “uniekheid” of “duidelikheid” of “opvallendheid” van die patroon te bepaal, beskryf en/of evalueer. Tweedens kan die ekstralinguistiese faktore wat 'n invloed mag hê op enige variasie in taalgebruik, ondersoek word. In hierdie studie word gepoog om te bepaal hoe duidelik 'n sekere linguistiese patroon (gebruik van kohesieskakels) na vore kom wanneer buitetalige faktore as veranderlikes gebruik word.

5.4 Navorsingsvrae

In hoofstuk 1 is daar twee hoofdoelwitte van die studie geïdentifiseer, naamlik:

- om te bepaal of spesifieke kohesieskakels patroonmatig met bepaalde stilistiese implikasies verbind kan word;
- om te bepaal in welke mate 'n individu se stilistiese voorkeur in terme van kohesiemerkers deur die reëls van 'n betrokke genre beïnvloed word.

Die primêre navorsingsvraag wat uit bogenoemde doelwitte voortvloei, kan soos volg geïdentifiseer word:

- Kan die gebruik van kohesieskakels in die werk van Dana Snyman en André le Roux 'n aanduiding van outeurskap gee?

Die sekondêre navorsingsvrae wat aandag sal geniet, is dus die volgende:

- Beïnvloed die stilistiese genrekonvensies van rubrieke en kortverhale Dana Snyman en André le Roux se gebruik van kohesieskakels?
- Is daar algemene kohesiepatrone wat as kenmerkend van onderskeidelik rubrieke en kortverhale beskou kan word?

Indien die oorsig in afdeling 5.2 en 5.3 in gedagte gehou word, is dit duidelik dat 'n korpusgebaseerde benadering 'n gepaste teoretiese raamwerk is om die bogenoemde navorsingsvrae te kan beantwoord. Volgens afdeling 5.2 kan daar gesoek word na patrone in tekste wat op 'n natuurlike wyse geskep is, wanneer gebruik gemaak word van 'n

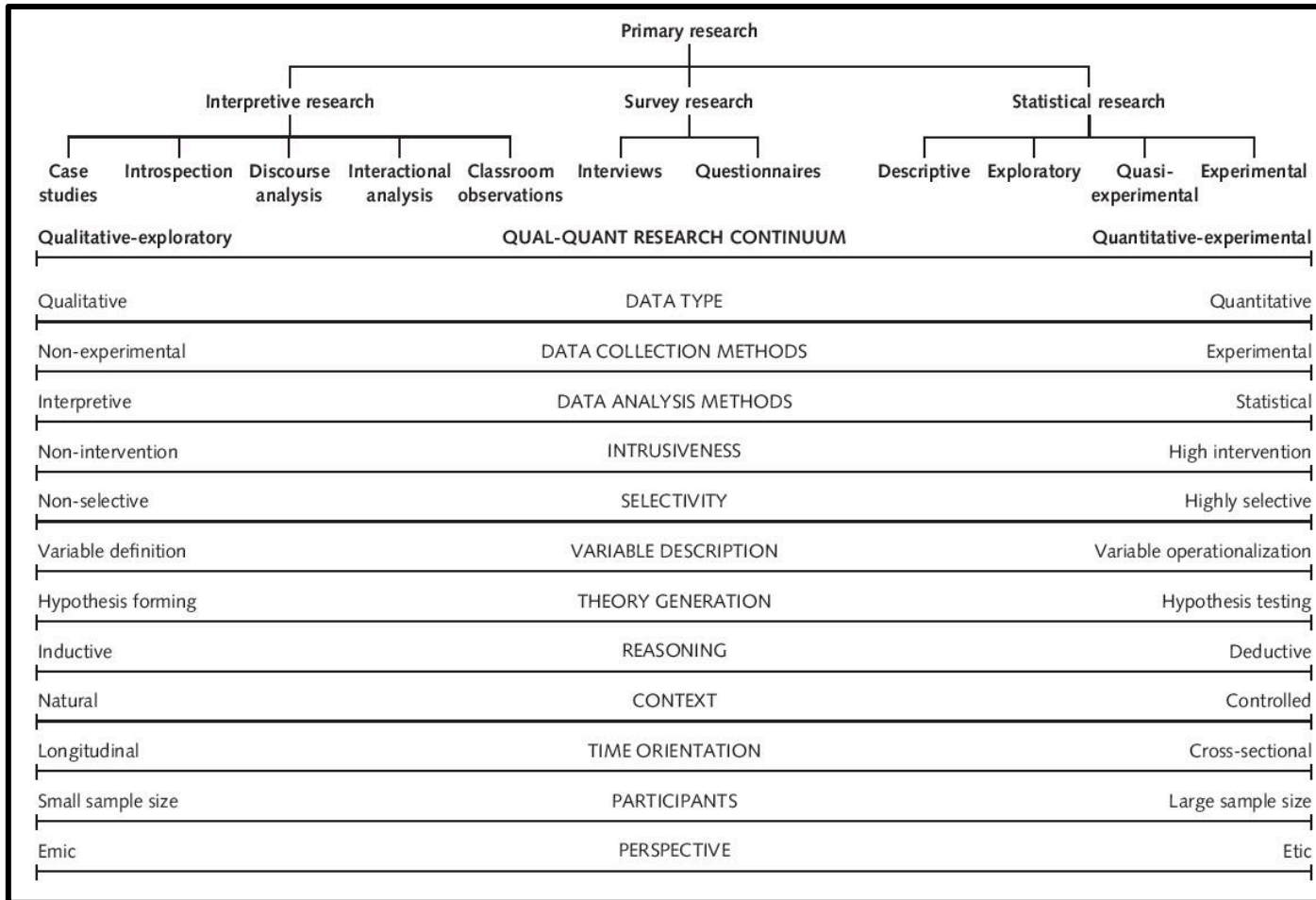
korpusgebaseerde benadering. Rekenaarsagteware kan ingespan word om hierdie waarnemings op te teken en te analiseer. Hierdie eienskappe kan duidelik nuttig wees wanneer bogenoemde navorsingvrae ondersoek word.

5.5 Navorsingsontwerp

In hierdie studie word korpuslinguistiek as metodologiese benadering tot teksanalise beskou. Volgens Meintjes (2014:138) toon hierdie benadering sterk ooreenkomste met die nie-eksperimentele, eenmalige dwarsnitopname-ontwerp wat deskriptief of beskrywend van aard is en “...gepaard [gaan] met die weergee van data-frekwensies en/of gedetailleerde beskrywing/-s”. In die huidige studie word op die mees elementêre vlak gepoog om die styl van skrywers te beskryf¹⁶.

Brown beskryf in sy hoofstuk oor navorsingsmetodologieë in die *The Handbook of Applied Linguistics* (2004:476-500) die uiteenlopende aard van taalkundige navorsing en sê dat die tradisionele beskrywing van navorsing, as byvoorbeeld óf kwalitatief óf kwantitatief, óf sinchronies óf diachronies, ensovoorts, onvoldoende is om taalkundige navorsing te beskryf. Hy verduidelik verder dat hierdie tradisionele beskrywings, hoewel in beginsel akkuraat, dalk eerder op ’n kontinuum tussen kwalitatiewe en kwantitatiewe navorsing behoort te lê. Hy noem twaalf sogenaamde karaktereenskappe van navorsing, alhoewel dit dalk meer akkuraat beskryf kan word as elemente van of veranderlikes in die navorsingsontwerp. Hy sê dat elk van hierdie eienskappe ook op ’n kontinuum is en dat die een kant van die kontinuum telkens in studies gebruik word wat kwalitatief van aard is en die ander kant, kwantitatief. Hy verduidelik dat alle toegepaste linguistiese studies in terme van elk van hierdie navorsingselemente iewers op die kontinuum geplaas kan word, en dat dit die navorsingsontwerp meer akkuraat kan beskryf as die tradisionele óf...óf-beskrywings (vergelyk figuur 5-1 uit Brown, 2004:490). In terme van hierdie uiteensetting, kan hierdie studie geklassifiseer word as statistiese navorsing wat beskrywend van aard is en dus aan die kwantitatiewe kant van die kontinuum lê.

¹⁶ Hoewel die tekste oor ’n tydperk van ongeveer 11 jaar verskyn het, is die doelwitte van hierdie studie nie om verandering oor hierdie tyd te beskryf nie, maar slegs die aard van die taalvariasie.



Figuur 5-1: Brown (2004) se voorstelling van die uiteenlopende aard van taalkundige navorsing

Enkele kwalitatiewe elemente kan egter verwag word aangesien dit nie suiwer op die kwalitatiewe kant van die kontinuum lê nie. Uit hierdie stel veranderlikes is daar veral drie elemente wat 'n groot invloed het op die aard van hierdie studie, naamlik die aard van redenering (induktief of deduktief; vergelyk afdeling 5.5.1), die tipe data (vergeelyk afdeling 5.5.2) en die metode van data-analise (vergeelyk afdeling 5.5.3). Die bespreking van hierdie studie se navorsingsontwerp en -metodologie word vervolgens aan die hand van hierdie drie elemente gedoen word.

5.5.1 Aard van redenering

Uit die literatuuroorsig het dit duidelik geword dat sekere aspekte van die teorieë wat in outeurskapstudies gebruik word nog vaag of onduidelik is, of nog nie deur empiriese data bewys is nie. Hierdie studie gebruik dus die teoretiese raamwerke van outeurskapstudies om die geldigheid daarvan in spesifieke omstandighede te ondersoek. Daar word sekere aannames gemaak op grond van die teorie in verband met outeurskapstudies en op grond van hierdie aannames word navorsingsvrae ontwikkel. Mouton en Marais (1990:105) verduidelik dat 'n induktiewe benadering juis op só 'n manier werk. Hulle sê dat die navorser in sulke studies "...algemene hipoteses of vermoedens (het) wat die navorsing breedweg lei" (Mouton & Marais, 1990:105). Die aannames wat hierdie studie lei, volg logies op mekaar. Onder die aannames en hipoteses wat hierdie studie lei, is dat:

1. 'n persoon se skryfstyl uitkenbaar is op grond van unieke tekstuele kenmerke,
2. 'n wye verskeidenheid tekstuele kenmerke 'n sekere skryfstyl kan karakteriseer,
3. kohesieskakels een van die tekstuele kenmerke is wat skryfstyl kan karakteriseer,
4. elke skrywer moontlik kohesieskakels op sy/haar unieke manier gebruik, en
5. 'n skrywer se skryfstyl herkenbaar kan wees op grond van sy/haar gebruik van kohesieskakels.

In terme van Brown (2008:280) se uiteensetting van die aard van 'n navorsingsontwerp in linguistiese studies, lê 'n induktiewe benadering soos die een wat in hierdie studie gevolg word, aan die kant van die kontinuum wat normaalweg met kwalitatiewe studies geassosieer word.

5.5.2 Tipe data

Die data wat in hierdie studie gebruik word, is kwalitatiewe data omdat dit uit tekste, meer spesifiek rubrieke en kortverhale, bestaan. Omdat dit 'n korpusgebaseerde studie is, bestaan die korpus uit 'n versameling tekste wat tydens die navorsing geanaliseer is. McEnery *et al.* verduidelik dat 'n korpus ontwerp moet word om 'n spesifieke taal of taalvariasie te verteenwoordig (2006:13). Hulle gaan voort deur te sê dat daar in breë terme twee tipes korpora geïdentifiseer kan word, naamlik algemene en gespesialiseerde korpusse. 'n Algemene korpus is ontwerp om “..as 'n basis [te dien] vir 'n beskrywing van 'n taal of taalvariëteit in geheel” (McEnery, 2006:15). In hierdie studie word 'n gespesialiseerde korpus egter gebruik omdat dit saamgestel is uit tekste van 'n spesifieke skrywer en spesifieke genres met die doel om slegs die taalgebruik van daardie skrywer te beskryf.

Ten spyte van die feit dat die data kwalitatief van aard is, is die navorsingsmetodologie daarop gemik om verskynsels in die kwalitatiewe data te kwantifiseer. Hierdie analiseproses word volledig in afdeling 5.6 beskryf. Dit is egter sinvol om op hierdie stadium te noem dat die rou data van die korpus gekodeer is om 'n stel kwantitatiewe metadata te skep¹⁷, wat uiteindelik statisties ontleed is.

5.5.2.1 Keuse van datastel

Vir hierdie studie is daar besluit om rubrieke en kortverhale van die populêre skrywers André le Roux en Dana Snyman te kies¹⁸. Hoewel daar verskeie redes vir taalvariasie is (vergelyk

¹⁷ In hierdie konteks, bestaan die metadata (1) uit inligting in terme van die posisie van die betrokke datapunt binne die teks waarin dit oorspronklik voorkom, en (2) uit inligting wat die kohesieskakel van die betrokke datapunt aandui.

¹⁸ Die kortverhale en rubrieke wat as datastel gebruik is, was en is beskikbaar in die publieke domein en in terme van die kopieregwet 89 van 1978 geld die volgende beginsel: “*Fair use provisions: Sections 12 of the Copyright Act 98 of 1978. Section 12(1) of Copyright Act 98 of 1978 provides for general exceptions to the protection of literary and musical works, i.e. “section. 12(1) Copyright shall not be infringed by any fair dealing with a literary or musical work - for the purpose of research or private study by, or the personal or private use of, the person using the work;” Section 12 (4) of the Copyright Act provides that “The copyright in a literary or musical work shall not be infringed by using such work, to the extent justified by the purpose, by way of illustration in any publication, broadcast or sound or visual record for teaching: Provided that such use shall be compatible with fair practice and that the source shall be mentioned, as well as the name of the author if it appears on the work.”*

afdeling 2.3.5.3) val die fokus in hierdie studie nie op die redes vir variasie tussen skrywers se skryfstyl nie, maar die op die aard van die variasie. Dit is dus nie belangrik om veranderlikes wat met die skrywer verband hou (soos geslag, ouderdom, opleiding, ensovoorts) in ag te neem wanneer die skrywers vir hierdie studie gekies is nie. Hoewel dit faktore is wat bydra tot stylvariasie, hoef dit nie konstant te wees ten einde die verskil in skryfstyl van skrywers te ondersoek nie. Indien die stylvoorkeure van die skrywer die veranderlike is, moet faktore wat ’n invloed het op die aard van die teks egter so konstant as moontlik wees.

Nadat die kortverhaalt tekste verkry is, is daar gepoog om rubriek tekste te vind wat so naby as moontlik aan die publikasiedatums van die kortverhaalbundels in vergelykbare publikasies (Afrikaanse koerante) verskyn het. Uiteindelik is rubrieke van André le Roux wat tussen 2000 en 2014 in die *Beeld*, *Die Volksblad* en *Die Burger* verskyn het en rubrieke van Dana Snyman wat tussen 2006 en 2012 in dieselfde twee koerante verskyn het, gekies. .

Die kortverhaalt tekste is vanuit die bundels ingeskandeer en die rubriek tekste is van die argiewe van die betrokke koerante in txt-formaat verkry.

5.5.2.1.1 *Beskrywing van die korpus*

Die korpus het bestaan uit ’n totaal van 40 tekste; 20 tekste per skrywer en 10 tekste per skrywer per genre. Die totale aantal lopende woorde in die korpus is 31 658. Die verdeling van hierdie woorde kan gesien word in tabel 5-1: Woordhoeveelhede per korpus. André le Roux se kortverhale bevat 8 977 woorde en sy rubrieke 7 249. Dana Snyman se kortverhale bevat 7 991 lopende woorde en sy rubrieke 7 441.

Tabel 5-1: Woordhoeveelhede per korpus

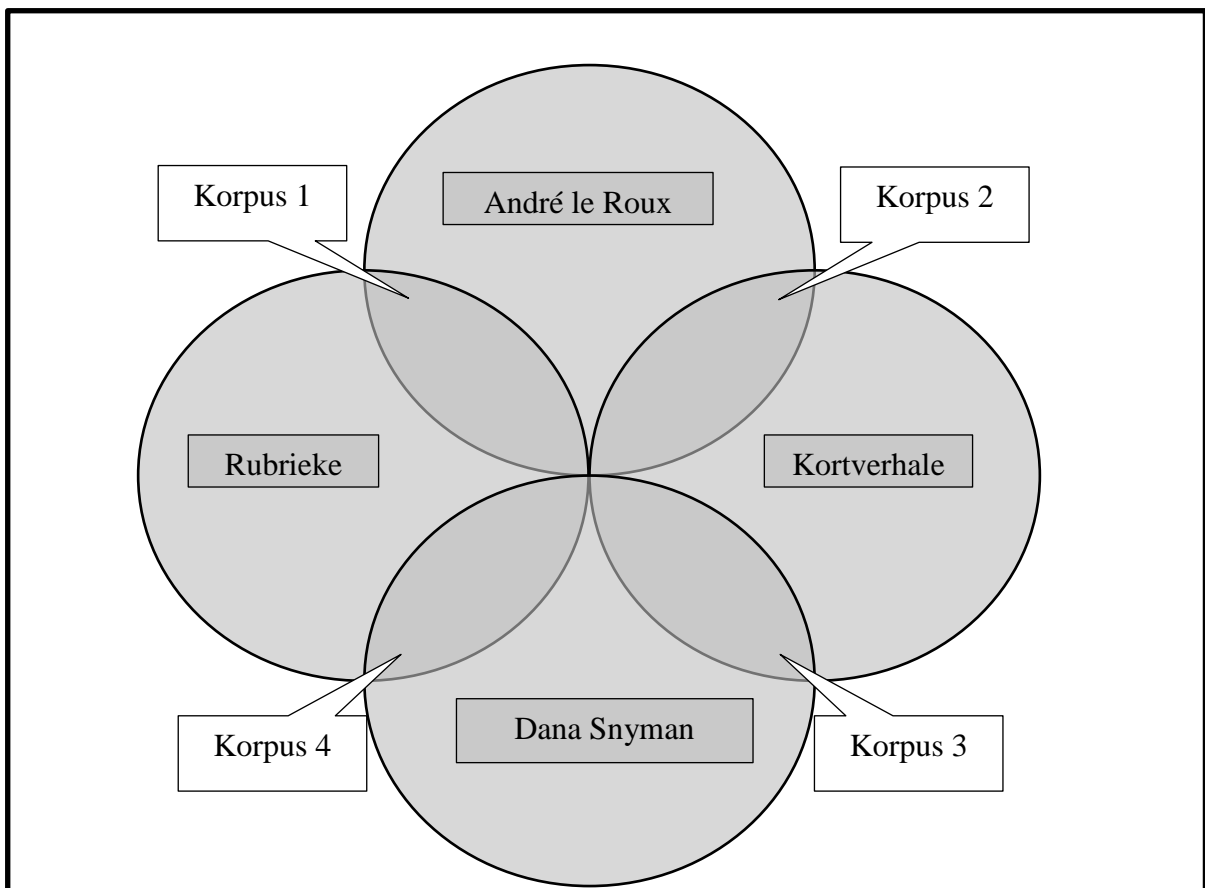
| Skrywer | Genre | Hoeveelheid woorde |
|------------------|-------------|--------------------|
| André le Roux | Kortverhale | 8 977 |
| | Rubrieke | 7 249 |
| Subtotaal | | 16 226 |

This publication complies with all the requirements for ‘fair use’, i.e. there is no commercial gain, sources are correctly cited, and the use of images is for illustration purposes only.”

| | | |
|------------------|-------------|--------|
| Dana Snyman | Kortverhale | 7 991 |
| | Rubrieke | 7 441 |
| Subtotaal | | 15 432 |
| Totaal | | 31 658 |

5.5.2.1.2 Korpusseleksie

Ten einde die navorsingsvrae te beantwoord, sal die vier korpusse met mekaar vergelyk word. Hierdie korpusse bestaan uit twee genres (kortverhalwe en rubrieke) van twee skrywers. Die veronderstelling kan gemaak word dat daar 'n aantal korrelasies is wat uit die aard van die korpusse gemaak kan word. Hierdie korrelasies kan met behulp van 'n grafiese voorstelling verduidelik word (vergeelyk figuur 5-2).



Figuur 5-2: Grafiese voorstelling van korpusseleksie

Hoewel die primêre navorsingsvraag op outeurstyl gefokus is, dien die gebruik van twee genres vir elke skrywer as 'n kontrole sowel as 'n bykomende veranderlike waarvolgens die gebruik van kohesiemerkers ondersoek kan word. Korrelasies in die gebruik van kohesiemerkers tussen korpus 1 en 2 (dieselfde skrywer, verskillende genres), kan dui op 'n moontlike stylvoorkeur in die gebruik van kohesiemerkers vir daardie spesifieke outeur *ongeag* die genre. Dieselfde geld vir 'n korrelasie tussen korpus 3 en korpus 4. Op hierdie manier dien die analise van twee genres as 'n metode om stylkenmerke wat dalk eie aan 'n spesifieke genre is en nie aan 'n skrywer nie, as moontlike stylkenmerke van 'n skrywer uit te skakel. Korrelasies in die gebruik van kohesiemerkers in korpus 2 en 3 en korpus 1 en 4 kan egter dui op stylkenmerke van die betrokke genre, ongeag die skrywer. Hoewel dit nie bydra tot die primêre navorsingsvraag in verband met outeurstyl nie, kan dit bydra tot beskrywingsmoontlikhede vir die styl of aard van die betrokke genre.

5.5.3 Metode van data-analise

Op Brown (2004:490) se kontinuum wat analisemetodes aandui (vergelyk figuur 5-1), verteenwoordig metodes wat interpretasie verg die kwalitatiewe aspek van data-analise en metodes wat statistiese metodes verg die kwantitatiewe aspek daarvan. Soos reeds genoem is (vergelyk afdeling 5.5.2), word daar in hierdie studie op 'n elementêre vlak gepoog om 'n kwalitatiewe eienskap van tekste te kwantifiseer sodat dit statisties meetbaar is. Die data-analise geskied dus in fases waartydens daar aanvanklik van interpretatiewe analisemetodes gebruik gemaak word en die metadata daarna kwantitatief geanaliseer word.

5.6 Analiseproses

Die doel van die aanvanklike analiseproses was om al die gevalle van kohesie (vergelyk afdeling 3.5) te dokumenteer sodat die voorkoms van hierdie skakels uiteindelik ondersoek kan word en patrone wat duidelik mag wees, statisties weergegee kan word. Om egter die voorkoms van kohesieskakels te dokumenteer en uiteindelik te analiseer, moet die skakels geïsoleer word omdat die voorkoms van die *skakel* en nie die voorkoms van 'n spesifieke leksikale item ondersoek word nie. Die analiseproses was dus daarop gemik om hierdie kohesieskakels te identifiseer en dan te isoleer ten einde die voorkoms daarvan te kan analiseer.

In hierdie studie is die analise in drie fases gedoen, waarvan twee interpretasie van die navorser geverg het en een rekenaarmatig was. Eers is die teks met die hand geëtiketteer waarna dit rekenaarmatig op verskeie maniere gesorteer is, om weer geklassifiseer te word. Hierdie klassifikasie tesame met die data rakende die posisie van die betrokke merker ten opsigte van die teks waarin dit voorkom, is aan statistiese ontleding onderwerp om vas te stel of daar beduidende verskille in die patroon/patrone of kombinasies is in die tekste waarin spesifieke kohesiemerkers voorkom.

5.6.1 Fase 1: Etiketterming

Eerstens is die data met die hand geëtiketteer. Dit is gedoen met die oog daarop om leksikale items te merk wat kohesief aan mekaar verbind is, ongeag die kategorie van kohesieskakel wat gebruik is. Die antesedent is telkens gemerk en alle daaropvolgende (of voorafgaande in die geval van kataforiese verwysing) items wat na die bepaalde antesedent verwys is met dieselfde etiket gemerk. 'n Voorbeeld hiervan kan in figuur 5-3 gesien word. In die voorbeeld kom 'Mev. Dos Santos' aan die begin van die teks voor met die etiket '<mev>' daarna. Alle leksikale items wat daarna voorkom en op 'n manier kohesief aan 'Mev. Dos Santos' verbind is, is dan met die etiket '<mev>' gemerk. In hierdie uittreksel, sluit dit die woorde 'sy', 'haar', 'Sy' en 'haar' in. Hierdie proses is vir elke etiket in al die tekste herhaal.

Mev. Dalva dos Santos <mev> <dossantos> bestudeer haar <mev> man <man> se kredietkaartrekening <krediet> en wha! kom sy <mev> onverwags af op 'n bedrag <krediet> betaalbaar <krediet> aan: House of Pleasure <house>. Sy <mev> weet dadelik haar <mev> man <man> is in groot moeilikheid <moeilikheid>. (Hy <man> weet dit <moeilikheid> nog nie.)

Figuur 5-3: Voorbeeld van etikettering van data

Die rede vir hierdie aanvanklike etikettering is tweeledig. Eerstens hoef die navorser op hierdie stadium slegs een analitiese besluit te neem, naamlik of die item op een of ander wyse kohesief aan 'n ander verbind is. Die besluit oor die wyse van verbinding word dus vir 'n latere stadium gelos. Tweedens bied dit ruimte vir differensiasie wanneer die metadata uiteindelik geanaliseer word. Etiketterming kan in 'n later stadium as 'n antesedent gemerk word, terwyl die res van die leksikale items wat met daardie antesedent skakel, elk as die gepaste kohesieskakel gemerk

kan word. In hierdie fase van die data-ontleding is voorlopige analises met studieleiers bespreek ten einde betroubaarheid te bevorder.

5.6.2 Fase 2: WordSmith-verwerking

In hierdie studie is die rekenaarprogrammatuur Oxford WordSmith Tools as navorsingsinstrument gebruik. WordSmith Tools is rekenaarsagteware wat dit moontlik maak om die voorkoms van spesifieke leksikale items of etikette *in hulle tekstuele konteks* te ondersoek. Dit bied moontlikhede om frekwensie op te spoor en kan konkordansies maak. In hierdie studie is die etikette wat in afdeling 5.6.1 bespreek is, in die oorspronklike tekste aangebring sodat daar in WordSmith sekere groeperings en konkordansies gemaak kan word op grond van die etikette en nie slegs die leksikale items in die tekste nie. Dit sou nie moontlik wees om hierdie groeperings en konkordansies rekenaarmatig uit die rou data te maak nie.

Vir hierdie fase van die analise is daar van die Concord-funksie in WordSmith gebruik gemaak. Hierdie funksie laat die gebruiker toe om al die voorkomste van enige woord, frase of etiket binne die konteks van die teks te sien. Die gebruiker voer die element as 'n soekfrase in en die voorkoms daarvan word binne die onmiddellike konteks gegee (vergelyk figuur 5-4). Daar kan hierbenewens ook 'n hele aantal soekfrases ingesleutel word waartydens die voorkoms van al die betrokke soekfrases deur WordSmith gewys kan word.

| N | Concordance | Set | Tag | Word # | Sent. # | Sent. Pos. | Para. # | Para. Pos. | Hea | Hea Sec | Sec | File | Date | % |
|-------|---|-----|-----|--------|---------|------------|---------|------------|-----|---------|-------------------|-------|--------|-----|
| 1 081 | , en met 'n vrou soos mev. Dos Santos <mev> so jy in die eerste plek nie | | | 322 | 31 | 21% | 0 | 36% | 0 | 36 | Tagged - 07/2016/ | 2016/ | Feb/23 | 36% |
| 1 082 | 'n vrou soos mev. Dos Santos sou y <man> in die eerste plek nie per | | | 325 | 31 | 37% | 0 | 36% | 0 | 36 | Tagged - 07/2016/ | 2016/ | Feb/23 | 37% |
| 1 083 | in die eerste plek nie per kredietkaart <krediet> betaal het nie. "Jy wil mos | | | 333 | 31 | 79% | 0 | 37% | 0 | 37 | Tagged - 07/2016/ | 2016/ | Feb/23 | 38% |
| 1 084 | eerste plek nie per kredietkaart betaal het nie. "Jy wil mos lê . . . " | | | 335 | 31 | 89% | 0 | 37% | 0 | 37 | Tagged - 07/2016/ | 2016/ | Feb/23 | 38% |
| 1 085 | nie per kredietkaart betaal het nie. "Jy <man> wil mos lê . . . " sê mev. Dos | | | 339 | 32 | 23% | 0 | 37% | 0 | 37 | Tagged - 07/2016/ | 2016/ | Feb/23 | 38% |
| 1 086 | betaal het nie. "Jy wil mos lê . . . <le> " sê mev. Dos Santos. "Voel só | | | 346 | 32 | 77% | 0 | 38% | 0 | 38 | Tagged - 07/2016/ | 2016/ | Feb/23 | 39% |
| 1 087 | wil mos lê . . . " sê mev. Dos Santos <mev> "Voel só , en só !" | | | 352 | 33 | 100% | 0 | 39% | 0 | 39 | Tagged - 07/2016/ | 2016/ | Feb/23 | 39% |
| 1 088 | . . . " sê mev. Dos Santos. "Voel só <so> , en só , en só !" "Maar ek wou | | | 355 | 34 | 36% | 0 | 39% | 0 | 39 | Tagged - 07/2016/ | 2016/ | Feb/23 | 40% |
| 1 089 | sê mev. Dos Santos. "Voel só , en só <so> , en só !" "Maar ek wou nie lê | | | 358 | 34 | 64% | 0 | 40% | 0 | 40 | Tagged - 07/2016/ | 2016/ | Feb/23 | 40% |
| 1 090 | Dos Santos. "Voel só , en só , en só <so> !" "Maar ek wou nie lê nie!" sê | | | 362 | 34 | 100% | 0 | 40% | 0 | 40 | Tagged - 07/2016/ | 2016/ | Feb/23 | 40% |
| 1 091 | "Voel só , en só , en só !" "Maar ek <man> wou nie lê nie!" sê mnr. Dos | | | 365 | 35 | 33% | 0 | 40% | 0 | 40 | Tagged - 07/2016/ | 2016/ | Feb/23 | 40% |
| 1 092 | só , en só , en só !" "Maar ek wou <wou> nie lê nie!" sê mnr. Dos Santos | | | 367 | 35 | 50% | 0 | 41% | 0 | 41 | Tagged - 07/2016/ | 2016/ | Feb/23 | 41% |
| 1 093 | , en só , en só !" "Maar ek wou nie lê <le> nie!" sê mnr. Dos Santos terwyl | | | 370 | 35 | 75% | 0 | 41% | 0 | 41 | Tagged - 07/2016/ | 2016/ | Feb/23 | 41% |
| 1 094 | ek wou nie lê nie!" sê mnr. Dos Santos <man> terwyl hy op die vloer lê en | | | 376 | 36 | 31% | 0 | 42% | 0 | 42 | Tagged - 07/2016/ | 2016/ | Feb/23 | 41% |
| 1 095 | lê nie!" sê mnr. Dos Santos terwyl y <man> op die vloer lê en keer. "Jy | | | 379 | 36 | 54% | 0 | 42% | 0 | 42 | Tagged - 07/2016/ | 2016/ | Feb/23 | 42% |
| 1 096 | terwyl hy op die vloer lê en keer. "Jy <man> wou nie! Maar hoekom het jy | | | 387 | 37 | 50% | 0 | 43% | 0 | 43 | Tagged - 07/2016/ | 2016/ | Feb/23 | 42% |
| 1 097 | hy op die vloer lê en keer. "Jy wou <wou> nie! Maar hoekom het jy dan ? | | | 389 | 37 | 83% | 0 | 43% | 0 | 43 | Tagged - 07/2016/ | 2016/ | Feb/23 | 43% |
| 1 098 | keer. "Jy wou nie! Maar hoekom het y <man> dan ? Broeis , nè? Broeis !" | | | 395 | 38 | 86% | 0 | 44% | 0 | 44 | Tagged - 07/2016/ | 2016/ | Feb/23 | 43% |
| 1 099 | nie! Maar hoekom het jy dan ? Broeis <broeis> , nè? Broeis !" "Maar ek sê | | | 398 | 39 | 60% | 0 | 44% | 0 | 44 | Tagged - 07/2016/ | 2016/ | Feb/23 | 43% |
| 1 100 | het jy dan ? Broeis , nè? Broeis <broeis> !" "Maar ek sê mos ek het | | | 402 | 40 | 100% | 0 | 44% | 0 | 44 | Tagged - 07/2016/ | 2016/ | Feb/23 | 44% |
| 1 101 | dan ? Broeis , nè? Broeis !" "Maar ek <man> sê mos ek het nie!" Die hele | | | 405 | 41 | 40% | 0 | 45% | 0 | 45 | Tagged - 07/2016/ | 2016/ | Feb/23 | 44% |
| 1 102 | , nè? Broeis !" "Maar ek sê mos ek <man> he nie!" Die hele ding het 'n | | | 409 | 41 | 80% | 0 | 45% | 0 | 45 | Tagged - 07/2016/ | 2016/ | Feb/23 | 45% |
| 1 103 | , berig die Brasiliaanse weekblad Istoé <v-w> in Rio. Mnr. Dos Santos , nou | | | 425 | 42 | 83% | 0 | 47% | 0 | 47 | Tagged - 07/2016/ | 2016/ | Feb/23 | 46% |
| 1 104 | die Brasiliaanse weekblad Istoé in Rio <rio> . Mnr. Dos Santos , nou self | | | 428 | 42 | 100% | 0 | 47% | 0 | 47 | Tagged - 07/2016/ | 2016/ | Feb/23 | 47% |
| 1 105 | Istoé in Rio. Mnr. Dos Santos <man> , nou self hoogs die hoenders | | | 432 | 44 | 11% | 0 | 48% | 0 | 48 | Tagged - 07/2016/ | 2016/ | Feb/23 | 47% |
| 1 106 | Santos , nou self hoogs die hoenders <hoender> in, het dadelik sy bank | | | 439 | 44 | 31% | 0 | 49% | 0 | 49 | Tagged - 07/2016/ | 2016/ | Feb/23 | 48% |
| 1 107 | hoogs die hoenders in, het dadelik sy <man> bank gebel en gevra hoe | | | 444 | 44 | 46% | 0 | 49% | 0 | 49 | Tagged - 07/2016/ | 2016/ | Feb/23 | 48% |
| 1 108 | die hoenders in, het dadelik sy bank <krediet> gebel en gevra hoe beland | | | 446 | 44 | 51% | 0 | 49% | 0 | 49 | Tagged - 07/2016/ | 2016/ | Feb/23 | 48% |
| 1 109 | gebel en gevra hoe beland 'n betaling <krediet> aan House of Pleasure op | | | 454 | 44 | 74% | 0 | 50% | 0 | 50 | Tagged - 07/2016/ | 2016/ | Feb/23 | 49% |
| 1 110 | 'n betaling aan House of Pleasure <house> op sy kredietkaart. Dit blyk | | | 459 | 44 | 89% | 0 | 51% | 0 | 51 | Tagged - 07/2016/ | 2016/ | Feb/23 | 50% |
| 1 111 | House of Pleasure op sy kredietkaart <krediet> Dit blyk toe die onkoste is | | | 463 | 44 | 100% | 0 | 51% | 0 | 51 | Tagged - 07/2016/ | 2016/ | Feb/23 | 50% |
| 1 112 | kredietkaart. Dit blyk toe die onkoste <krediet> s in der waarheid | | | 469 | 45 | 26% | 0 | 52% | 0 | 52 | Tagged - 07/2016/ | 2016/ | Feb/23 | 51% |
| 1 113 | wag daarvoor . . . mevrou Dos Santos <mev> ! Jy kon mnr. Dos Santos | | | 489 | 45 | 100% | 0 | 54% | 0 | 54 | Tagged - 07/2016/ | 2016/ | Feb/23 | 53% |
| 1 114 | Dos Santos ! Jy kon mnr. Dos Santos <man> ontik met 'n hoenderveer. Vir | | | 495 | 47 | 44% | 0 | 55% | 0 | 55 | Tagged - 07/2016/ | 2016/ | Feb/23 | 54% |
| 1 115 | . Dos Santos ontik met 'n hoenderveer <hoender> . Vir 'n oomblik het hy met | | | 500 | 47 | 100% | 0 | 55% | 0 | 55 | Tagged - 07/2016/ | 2016/ | Feb/23 | 54% |
| 1 116 | 'n hoenderveer. Vir 'n oomblik het y <man> met nuwe oë na sy vrou gekyk, | | | 506 | 48 | 14% | 0 | 56% | 0 | 56 | Tagged - 07/2016/ | 2016/ | Feb/23 | 55% |
| 1 117 | oomblik het hy met nuwe oë na sy vrou <mev> gekyk, maar hy kon dit glad nie | | | 513 | 48 | 29% | 0 | 57% | 0 | 57 | Tagged - 07/2016/ | 2016/ | Feb/23 | 55% |

Figuur 5-4: Voorbeeld van etikette binne die sintaktiese konteks WordSmith.

As gevolg van die aard van die linguïstiese elemente wat ondersoek word, is daar egter twee maniere waarop die voorkoms van al die kohesieskakels deur WordSmith geïsoleer kon word. Leksikale elemente wat verwysing, leksikale kohesie en substitusie verteenwoordig, kan nie in 'n lys saamgestel en gesoek word nie omdat daar 'n oneindige hoeveelheid leksikale items is wat deur hierdie skakels kohesie in 'n teks kan bewerkstellig. Daar kan dus nie soekfrases ingesleutel word om die voorkoms van linguïstiese elemente wat hierdie skakels verteenwoordig, op te spoor nie. Om hierdie verwysing, substitusie en drie kohesieskakels te isoleer, is daar dus op grond van die etikette 'n konkordansielys opgestel. Wanneer die voorkoms van al die etikette gesoek is, is daar 'n lys saamgestel wat die etikette of dan gevalle van kohesie lys, al is daar nog nie bygevoeg op watter wyse die kohesie bewerkstellig is nie. Elemente wat deur middel van verwysing, leksikale kohesie en substitusie kohesief aan 'n ander verbind is, is dus dan geïsoleer al is daar nog nie onderskeid tussen hierdie tipes getref nie. Daardie klassifikasie is in fase drie van die analise gedoen (vergelyk afdeling 5.6.3).

Die tweede wyse waarop kohesieskakels geïsoleer kan word, hou verband met die feit dat verbandsmerkers en ellips (soos dit in hierdie studie ondersoek word, vergelyk afdeling 3.5.3 en Bylaag A onderskeidelik) altyd met sekere linguistiese items geïdentifiseer word. Daar is byvoorbeeld 'n voorafbepaalde lys verbandsmerkers (deur Jordaan, 2014:43-80) opgestel. Verder word ellips binne 'n neweskikkende sinskonstruksie ondersoek en kan neweskikkende voegwoorde opgespoor word en die konteks ondersoek word, om te bepaal of daar ellips in die betrokke sin voorkom. Indien hierdie linguistiese items dus as soekfrase vir 'n konkordansielys ingevoer word, word 'n lys geskep met die voorkoms van al die betrokke woorde en moet dit slegs gekategoriseer word. Deur die konkordansielyste van bogenoemde twee prosesse bymekaar te voeg, is die gevalle waar kohesie voorgekom het almal geïsoleer en hulle kon gekategoriseer word.

5.6.3 Fase 3: Kategorisering van kohesief verwante items

In die laaste fase van die teksanalise, is die konkordansielys soos afdeling 5.6.2 beskryf is, gebruik. Daar is weer van effens verskillende metodes gebruik gemaak om uiteindelik die voorkoms van al die kohesieskakels op een lys te karteer.

5.6.3.1 Kategorisering van verwysing, leksikale kohesie en substitusie

Die lys wat hiervoor gebruik is, het die kere wat 'n kohesiemerker in 'n spesifieke korpus voorgekom het ('instances of cohesion'), bevat (gemerk met etikette) en al was dit nog nie gekategoriseer nie, was dit geïsoleer. Saam met die konkordansielys, verskaf WordSmith vir elkeen van die trefpunte addisionele statistiese inligting – veral in verband met die posisie van die betrokke voorkoms ten opsigte van die teks waarin dit voorkom (sien figuur 5-5). Verdere inligting wat op die lys voorkom, is onder meer die oorspronklike lêer waaruit die betrokke trefslag kom, se naam.

| | | | | | | | | | | | | |
|-----|---|----|---|---------|---|-------|--|---|-------|------------|-----------|-------|
| 6 | Mev. Dalva dos Santos <mev> bestudeer haar man se kredietkaartrekening en whal kom s | 5 | 1 | 16.13% | 0 | 0.44% | | 0 | 0.44% | Kd ALR.txt | #09 00:00 | 0.00% |
| 15 | Mev. Dalva dos Santos <dossantos> bestudeer haar man se kredietkaartrekening en whal | 6 | 1 | 19.35% | 0 | 0.55% | | 0 | 0.55% | Kd ALR.txt | #09 00:00 | 1.00% |
| 12 | Mev. Dalva dos Santos bestudeer haar <doss> man se kredietkaartrekening en whal kom sy onerwags af | 9 | 1 | 29.03% | 0 | 0.88% | | 0 | 0.88% | Kd ALR.txt | #09 00:00 | 1.00% |
| 19 | Mev. Dalva dos Santos bestudeer haar man <man> se kredietkaartrekening en whal kom sy onerwags af op | 11 | 1 | 35.48% | 0 | 1.10% | | 0 | 1.10% | Kd ALR.txt | #09 00:00 | 1.00% |
| 37 | alva dos Santos bestudeer haar man se kredietkaartrekening <krediet> en whal kom sy onerwags af op 'n bedrag betaalbaar | 14 | 1 | 45.16% | 0 | 1.43% | | 0 | 1.43% | Kd ALR.txt | #09 00:00 | 2.00% |
| | er haar man se kredietkaartrekening en whal kom sy <sy> onerwags af op 'n bedrag betaalbaar aan: House of Pleasure. Sy weet d | 18 | 1 | 66.32% | 0 | 1.94% | | | | | | |
| 61 | dietkaartrekening en whal kom sy onerwags af op 'n bedrag <krediet> betaalbaar aan: House of Pleasure. Sy weet dadelik | 24 | 1 | 77.42% | 0 | 2.54% | | 0 | 2.54% | Kd ALR.txt | #09 00:00 | 3.00% |
| 57 | kening en whal kom sy onerwags af op 'n bedrag betaalbaar <mev> aan: House of Pleasure. Sy weet dadelik haar man i | 26 | 1 | 83.87% | 0 | 2.76% | | 0 | 2.76% | Kd ALR.txt | #09 00:00 | 3.00% |
| 83 | nerwags af op 'n bedrag betaalbaar aan: House of Pleasure <house>. Sy weet dadelik haar man is in groot moeilikheid. | 31 | 1 | 100.00% | 0 | 3.31% | | 0 | 3.31% | Kd ALR.txt | #09 00:00 | 4.00% |
| 81 | ags af op 'n bedrag betaalbaar aan: House of Pleasure. Sy <mev> weet dadelik haar man is in groot moeilikheid. (Hy we | 33 | 2 | 21.43% | 0 | 3.53% | | 0 | 3.53% | Kd ALR.txt | #09 00:00 | 4.00% |
| 80 | g betaalbaar aan: House of Pleasure. Sy weet dadelik haar <mev> man is in groot moeilikheid. (Hy weet dit nog nie.) Ho | 37 | 2 | 50.00% | 0 | 3.97% | | 0 | 3.97% | Kd ALR.txt | #09 00:00 | 4.00% |
| 105 | aalbaar aan: House of Pleasure. Sy weet dadelik haar man <man> is in groot moeilikheid. (Hy weet dit nog nie.) House | 39 | 2 | 64.29% | 0 | 4.19% | | 0 | 4.19% | Kd ALR.txt | #09 00:00 | 5.00% |
| 102 | leasure. Sy weet dadelik haar man is in groot moeilikheid <moeilikheid>. (Hy weet dit nog nie.) House of Pleasure is e | 44 | 2 | 100.00% | 0 | 4.74% | | 0 | 4.74% | Kd ALR.txt | #09 00:00 | 5.00% |
| 124 | e. Sy weet dadelik haar man is in groot moeilikheid. (Hy <man> weet dit nog nie.) House of Pleasure is een van daardie | 46 | 3 | 37.50% | 0 | 4.96% | | 0 | 4.96% | Kd ALR.txt | #09 00:00 | 6.00% |
| 135 | et dadelik haar man is in groot moeilikheid. (Hy weet dit <moeilikheid> nog nie.) House of Pleasure is een van daardie | 49 | 3 | 75.00% | 0 | 5.29% | | 0 | 5.29% | Kd ALR.txt | #09 00:00 | 6.00% |

Figuur 5-5: Voorbeeld van Microsoft Excel-lêer waarin die voorkomste van kohesie aangedui is tesame met ander statistiese inligting

Om die kategoriseringsproses makliker en meer akkuraat te maak, is die volledige konkordansielys op 'n spesifieke manier gesorteer. Die lys is eerstens volgens oorspronklike lêernaam gesorteer. Dit verseker dat al die skakels wat uit dieselfde teks kom, saamgegroeper is. Binne hierdie sortering is daar volgens nog twee vlakke gesorteer, naamlik die naam van die etiket of frase waarvolgens die oorspronklike konkordansie gedoen is, en die posisie in die teks waar die skakel voorkom (van laag na hoog). Hierdie sortering het in die geval van verwysing, leksikale kohesie en substitusie verseker dat skakels wat na dieselfde antesedent verwys, saam voorkom (omdat dit met dieselfde etiket gemerk is) en dat dit in die volgorde voorkom waarin dit binne die oorspronklike teks voorgekom het (sien figuur 5-6).

| | | | | | | | |
|-----|--|---------------|----|---|---------|---|-------|
| 12 | Mev. Dalva dos Santos bestudeer haar <doss> man se kredietkaartrekening en whal kom sy onerwags af | <doss> | 9 | 1 | 29.03% | 0 | 0.88% |
| 15 | Mev. Dalva dos Santos <dossantos> bestudeer haar man se kredietkaartrekening en whal | <dossantos> | 6 | 1 | 19.35% | 0 | 0.55% |
| 83 | nerwags af op 'n bedrag betaalbaar aan: House of Pleasure <house>. Sy weet dadelik haar man is in groot moeilikheid. | <house> | 31 | 1 | 100.00% | 0 | 3.31% |
| 37 | alva dos Santos bestudeer haar man se kredietkaartrekening <krediet> en whal kom sy onerwags af op 'n bedrag betaal | <krediet> | 14 | 1 | 45.16% | 0 | 1.43% |
| 61 | dietkaartrekening en whal kom sy onerwags af op 'n bedrag <krediet> betaalbaar aan: House of Pleasure. Sy weet dade | <krediet> | 24 | 1 | 77.42% | 0 | 2.54% |
| 57 | kening en whal kom sy onerwags af op 'n bedrag betaalbaar <krediet> aan: House of Pleasure. Sy weet dadelik haar me | <krediet> | 26 | 1 | 83.87% | 0 | 2.76% |
| 19 | Mev. Dalva dos Santos bestudeer haar man <man> se kredietkaartrekening en whal kom sy onerwags af op | <man> | 11 | 1 | 35.48% | 0 | 1.10% |
| 105 | taalbaar aan: House of Pleasure. Sy weet dadelik haar man <man> is in groot moeilikheid. (Hy weet dit nog nie.) House | <man> | 39 | 2 | 64.29% | 0 | 4.19% |
| 124 | e. Sy weet dadelik haar man is in groot moeilikheid. (Hy <man> weet dit nog nie.) House of Pleasure is een van daardie | <man> | 46 | 3 | 37.50% | 0 | 4.96% |
| 6 | Mev. Dalva dos Santos <mev> bestudeer haar man se kredietkaartrekening en whal kom s | <mev> | 5 | 1 | 16.13% | 0 | 0.44% |
| 81 | ags af op 'n bedrag betaalbaar aan: House of Pleasure. Sy <mev> weet dadelik haar man is in groot moeilikheid. (Hy we | <mev> | 33 | 2 | 21.43% | 0 | 3.53% |
| 80 | g betaalbaar aan: House of Pleasure. Sy weet dadelik haar <mev> man is in groot moeilikheid. (Hy weet dit nog nie.) Ho | <mev> | 37 | 2 | 50.00% | 0 | 3.97% |
| 102 | leasure. Sy weet dadelik haar man is in groot moeilikheid <moeilikheid>. (Hy weet dit nog nie.) House of Pleasure is e | <moeilikheid> | 44 | 2 | 100.00% | 0 | 4.74% |
| 135 | et dadelik haar man is in groot moeilikheid. (Hy weet dit <moeilikheid> nog nie.) House of Pleasure is een van daardie | <moeilikheid> | 49 | 3 | 75.00% | 0 | 5.29% |
| 97 | Mev. Dalva dos Santos bestudeer haar man se kredietkaartrekening en whal kom sy onerwags af op 'n bedrag bet | en | 15 | 1 | 48.39% | 0 | 1.54% |
| 60 | artrekening en whal kom sy onerwags af op 'n bedrag betaalbaar aan: House of Pleasure. Sy weet dadelik haar man is i | of | 29 | 1 | 93.55% | 0 | 3.09% |

Figuur 5-6: Voorbeeld van sortering van etikette binne Microsoft Excel

Ná bogenoemde stappe voltooi is, kon daar duidelik gesien word watter woord/frase kohesief aan watter antesedent verbind is en kon die navorser telkens slegs besluit op watter manier dit verbind is. Met ander woorde, die konteks is so gefiltreer dat slegs die konteks wat nodig is om 'n spesifieke gevalle¹⁹ van kohesievoorkoms ("instance of cohesion") te kategoriseer, telkens beskikbaar is.

¹⁹ Beide die terme 'voorkoms/voorkomste' en 'geval/gevalle' gaan afwisselend gebruik word (na gelang van die konteks) vir die vertaling van "instances of cohesion"

Ná die loodsanalises is gesien dat die identifikasie van kollokasies moontlik probleme kan gee (vergelyk ook die bespreking in afdeling 3.5.4.2). Dit wil voorkom asof kollokasie tussen kohesie en koherensie lê in die sin dat dit aan die een kant deel uitmaak van die oppervlakstruktuur van die teks en as't ware in 'n mate te doen het met 'n sekere betekenisveld en aan die ander kant sterk op die leser se ekstralinguistiese kennis steun. Dit is juis hierdie ekstralinguistiese kennis wat gebruik word wanneer koherensie in tekste bepaal word en dit maak die onderskeid tussen semanties verwante woorde, kollokatief verwante woorde en begrippe wat koherensie bewerkstellig moeilik. Dit kan die beste geïllustreer word aan die hand van voorbeelde uit die korpus.

Semanties verwante woorde

- 'Kaapse', 'Kapenare' in die teks, *Kaap toe* deur Dana Snyman
- 'kêrels' en 'mans' in die teks, *Fred sien 'n plan* deur Dana Snyman
- 'gepleit', 'geraas' en 'gedreig' in die teks, *Na 'n ou-ou kortverhaal van Roald Dahl* deur André le Roux

Kollokatief verwante woorde

- 'mortier', 'kontaksituasie' en 'ontploff' in die teks *Solitary man* deur André le Roux
- 'toets', 'All Black-skrumskakel', 'Nuweland' en 'tackle' in die teks, *Soos 'n vleuel op pad na die doellyn* deur Dana Snyman
- 'spoorwegstoele', 'stasie' en 'sitkamerwa' in die teks, *Bo-op die lamppaal sweef die gees van Olive Schreiner* deur André le Roux.

Begrippe wat koherensie bewerkstelling²⁰

- 'oorlog' en 'Wes-Zambië' in die teks *Solitary man* deur André le Roux
- 'Frik du Preez' en 'ge-tackle' in die teks, *Soos 'n vleuel op pad na die doellyn* deur Dana Snyman

²⁰ Voorbeelde word geneem uit die ooreenstemmende tekste van die voorafgaande afdeling 'Kollokatief verwante woord'. Begrippe wat koherensie bewerkstellig, is nie in hierdie studie gemerk nie aangesien die voorkoms van kohesieskakels ondersoek is.

- ‘gedrink’ (as antesedent), ‘Hulle speel ’n wals’ en ‘Almal kyk op en beduie met koue wysvingers’ in die teks, *Bo-op die lamppaal sweef die gees van Olive Schreiner* deur André le Roux.

Dit was die beoordeling van die navorser dat daar te veel ekstralinguistiese kennis benodig word om die verwantskap tussen die woorde/frases in die bogenoemde voorbeelde (“Begrippe wat koherensie bewerkstellig”) te verstaan en dat die interpretasie dus net nie deur middel van kohesie bewerkstellig word nie. In die teks waarin die laasgenoemde voorbeeld voorkom, is daar etlike frases in die teks wat op die bedwelmdede toestand van die karakters dui. Dit sou deur die meeste lesers so geïnterpreteer word, maar die navorser is nie van mening dat die frases kollokatief aan mekaar verbind is nie.

5.6.3.2 Kategorisering van verbandsmerkers

Vir die kategorisering van die verbandsmerkers is Jordaan (2014) se lys van Afrikaanse verbandsmerkers gebruik (vergelyk afdeling 3.5.5). Hoewel sy nie aandring daarop dat die lys volledig is nie, is dit ’n baie omvattende lys wat ingespan kan word om die analise in hierdie studie af te baken. Al die woorde op hierdie lys is as soekfrases gebruik vir die konkordansie en dus is elke voorkoms van al die woorde op die lys op die konkordansielys ingesluit. Daar is in afsonderlike soektogte na die vier subkategorieë (aaneenskakelend, teenstellend, redegewend, tydsaanduidend) van verbandsmerkers gesoek. Omdat hierdie frases so goed geïsoleer kon word, is aanvanklik aanvaar dat al die voorkomste van ’n betrokke tipe verbandsmerker se woorde noodwendig daardie verband aandui. Die klassifikasie is egter nagegaan en ’n aantal uitsonderings is gekry. Die meeste van die woorde op die konjunksiewoordelys kom slegs as verbandsmerkers voor. Daar was egter homonieme onder wat na aanleiding van die tekstkonteks(te) in ’n ander betekenisverband sou kon staan. Voorbeelde word in tabel 5-2 gegee.

Tabel 5-2: Voorbeelde van woorde wat deur outomatiese sifting as moontlike verbandsmerkers geïdentifiseer is, maar ’n ander funksie vervul.

| Voorbeeld | Woordsoort of funksie in sin |
|---|----------------------------------|
| Al | |
| Ek kook dit nog vir vyf minute <u>al</u> sê die resepslegs vir tien minute. | Konjunksiemerker: Teenstellend |
| Ek het <u>al</u> drie keer gekyk | Bywoord |
| Toe | |
| <u>Toe</u> gaan ons eet. | Konjunksiemerker: Tydsaanduidend |
| Maak die deur <u>toe</u> . | Byvoeglike naamwoord |
| Tog | |
| <u>Tog</u> het hy nie geweet dat... | Konjunksiemerker: Teenstellend |
| Help <u>tog</u> ! | Bywoord |
| Tot | |
| Wat <u>tot</u> oormore duur | Konjunksiemerker: Tydsaanduidend |
| Van skoonmaakmiddels, <u>tot</u> paraffien | Voorsetsel |
| Voor | |
| <u>Voor</u> ons ry, moet ons eers.... | Konjunksiemerker: Tydsaanduidend |
| Ek staan <u>voor</u> jou. | Voorsetsel |

5.6.3.3 Kategorisering van ellips

As gevolg van die spesifieke aard van die ellips wat in hierdie studie ondersoek word, kon 'n baie soortgelyke metode gevolg word om die voorkoms van ellips te isoleer as wat gebruik is vir die verbandsmerkers. In hierdie studie is slegs ellips wat verband hou met sinne in aaneenskakelende verband wat met 'en' en 'maar' verbind is, ondersoek (vergelyk afdeling 3.5.3). Dus kon al die voorkomste van 'en' en 'maar' deur middel van 'n konkordansielys geïsoleer word, en kon daar telkens slegs besluit word of daar in die konteks van daardie woord ellips voorkom al dan nie.

Die bogenoemde drie afdelings van hierdie fase is afsonderlik gedoen en uiteindelik op een lys met mekaar gekonsolideer ten einde 'n lys te kry waarop al die voorkomste van kohesiemerkers wat in hierdie studie ondersoek word in al die tekste van 'n spesifieke korpus gekarteer is. In hierdie lys word onder meer die onmiddellike konteks waarin die woord oorkom, watter kohesieskakel verteenwoordig word deur die spesifieke woord/frase en waar in die teks die skakel voorkom, aangedui. Die posisie van die woord/etiket is deur WordSmith bepaal by die

opstel van die konkordansielys. Indien ’n mens hierdie lys dus sorteer om telkens al die voorkomste van ’n spesifieke tipe kohesieskakel te wys, is hierdie skakel se voorkoms en posisie oor al die tekste in die betrokke korpus geïsoleer, ongeag van antesedent of van die woord/frase wat hierdie kohesie bewerkstellig het.

Die data wat uiteindelik aan statistiese analise onderwerp gaan word, is aangedui in ’n tabel met onder meer die volgende inligting: die werklike tekstuele konteks van die oorspronklike etiket en die naam van die etiket²¹. Hierdie kolomme is a) vir die nagaan van die analise en b) vir konteks in die oorspronklike data ingesluit en word nie gebruik in die interpretasie van die uiteindelige statistiese data nie. Verder is daar kolomme waarin elk van die etikette of ‘voorkomste van kohesie’ gekategoriseer is volgens die kategorieë van kohesieskakels. Laastens is die data wat WordSmith aanvanklik verskaf het aangaande die posisie van die skakel in die teks steeds aangeheg. Tydens die statistiese analise sal die wisselwerking tussen die kategorieë en die posisie waarin skakels van spesifieke kategorieë tipies verskyn, ondersoek word. In figuur 5-7 kan gesien word hoe die kategorisering van die onderskeie gevalle van kohesievoorkoms geëtiketteer is.

| | | | | | | |
|-----|---|-------------|------------|-------------------|-----------------|-----------|
| 6 | Mev. Dalva dos Santos <mev> bestudeer haar man se kredietkaartrekening en whal kom s | <mev> | Antesedent | | | |
| 15 | Mev. Dalva dos Santos <dossantos> bestudeer haar man se kredietkaartrekening en whal | <dossantos> | Antesedent | | | |
| 12 | Mev. Dalva dos Santos bestudeer haar <dos> man se kredietkaartrekening en whal kom sy onerwags af | <dos> | | Verwysing | Bep | |
| 19 | Mev. Dalva dos Santos bestudeer haar man <man> se kredietkaartrekening en whal kom sy onerwags af op | <man> | Antesedent | | | |
| 37 | alva dos Santos bestudeer haar man se kredietkaartrekening <krediet> en whal kom sy onerwags af op ’n bedrag betaal | <krediet> | Antesedent | | | |
| 97 | Mev. Dalva dos Santos bestudeer haar man se kredietkaartrekening en whal kom sy onerwags af op ’n bedrag beta | en | | Verbandsmerker | Aaneenskakelend | |
| 61 | dietkaartrekening en whal kom sy onerwags af op ’n bedrag <krediet> betaalbaar aan: House of Pleasure . Sy weet dade | <krediet> | | Leksikale kohesie | Kollokasie | |
| 57 | kening en whal kom sy onerwags af op ’n bedrag betaalbaar <krediet> aan: House of Pleasure . Sy weet dadelik haar ma | <krediet> | | Leksikale kohesie | Kollokasie | |
| 60 | artrekening en whal kom sy onerwags af op ’n bedrag betaalbaar aan: House of Pleasure . Sy weet dadelik haar ma | of | | Verbandsmerker | Teenstellend | |
| 83 | nerwags af op ’n bedrag betaalbaar aan: House of Pleasure <house> . Sy weet dadelik haar man is in groot moeilikh | <house> | Antesedent | | | |
| 81 | ags af op ’n bedrag betaalbaar aan: House of Pleasure . Sy <mev> weet dadelik haar man is in groot moeilikh | <mev> | | Verwysing | Bep | |
| 80 | g betaalbaar aan: House of Pleasure . Sy weet dadelik haar <mev> man is in groot moeilikh | <mev> | | Verwysing | Bep | |
| 105 | haalbaar aan: House of Pleasure . Sy weet dadelik haar man <man> is in groot moeilikh | <man> | | Leksikale kohesie | Semanties | Herhaling |
| 102 | leasure . Sy weet dadelik haar man is in groot moeilikh <moeilikh> . (Hy weet dit nog nie.) House of Pleasure is e | <moeilikh> | Antesedent | | | |
| 124 | e . Sy weet dadelik haar man is in groot moeilikh . (Hy <man> weet dit nog nie.) House of Pleasure is een van daardie | <man> | | Verwysing | Bep | |
| 135 | et dadelik haar man is in groot moeilikh . (Hy weet dit <moeilikh> nog nie.) House of Pleasure is een van daardie | <moeilikh> | | Verwysing | Bep | |

Figuur 5-7: Voorbeeld van kategorisering van kohesieskakels in Microsoft Excel

5.7 Statistiese analysesagteware

Vir die statistiese analise in hierdie studie, is van die sagteware *IBM Statistics SPSS 25* en die webplatform *Social Science Statistics* gebruik gemaak. Volgens die vervaardigers van SPSS Statistics is hierdie sagteware die wêreld se voorste platform vir ’n verskeidenheid statistiese analyses. Dit is sagteware waarmee “ad-hoc analise, hipotesetoetsing en voorspellende analise” gedoen kan word (IBM SPSS Statistics, 2017). Die webplatform *Social Science Statistics* is ’n

²¹ Nie die naam van die tipe kohesieskakel nie.

gratis hulpbron wat statistiese hulpmiddels spesifiek aan navorsers in die sosiale wetenskappe bied. Die omvang van die berekeninge wat hierdie platform bied, is heelwat kleiner as SPSS Statistics, maar bied 'n meer gebruikersvriendelike ervaring met enkele statistiese metodes wat gereeld deur diegene in die sosiale wetenskappe gebruik word (Social Science Statistics, 2019).

5.8 Statistiese analisemetode

Soos reeds vermeld, is daar van statistiese metodes gebruik gemaak om die data in hierdie studie te analiseer. Aangesien die studie tot 'n mate hipotesetoetsend is (vergelyk afdeling 5.5), is daar nie 'n standaard statistiese analisemetode wat normaalweg vir hierdie tipe studie gebruik word nie. Daar is aanvanklik besluit om van die chi-kwadraattoets gebruik te maak om te bepaal of verskille tussen die voorkomste van kohesieskakels statisties beduidend is. Nadat hierdie toets gedoen is en die resultate opgeskryf is, was die navorser egter van mening dat dit nie voldoende is om die navorsingsvrae te kan beantwoord nie. Daar is gesoek na 'n meer genuanseerde statistiese ontleding om die data-ontleding verder toe te lig. Binêre logistiese regressie-analise (*binary logistic regression analysis*) is die gekose metode. Dit is 'n analisemetode wat 'n aanduiding gee van die mate waartoe sekere veranderlikes voorspellingswaarde het ten opsigte van 'n sekere uitkoms. In die geval van hierdie studie is die moontlike veranderlikes die kohesieskakels, en die skrywers dien as moontlike uitkomst.

5.8.1 Chi-kwadraattoetse

Om te bepaal of daar statisties beduidende verskille tussen die voorkoms en posisie van kohesieskakels in die tekste voorkom, is daar van Pearson se chi-kwadraattoets gebruik gemaak. Hierdie toets word gebruik om 'n reeks data in een kategorie te vergelyk met 'n reeks data in 'n ander kategorie (Jordaan, 2014:141). Dit bepaal tot watter mate die twee of meer stelle data met mekaar verband hou (SPSS, 2018). Met die afloop van die toets kan aangedui word tot watter mate die reeks data in die een kategorie afhanklik is van die data in die ander kategorie. In die geval van hierdie studie, beteken dit dat daar byvoorbeeld gesê kan word tot watter mate die voorkoms van 'n sekere kohesieskakel in die twee genres van 'n sekere skrywer met mekaar ooreenkom, en by implikasie, tot watter mate die verspreiding van kohesieskakels in die twee genres waarskynlik afhanklik is van die ander veranderlike, naamlik die skrywer.

Indien hierdie toetse 'n p-waarde kleiner as 0,05 wys, beteken dit dat die data in die een kategorie waarskynlik afhanklik is van die data in die ander kategorie.

5.8.2 Binêre logistiese regressie-analise

Logistiese regressie-analise word gebruik wanneer daar bepaal wil word tot watter mate een of meer onafhanklike veranderlikes 'n voorspeller is van een van twee moontlike uitkomst, met ander woorde, die afhanklike veranderlike (of uitkoms) moet slegs uit twee kategorieë bestaan (Sieben & Linssen, 2009:1). Hierdie tipe analise kan gebruik word wanneer jy die "...aan- of afwesigheid van 'n eienskap of uitkoms [wil voorspel] (*afhanklike veranderlikes*), gebaseer op 'n stel voorspeller veranderlikes (*onafhanklike veranderlikes*)" (SPSS, eie vertaling, eie klem). Die toepassing van hierdie metode kan aan die hand van 'n voorbeeld verduidelik word. Indien 'n navorser wil bepaal watter faktore die belangrikste voorspellers is daarvan of 'n student sy/haar studies binne die voorgeskrewe tyd sal voltooi, kan van hierdie metode gebruik gemaak word. Die afhanklike veranderlike bied dan een van twee moontlike kategorieë vir uitkomst (JA, die student het sy/haar studies voltooi of NEE, die student het nie sy/haar studies voltooi nie) en die onafhanklike veranderlike kan enige veranderlike wees wat die navorser wil ondersoek; vorige akademiese prestasie, die student se sosio-ekonomiese agtergrond, demografiese inligting of enige iets anders wat die navorser dink relevant kan wees. Daar kan dan op grond van die resultate van hierdie analise, bepaal word tot watter mate bv. hoër sosio-ekonomiese status 'n student se kans om sy/haar studies in die voorgeskrewe tyd te voltooi, beïnvloed. In hierdie studie is daar in 'n gegewe regressie-analise een van twee moontlike stelle afhanklike veranderlikes. Die een is die onderskeie skrywers (JA, hierdie veranderlikes dui op 'n teks van Dana Snyman of NEE, hierdie veranderlikes dui nie op 'n teks van Dana Snyman nie en dus op 'n teks van André le Roux). Die tweede moontlike afhanklike veranderlike is tekstipe (JA, hierdie veranderlikes dui op die waarskynlikheid dat die teks 'n rubriek is, of NEE, dit is waarskynlik nie 'n rubriek nie, dus 'n kortverhaal).

Wanneer 'n binêre logistiese regressie-analise gedoen is, bevat die resultate 'n hele aantal komponente (nie net bv. $p < 0,05$ of $p > 0,05$ soos in die geval van die chi-kwadraattoets nie). Een van hierdie komponente is 'n nulmodel. In 'n regressie-analise dien die onafhanklike veranderlikes as 'n model om die onafhanklike veranderlike te voorspel. Suksesvolle modelle sal met groter akkuraatheid 'n spesifieke uitkoms kan voorspel, terwyl minder suksesvolle

modelle nie so gereeld die korrekte uitkoms kan voorspel nie. Wanneer 'n binêre regressie-analise gedoen word, word die mate van akkuraatheid teen die sg. nulmodel getoets. Die nulmodel wys op die waarskynlikheid dat die onafhanklike veranderlikes geensins voorspellingswaarde het nie; dus, 'n raaiskoot tussen die twee moontlike uitkomste wat deur die afhanklike veranderlikes voorgestel word. Die model wat aan statistiese analise onderwerp word, moet dus 'n beter kans as 50/50 hê, indien daar enige voorspellingswaarde aan geheg wil word. Die akkuraatheid van beide die nulmodel en die model wat ondersoek word, word met 'n persentasie aangedui.

Indien 'n studie op soek is na 'n spesifieke veranderlike of veranderlikes wat 'n invloed op die voorspellingswaarde van 'n uitkoms is, val die fokus normaalweg op die resultate wat in SPSS as "Exp(B)" gemerk is. Die statistiese inligting in hierdie resultate gee 'n aanduiding van die waarskynlikheid dat die voorkoms van hierdie betrokke veranderlike een uitkoms bo 'n ander kan verseker. In 'n studie oor faktore van studentesukses kan daar twee moontlikhede wees ten opsigte van die uitkoms, naamlik, die student was *suksevol* of die student was *onsuksevol* en 'n aantal moontlikhede ten opsigte van die veranderlikes, bv. demografiese oorwegings, vorige akademiese prestasie, sosio-ekonomiese agtergrond, ensovoorts. Na afloop van 'n logistiese regressie-analise, kan daar dus gesê word tot watter mate hoër skoolpunte (teenoor laer skoolpunte) jou kans op sukses verhoog. Die hoë of lae punte wat 'n spesifieke student voorheen behaal het, kan dus beskryf word as iets wat sy/haar kans op studentesukses met een of twee of drie keer verhoog. In hierdie studie, kan dieselfde egter nie van 'n enkele voorkoms van 'n kohesiemerker gesê word nie. 'n Enkele voorkoms van bv. verwysing is nie genoeg om die kans dat die voorkoms in 'n spesifieke skrywer se teks voorkom te verhoog nie; die kollektiewe voorkoms van hierdie merkers kan egter 'n aanduiding wees. Daarom sal daar nie in hierdie bespreking van die binêre regressie-analise op hierdie Exp(B) gefokus word nie, maar op aanduidings wat vir die model in geheel geld. Dit beteken dat daar nie gesê gaan word dat byvoorbeeld die aanwesigheid van 'n enkele voorkoms van kohesie, die kans dat die teks deur Dana Snyman geskryf is, met 'n sekere hoeveelheid verhoog/verlaag nie. Wat wel gesê kan word, is die kans dat 'n spesifieke stel voorkomste van sekere kohesieskakels (in 'n teks) die kans dat die teks deur 'n spesifieke outeur geskryf is, met 'n sekere hoeveelheid kan verhoog of verlaag.

5.8.3 Beskrywing van datastel vir statistiese analise

Vir die statistiese analise was daar verskeie ondersoekmoontlikhede beskikbaar op grond van die taalkundige analises wat gedoen is en die data wat dit geskep het. Vir elke geval van kohesievoorkoms was daar nou data beskikbaar ten opsigte van die tipe kohesiewe element, die tipe teks waarin dit voorkom, wie die skrywer is en waar die element in die teks voorkom (posisie binne 'n sin en posisie binne 'n paragraaf). Die posisie van die element is op twee maniere aangetoon, naamlik vir hul posisie indien die sin/paragraaf in kwarte verdeel sou word (Posisie 4) en indien die sin/paragraaf in tiendes verdeel sou word (Posisie 10).

Soos in afdeling 5.6 bespreek is, is die tipes kohesie aangedui nadat die data aanvanklik geëtiketteer is en in WordSmith Tools sorteer is. Data wat vir die analises beskikbaar was, was 1) die tipe kohesieskakel wat op daardie spesifieke plek in die teks teenwoordig was, 2) die posisie waarin daardie kohesieskakel binne die sin bekleed en 3) die posisie wat die kohesieskakel in die paragraaf bekleed. In die geval van die klassifikasie van die skakels, leksikale kohesie, verbandsmerkers, en verwysing was daar ook 'n verdere klassifikasie om die aard van die skakel verder te beskryf. Vir die posisie wat die kohesieskakels in die sinne bekleed, is die sinne in kwarte gedeel (volgens aantal woorde in die sin) en die posisie deur getalle 1 tot 4 aangedui. Vir paragrawe, is 'n soortgelyke prosedure gevolg, maar die paragrawe is in tiendes verdeel en getalle een tot tien is gebruik om aan te dui waar die kohesieskakel voorkom. Vir 'n uiteensetting van die bogenoemde verduideliking, vergelyk *tabel 5-3*.

Om te illustreer hoe hierdie posisies in die werklike tekste lyk, kan na die paragrawe in tabel 5-4 gekyk word. Dit is opeenvolgende paragrawe wat voorkom in die teks, *1 Desember 2006 op 'n sypaadjie in Sandton* deur André le Roux²². In die tabel, word die tipe kohesieskakel en die sins- en paragraafposisie van die skakel aangedui. Die woord 'bloei' in die tweede paragraaf, is dus in die tweede sinsposisie (of die tweede kwart van die sin - die derde woord van agt), maar in die vierde paragraafposisie (of die vierde kwart van die paragraaf – die 32ste woord van 38).

Tabel 5-3: Uiteensetting van die tipes data wat vir die statistiese ontleding beskikbaar was

²² Ter wille van duidelikheid, word al die kohesieskakels wat in hierdie paragrawe geïdentifiseer is, nie in die voorbeeld gemerk nie. Die voorbeeld is slegs om die aanduidingswyse van die posisies van kohesieskakels in die teks te verhelder. Hierdie twee paragrawe was die eerste twee in die teks, en sommige items is as 'antesedent' gemerk.

| | Tipe kohesieskakel | Posisie van kohesieskakel in sinsverband | Posisie van kohesieskakel in paragraafverband |
|---|--|--|--|
| Primêre klassifikasie | <p>Elke voorkoms van 'n kohesieskakel is geklassifiseer as:</p> <ul style="list-style-type: none"> • Ellips, • Leksikale Kohesie, • Substitusie, • Verbandsmerker, of • Verwysing | Getalle 1-4 is gebruik om aan te dui of die skakel in die 1ste, 2de, 3de of 4de kwart van die sin voorkom. | Getalle 1-10 is gebruik om aan te dui in watter tiende van die paragraaf die skakel voorkom. |
| Addisionele klassifikasie²³ | <p>Leksikale kohesie is verder geklassifiseer as:</p> <ul style="list-style-type: none"> • kollokasie, of • semanties <p>Indien 'n kohesieskakel as Leksikale Kohesie > Semanties geklassifiseer is, kon dit verder ook as herhaling geklassifiseer word.</p> <p>Verbandsmerkers is verder geklassifiseer as:</p> <ul style="list-style-type: none"> • Aaneenskakelend, • Redegewend, • Teenstellend, of • Tydsaanduidend <p>Verwysing is verder geklassifiseer as:</p> <ul style="list-style-type: none"> • Eienaam • Voornaamwoord (vnw.) | | |

²³ Sien volledige bespreking van addisionele klassifikasies in Afdelings 3.5.1-3.5.5.

Tabel 5-4: Illustrasie van die aanduiding van die posisie van die kohesieskakels in die tekste

| Voorbeeldteks | | | |
|---|---------------------------|--------------------|-------------------------|
| <i>Motors ry vir 'n oomblik stadiger. 'n Man lê op die <u>sypaadjie</u> in <u>Katherinestraat</u> met <u>bloed</u> in <u>sy mond</u>.</i> | | | |
| <i>Dié in die motors is dankbaar <u>hulle</u> hoef nie stil te hou om hulp te verleen nie. Daar is reeds omstanders en iemand sou wel 'n ambulans gebel het. As iemand <u>bloei</u> is daar die moontlikheid van <u>VIGS</u>.</i> | | | |
| Leksikale item | Tipe kohesieskakel | Sinsposisie | Paragraafposisie |
| Motors | Antesedent | 1 | 1 |
| 'n Man | Antesedent | 2 | 2 |
| sypaadjie | Antesedent | 2 | 2 |
| Katherinestraat | Eienaamverwysing | 3 | 3 |
| bloed | Antesedent | 4 | 4 |
| sy | Voornaamwoordverwysing | 4 | 4 |
| Dié in die motors | Antesedent | 1 | 1 |
| hulle | Voornaamwoordverwysing | 2 | 1 |
| Daar | Voornaamwoordverwysing | 1 | 2 |
| ambulans | Antesedent | 4 | 3 |
| bloei | Leksikale Kohesie | 2 | 4 |
| VIGS | Eienaamverwysing | 4 | 4 |

Al die bogenoemde klassifikasies is gemaak op grond van die teoretiese beskrywings van die betrokke kohesieskakel wat in afdeling 3.5 bespreek is.

Uiteraard was die bogenoemde data beskikbaar vir tekste van beide skrywers, in beide genres. Enige kombinasie van veranderlikes sou dus voorgestel of geanaliseer word. Indien die navorsingsvrae in ag geneem word (vergelyk 1.3) is die mees sinvolle manier om die data te interpreteer en aan te bied, om eers op die onderskeie skrywers te fokus en eers daarna op die genres. Daar gaan dus deurgaans gepoog word om beskrywende en ander statistiese data op hierdie wyse aan te bied.

5.8.4 Doelwitte van die statistiese analise

Die doelwit van die statistiese analise (chi-kwadraat en logistiese regressie-analise) is hoofsaaklik om die handanalises te beskryf. Dit is egter net 'n hulpmiddel en soos in hoofstuk 6 gesien sal word, gee dit nie 'n enkele eenvoudige antwoord tot die vraagstuk nie. Die uitdagings van die multidissiplinêre aard van 'n studie soos hierdie (vergelyk afdeling 2.3.3.1)

het tot gevolg dat die statistiese analise slegs aanduidings van beduidendheid en effekgroottes gee en die resultate steeds tot 'n groot mate deur die navorser geïnterpreteer moet word. Soos reeds genoem (in afdeling 2.3.4.2), is die gevaar van die 'linguistiese vingerafdruk'-metafoer, die indruk dat 'n stel veranderlikes in 'n analiseproses ingelees word, en daar met 'n antwoord na die proses vorendag gekom word. Dieselfde gevaar bestaan in terme van die interpretasie van die statistiese analise van hierdie studie.

5.9 Samevatting

In hierdie hoofstuk is die metode wat gevolg is ten einde die navorsingsvrae te beantwoord, uiteengesit. Daar is aanvanklik verduidelik dat hierdie ondersoek van 'n korpusgebaseerde metode gebruik maak en die wesensaard en beperkinge van hierdie metode is bespreek. Hierna is die navorsingsontwerp en analiseproses wat gevolg is, uiteengesit. Laastens is aandag gegee aan die statistiese analisemetodes wat vir dataverwerking gebruik gaan word. Daar is verduidelik watter rekenaarsagteware gebruik gaan word, watter statistiese metodes gevolg gaan word, en hoe die datastel lyk wat aan statistiese analise onderwerp gaan word. In hoofstuk 6 word die bevindinge van die statistiese analise bespreek.

6 Data-analise en interpretasie

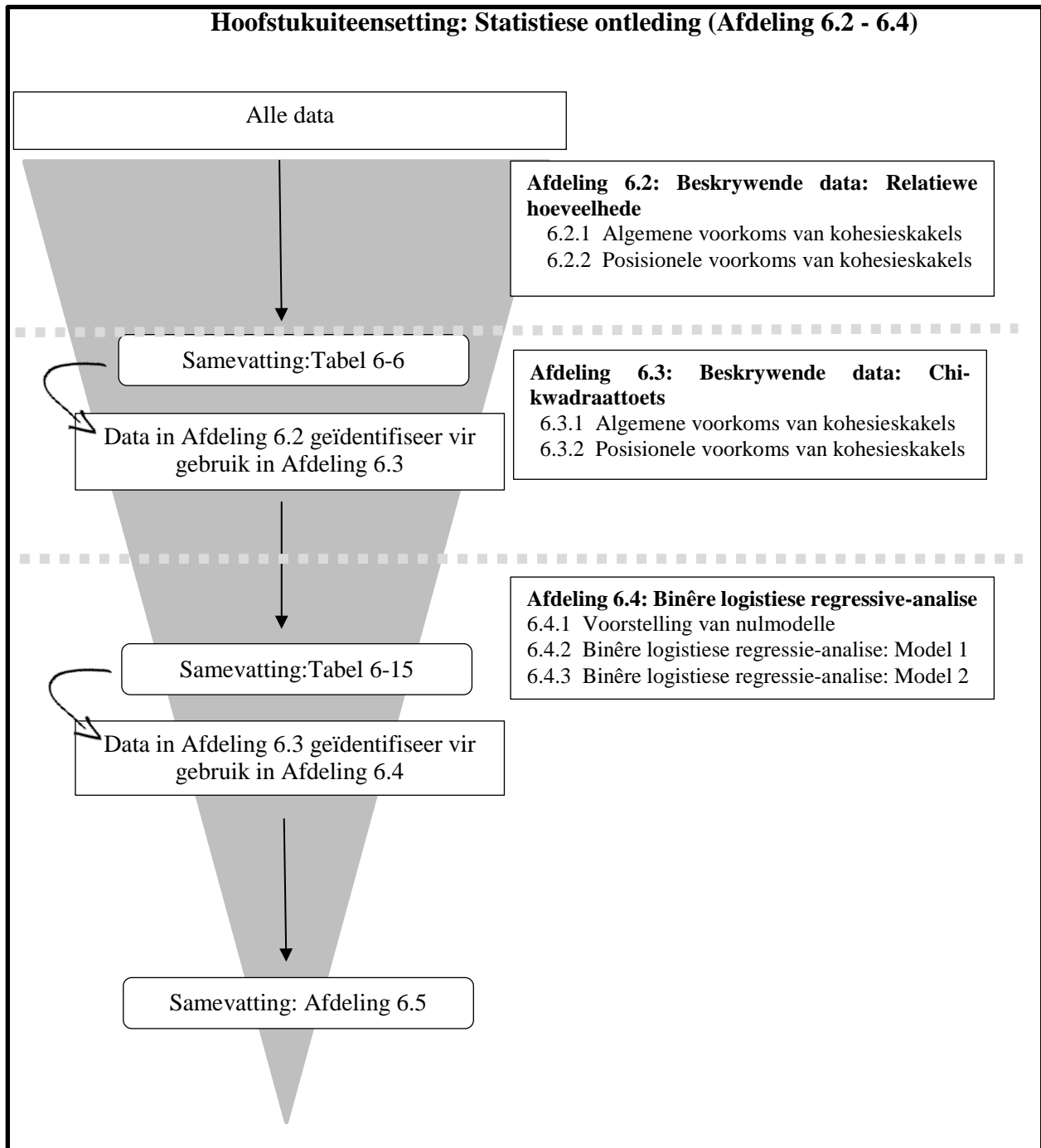
6.1 Inleiding

Soos in afdeling 2.3.1 bespreek is, is die interdisiplinêre aard van die veld van outeurskapidentifikasie een van sy grootste uitdagings. Meestal word kundigheid op minstens drie velde vereis, naamlik linguistiek, rekenaarprogrammering en statistiek. Wanneer dit by die analise en interpretasie van die data vir hierdie studie kom, was die navorser ook deeglik bewus van die bogenoemde uitdaging. Hoewel daar in hoofstuk 2 'n meer volledige beskrywing van die uitdagings asook aanbevelings vir toekomstige studies is, word daar tans volstaan by die volgende omskrywing en vooruitskouing van hierdie hoofstuk.

As gevolg van die omvang van die beskikbare data, is dit nie haalbaar of sinvol om al die data aan verskeie statistiese berekeninge te onderwerp nie; selfs al is dit vir diagnostiese doeleindes. Ná 'n poging om so veel as moontlik van die data aan 'n aantal (moontlik) relevante statistiese berekeninge te onderwerp, is uiteindelik besluit om die data stapsgewys deur middel van statistiese data-analisemetodes, data af te baken en in die proses elemente van variasie in die skrywers se skryfstyl te identifiseer. Die proses wat gevolg is, word in figuur 6-1 uiteengesit. Die inligting in hierdie diagram dien ook as 'n hoofstukuiteensetting van die oorhoofse afdelings en hul verhouding tot mekaar.

Die stapsgewyse benadering waarna in die bostaande paragraaf verwys is, word deur die diagram in figuur 6-1 geïllustreer. Om mee te begin, word beskrywende data gebruik om te bepaal watter aspek van kohesieskakelvoorkoms moontlik tussen korpusse varieer. Die eerste voorstelling van die beskrywende data (Afdeling 6.2), is die voorstelling van genormaliseerde hoeveelhede²⁴ van die gevalle van kohesievoorkoms in die onderskeie korpusse. Vir hierdie afdeling word al die beskikbare data en kategorieë in ag geneem. Aan die einde van hierdie afdeling is 'n tabel (tabel 6-6) waarin die belangrikste afleidings na afloop van die bestudering van die beskrywende data opgesom is. Hierdie opsomming dien as basis vir die analises in die daaropvolgende afdeling (Afdeling 6.3). In afdeling 6.3 word die data wat in afdeling 6.2

²⁴ Die konsep 'genormaliseerde hoeveelhede' word in Afdeling 6.2 uiteengesit.



Figuur 6-1: Hoofstukuiteensetting: Statistiese ontleding (Afdeling 6.2 - 6.4)

geïdentifiseer en in tabel 6-6 uiteengesit is, gebruik. Die chi-kwadraattoets²⁵ word op hierdie stappe data uitgevoer ten einde verdere statistiese afleidings oor die voorkoms van kohesieskakels in die onderskeie skrywers se werk te maak. Die belangrikste bevindinge van

²⁵ n Bespreking van die chi-kwadraattoets as analyse-instrument kan in Afdeling 5.8.1 gevind word.

hierdie statistiese analise (die chi-kwadraattoets) word in tabel 6-15 uiteengesit en hierdie inligting word gebruik vir die binêre logistiese regressie-analise²⁶ wat in afdeling 6.4 aan die beurt kom. Na hierdie afdeling word 'n opsomming van die afleidings na aanleiding van die regressie-analise gegee, waarna die navorsingsvrae in hoofstuk 7 beantwoord word.

Deur die bogenoemde benadering te volg, word data wat nie as uniek tot 'n skrywer en/of genre gevind word nie geleidelik uitgefilter en afleidings word stapsgewys gemaak. Basiese statistiese metodes (soos beskrywende statistiek) wat in hierdie geval op die hele datastel toegepas kan word, word eerste gebruik en meer doelgerigte metodes (soos binêre logistiese regressie-analise) word op seleksies van die data toegepas ten einde meer gefokusde resultate te verkry.

6.2 Beskrywende data: Relatiewe hoeveelhede

6.2.1 Algemene voorkoms van kohesieskakels

Soos in tabel 6-1 blyk, is daar in die hele korpus 'n totaal van 9 098 kohesieskakels gemerk en geëtiketteer. Hiervan het 4 654 voorgekom in die tekste van André le Roux en 4 444 in die tekste van Dana Snyman. Die tipe kohesieskakel wat die meeste voorgekom het, is leksikale kohesie, wat 3 235 keer voorgekom het. Daarvan was 1 594 voorkomste in die tekste van André le Roux en 1 641 in die tekste van Dana Snyman. Verwysing het die tweede meeste voorgekom (3163 voorkomste); 1 749 in die werk van Le Roux en 1 414 in die werk van Snyman. Die kohesieskakel met die derde hoogste voorkoms was verbandsmerkers. Uit die 2 195 totale verbandsmerkers, het 1 051 uit André le Roux se tekste gekom en 1 144 uit Dana Snyman s'n. Die voorkoms van ellips en substitusie as kohesieskakels is merkbaar laer met onderskeidelik 460 en 45 voorkomste. Daarvan het 235 van die voorkomste van ellips in André le Roux se tekste voorgekom en 225 in Dana Snyman se tekste. Persentasies word onder die geobserveerde hoeveelhede aangedui en is telkens bereken as 'n persentasie van al die kohesieskakels in die hele datastel. Indien die totale hoeveelheid geïdentifiseerde kohesieskakels in ag geneem word

²⁶ 'n Bespreking van die binêre logistiese regressie-analise as statistiese metode kan in Afdeling 5.8.2 gevind word.

(9 098) beteken die bogenoemde verduideliking in effek dat 90,98 voorkomste van 'n kohesieskakel aan een persentasiepunt gelyk is.

Tabel 6-1: Voorkoms van kohesieskakels in tekste van onderskeie skrywers

| Kohesieskakel | André le Roux | | | Dana Snyman | | | Totaal |
|--------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | K/verhale | Rubrieke | Totaal | K/verhale | Rubrieke | Totaal | |
| Ellips | 158 1,7% | 77 0,8% | 235 2,6% | 124 1,4% | 101 1,1% | 225 2,5% | 460 5,1% |
| Leksikale kohesie | 933 10,3% | 661 7,3% | 1 592 17,5% | 832 9,1% | 809 8,9% | 1 641 18,0% | 3 235 35,6% |
| Substitusie | 9 0,1% | 16 0,2% | 25 0,3% | 9 0,1% | 11 0,1% | 20 0,2% | 45 0,5% |
| Verbandsmerker | 535 5,9% | 516 5,7% | 1 051 11,6% | 593 6,5% | 551 6,1% | 1 144 12,6% | 2 195 24,1% |
| Verwysing | 956 10,5% | 793 8,7% | 1749 19,2% | 675 7,4% | 739 8,1% | 1 414 15,5% | 3 163 34,8% |
| Totaal | 2 591 28,5% | 2 063 22,7% | 4 654 51,2% | 2 233 24,5% | 2 211 24,3% | 4 444 48,8% | 9 098 100% |

Op hierdie stadium is dit sinvol om oor genormaliseerde hoeveelhede 'n opmerking te maak. Uiteraard bestaan die inligting in tabel 6-1 slegs uit die geobserveerde voorkomste van geëtiketteerde kohesieskakels in die onderskeie korpusse. Hierdie getalle en berekeninge wat hierdie getalle gebruik, behoort onder meer geïnterpreteer te word aan die hand van die korpusgroottes van die onderskeie korpusse waarin hulle voorkom. Met ander woorde, daar kan nie eenvoudig na die hoeveelheid voorkomste gekyk word sonder om in ag te neem hoe groot die korpus is waarin hierdie hoeveelheid van die betrokke skakel verskyn het nie. Normaalweg word hierdie probleem oorbrug deur van genormaliseerde hoeveelhede gebruik te maak. Dit is hoeveelhede wat verwerk is om bv. 'n hoeveelheid spesifieke voorkomste van 'n verskynsel per x-aantal woorde in daardie spesifieke korpus uit te druk. Eers dan kan hoeveelhede werklik vergelyk word. tabel 6-2 wys die genormaliseerde hoeveelhede kohesieskakels per 100 woorde in elke korpus (elke genre van elke skrywer). Die laaste horisontale ry in die tabel verwys na die hoeveelheid woorde in elke korpus soos in tabel 5-1 aangedui is; dus, nie die hoeveelheid gemerkte kohesieskakels nie. Hierdie inligting dui aan dat 'n gemiddeld van 28,74 woorde per 100 (oor al vier subkorpusse heen) as 'n kohesieskakel gemerk is. Ongeveer 30% van alle woorde wat in hierdie korpus gebruik is, is dus op een of

ander wyse aan 'n ander woorde/konsep verbind. Die getalle in die “gemiddeld”-kolom is die hoeveelheid woorde per 100 woorde wat as 'n spesifieke kohesiemerker in die hele korpus verskyn het. Afwykings van hierdie hoeveelheid dui moontlike afwykings in die tekste van óf die betrokke skrywer óf die betrokke genre aan.

Tabel 6-2: Genormaliseerde hoeveelhede kohesieskakels per 100 woorde in die korpus

| Kohesieskakel | André le Roux | | | Dana Snyman | | | Totale korpus |
|---|--------------------|--------------------|-------------------------|--------------------|--------------------|-------------------------|----------------------|
| | K/verhale | Rubrieke | Gemiddeld ²⁷ | K/verhale | Rubrieke | Gemiddeld ²⁸ | Gemiddeld |
| Ellips | 1,76 | 1,06 | 1,45 | 1,55 | 1,36 | 1,46 | 1,45 |
| Leksikale kohesie | 10,39 | 9,12 | 9,81 | 10,41 | 10,87 | 10,63 | 10,22 |
| Substitusie | 0,10 | 0,22 | 0,15 | 0,11 | 0,15 | 0,13 | 0,14 |
| V/merker | 5,96 | 7,12 | 6,48 | 7,42 | 7,40 | 7,41 | 6,93 |
| Verwysing | 10,65 | 10,94 | 10,78 | 8,45 | 9,93 | 9,16 | 9,99 |
| Totaal | 28.86 | 28.46 | 28.68 | 27.94 | 29.71 | 28.80 | 28.74 |
| Hoeveelheid woorde per korpus²⁹ | 8 977 (28%) | 7 249 (23%) | 16 226 (51%) | 7 991 (25%) | 7 441 (24%) | 15 432 (49%) | 31 658 (100%) |

Aangesien die doel van die bespreking van die beskrywende data is om areas te identifiseer waar daar moontlik afwykings gaan wees, gaan alle data in hierdie afdeling as genormaliseerde hoeveelhede aangedui word, tensy anders vermeld. tabelle met die geobserveerde hoeveelhede, totale en persentasies, kan in Bylaag A nagegaan word.

Indien die data in tabel 6-2 deursoek word vir potensiële uniekhede tussen skrywers en/of genres, is dit sinvol om eers na totale en gemiddeldes te kyk voor na spesifieke skakels gekyk

²⁷ Dui die gemiddelde hoeveelheid vir die betrokke skrywer se tekste, ongeag genre aan. So kan die afwyking (al dan nie) tussen die genres van 'n spesifieke skrywer makliker raakgesien word.

²⁸ Dui die gemiddelde hoeveelheid vir die betrokke skrywer se tekste, ongeag genre aan. So kan die afwyking (al dan nie) tussen die genres van 'n spesifieke skrywer makliker raakgesien word.

²⁹ Die persentasie tussen hakies in hierdie ry data is 'n aanduiding van die persentasie wat die betrokke korpus van die totale korpus van 40 tekste uitmaak. Elke subkorpus (per genre, per skrywer) bestaan uit 10 tekste. Hierdie persentasies kan later nuttig wees in die interpretasie van verdere statistiese analise.

word. Eerstens kan geobserveer word dat die subkorpuse van die twee skrywers relatief soortgelyk in grootte is met André le Roux se tekste wat 51% van die totale hoeveelheid woorde van die algemene korpus beslaan en Dana Snyman se tekste wat 49% van die totale hoeveelheid woorde beslaan. In terme van die hoeveelheid kohesieskakels wat elke skrywer per 100 woorde gebruik het, is die gemiddeld tussen elk van die twee genres vir elk van die skrywers ook soortgelyk, met 28,68 kohesieskakels per 100 woorde in die tekste van André le Roux en 28,80 kohesieskakels per 100 woorde in die tekste van Dana Snyman. Wanneer daar egter na die onderskeie genres per skrywer gekyk word, is daar effense variasie. Terwyl André le Roux 'n soortgelyke hoeveelheid kohesieskakels per 100 woorde in sy kortverhale en rubrieke gebruik (onderskeidelik 28,86 en 28,46), gebruik Dana Snyman minder kohesieskakels in sy kortverhale as in sy rubrieke (27,94 teenoor 29,71).

Wanneer na die detail van Dana Snyman se kohesiegebruik gekyk word, word die meeste variasie veroorsaak deur die meer gereelde gebruik van verwysing in rubrieke as in kortverhale (9,93 teenoor 8,45 woorde voorkomste per 100 woorde). Die voorkoms van die ander kohesieskakels in die tekste van Dana Snyman is soortgelyk ten opsigte van die verskeie genres. André le Roux se verdeling van kohesieskakels tussen genres is egter meer dinamies, ten spyte van die feit dat die algehele hoeveelheid per genre basies dieselfde is. Die twee kohesieskakels wat grootliks hierdie verskynsel illustreer, is leksikale kohesie en verbandsmerkers. André le Roux maak meer gereeld van leksikale kohesie in sy kortverhale as in sy rubrieke gebruik (onderskeidelik 10,39 teenoor 9,12 voorkomste per 100 woorde) maar minder gereeld van verbandsmerkers (onderskeidelik 5,96 voorkomste per 100 woorde in kortverhale teenoor 7,12 in rubrieke). 'n Soortgelyke tendens as wat hierbo beskryf is (groter verskille in André le Roux se toepassing van kohesieskakels as in Dana Snyman s'n tussen die genres) is ook in ellips en substitusie waargeneem, maar as gevolg van die relatiewe klein hoeveelheid voorkomste van hierdie kohesieskakels blyk die invloed daarvan nie groot te wees nie.

Voorlopig kan dus opgemerk word dat Dana Snyman, met die uitsondering van die gebruik van verwysing, kohesieskakels in relatief soortgelyke ratio's gebruik, ongeag of hy 'n kortverhaal of 'n rubriek skryf. Hierteenoor blyk dit dat André le Roux sy gebruik van veral leksikale kohesie en verbandsmerkers aanpas na gelang van die genre waarin hy skryf.

6.2.1.1 Algemene voorkoms van subkategorieë

Soos daar in tabel 5-3 (vergelyk afdeling 5.8.3) uiteengesit is, is daar vir leksikale kohesie, verbandsmerkers en verwysing, data ten opsigte van subkategorieë beskikbaar. Bylaag B illustreer die geobserveerde voorkomste ten opsigte van hierdie kategorieë³⁰.

tabel 6-3 wys die genormaliseerde hoeveelheid voorkomste van die onderskeie subkategorieë van kohesieskakels in die onderskeie korpusse. Om data te identifiseer wat dalk op uniekhede dui, word tabel 6-3 deursoek vir gevalle waar die hoeveelhede voorkomste per 100 woorde per skrywer of per genre verskillend is. Die statistiese beduidendheid van hierdie verskille sal in afdeling 6.3 en 6.4 bespreek word.

In terme van leksikale kohesie lyk dit asof die aantal kere wat André le Roux beide tipes leksikale kohesie in rubrieke gebruik, afwyk van die patroon waarvolgens dit in die ander subkorpusse voorkom. Die voorkoms van 'Leksikale Kohesie: Kollokasie' is die hoogste in die rubrieke van André le Roux (3,32 voorkomste per 100 woorde), terwyl die voorkoms in die ander drie korpusse relatief naby aan mekaar is (2,16 in die kortverhale van André le Roux en onderskeidelik 2,16 en 2,51 in die kortverhale en rubrieke van Dana Snyman). Die voorkoms van 'Leksikale Kohesie: Semanties' in die rubrieke van André le Roux, is weer ooglopend laer (5,81) as in die ander drie subkorpusse. Die verskil tussen die hoogste en die laagste voorkoms in die ander drie sub-korpusse is maar 0,11 voorkomste per 100 terwyl die 5,81 voorkomste per 100 woorde in Le Roux se rubrieke 2,41 voorkomste per 100 woorde minder is as die naaste ander subkorpus.

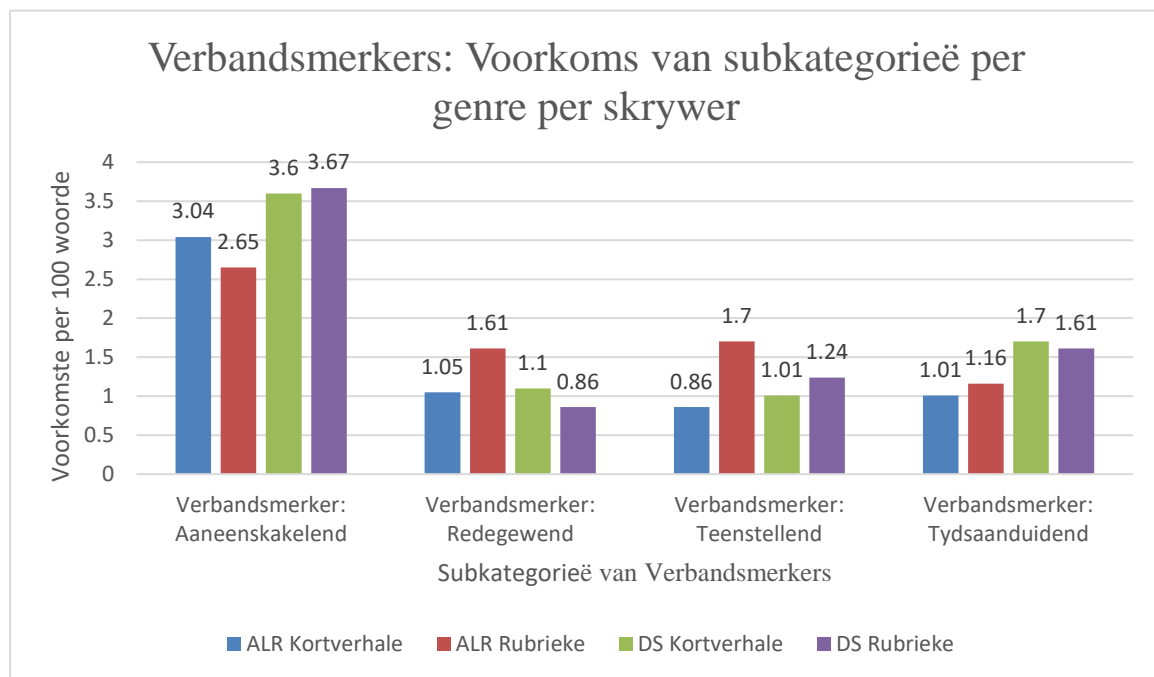
Die verdeling van die subkategorieë van verbandsmerker is moeilik om te interpreteer vanuit die tabel omdat daar vier subkategorieë is en daarom is dit sinvol om vir die doel van die bespreking 'n grafiek van daardie datastel voor te hou. figuur 6-2 stel hierdie gegewens voor.

³⁰ Hoeveelhede is genormaliseer volgens die hoeveelheid voorkomste per 100 woorde in die betrokke subkorpus, bv. kortverhale van Dana Snyman of rubrieke van André le Roux.

Tabel 6-3: Genormaliseerde data vir hoeveelhede van verskillende tipes Leksikale Kohesie, Verbandsmerkers en Verwysing per 100 woorde in die korpus

| Kohesieskakel | André le Roux | | | Dana Snyman | | | Totaal |
|----------------------------------|---------------------|---------------------|----------------------|---------------------|---------------------|----------------------|-----------------------|
| | K/verhale | Rubrieke | Totaal | K/verhale | Rubrieke | Totaal | |
| Leksikale kohesie: Kollokasie | 2,16 | 3,32 | 2,68 | 2,16 | 2,51 | 2,33 | 2,51 |
| Leksikale kohesie: Semanties | 8,22 | 5,81 | 7,14 | 8,22 | 8,33 | 8,28 | 7,69 |
| Totaal: Leksikale kohesie | 10,38 | 9,10 | 9,81 | 10,39 | 10,85 | 10,61 | 10,21 |
| Verbandsmerker: Aaneenskakelend | 3,04 | 2,65 | 2,87 | 3,60 | 3,67 | 3,64 | 3,24 |
| Verbandsmerker: Redegewend | 1,05 | 1,61 | 1,30 | 1,10 | 0,89 | 1,00 | 1,15 |
| Verbandsmerker: Teenstellend: | 0,86 | 1,70 | 1,23 | 1,01 | 1,24 | 1,12 | 1,18 |
| Verbandsmerker: Tydsaanduidend | 1,01 | 1,16 | 1,08 | 1,70 | 1,61 | 1,66 | 1,36 |
| Totaal: Verbandsmerker | 5,96 | 7,12 | 6,48 | 7,42 | 7,40 | 7,41 | 6,93 |
| Verwysing: Vnw | 10,07 | 9,02 | 9,60 | 7,56 | 8,74 | 8,13 | 8,88 |
| Verwysing: Eienaam | 0,53 | 1,89 | 1,14 | 0,88 | 1,20 | 1,03 | 1,09 |
| Totaal: Verwysing | 10,60 | 10,91 | 10,74 | 8,43 | 9,93 | 9,16 | 9,97 |
| Hoeveelheid woorde per korpus | 8 977 28% | 7 249 23% | 16 226 51% | 7 991 25% | 7 441 24% | 15 432 49% | 31 658 100% |

In figuur 6-2 is dit duidelik dat die rubrieke van André le Roux (soos in die geval van leksikale kohesie) in al die subkategorieë behalwe tydsaanduidende verbandsmerkers, in 'n mindere of 'n meerdere mate afwyk van die tendens vir die ander verbandsmerkers (aaneenskakelende verbandsmerkers, 2,65 voorkomste per 100 woorde; redegewende verbandsmerkers, 1,61 voorkomste per 100 woorde; teenstellende verbandsmerkers, 1,7 voorkomste per 100 woorde). In die geval van teenstellende verbandsmerkers, is die voorkoms in die kortverhale van André le Roux die laagste terwyl dit in sy rubrieke die hoogste is. In die geval van aaneenskakelende verbandsmerkers is daar tipies minder voorkomste in tekste in beide genres van Le Roux as wat daar in dié van Snyman is. Die gebruik van redegewende verbandsmerkers is met die uitsondering van André le Roux se rubrieke, soortgelyk. In die geval van tydsaanduidende verbandsmerkers, is die voorkomste in beide genres van André le Roux laer as in die tekste van Dana Snyman.



Figuur 6-2: Uiteensetting van voorkomste van subkategorieë van Verbandsmerkers in die hele datastel

Wanneer daar na die onderafdeling van die voorkomste van tipes verwysing gekyk word, kan waargeneem word dat André le Roux in beide kortverhale en rubrieke tekste meer van voornaamwoordverwysing gebruik maak as Dana Snyman (10,07 en 9,02 teenoor 7,56 en 8,74 voorkomste per 100 woorde). In die afdeling van eiernaamverwysing is dit opmerklik dat die kortverhale van André le Roux die minste voorkomste per 100 woorde bevat (0,53), maar sy

rubrieke die meeste (1,89). Hierdie verskynsel strook met die voorkoms van sommige van die ander skakels tussen die genres van die skrywers waar die tekste van Dana Snyman meer soortgelyk is en dié van André le Roux, meer verskillend.

Uit die beskrywende data ten opsigte van die voorkoms van die vyf kohesieskakels en sommige onderafdelings daarvan, is daar 'n aantal moontlike afwykings op te merk wat met verdere statistiese analise ontleed sou kon word. Hierdie ontleding word in afdeling 6.3 en 6.4 bespreek. Die data wat lyk asof dit die meeste van die patroon afwyk en dus vir statistiese beduidendheid ondersoek sal word, sluit die volgende in:

- Die ratio's van leksikale kohesie en verbandsmerkers van die onderskeie skrywers
- Die ratio's van subkategorieë van leksikale kohesie van die onderskeie skrywers
- Die ratio's van subkategorieë van verbandswoorde (sonder tydsaanduidende verbandswoorde) van die onderskeie skrywers
- Die ratio's van subkategorie van verwysing van onderskeie skrywers

Die bogenoemde afwykings, en dié wat in die volgende afdelings (6.2.2 en 6.2.3) bespreek gaan word, is aan die einde van afdeling 6.2.3 in tabel 6-6 saamgevat. Dit inligting in hierdie tabel dien as vertrekpunt vir die daaropvolgende analises.

6.2.2 Posisionele voorkoms van kohesieskakels: Sinsvlak

'n Soortgelyke proses as in die vorige afdeling kan gevolg word wanneer die data vir die posisie van kohesieskakels bespreek word. As gevolg van die hoeveelheid data wat in terme van die posisies van die kohesieskakels beskikbaar is, is die data wat in die volgende paragrawe ter sake is, in grys in die tabel 6-4 gemerk. Let ook op dat die kohesieskakel *substitusie* uit hierdie datastel weggelaat is. Die algemene voorkoms van substitusie is reeds in afdeling 6.2 bespreek, maar vanweë die lae voorkoms daarvan, is dit nie sinvol om dit verder in die posisionele voorkoms ook te verdeel nie. Die data met die geobserveerde hoeveelhede kan onder Bylaag A bestudeer word. Dieselfde inligting word egter as genormaliseerde data in tabel 6-4 voorgestel en dit is die data wat gebruik gaan word om te probeer bepaal waar daar moontlik afwykings mag wees.

Tabel 6-4: Genormaliseerde data ten opsigte van die voorkoms van kohesieskakels op sinsvlak

| | André le Roux | | | Dana Snyman | | | Totale korpus |
|-------------------------------|---------------|--------------|---------------|--------------|--------------|---------------|---------------|
| | K/Verhale | Rubrieke | Gemiddeld | K/Verhale | Rubrieke | Gemiddeld | Gemiddeld |
| Ellips | 1.76 | 1.06 | 1.45 | 1.55 | 1.36 | 1.46 | 1.45 |
| Posisie 1 | 0.16 | 0.12 | 0.14 | 0.20 | 0.12 | 0.16 | 0.15 |
| Posisie 2 | 0.40 | 0.26 | 0.34 | 0.49 | 0.46 | 0.47 | 0.40 |
| Posisie 3 | 0.75 | 0.30 | 0.55 | 0.44 | 0.43 | 0.43 | 0.49 |
| Posisie 4 | 0.46 | 0.37 | 0.42 | 0.43 | 0.35 | 0.39 | 0.40 |
| L. Kohesie | 10.39 | 9.12 | 9.82 | 10.41 | 10.87 | 10.63 | 10.22 |
| Posisie 1 | 1.01 | 0.99 | 1.00 | 1.51 | 1.13 | 1.33 | 1.16 |
| Posisie 2 | 2.66 | 2.03 | 2.38 | 2.62 | 2.67 | 2.64 | 2.51 |
| Posisie 3 | 2.50 | 2.32 | 2.42 | 2.42 | 2.62 | 2.51 | 2.46 |
| Posisie 4 | 4.22 | 3.78 | 4.02 | 3.87 | 4.45 | 4.15 | 4.08 |
| V/merkers | 5.96 | 7.12 | 6.48 | 7.42 | 7.40 | 7.41 | 6.93 |
| Posisie 1 | 1.24 | 2.04 | 1.60 | 1.88 | 1.84 | 1.86 | 1.72 |
| Posisie 2 | 1.54 | 1.78 | 1.65 | 2.34 | 2.27 | 2.31 | 1.97 |
| Posisie 3 | 2.02 | 1.74 | 1.89 | 1.90 | 2.07 | 1.98 | 1.94 |
| Posisie 4 | 1.17 | 1.56 | 1.34 | 1.30 | 1.22 | 1.26 | 1.30 |
| Verwysing | 10.65 | 10.94 | 10.78 | 8.45 | 9.93 | 9.16 | 9.99 |
| Posisie 1 | 2.37 | 2.10 | 2.25 | 2.24 | 2.30 | 2.27 | 2.26 |
| Posisie 2 | 3.23 | 3.37 | 3.29 | 2.14 | 2.85 | 2.48 | 2.90 |
| Posisie 3 | 2.50 | 2.66 | 2.57 | 1.85 | 2.06 | 1.95 | 2.27 |
| Posisie 4 | 2.55 | 2.81 | 2.67 | 2.21 | 2.73 | 2.46 | 2.57 |
| Hoeveelheid woorde per korpus | 8 977 | 7 249 | 16 226 | 7 991 | 7 441 | 15 432 | 31 658 |
| Persentasie van totale korpus | 28% | 23% | 51% | 25% | 24% | 49% | 100% |

tabel 6-4 wys die genormaliseerde hoeveelheid voorkoms van die onderskeie kohesieskakels in die onderskeie korpusse ten opsigte van die posisie wat hul op sinsvlak voorgekom het. Indien daar na afwykings in die patroon van voorkoms van 'n spesifieke kohesieskakel in óf 'n spesifieke skrywer se tekste óf 'n spesifieke genre in tabel 6-4 gesoek word, blyk dit dat daar 'n aantal reekse data wat moontlike afwykings toon, is. Data wat moontlik van 'n norm afwyk en verder bespreek gaan word, is in grys op die tabel gemerk. Vir leksikale kohesie blyk daar in sinsposisie twee en vier afwykings te wees terwyl daar vir die voorkoms van verbandsmerkers in sinsverband in posisie twee 'n afwyking opgelet kan word. Verder is daar ook vir verwysing 'n merkbare afwyking in die voorkoms in die tweede sinsposisie. Die aard van hierdie afwykings word vervolgens bespreek.

Indien die voorkoms van leksikale kohesie in sinsverband ontleed word, kan gesien word dat daar 'n amper identiese voorkoms per 100 woorde is tussen die tekste van beide skrywers, behalwe vir die rubrieke van André le Roux. In Dana Snyman se kortverhale en rubrieke kom leksikale kohesie onderskeidelik 2,62 en 2,67 keer per 100 woorde voor en in André le Roux se kortverhale, 2,66 keer. Hierteenoor is die voorkoms van leksikale kohesie in die tweede sinsposisie heelwat laer (2,03 keer per 100 woorde) in André le Roux se rubrieke. Sy kortverhale toon egter 'n baie soortgelyke voorkoms.

Die voorkoms van leksikale kohesie in die vierde sinsposisie toon nie 'n duidelike patroon nie; trouens, elke voorkoms blyk onafhanklik van enige veranderlikes te wees. André le Roux se kortverhale wys 'n hoër voorkoms van leksikale kohesie in die vierde sinsposisie as in sy rubrieke (4,22 per 100 woorde teenoor 3,78 per 100 woorde), terwyl die voorkoms in Dana Snyman se tekste die teenoorgestelde wys. In sy rubrieke kom leksikale kohesie 4,45 keer per 100 woorde voor terwyl dit slegs 3,87 keer per 100 woorde in sy kortverhale voorkom.

In terme van verbandsmerkers is daar 'n merkbaar hoër algemene voorkoms in die tweede sinsposisie van Dana Snyman se tekste. Dit strook met die algeheel hoër voorkoms van dié skakel in sy tekste (tabel 6-2). Die grootste deel van hierdie algehele hoër voorkoms, word deur die voorkoms in die tweede posisie uitgemaak met 'n gemiddelde voorkoms van 2,31 keer per 100 woorde (in die tweede sinsposisie) in Dana Snyman se tekste teenoor 1,65 keer per 100 woorde in (die tweede sinsposisie van) die tekste van André le Roux. Indien na die voorkoms binne die onderskeie genres gekyk word, is die voorkoms in André le Roux se kortverhale effens laer as in sy rubrieke (1,54 voorkoms per 100 woorde teenoor 1,78 voorkoms per

100 woorde) terwyl Dana Snyman se tekste die teenoorgestelde toon. In sy tekste is die voorkoms weer hoër in die kortverhale as in die rubrieke (2,34 keer per 100 woorde teenoor 2,27 per 100 woorde). In beide gevalle is die verskil tussen die voorkoms in die twee genres egter betreklik klein.

Indien daar na die verdeling van die voorkoms van verwysing op sinsvlak gekyk word, kan 'n mens sien dat André le Roux oor die algemeen meer van verwysing gebruik maak as Dana Snyman (sien tabel 6-2). Die meeste van hierdie hoë voorkomste word in die tweede en derde sinsposisie waargeneem waar die gemiddelde voorkoms in André le Roux se tekste telkens hoër is as vir Dana Snyman se tekste (Posisie 2: 3,29 per 100 woorde teenoor 2,48 per 100 woorde; Posisie 3: 2,57 per 100 woorde teenoor 1,95 per 100 woorde). In beide posisies, en in die tekste van beide skrywers, is die voorkoms in rubrieke hoër as in kortverhale - in die meeste gevalle betreklike min, maar in sinsposisie 2 vir die tekste van Dana Snyman, met 'n hoër voorkoms (2,85 per 100 woorde teenoor 2,14 per 100 woorde).

In terme van die posisie waarin die onderskeie kohesieskakels op sinsvlak voorkom, is daar dus 'n aantal afwykings van die algemene patroon. Hoewel die aard van die afwykings telkens verskil, kan dit op hierdie stadium aangedui word as data wat gebruik kan word in verdere statistiese analise om die grootte of beduidendheid van die afwyking te probeer bepaal. Hierdie moontlike afwykings sluit in:

- Leksikale kohesie in sinsposisie 2 en 4
- Verbandsmerkers in sinsposisie 2
- Verwysing in sinsposisie 2 en 3

In hierdie stadium kan ook opgelet word dat daar telkens afwykings in die tweede sinsposisie is. Indien daar dus vir aanduiders van uniekheid gesoek word, sou die voorkoms van die kohesieskakels wat in die tweede sinsposisie is, dalk per skrywer verskil. Hierdie veronderstelling sal in die verdere statistiese analise ondersoek word vir beduidendheid (vergelyk afdeling 6.3.2).

6.2.3 Posisionele voorkoms van kohesieskakels: Paragraafvlak

In terme van die posisie van verbandsmerkers (of enige skakel) op paragraafvlak, is die uitdaging vir die ontleder om 'n paragraaf in 'n sinvolle aantal dele te verdeel. Die gevaar is dat 'n té groot aantal verdelings nutteloos sal wees in 'n bespreking. Die verskil tussen byvoorbeeld posisie 9 uit 15 en posisie 10 of 11 uit 15, kan onbeduidend raak. Daar is gekies om die paragrawe in tiendes te verdeel en na die posisionele voorkoms van die onderskeie kohesieskakels per skrywer en per genre te kyk, soos daar met die sinsposisie gedoen is. Selfs met die tiende verdeling was daar reeds uitdagings om die data sinvol te interpreteer. Die grootste uitdaging is dat die voorkomste per paragraafposisie so klein word, dat dit moeilik is om te onderskei waar daar afwykings is. Om die bespreking gefokus te hou, is besluit om hoofsaaklik na afwykings in die gemiddelde voorkoms per skrywer te kyk en nie na posisionele data ook nie (in ooreenstemming met die primêre navorsingsvraag). Opvallende data-afwyking ten opsigte van genreverdeling sal egter ook genoem word. Die voorkoms in die paragrawe toon in die algemeen egter minder variasie tussen beide skrywers en genres. Tabel 6-5 wys die genormaliseerde hoeveelheid voorkomste van die onderskeie kohesieskakels ten opsigte van hul posisie binne die paragrawe van die onderskeie korpuse. Vir die bespreking van hierdie data, is die relevante inligting in tabel 6-5 in grys gemerk en is substitusie uitgelaat as gevolg van die betreklike lae algehele voorkoms van hierdie kohesieskakel

Tabel 6-5: Genormaliseerde data ten opsigte van die voorkoms van kohesieskakels op paragraafvlak

| | André le Roux | | ALR Totaal | ALR Gemiddeld | Dana Snyman | | DS Totaal | DS Gemiddeld | Totaal |
|--------------------------|---------------|-------------|--------------|---------------|-------------|-------------|--------------|--------------|--------------|
| | K/Verhale | Rubrieke | | | K/Verhale | Rubrieke | | | |
| Ellips | 1.74 | 0.85 | 2.58 | 1.29 | 1.36 | 1.11 | 2.47 | 1.24 | 5.06 |
| 1 | 0.22 | 0.05 | 0.27 | 0.14 | 0.14 | 0.10 | 0.24 | 0.12 | 0.52 |
| 2 | 0.16 | 0.09 | 0.25 | 0.13 | 0.13 | 0.10 | 0.23 | 0.12 | 0.48 |
| 3 | 0.24 | 0.04 | 0.29 | 0.14 | 0.14 | 0.11 | 0.25 | 0.13 | 0.54 |
| 4 | 0.14 | 0.05 | 0.20 | 0.10 | 0.15 | 0.11 | 0.26 | 0.13 | 0.46 |
| 5 | 0.11 | 0.10 | 0.21 | 0.10 | 0.12 | 0.10 | 0.22 | 0.11 | 0.43 |
| 6 | 0.16 | 0.10 | 0.26 | 0.13 | 0.14 | 0.18 | 0.32 | 0.16 | 0.58 |
| 7 | 0.16 | 0.11 | 0.27 | 0.14 | 0.10 | 0.04 | 0.14 | 0.07 | 0.42 |
| 8 | 0.12 | 0.12 | 0.24 | 0.12 | 0.16 | 0.12 | 0.29 | 0.14 | 0.53 |
| 9 | 0.26 | 0.08 | 0.34 | 0.17 | 0.11 | 0.15 | 0.26 | 0.13 | 0.60 |
| 10 | 0.14 | 0.10 | 0.24 | 0.12 | 0.15 | 0.10 | 0.25 | 0.13 | 0.49 |
| Leksikale Kohesie | 10.26 | 7.27 | 17.52 | 8.76 | 9.14 | 8.89 | 18.04 | 9.02 | 35.56 |
| 1 | 0.90 | 0.59 | 1.49 | 0.75 | 0.73 | 0.56 | 1.29 | 0.64 | 2.78 |
| 2 | 1.51 | 0.69 | 2.20 | 1.10 | 0.89 | 0.69 | 1.58 | 0.79 | 3.78 |
| 3 | 1.22 | 0.82 | 2.04 | 1.02 | 0.82 | 0.95 | 1.77 | 0.88 | 3.81 |
| 4 | 1.03 | 0.63 | 1.66 | 0.83 | 1.00 | 0.98 | 1.98 | 0.99 | 3.64 |
| 5 | 1.19 | 0.60 | 1.79 | 0.90 | 1.00 | 0.90 | 1.90 | 0.95 | 3.69 |
| 6 | 0.96 | 0.75 | 1.70 | 0.85 | 0.87 | 1.06 | 1.92 | 0.96 | 3.63 |
| 7 | 1.09 | 0.76 | 1.85 | 0.92 | 0.97 | 0.90 | 1.87 | 0.93 | 3.72 |
| 8 | 0.87 | 0.76 | 1.63 | 0.81 | 0.98 | 0.99 | 1.97 | 0.98 | 3.59 |
| 9 | 0.76 | 0.95 | 1.70 | 0.85 | 0.87 | 1.09 | 1.96 | 0.98 | 3.66 |
| 10 | 0.74 | 0.71 | 1.45 | 0.73 | 1.02 | 0.78 | 1.80 | 0.90 | 3.25 |

Tabel 6 5: Genormaliseerde data ten opsigte van die voorkoms van kohesieskakels op paragraafvlak (vervolg)

| | | | | | | | | | |
|---------------------------------------|--------------|--------------|--------------|-------------|--------------|--------------|--------------|-------------|--------------|
| Verbandsmerker | 5.88 | 5.67 | 11.55 | 5.78 | 6.52 | 6.06 | 12.57 | 6.29 | 24.13 |
| 1 | 0.59 | 0.47 | 1.07 | 0.53 | 0.52 | 0.44 | 0.96 | 0.48 | 2.02 |
| 2 | 0.49 | 0.58 | 1.08 | 0.54 | 0.62 | 0.47 | 1.09 | 0.54 | 2.17 |
| 3 | 0.60 | 0.52 | 1.12 | 0.56 | 0.62 | 0.52 | 1.13 | 0.57 | 2.25 |
| 4 | 0.65 | 0.55 | 1.20 | 0.60 | 0.66 | 0.55 | 1.21 | 0.60 | 2.41 |
| 5 | 0.59 | 0.60 | 1.20 | 0.60 | 0.69 | 0.60 | 1.30 | 0.65 | 2.50 |
| 6 | 0.58 | 0.59 | 1.18 | 0.59 | 0.74 | 0.70 | 1.44 | 0.72 | 2.62 |
| 7 | 0.54 | 0.60 | 1.14 | 0.57 | 0.65 | 0.59 | 1.24 | 0.62 | 2.39 |
| 8 | 0.62 | 0.67 | 1.29 | 0.64 | 0.76 | 0.75 | 1.51 | 0.75 | 2.79 |
| 9 | 0.64 | 0.59 | 1.23 | 0.62 | 0.67 | 0.82 | 1.49 | 0.75 | 2.73 |
| 10 | 0.57 | 0.48 | 1.06 | 0.53 | 0.60 | 0.60 | 1.21 | 0.60 | 2.26 |
| Verwysing | 10.51 | 8.72 | 19.22 | 9.61 | 7.42 | 8.12 | 15.54 | 7.77 | 34.77 |
| 1 | 1.20 | 1.00 | 2.20 | 1.10 | 0.80 | 0.74 | 1.54 | 0.77 | 3.74 |
| 2 | 1.03 | 0.97 | 2.00 | 1.00 | 0.74 | 0.79 | 1.53 | 0.76 | 3.53 |
| 3 | 1.24 | 0.84 | 2.08 | 1.04 | 0.79 | 0.78 | 1.57 | 0.79 | 3.65 |
| 4 | 1.20 | 0.91 | 2.11 | 1.06 | 0.78 | 0.88 | 1.66 | 0.83 | 3.77 |
| 5 | 0.99 | 0.97 | 1.96 | 0.98 | 0.68 | 0.80 | 1.48 | 0.74 | 3.44 |
| 6 | 1.12 | 0.84 | 1.96 | 0.98 | 0.68 | 0.86 | 1.54 | 0.77 | 3.50 |
| 7 | 0.93 | 0.84 | 1.77 | 0.88 | 0.76 | 0.75 | 1.51 | 0.75 | 3.28 |
| 8 | 0.96 | 0.99 | 1.95 | 0.97 | 0.67 | 0.84 | 1.51 | 0.75 | 3.45 |
| 9 | 0.81 | 0.66 | 1.47 | 0.74 | 0.74 | 0.77 | 1.51 | 0.75 | 2.98 |
| 10 | 1.02 | 0.71 | 1.74 | 0.87 | 0.78 | 0.92 | 1.70 | 0.85 | 3.44 |
| Hoeveelheid skakels per korpus | 2 591 | 2 063 | 4 654 | | 2 233 | 2 211 | 4 444 | | 9 098 |

Aangesien daar so baie data is, is dit belangrik om die algehele voorkomste in gedagte te hou. Indien André le Roux byvoorbeeld 'n algemene laer voorkoms van verbandsmerkers in sy tekste het (soos wat die geval is), bevestig 'n laer voorkoms in die onderskeie paragraafposisies net die algemene tendens en dui dit nie noodwendig op 'n afwyking nie. Indien daar egter te midde van 'n algemene hoër of laer voorkoms van 'n sekere kohesieskakel die teenoorgestelde in 'n sekere paragraafposisie of genre gebeur, dui dit egter op 'n afwyking. Met hierdie raamwerk in gedagte, kan die bespreking van die data in *tabel 6-5* meer sinvoller.

In *tabel 6-5*, is daar twee areas geïdentifiseer wat veral potensieel variasie op die algemene patroon toon, naamlik verwysing en ellips.

Die voorkoms van ellips in die tekste van die onderskeie skrywers is soortgelyk in hoeveelheid. Die genormaliseerde waardes dui aan dat 2,58 van elke 100 gevalle van kohesievoorkoms in André le Roux se tekste ellips is. Die voorkoms in Dana Snyman se tekste, is 2,47 keer per 100 gevalle van kohesievoorkoms. Per genre, kan egter gesien word dat hierdie soortgelyke totale voorkomste, nie eweredig tussen die tekste van verskillende genres van 'n skrywer versprei is nie, met die tekste van André le Roux wat groter variasie tussen genres toon (1,74 vir André le Roux se kortverhale en 0,85 vir sy rubrieke teenoor 1,36 en 1,11 vir Dana Snyman se kortverhale en rubrieke onderskeidelik). In daardie lig kan genoem word dat die verspreiding in paragraafposisie vyf en agt vir dieselfde rede van die algemene tendens afwyk (vergelyk grysgekleurde rye in *tabel 6-5*). Hierdie rede is dat dit nie die algemene verskil tussen die voorkoms in verskillende genres van André le Roux se tekste weerspieël nie, maar dat die voorkoms in hierdie twee posisies, vir beide genres van André le Roux se tekste, soortgelyk is aan mekaar, maar ook aan die voorkoms van Dana Snyman se tekste.

Die enigste ander skakel wat moontlik 'n afwyking toon, is verwysing. In paragraafposisie 8 is daar 'n afwyking van die algemene tendens. Volgens die algemene tendens moet daar meer voorkomste van verwysing in die kortverhale van André le Roux wees, as in die rubrieke. In hierdie posisie is die voorkoms in tekste van hierdie twee genres van die betrokke egter soortgelyk. Trouens, daar is 'n effense laer voorkoms in die kortverhale.

Soos gesien kan word, lyk dit op die oog af dat die verskille op paragraafvlak baie klein is en ook nie noodwendig sal help met die beantwoording van die navorsingsvraag nie. Soos met die voorkoms van kohesieskakels in sinsverband, lyk dit of André le Roux soms die manier wat

hy verbandsmerkers in paragraafverband gebruik, meer aanpas tussen die twee genres as wat Dana Snyman syne aanpas. Die navorser is van mening dat daar in terme van die voorkoms van kohesieskakels op paragraafvlak, met die bogenoemde interpretasie volstaan kan word en dat hierdie diagnostiese analise nie verdere statistiese verwerking noodsaak nie. Daar sal egter weer na hierdie resultate in die algemene bevindinge verwys word (sien tabel 6-15).

tabel 6-6 stel 'n opsomming voor van die resultate wat in hierdie afdeling bespreek is. Na onderskeidelik die chi-kwadraattoets en die binêre regressie-analise gaan daar telkens 'n kolom bygevoeg word om boek te hou van die resultate van die verdere statistiese analyses.

Tabel 6-6: Opsomming van moontlike afwykings wanneer die hoeveelheid voorkomste van kohesieskakels in ag geneem word

| | Watter data lyk of dit afwyk? | Aantal gevalle van kohesievoorkoms |
|---|--|---|
| Algemene voorkoms van kohesieskakels | Die ratio's van leksikale kohesie, verbandsmerkers en verwysing van die onderskeie skrywers | André le Roux se kohesieskakelgebruik per genre verskil. Die twee kohesieskakels wat grootliks hierdie verskynsel illustreer, is leksikale kohesie en verbandsmerkers. André le Roux maak meer gereeld van leksikale kohesie in sy kortverhale as in sy rubrieke gebruik, maar minder gereeld van verbandsmerkers. Dana Snyman se gebruik tussen verskillende genres is soortgelyk. |
| | Die ratio's van subkategorieë van leksikale kohesie van die onderskeie skrywers | In terme van leksikale kohesie lyk dit asof die hoeveelheid wat André le Roux beide tipes leksikale kohesie in rubrieke gebruik, afwyk van die patroon waarvolgens dit in die ander subkorpuse voorkom. |
| | Die ratio's van subkategorieë van verbandswoorde (sonder tydsaanduidende verbandswoorde) van onderskeie skrywers | André le Roux se rubrieke wyk in al die subkategorieë behalwe tydsaanduidende verbandsmerkers, in 'n mindere of 'n meerdere mate af van die tendens vir die ander verbandsmerkers. In die geval van tydsaanduidende verbandsmerkers, is die voorkomste in beide genres van André le Roux laer as in die tekste van Dana Snyman. |
| | Die ratio's van subkategorie van verwysing van onderskeie skrywers | André le Roux gebruik in beide kortverhale en rubrieke meer voornaamwoordverwysing as Dana Snyman. In die afdeling van eienaamverwysing is dit opmerklik dat die kortverhale van André le Roux die minste voorkomste per 100 woorde bevat, maar sy rubrieke die meeste. |

| | | |
|---|---|--|
| Posisie van kohesieskakels in sinsverband | Leksikale kohesie: Sinsposisie 2 en 4 | Indien die voorkoms van leksikale kohesie in sinsverband ontleed word, kan gesien word daar 'n amper identiese voorkoms per 100 woorde is tussen die tekste van beide skrywers, behalwe vir die rubrieke van André le Roux (die voorkoms is laer in die tweede sinsposisie). Geen patroon lyk duidelik vir die vierde sinsposisie nie. |
| | Verbandsmerker: Sinsposisie 2 | In terme van verbandsmerkers is daar 'n merkbaar hoër algemene voorkoms in die tweede sinsposisie van Dana Snyman se tekste. Dit strook met die algemene hoër voorkoms van dié skakel in sy tekste. Die grootste deel van hierdie algehele hoër voorkoms, word egter deur die voorkoms in die tweede posisie uitgemaak |
| | Verwysing: Sinsposisie 2 en 3 | Indien daar na die verdeling van die voorkoms van Verwysing op sinsvlak gekyk word, kan 'n mens sien dat André le Roux oor die algemeen meer van Verwysing gebruik maak as Dana Snyman (sien tabel 6-3). Die grootste hoeveelheid van hierdie verhoogde voorkoms word in die tweede en derde sinsposisie waargeneem waar die gemiddelde voorkoms in André le Roux se tekste telkens hoër is as vir Dana Snyman se tekste |
| | Algemene opmerking: Tweede sinsposisie lyk soos 'n posisie waar moontlik variasie plaasvind | |
| Posisie van kohesieskakels in paragraafverband | Ellips: Paragraafposisie 5 en 8 | In daardie lig kan genoem word dat die verspreiding in paragraafposisie vyf en agt vir dieselfde rede van die algemene tendens afwyk. Hierdie rede is dat dit nie die algemene verskil tussen die voorkoms in verskillende genres van André le Roux se tekste weerspieël nie, maar dat die voorkoms in hierdie twee posisies, vir beide genres van André le Roux se tekste, soortgelyk is aan mekaar, maar ook aan die voorkoms van Dana Snyman se tekste. |
| | Verwysing: Paragraafposisie 8 | Volgens die algemene tendens moet daar meer voorkoms van Verwysing in die kortverhale van André le Roux wees, as in die rubrieke. In hierdie posisie is die voorkoms in tekste van hierdie twee genres van die betrokke egter soortgelyk. Trouens daar is 'n effense laer voorkoms in die kortverhale. |
| | Algemene opmerking: Die navorser is van mening dat daar in terme van die voorkoms van kohesieskakels op paragraafvlak, met die aanvanklike analise van genormaliseerde hoeveelhede volstaan kan word en dat hierdie diagnostiese analise nie verdere statistiese verwerking ondersteun nie. | |

6.3 Beskrywende data: Chi-kwadraattoets

Soos reeds in die inleiding van hierdie hoofstuk vermeld is (Afdeling 6.1), is daar na afloop van die ontleding van die hoeveelheid gevalle van kohesievoorkoms, nog statistiese ontleding op sekere dele van die datastel gedoen. Die eerste is die uitvoer van 'n chi-kwadraattoets en die tweede, die uitvoer van 'n logistiese regressie-analise. Soos in afdeling 5.8.1 beskryf, is die doel van 'n chi-kwadraattoets om te bepaal of daar 'n statisties beduidende verskil tussen twee

datastelle is. Dit is belangrik om daarop te let dat 'n statisties beduidende verskil nie noodwendig op 'n prakties beduidende verskil dui nie (dus sal twee datastelle wat statisties beduidend is, nie noodwendig ooglopend van mekaar verskil nie). Hierby is die chi-kwadraattoets sensitief vir groot hoeveelhede data. 'n Statisties beduidende verskil word makliker verkry, hoe groter die datastel is. In 'n datastel van 10 000 datapunte, kan 'n afwyking van 1% tussen veranderlikes op 'n statisties beduidende verskil dui, aangesien 1% van 10 000, 100 voorkomste van 'n sekere veranderlike is en 100 voorkomste as 'n groot verskil gesien kan word. Daarom word die chi-kwadraattoets tesame met ander statistiese bewerkings geïnterpreteer. Die opsomming wat tabel 6-6 gegee word, sal as riglyn vir die uitvoer van die chi-kwadraattoets gebruik word.

6.3.1 Algemene voorkoms van kohesieskakels

'n Chi-kwadraattoets is gedoen met die data wat op die oog af lyk of dit afwykings toon van 'n patroonmatige voorkoms. Die rede vir die chi-kwadraattoets is om die navorser in staat te stel om 'n meer genuanseerde interpretasie van die statistiese bevindinge te gee. Soos reeds voorheen genoem is, word geen van die statistiese beskrywings op sy eie gebruik om 'n interpretasie aangaande 'n outeur se styl te maak nie, maar word 'n aantal bewerkings gebruik om die interpretasies toe te lig.

Die webplatform *Social Science Statistics* stel die navorser in staat om 'n chi-kwadraattoets uit te voer wat ook 'n chi-kwadratstatistiek aan elke veranderlike toeken en nie net aan die reeks veranderlikes nie. Dit is sinvol om kennis te neem dat chi-kwadraattoets altyd met rou data of geobserveerde hoeveelhede uitgevoer word. Dit word genoem aangesien die data in die vorige afdeling as genormaliseerde hoeveelhede aangebied is. Dit is nie 'n probleem nie, aangesien die aard van die twee stelle berekeninge (genormaliseerde hoeveelhede en chi-kwadratstatistiek) in hul beskrywingsmoontlikhede verskil.

Die eerste stel chi-kwadraattoets wat uitgevoer is, is uitgevoer om te bepaal of die algemene voorkoms van die kohesieskakels in die twee skrywers se tekste, statisties beduidend van mekaar verskil. tabel 6-7 toon die resultate van hierdie toets. Daar kan gesien word dat die twee reekse data (André le Roux s'n en Dana Snyman s'n), statisties beduidend van mekaar verskil aangesien die p-waarde vir die reeks kleiner is as 0,05. Wat egter ook van belang is, is die

getalle wat tussen blokhakies aangedui is. Dit is die sogenaamde kritiese waardes wat die chi-kwadraat aan van daardie spesifieke kohesieskakel aandui. Dit dui dus aan of die voorkoms van daardie spesifieke kohesieskakel as 'n enkele veranderlike binne die reeks statisties beduidend verskil na gelang van die skrywer. Vir die voorkoms van vyf veranderlikes (of $df^{31}=4$; soos hier die geval is) is die kritiese waarde vir die chi-kwadraat wat 'n p-waarde laer as 0,05 sal veroorsaak, 9,40. Alle chi-kwadraathoeveelhede hoër as daardie getal, sal 'n p-waarde laer as 0,05 oplewer en dus op statistiese beduidendheid dui. In hierdie geval kan 'n mens sien dat verwysing as veranderlike op sy eie, statisties beduidend verskil na gelang van watter een van die twee skrywers die betrokke teks geskryf het. Hoewel leksikale kohesie en verbandsmerkers se chi-kwadraat aansienlik hoër is as ellips en substitusie s'n, is die chi-kwadraat nie hoog genoeg om op hul eie op 'n statisties beduidende verskil tussen die skrywers se tekste te dui nie. Dit is egter sinvol om te noem dat die afsnypunt van $p<0,05$ algemeen aanvaar word om te beteken dat daar met 95% sekerheid gesê kan word dat die verskil tussen twee voorkomste van 'n veranderlike (in hierdie geval 'n spesifieke kohesieskakel) toegeskryf kan word aan die betrokke onafhanklike veranderlike (in hierdie geval die betrokke skrywer). Daar is ook chi-kwadraathoeveelhede vir 'n laer vlak van sekerheid. Vir $p<0,1$, en dus 90% sekerheid, is die kritiese waarde 4,602 wat omtrent is waar verbandsmerkers se chi-kwadraathoeveelheid lê. So hoewel daar met meer as 95% sekerheid gesê kan word dat die voorkoms van Verwysing in die tekste afhang van die teksproduseerder, kan daar met 90% sekerheid gesê word dat die voorkoms van verbandsmerkers ook van die teksproduseerder afhang.

Tabel 6-7: Chi-kwadraattoets: Algemene voorkoms van die kohesieskakels in die twee skrywers se tekste

| | Ellips | L/kohesie | Sub. | V/merkers | Verwysing | Totale | Chi-kwadraat |
|---------------|------------|--------------|-----------|-------------|--------------|------------------------------|--|
| ALR | 235 [0.00] | 1594 [2.24] | 25 [0.17] | 1051 [4.60] | 1749 [10.61] | 4654 | Chi-kwadraat: 36.0487 p-waarde: 0.00001 |
| DS | 225 [0.00] | 1641 [2.34] | 20 [0.18] | 1144 [4.81] | 1414 [11.11] | 4444 | |
| Totaal | 460 | 3 235 | 45 | 2195 | 3163 | 9 098 (groototaal) | |

³¹ df: "degrees of freedom"

6.3.1.1 Algemene voorkoms van subkategorieë van kohesieskakels

Die volgende stel chi-kwadraattoets is uitgevoer om te bepaal of die voorkoms van die subkategorieë binne leksikale kohesie statisties beduidend per skrywer van mekaar verskil. tabel 6-8 gee die resultate van daardie toets. Soos in die vorige analise, die chi-kwadraatstatistiek (chi-kwadraat=12,3669) 'n p-waarde van kleiner as 0,05 ($p=0,000437$) vir die kombinasie van veranderlikes. In hierdie geval is die chi-kwadraatstatistiek wat aan leksikale kohesie: kollokasie toegeken word, ook statisties beduidend aangesien die chi-kwadraatstatistiek daarvoor as 4,73 en 4,6 aangedui word, die kritiese waarde vir $p<0,05$, is 3,841. Aangesien die statistieke vir leksikale kohesie:kollokasie groter is as die kritiese waarde, dui dit op statistiese beduidendheid vir daardie veranderlike as individuele veranderlike, ongeag van die reeks

Tabel 6-8: Chi-kwadraattoets: Algemene voorkoms van subkategorie van Leksikale Kohesie in die tekste van die onderskeie skrywers

| | LK:Koll | LK:Sem | Totaal | Chi-kwadraat |
|---------------|------------|--------------|----------------------------|---|
| ALR | 435 [4.73] | 1157 [1.54] | 1592 | Chi-kwadraat: 12,3669 p-waarde: 0.000437 |
| DS | 360 [4.60] | 1277 [1.50] | 1637 | |
| Totaal | 795 | 2 434 | 3 229 (Groottotaal) | |

Die derde chi-kwadraatanalise wat uitgevoer is, is op die subkategorieë van verbandwoorde uitgevoer. Tydsaanduidende verbandwoorde is in die berekeninge uitgelaat aangesien die genormaliseerde hoeveelhede nie gedui het op 'n afwyking in daardie subkategorie nie. Die data in tabel 6-9 wys die resultate van hierdie chi-kwadraattoets. Vir hierdie reeks het die chi-kwadraattoets 'n p-waarde vir 0,000541 opgelewer wat dui op 'n statisties beduidende verskil tussen die twee reekse data. Die kritiese waarde vir die chi-kwadraatstatistiek om 'n p-waarde van kleiner as 0,05 vir hierdie hoeveelheid veranderlikes op te lewer, is 5,99. Geen van die hoeveelhede wat tussen blokhakies aangetoon is, is meer as dit nie; dus kan geen van die subkategorieë op hul eie 'n statisties beduidende aanduiding van outeurskap gee nie.

Tabel 6-9: Chi-kwadraattoets: Algemene voorkoms van subkategorieë van Verbandswoorde in die tekste van die onderskeie skrywers

| | Aaneensk | Redegewend | Teenstellend | Totaal | Chi-kwadraat |
|---------------|------------|------------|--------------|-------------------------------------|---|
| ALR | 465 [2.90] | 211 [3.84] | 200 [0.71] | 876 | Chi-kwadraat: 15,0441 p-waarde: 0,000541 |
| DS | 531 [2.96] | 154 [3.92] | 173 [0.72] | 858 | |
| Totaal | 996 | 365 | 373 | 1 734 (Groototaal) | |

Die laaste chi-kwadraattoets wat in terme van die algemene voorkoms van kohesieskakels gedoen is, is uitgevoer met die data ten opsigte van die subkategorieë van verwysing, naamlik voornaamwoord- en eienaamverwysing. In die voorafgaande afdeling is opgelet dat daar veral binne tekste van verskillende genres van André le Roux, afwyking getoon is ten opsigte van hierdie skakels. Daarom is daar 'n ander benadering tot die chi-kwadraat veranderlikes geneem ten einde te bepaal of die onderskeid tussen genres wel statisties beduidend is, soos wat in die genormaliseerde data gesuggereer is. Daar is dus nie bepaal of hierdie twee subkategorieë beduidend verskil vir die twee skrywers nie, maar wel of daar binne beide genres van beide skrywers variasie is in die gebruik van hierdie twee subkategorieë van verwysing.

Tabel 6-10: Chi-kwadraattoets: Voorkoms van subkategorieë van Verwysing in tekste van die onderskeie genres vir die onderskeie skrywers

| | ALR Kortverhale | ALR Rubrieke | DS Kortverhale | DS Rubrieke | Totaal | Chi-kwadraat |
|---------------------------|-----------------|--------------|----------------|-------------|-------------------------------------|---|
| Verwysing: vnw | 904 [2.26] | 854 [1.89] | 604 [0.04] | 850 [0.00] | 3212 | Chi-kwadraat: 43,344 p-waarde: 0,00001 |
| Verwysing: eienaam | 48 [21.11] | 137 [17.65] | 70 [0.35] | 89 [0.04] | 344 | |
| Totaal | 952 | 991 | 674 | 939 | 3 556 (Groototaal) | |

tabel 6-10 dui die resultate aan van die chi-kwadraattoets vir die voorkoms van voornaamwoord en eienaamverwysing binne die tekste van die onderskeie genres van die onderskeie skrywers. Soos gesien kan word, verskil die twee reekse data statisties beduidend van mekaar (chi-kwadraat: 43,344; p-waarde=0,00001). Wanneer na die chi-kwadraatstatistieke vir die onderskeie afdelings gekyk word, kan gesien word dat die grootste verskil lê in die voorkoms van eienaamverwysing in André le Roux se kortverhale (chi-kwadraat 21,11 en 17,65). Die kritiese waarde vir 'n chi-kwadraatstatistiek om 'n kleiner p-waarde as 0,05 te lewer vir 'n

datastel van hierdie grootte is 7,815. Hierdie twee voorkomste is amper alleen verantwoordelik vir die hoë algemene chi-kwadraatstatistiek vir hierdie reeks.

6.3.2 Posisionele voorkoms van kohesieskakels: Sinsverband

Na afloop van die bespreking van die aanvanklike beskrywende data in afdeling 6.2.2, is daar 'n aantal voorstelle gemaak vir datareekse wat aan verdere statistiese ontleding onderwerp kan word om die statistiese beduidendheid daarvan te bepaal. Die eerste analise is vir die voorkoms van leksikale kohesie in sinsverband, met spesiale aandag aan die voorkoms daarvan in die tweede en vierde sinsposisie. tabel 6-11 dui die resultate van hierdie chi-kwadraattoets aan.

Tabel 6-11: Chi-kwadraattoets: Voorkoms van Leksikale Kohesie in die sinne van tekste van die onderskeie skrywers

| | Posisie 1 | Posisie 2 | Posisie 3 | Posisie 4 | Totaal | Chi-kwadraat |
|---------------|------------|------------|------------|-------------|--------------------------------------|--|
| ALR | 163 [1.85] | 386 [0.07] | 392 [0.15] | 653 [0.40] | 1594 | Chi-kwadraat: 4,8725 p-waarde: 0,181379 |
| DS | 205 [1.80] | 408 [0.07] | 388 [0.15] | 640 [0.39] | 1641 | |
| Totaal | 368 | 794 | 780 | 1293 | 3 235 (Groottotaal) | |

In hierdie tabel kan gesien word dat die chi-kwadraatstatistiek vir hierdie reeks data, 'n p-waarde van 0,181379 oplewer, wat nie op 'n statisties beduidende verskil dui nie. Dit is vir wanneer die algemene verspreiding in sinsverband in André le Roux se tekste met Dana Snyman se tekste vergelyk word. In afdeling 6.2.2 is daar genoem dat die variasie dalk verklaar kan word as genreverwant en daarom is 'n addisionele chi-kwadraattoets uitgevoer wat die geïdentifiseerde posisies (sinsposisie 2 en 4) se verspreiding vir die onderskeie genres van die onderskeie tekste toets. Ook hierdie toets het statisties onbeduidende resultate opgelewer soos in tabel 6-12 gesien kan word, Die chi-kwadraatwaarde is 3,072 wat tot 'n p-waarde van 0,380658 lei. Die twee reekse data verskil dus nie statisties beduidend van mekaar nie.

Tabel 6-12: Chi-kwadraattoets: Voorkoms van leksikale kohesie in die tweede en vierde sinsposisie in tekste van die onderskeie genres van die onderskeie skrywers

| | ALR KV | ALR Rub | DS KV | DS Rub | Totaal | Chi-kwadraat |
|-----------------------------|------------|------------|------------|------------|----------------------------|---|
| Leksikale kohesie P2 | 239 [0.06] | 147 [1.08] | 209 [0.72] | 199 [0.03] | 794 | Chi-kwadraat: 3,072 p-waarde: 0,380658 |
| Leksikale kohesie P4 | 379 [0.04] | 274 [0.66] | 309 [0.44] | 331 [0.02] | 1293 | |
| Totaal | 618 | 421 | 518 | 530 | 2 087 (Groottotaal) | |

Die tweede chi-kwadraattoets wat gedoen is, is gedoen vir die voorkoms van verbandsmerkers in sinsverband vir die tekste van die onderskeie skrywers. tabel 6-13 sit die resultate van hierdie toets uiteen. Soos gesien kan word, wys die p-waarde ($p=0,009252$) dat die twee reekse data beduidend van mekaar verskil. Die kritieke waarde vir $p<0,05$ vir hierdie datastel is 11,513. Waardes groter as dit, gee 'n p-waarde van kleiner as 0,05. In die lig hiervan kan ook nou geïnterpreteer word dat die voorkoms van verbandsmerkers in geen enkele posisie 'n statisties beduidende verskil tussen die tekste van die twee skrywers bied nie, hoewel die twee reekse in hul geheel statisties beduidend van mekaar verskil.

Tabel 6-13: Chi-kwadraatstatistiek: Voorkoms van Verbandsmerkers in die vier sinsposisies van die onderskeie skrywers

| | Pos 1 | Pos 2 | Pos 3 | Pos 4 | Totaal | Chi-kwadraat |
|---------------|------------|------------|------------|------------|----------------------------|--|
| ALR | 259 [0.02] | 267 [3.28] | 307 [0.62] | 218 [2.07] | 1051 | Chi-kwadraat: 11,513 p-waarde: 0.009252 |
| DS | 287 [0.02] | 356 [3.02] | 306 [0.57] | 195 [1.90] | 1144 | |
| Totaal | 546 | 623 | 613 | 413 | 2 195 (Groottotaal) | |

Die laaste chi-kwadraattoets wat vir die voorkoms van kohesieskakels in sinsverband gedoen is, is vir Verwysing gedoen. Hier is weer gekyk of die algemene voorkoms van Verwysing in die onderskeie sinsposisies, statisties beduidend van mekaar verskil al dan nie. tabel 6-14 wys dat die reeks wel statisties beduidend van mekaar verskil (chi-kwadraat=12,0299; p-waarde=0,007281). Die kritiese waarde vir 'n datastel met hierdie hoeveelheid veranderlikes is 7,815. Geen chi-kwadraatstatistieke vir die voorkoms in die onderskeie posisies gee individueel 'n waarde groter as die kritieke waarde nie, en daarom is geen van die voorkomste binne 'n enkele posisie statisties beduidend van tekste van die spesifieke skrywer nie, al verskil die twee reekse statisties beduidend van mekaar.

Tabel 6-14: Chi-kwadraattoets: Voorkoms van Verwysing in die sinne in tekste van die onderskeie genres van die onderskeie skrywers

| | Pos 1 | Pos 2 | Pos 3 | Pos 4 | Totaal | Chi-kwadraat |
|---------------|------------|------------|------------|------------|--------------------------------------|---|
| ALR | 365 [2.33] | 534 [1.43] | 417 [1.01] | 433 [0.61] | 1749 | Chi-kwadraat: 12,0299 p-waarde: 0,007281 |
| DS | 350 [2.88] | 383 [1.77] | 301 [1.24] | 380 [0.75] | 1414 | |
| Totaal | 715 | 917 | 718 | 813 | 3 163 (Groottotaal) | |

6.3.3 Opsomming: Beskrywende data

Aan die einde van afdeling 6.2 is daar in tabel 6-6 'n opsomming gegee van die resultate van die data wat in daardie afdeling bespreek is. In tabel 6-15 word die opsomming van die bespreking in die voorafgaande afdeling by die inligting in tabel 6-6 gevoeg. Enkele voorbeelde word by relevante bevindinge gegee ter illustrasie.

Tabel 6-15: Opsomming van resultate van beskrywende data

| | Watter data lyk of dit afwyk? | Hoeveelheid voorkomste | Chi-kwadraattoetse |
|--------------------------------------|--|---|---|
| Algemene voorkoms van kohesieskakels | Die ratio's van leksikale kohesie, verbandsmerkers en verwysing van die onderskeie skrywers | André le Roux se kohesieskakelgebruik per genre verskil. Die twee kohesieskakels wat grootliks hierdie verskynsel illustreer is leksikale kohesie en verbandsmerkers. André le Roux maak meer gereeld van leksikale kohesie in sy kortverhale as in sy rubrieke gebruik maar minder gereeld van verbandsmerkers. Dana Snyman se gebruik tussen verskillende genres is soortgelyk. | Stel data wat al 5 kohesieskakels insluit, verskil statisties beduidend van mekaar. Binne die reeks self, verskil verwysing statisties beduidend na gelang van die skrywer ($p < 0,05$) en verbandsmerkers verskil met 'n p-waarde van om en by 0,01. Leksikale kohesie op sy eie verskil nie statisties beduidend nie. |
| | Die ratio's van subkategorieë van leksikale kohesie van die onderskeie skrywers | In terme van leksikale kohesie lyk dit asof die hoeveelheid wat André le Roux beide tipes leksikale kohesie in rubrieke gebruik, afwyk van die patroon waarvolgens dit in die ander subkorpuse voorkom. | Die reeks wat die twee subkategorieë insluit, verskil statisties beduidend na gelang van skrywer ($p < 0,05$). Leksikale kohesie: kollokasie, verskil op sy eie ook statisties beduidend na gelang van skrywer. |
| | Die ratio's van subkategorieë van verbandswoorde (sonder tydsaanduidende verbandswoorde) van onderskeie skrywers | André le Roux se rubrieke wyk in al die subkategorieë behalwe tydsaanduidende verbandsmerkers, in 'n mindere of 'n meerdere mate af van die tendens vir die ander verbandsmerkers. In die geval van tydsaanduidende verbandsmerkers, is die voorkomste in beide genres van André le Roux laer as in die tekste van Dana Snyman. | Die reeks van die 4 subkategorieë verskil statisties beduidend na gelang van die skrywer ($p < 0,05$). Geen Verbandswoord verskil egter individueel statisties beduidend nie. |

| | | | |
|--|---|--|--|
| | Die ratio's van subkategorieë van verwysing van onderskeie skrywers | André le Roux gebruik in beide kortverhale en rubrieke tekste meer voornaamwoordverwysing as Dana Snyman. In die afdeling van eienaamverwysing is dit opmerklik dat die kortverhale van André le Roux die minste voorkomste per 100 woorde bevat, maar sy rubrieke die meeste. | Die reeks van die twee subkategorieë verskil statisties beduidend. Verwysing: Eienaam verskil veral in die voorkomste in die onderskeie genres van André le Roux se tekste (ook statisties beduidend; $p < 0,05$) |
| Posisie van kohesiekakels in sinsverband | Leksikale kohesie: Sinsposisie 2 en 4 | <p>Indien die voorkoms van leksikale kohesie in sinsverband ontleed word, kan gesien word daar 'n amper identiese voorkoms per 100 woorde is tussen die tekste van beide skrywers, behalwe vir die rubrieke van André le Roux (die voorkoms is laer in die tweede sinsposisie). Geen patroon lyk duidelik vir die vierde sinsposisie nie.</p> <p>Voorbeelde uit oorspronklike tekste:</p> <p><u>Leksikale kohesie (sinsposisie 2):</u></p> <p><i>Net 'n luggie stoot op van die see.</i></p> <ul style="list-style-type: none"> • Uit “Hulle kan gaan vrek” van André le Roux (kortverhaal) • Die antesedent “lug” het in 'n vorige sin voorgekom. <p><u>Leksikale kohesie (sinsposisie 4):</u></p> <p>Dis een ding van die, sê hy, hulle is nie so statusbehep as die Gautengers nie.</p> <ul style="list-style-type: none"> • Uit “Kaap toe” van Dana Snyman (kortverhaal) • Die antesedent “Gauteng” het in die vroeër in die teks, voorgekom. | Algemene reekse per skrywer verskil nie statisties beduidend van mekaar nie. 'n Verdeling wat slegs die twee sinsposisies en die vier subkorpuse vir leksikale kohesie in ag neem, verskil ook nie statisties beduidend nie. |

| | | |
|---|--|---|
| <p>Verbandsmerkers: Sinsposisie 2</p> | <p>In terme van verbandsmerkers is daar 'n merkbaar hoër algemene voorkoms in die tweede sinsposisie van Dana Snyman se tekste. Dit strook met die algemene hoër voorkoms van dié skakel in sy tekste. Die grootste deel van hierdie algehele hoër voorkoms, word egter deur die voorkoms in die tweede posisie uitgemaak.</p> <p>Voorbeelde uit oorspronklike tekste:</p> <p>Verbandsmerkers (sinsposisie 2):</p> <p><i>Vloog Theron word tronk toe gestuur, <u>maar</u> die verwante klag n teen Doktor word laat vaar.</i></p> <ul style="list-style-type: none"> • Uit 'n rubriek deur Dana Snyman wat op 7 November 2012 in Die Volksblad verskyn het. <p><i>En imagine 'n jaar nadat hulle ge-disappear het, het ek nog afbetaal.</i></p> <ul style="list-style-type: none"> • Uit die kortverhaal “Fabian van der Merwe will move, deur Dana Snyman. | <p>Reeks verskil statisties beduidend van mekaar. Die voorkoms in geen enkele posisie verskil egter statisties beduidend binne die reeks nie.</p> |
| <p>Verwysing: Sinsposisie 2 en 3</p> | <p>Indien daar na die verdeling van die voorkoms van Verwysing op sinsvlak gekyk word, kan 'n mens sien dat André le Roux in die algemeen meer van Verwysing gebruik maak as Dana Snyman (sien tabel 6-3). Die grootste hoeveelheid van hierdie verhoogde voorkoms word in die tweede en derde sinsposisie waargeneem waar die</p> | <p>Reeks verskil statisties beduidend van mekaar. Die voorkoms in geen enkele posisie verskil egter statisties beduidend binne die reeks nie.</p> |

| | | | |
|--------------------------------------|---|--|--------------------------|
| | | <p>gemiddelde voorkoms in André le Roux se tekste telkens hoër is as vir Dana Snyman se tekste.</p> <p>Voorbeelde uit oorspronklike tekste:</p> <p><u>Verwysing (sinsposisie 2):</u></p> <p><i>Ons sak in ons stoele terug, <u>ek</u> en Pa, en sit stil stil na die foto en kyk.</i></p> <ul style="list-style-type: none"> • Uit die kortverhaal “Familiefoto” deur Dana Snyman <p><u>Verwysing (sinsposisie 3):</u></p> <p><i>Van chalet tot chalet met <u>my</u> aktetas voor jou deur.</i></p> <ul style="list-style-type: none"> • Uit ’n rubriek in deur André le Roux wat in Beeld op 8 Januarie 2002 verskyn het. deur Andre le Roux | |
| | Algemene opmerking: Tweede sinsposisie lyk soos ’n posisie waar moontlik variasie plaasvind | | |
| Posisie van kohesieskakels in | <p>Ellips: Paragraafposisie 5 en 8</p> | <p>In daardie lig kan genoem word dat die verspreiding in paragraafposisie vyf en agt vir dieselfde rede van die algemene tendens afwyk. Hierdie rede is dat dit nie die algemene verskil tussen die voorkoms in verskillende genres van André le Roux se tekste weerspieël nie, maar dat die voorkoms in hierdie twee posisies, vir beide genres van André le Roux se tekste, soortgelyk is aan mekaar, maar ook aan die voorkoms van Dana Snyman se tekste.</p> | Geen Chi-kwadraatanalise |

| | | | |
|--|---|--|---------------------------------|
| | <p>Verwysing: Paragraafposisie 8</p> | <p>Volgens die algemene tendens moet daar meer voorkomste van verwysing in die kortverhale van André le Roux wees, as in die rubrieke. In hierdie posisie is die voorkoms in tekste van hierdie twee genres van die betrokke egter soortgelyk. Trouens, daar is 'n effense laer voorkoms in die kortverhale.</p> | <p>Geen Chi-kwadraatanalise</p> |
| | <p>Algemene opmerking: Die navorser is van mening dat daar in terme van die voorkoms van kohesieskakels op paragraafvlak, met die aanvanklike analise van genormaliseerde hoeveelhede volstaan kan word en dat hierdie diagnostiese analise nie verdere statistiese verwerking regverdig nie.</p> | | |

6.4 Binêre logistiese regressie-analise

Die derde tipe statistiese analise wat gebruik is, is 'n binêre logistiese regressie-analise. 'n Volledige uiteensetting van hierdie statistiese metode is in afdeling 5.8.2 gegee. Daar kan nou volstaan word deur te sê dat hierdie tipe analise gebruik word wanneer een van twee moontlike uitkomstes op grond van 'n veranderlike of veranderlikes voorspel wil word. In die geval van hierdie studie, wil ondersoek word tot watter mate 'n sekere stel veranderlikes kan voorspel of 'n teks deur André le Roux of deur Dana Snyman geskryf is.

Die beperkinge wat in afdeling 5.8.2 bespreek is, is ook hier ter sake en kan gedeeltelik geïllustreer word deur die statistiek wat volgens die nulmodel beskikbaar is. Die nulmodel is die model wat deur die sagteware *SPSS Statistics* voorgestel word as een waarin die veranderlikes nie in ag geneem word nie. Dus wys dit die waarskynlikheid dat jy die uitkomstes reg sou raai sonder om spesifieke veranderlikes in ag te neem. Aangesien daar in hierdie studie se datastel omtrent ewe veel datapunte in totaal vir elke skrywer is (met ander woorde 'n soortgelyke aantal ewelheid voorkomste van kohesieskakels) gaan hierdie persentasie meestal naby 50% lê. Indien jy na 'n teks kyk en na geen veranderlike kyk nie, en raai wie die teks geskryf het, gaan jou kans naby 50% wees om reg te raai. Hierdie statistiek word in die volgende afdeling bespreek, waarna die statistiek van die onderskeie modelle met spesifieke veranderlikes wat as moontlike voorspellers voorgestel is (in afdeling 6.2 en 6.3), aangebied sal word.

Op grond van die analyses wat in afdeling 6.2 en 6.3 gedoen is, is daar twee modelle wat moontlik statisties beduidende uitslae as voorspeller van die betrokke teks se skrywer sou kon gee. Hierdie twee modelle is:

1. die algemene voorkoms van die kohesieskakels (substitusie uitgesluit), en
2. die algemene voorkoms van die subkategorieë van die kohesieskakels

Die resultate van hierdie twee analyses word vervolgens bespreek.

6.4.1 Binêre logistiese regressie-analise: Voorstelling van nulmodelle

Die voorstelling van basiese statistiek op grond van die nulmodelle is dubbeldoelig. Eerstens gee dit 'n verwysingspunt om die meer spesifieke modelle teen te meet, maar soos gesien sal word, illustreer dit ook die problematiek van hierdie statistiese metode vir hierdie studie.

In tabel 6-16 word die akkuraatheid van die onderskeie nulmodelle, die statistiese beduidendheid van die model (of die model 'n statisties beduidende voorspeller van 'n sekere uitkoms is) en die hoeveelheid datapunte wat vir elke model ter sake is, saamgevat. Vir die data wat in die eerste model ter sake is, gee die nulmodel 'n voorspellingswaarde van 51,2%. Dit beteken prakties dat jy 51,2% kans het om reg te raai (wie die skrywer is) indien jy geen kohesieskakels of ander veranderlike in ag neem nie. Hierdie getal word bepaal op grond van die aantal algemene voorkomste per skrywer as persentasie van die hele datastel. Vir die tweede model is die voorspellingswaarde dieselfde. Die rede hiervoor is dat die data in hierdie twee datastelle grootliks dieselfde is in die *aantal kere* wat elke skrywer die onafhanklike veranderlike (kohesieskakel of subkategorie van kohesieskakel) gebruik; al is die onafhanklike veranderlike in elke datastel verskillend.

'n Tipiese uitdaging van hierdie tipe analise, kom egter voor by die statisties beduidendheid van die nulmodelle. Soos gesien kan word, bied elkeen van die nulmodelle 'n uitkoms wat as statisties beduidend geklassifiseer kan word ($\text{sig.} < 0.05$). Dit beteken dat die model, selfs sonder die veranderlikes, 'n voorspelling van die uitkoms kan gee, wat 'n statisties beduidend beter voorspelling is as om net te raai, al is die voorspelling slegs 51,1%. Die rede vir hierdie verskynsel is dieselfde as wat in die geval van die chi-kwadraattoetse die geval was. Omdat die datastel so groot is ($n=8593$, $n=8576$ en $n=5351$), kan selfs 'n variasie wat op die oog af klein lyk as statisties beduidend beskou word omdat dit deur baie datapunte beïnvloed moet word vir 'n variasie van selfs net een persentasiepunt.

Tabel 6-16: Voorstelling van nulmodelle van onderskeie logistiese regresse-analises

| Model | Persentasie akkuraatheid | Statisties beduidendheid van nulmodel ($p < 0,05$) | n |
|---|--------------------------|--|-------|
| 1. Effek van algemene voorkoms van kohesiemerkers as voorspeller van skrywer | 51,2 % | 0,028 | 9 098 |
| 2. Effek van die voorkoms van die subkategoriee van leksikale kohesie, verbandswoorde en verwysing as voorspeller van skrywer | 51,2% | 0,028 | 9 098 |

6.4.2 Binêre logistiese regresse-analise: Model 1

Tabel 6-17 stel die resultate voor van die logistiese regresse-analise van die algemene voorkoms van leksikale kohesie, verbandsmerkers en verwysing. Die tabel is in die formaat

wat *SPSS Statistics* resultate vir hierdie statistiese toets gee. Daar is bykomende inligting onder die tabel ingevoeg wat ook deur *SPSS Statistics* gegee word.

Indien 'n studie op soek is na 'n spesifieke veranderlike of veranderlikes wat 'n invloed op die voorspellingswaarde van 'n uitkoms is, val die fokus normaalweg op die data in die kolom, 'Exp(B)'. Die statistiese aanduidings in hierdie kolom gee 'n aanduiding van die waarskynlikheid dat die voorkoms van hierdie betrokke veranderlike een uitkoms bo 'n ander kan verseker. Indien 'n studie byvoorbeeld gedoen word waarin daar gepoog word om studentesukses te voorspel, word deur 'n aantal veranderlikes in ag te neem. Indien daar byvoorbeeld twee moontlikhede ten opsigte van die uitkoms, naamlik, suksesvol of onsuksesvol en 'n aantal moontlikhede ten opsigte van die veranderlikes, byvoorbeeld demografiese oorwegings, vorige akademiese prestasie, sosio-ekonomiese agtergrond, ensovoorts. Die hoë of lae punte wat 'n spesifieke student voorheen behaal het, kan dus beskryf word as iets wat sy/haar kans op studentesukses met byvoorbeeld een of twee of drie keer verhoog. In hierdie studie, kan dieselfde egter nie van 'n enkele voorkoms van 'n kohesiemerker gesê word nie. 'n Enkele voorkoms van byvoorbeeld verwysing is nie genoeg om die kanse dat die voorkoms in 'n spesifieke skrywer se teks voorkom te verhoog nie, die kollektiewe voorkoms van hierdie merkers, egter wel. Daarom sal daar nie in hierdie bespreking van die binêre regressie-analise op hierdie Exp(B) gefokus word nie, maar op aanduidings wat vir die datastel in geheel geld.

Eerstens val die fokus op modelakkuraatheid. In tabel 6-16 is aangedui dat die nulmodel 51,2% akkuraat is. In tabel 6-17 kan gesien word dat die model se akkuraatheid tot 52,7% styg wanneer die betrokke veranderlikes in ag geneem word. Die voorspelling per skrywer lyk egter effens anders. Vir hierdie datastel is hierdie model meer suksesvol as voorspeller van Dana Snyman as die teksproduseerder as vir André le Roux. Volgens hierdie model kan Dana Snyman se tekste 62,7% van die tyd van André le Roux s'n onderskei word, terwyl die persentasie vir André le Roux se tekste, slegs 43,2% is.

Ondanks die betreklike klein verbetering op die modelakkuraatheid van die nulmodel, word hierdie model steeds as 'n statisties beduidende verbetering op die nulmodel gesien (Model chi-kwadraat: 53,770; $p < 0,05$). Die Nagelkerke pseudo R^2 -waarde dui aan hoeveel van die verbetering van modelakkuraatheid waarskynlik aan die veranderlikes in die model te danke is. Die waarde van 0.005 kan voorgestel word as 'n variasie van 0,5%.

Tabel 6-17: Logistiese regressie-analise van die algemene voorkoms van leksikale kohesie, verbandsmerkers en verwysing: SPSS analise

| Veranderlikes in die Vergelyking | | | | | | | |
|---|-------------------|-------|------|-------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Stap 1 ^a | Leksikale Kohesie | .088 | .096 | .854 | 1 | .355 | 1.093 |
| | Verbandsmerkers | .144 | .099 | 2.132 | 1 | .144 | 1.155 |
| | Verwysing | -.153 | .096 | 2.549 | 1 | .110 | .858 |
| | Konstante | -.020 | .190 | .011 | 1 | .916 | .980 |
| Modelakkuraatheid = 52,7% Akkuraatheid per skrywer: Dana Snyman: 62,7% André le Roux: 43,2% Model chi-kwadraat = 53,770 ; p = 0,000 Pseudo R ² (Nagelkerke) = 0.005 N = 9098 | | | | | | | |

6.4.3 Binêre logistiese regressie-analise: model 2

Tabel 6-18 stel die resultate voor van die regressie-analise op die onderafdelings wat binne leksikale kohesie, verbandsmerkers en verwysing geklassifiseer is. Die tabel is in die formaat hoe die statistiese verwerkingsagteware, *SPSS Statistics*, die resultate na afloop van die analise verskaf. Daar is bykomende inligting wat ook deur *SPSS Statistics* gegee word, wat onderaan die tabel ingevoeg is. Vir die doel van hierdie bespreking, word grootliks op hierdie inligting gefokus. afdeling 6.4.1 gee die rasionaal vir hierdie werkswyse.

Eerstens val die fokus op modelakkuraatheid. In tabel 6-16 is aangedui dat die nulmodel vir hierdie stel veranderlikes 51,2% akkuraat is. In tabel 6-18 kan gesien word dat die model se akkuraatheid tot 54,7% styg wanneer die betrokke subkategorieë as model gebruik word om voorspellings oor die skrywer van 'n sekere teks te maak. Dit is effens meer as die 52,7% suksessyfer van die model wat slegs die oorhoofse kategorieë van kohesieskakels bevat (vergelyk afdeling 6.4.2). Net soos met die vorige model, is daar egter meer variasie in die akkuraatheid waarvolgens die model tekste van een van die twee skrywers voorspel. Terwyl die vorige model Dana Snyman se tekste meer akkuraat kon voorspel, voorspel hierdie model André le Roux se tekste meer akkuraat. Volgens hierdie model kan André le Roux se tekste

61,3% van die tyd van Dana Snyman s'n onderskei word, terwyl die persentasie vir Dana Snyman se tekste 47,1% is.

Soos in die geval van die vorige model, kan hierdie model ten spyte van die oënskynlike klein verbetering op die akkuraatheid van die nulmodel ook as 'n statisties beduidende verbetering daarvan gesien word (chi-kwadraat: 79,770; $p < 0,05$). Die Nagelkerke pseudo R^2 -waarde dui aan dat 1,1% van die verbetering van modelakkuraatheid, waarskynlik aan die veranderlikes in die model te danke is.

Tabel 6-18: Logistieke regressie-analise van die algemene voorkoms van die subkategorieë binne leksikale kohesie, verbandsmerkers en verwysing

| Veranderlikes in die Vergelyking | | | | | | | |
|----------------------------------|---|-------|------|--------|----|------|--------|
| | | B | S.E. | Wald | df | Sig. | Exp(B) |
| Stap 1 ^a | Verwysing: Bepaald | -.597 | .105 | 32.275 | 1 | .000 | .550 |
| | Verbandswoorde: Teenstellend | -.525 | .143 | 13.532 | 1 | .000 | .591 |
| | Verbandswoorde: Aaneenskakelend | -.193 | .116 | 2.740 | 1 | .098 | .825 |
| | Verbandswoorde: Redegewend | -.695 | .144 | 23.183 | 1 | .000 | .499 |
| | Leksikale Kohesie: Semanties | -.282 | .106 | 7.043 | 1 | .008 | .754 |
| | Leksikale Kohesie: Kollokasie | -.570 | .121 | 22.078 | 1 | .000 | .566 |
| | Verwysing: Eienaam | -.557 | .147 | 14.462 | 1 | .000 | .573 |
| | Konstante | 3.039 | .626 | 23.544 | 1 | .000 | 20.887 |
| | Modelakkuraatheid = 54,7% Dana Snyman: 47,1% André le Roux: 61,3% Model chi-kwadraat = 79,308 ; $p = 0,000$ Pseudo R^2 (Nagelkerke) = 0.011 N = 9098 | | | | | | |

6.5 Samevatting

In hierdie hoofstuk is die bevindinge van die verskillende statistiese analises van die data gerapporteer. Daar is deurgaans gepoog om by te hou met die implikasies wat die statistiese toetsresultate vir die beantwoording van die navorsingsvrae in hierdie studie inhou. Die volledige beantwoording van die navorsingsvrae kom onder meer aan bod in hoofstuk 7. Wat wel reeds gesê kan word, is dat die voorkoms van kohesieskakels soos dit in hierdie studie

ondersoek is, grootliks ooreenstem tussen die vier korpusse. Die meeste tekste in 'n spesifieke taal, behoort trouens ooreenkomste in kohesieskakels te bevat. Dit maak dit moeilik om wel moontlike variasie op hierdie manier te identifiseer. Volledige bevindinge en aanbevelings volg in die volgende hoofstuk.

7 Samevatting, beperkinge en aanbevelings

7.1 Inleiding

Hierdie studie het ondersoek ingestel na die skryfstyl van twee populêre Afrikaanse skrywers, André le Roux en Dana Snyman. Daar is spesifiek gefokus op die wyse waarop hulle kohesiemerkers in hul tekste gebruik. In hierdie hoofstuk word 'n samevatting van die studie gegee deur spesifiek te fokus op die doelwitte en doelstellings, die literatuurstudie en empiriese ondersoek en die beantwoording van die navorsingsvrae. Hierna word melding gemaak van die beperkinge van die studie sowel as aanbevelings vir verdere navorsing.

7.2 Doelstellings en doelwitte van die studie

Die oorhoofse doelstelling van die studie was om te bepaal tot watter mate daar 'n voorkeur vir die gebruik van bepaalde kohesieskakels in die tekste van André le Roux en Dana Snyman sigbaar is. Linguistiese elemente word dikwels ondersoek in outeurskapidentifikasiestudies en in hierdie studie is daar ondersoek ingestel na die gebruik van kohesieskakels. Hierdie doelstelling kan sistematies uiteengesit word, deur die doelwitte wat in hoofstuk 1 uiteengesit is. Die ondersoek het ten doel gehad om te bepaal:

- of spesifieke kohesieskakels patroonmatig met 'n bepaalde skrywer se skryfstyl verbind kan word;
- in welke mate 'n individu se stilistiese voorkeur in terme van kohesiemerkers deur die reëls van 'n betrokke genre beïnvloed word.

7.3 Literatuurstudie

Voordat die empiriese ondersoek van die studie kon plaasvind, is 'n deeglike studie van beskikbare literatuur gedoen om te dien as teoretiese grondslag. Hierdie afdeling het oor drie hoofstukke gestrek (hoofstuk 2, 3 en 4) ten einde drie akademiese studieveld wat vir hierdie studie relevant is, te omskryf. Hierdie oorsig het styl, stilistiek en outeurskapidentifikasie, die studieveld van tekslinguistiek en teorieë oor kohesie en kohesieskakels sowel as literatuur in verband met genre en tekstipe ingesluit.

Die belangrikste tema wat uit hoofstuk 2 duidelik geword het, is die belang van empiriese navorsing in die veld van outeurskapidentifikasie. Die literatuur het telkens daarop gedui dat 'n teksproduseerder 'n identifiseerbare styl het, maar dat metodes wat tot dusver gebruik is om die styl te meet of te kwantifiseer, nie geldige en dupliseerbare resultate gelewer het nie. In hoofstuk 3 is die onderskeie kohesieskakels uiteengesit, hoofsaaklik met die doel om die omvang van elk af te baken, sodat daar konsekwente beoordelings kan plaasvind in die analiseproses. Die aard van die terme *genre* en *tekstipe* is in hoofstuk 4 bespreek, sowel as die wesenswaard van rubrieke en kortverhale. Uit hierdie bespreking het dit duidelik geword dat hierdie twee genres met die eerste oogopslag hoofsaaklik in terme van inhoud of temas verskil. Dit sou interessant wees om te probeer bepaal of daar ook tekstuele of grammatikale verskille (soos die voorkoms van kohesieskakels) tussen die twee genres is.

7.4 Die empiriese ondersoek

In hoofstuk 5 is die navorsingsmetodologie uiteengesit. Daar is eerstens verduidelik dat 'n korpusgebaseerde benadering gevolg is ten einde die navorsingsvrae te beantwoord. Daar is ook verduidelik dat die data kwalitatief van aard is, maar dat dit uiteindelik kwantitatief geanaliseer sou word. Dit is bereik deur 'n driefase analiseproses. Die uiteindelijke kwantitatiewe data is deur middel van statistiese metodes geanaliseer. Beskrywende statistiek, insluitende die gebruik van chi-kwadraattoetse sowel as 'n binêre logistiese regressie-analise, is gebruik.

Die drie fases van die analiseproses is so ontwerp om uit die kwalitatiewe data (kortverhaal- en rubriektekste) kwantitatiewe data te kry ten einde statistiese verwerkings daarvan te doen. Tydens die eerste fase is woorde of frases wat op 'n manier kohesief aan 'n ander woord of frase verbind is, gemerk. Hierdie etiketteringsproses het slegs aangedui aan watter antesedent die betrokke woord of frase verwant is en het nog nie gepoog om die tipe kohesieskakel aan te dui nie. Hierna is die tekste omgesit in txt-formaat en in WordSmith 7 ingevoer en elemente wat geëtiketteer is, is geïsoleer. Eers nadat die gevalle van kohesievoorkoms geïsoleer is, is die voorkomste daarvan gekategoriseer volgens die tipe kohesieskakel. Hierdie kategorieë, tesame met die data wat deur WordSmith verskaf is ten opsigte van posisie in die teks, is gebruik vir die statistiese ontleding wat in hoofstuk 6 bespreek is.

7.5 Beantwoording van navorsingsvrae

Na afloop van die statistiese ontleding, blyk dit dat die antwoorde op die navorsingsvrae meer kompleks is as wat aanvanklik geantisipeer is. Dit strook egter met die literatuur in outeurskapstudies (vergelyk hoofstuk 2) asook die hipotesetoetsende aard van die studie en bied moontlikhede vir verdere navorsing.

Die navorsingsvrae wat in afdeling 1.3 van hierdie studie uiteengesit is, het uit 'n primêre en twee sekondêre navorsingsvrae bestaan.

1. Primêre navorsingsvraag: Kan die gebruik van kohesieskakels in die werk van Dana Snyman en André le Roux 'n aanduiding van outeurskap gee?
2. Sekondêre navorsingsvrae:
 - a) Beïnvloed die stilistiese genrekonvensies van rubrieke en kortverhale Dana Snyman en André le Roux se gebruik van kohesieskakels?
 - b) Is daar algemene kohesiepatrone wat as kenmerkend van onderskeidelik rubrieke en kortverhale beskou kan word?

7.5.1 Navorsingsvraag 1

Die gebruik van kohesieskakels in die werk van Dana Snyman en André le Roux kan tot 'n mate 'n aanduiding van outeurskap gee. Daar is egter 'n aantal inkonsekwentheid wat veroorsaak dat die kohesieskakels nie as oorhoofse model gebruik kan word om tekste van beide skrywers ewe suksesvol en met relatiewe akkuraatheid te kan klassifiseer nie. Hierdie uitdagings sluit die volgende in:

- In die korpusse wat in hierdie studie ondersoek is, word kohesieskakels die meeste van die tyd op 'n relatief soortgelyke manier gebruik, ongeag skrywer of genre.
- Daar is dus nie 'n voorafbepaalde manier waarop moontlike variasie op 'n outomatiese of rekenaarmatige manier ondersoek kan word om óf die skrywer óf die genre te probeer identifiseer nie.
- Die feit dat die gebruik van so 'n groot aantal kohesieskakels ooreenstem, veroorsaak dat die statistiese metodes relatiewe klein effekgroottes oplewer omdat die faktore wat wel afwyk baie moeilik is om te isoleer.

Wat wel op grond van die data en die skrywers se skryfstyl gesê kan word, is dat:

- die voorkoms van leksikale kohesie, verbandsmerkers en verwysing (binne hierdie korpora) die duidelikste aanduiders (in terme van kohesieskakelgebruik) van die styl van 'n spesifieke skrywer of genre is,
- die subkategorieë van die bogenoemde kohesieskakels groter voorspellingswaarde as die oorhoofse kategorieë self het,
- die aantal kere wat 'n spesifieke veranderlike in die tekste van die onderskeie skrywers voorkom, 'n beter voorspeller van 'n teks van die betrokke skrywer is, as die posisie waarin daardie skakel op sinsvlak en/of paragraafvlak voorkom.
- Dana Snyman se kortverhale en rubrieke meer soortgelyk in terme van kohesieskakelgebruik is as André le Roux s'n,
- André le Roux ten opsigte van sekere kohesieskakels sy styl waarneembaar aanpas na gelang van die genre. Die duidelikste voorbeeld hiervan is eienaamverwysing.

7.5.2 Navorsingsvraag 2a

André le Roux pas sy styl meer aan na gelang van die genre as Dana Snyman (vergelyk afdeling 6.4). Of dit die genre-konvensie is wat dit veroorsaak of die skrywer se persoonlike doelwitte wanneer hy binne die betrokke genre skryf, is egter nie uit die data sigbaar nie.

7.5.3 Navorsingsvraag 2b

Die data ondersteun nie afwykings van 'n kohesieskakel slegs binne 'n sekere genre nie. Wanneer daar 'n spesifieke kohesieskakel is wat verskillend in een genre as 'n ander genre van 'n spesifieke skrywer voorkom, is dit nie dieselfde vir die twee genres van die ander skrywer nie. Die afwykings op die vlak van genres is dus nie veralgemeenbaar vir hierdie datastel nie.

7.6 Beperkinge van die studie

Die twee belangrikste beperkinge hou verband met die analise van kohesieskakelvoorkoms, asook statistiese analise. Daar is 'n element van subjektiwiteit met die identifikasie van sommige kohesieskakels (soos leksikale kohesie) en daarom is dit moontlik dat dit analise van verskillende navorsers variasie sou kon toon.

Die statistiese analisemetodes is waarskynlik die grootste beperking van die studie. Die belangrikste kwessie is dat die gepaste statistiese analise vir die toets van die voorkoms of verspreiding van linguistiese verskynsels binne 'n teks, steeds onduidelik is. Hierdie uitdaging is reeds in die literatuuroorsig (in afdeling 2.3.3) voorspel. Selfs nouer samewerking met statistici en uitgebreide kennis van die statistiese analise deur die navorser, sou in die toekoms die effek van hierdie beperking kon minimaliseer.

7.7 Implikasies van die huidige ondersoek

Aangesien die studie tot 'n groot mate hipotesetoetsend was en daar met verskeie aspekte van die ondersoekproses geëksperimenteer is (analisemetodes, keuse van linguistiese veranderlike, statistiese verwerkingsmetodes, ensovoorts) is daar 'n aantal afleidings wat na afloop van hierdie ondersoek gemaak kan word. Dit sluit die volgende in:

- a) Daar is 'n mate van variasie in die wyse waarop die spesifieke twee skrywers kohesiemiddele in hul skryfwerk inspan. Die juisste manier om hierdie variasie te meet, is egter nie seker nie.
- b) 'n Groot aantal van die kohesiemerkervoorlooms vir die onderskeie tekste is soortgelyk ('n mens sou dalk kon sê standaard). Om werklik die variasie te kan meet, sal daar óf tydens die analiseproses óf met die keuse van statistiese ontleding, keuses gemaak moet word om die variasie te kan isoleer ten einde die aard omvang daarvan te kan ondersoek. As die oorgrote meerderheid (argumentsonthouwe 90%) van die kohesiegebruik tussen skrywers soortgelyk is, maar die variasie in die oorblywende 10% voorkom, moet die metodologiese besluite en keuses van statistiese ontleding so wees, dat dit die 10% van die 90% kan skei alvorens die aard van die variasie bepaal kan word. Bogenoemde implikasie strook met die stilistiese benadering tot outeurskapstudies waartydens die linguïst 'n dokument analiseer deur eerstens onreëlmatighede te merk en dít dan te gebruik as basis vir interpretasie en nie die standaardgebruik van verskynsels ook in ag neem nie (hierdie benadering is volledig bespreek in afdeling 2.2 en 2.3).
- c) In die lig van die voorafgaande paragraaf (7.7. b), kan 'n benadering wat aanvanklik 'n stilistiese werkswyse het waarna bevindinge/afleidings gebruik word as basis vir statistiese ontleding, sinvol wees.
- d) Die statistiese analisemetodes wat in hierdie studie gebruik is, is 'n gedeeltelike, maar nuttige basis vir verdere ondersoek vir die linguistiese analise. Die metode se

inagneming van al die moontlike data (sien paragraaf 7.7. b) veroorsaak uitdagings in die interpretasie van effekgrootte van voorspellings.

7.8 Aanbevelings vir verdere navorsing

Deur die loop van die studie en na afloop van die data-analise en -bespreking is daar 'n aantal verwante temas binne die veld geïdentifiseer waarvoor verdere navorsing onderneem kan word. Hierdie temas sluit die volgende in:

- a) Die datastel sou uitgebrei en/of verander kon word om te bepaal of daar 'n groter mate van variasie is in die voorkoms van kohesieskakels in tekste van 'n ander genre of skrywer. Veral die gebruik van 'n ander genre, wat op die oog af baie van kortverhale of rubrieke verskil (bv. akademiese artikels of poësie) sou interessante resultate kon oplewer.
- b) Met die inset van 'n statistikus (erkenningstelsel), sou dieselfde data aan ander statistiese analisemetodes onderwerp kon word, ten einde te bepaal of daar metodes is wat meer bruikbaar is vir die doel van studies soos hierdie (sien paragraaf 7.7. d vir konteks).
- c) 'n Ander linguïstiese element (as kohesieskakels) sou op dieselfde wyse ondersoek kon word, ten einde te bepaal of dit 'n duideliker merker van outeurskap is.
- d) 'n Soortgelyke studie sou in 'n ander taal soos Engels onderneem kan word, om te bepaal tot watter mate die voorkoms van kohesieskakels in verskillende tale varieer. Op hierdie manier sou ook ondersoek kon word of die aard en mate van variasie 'n outeurspesifieke of 'n taalspesifieke verskynsel is. Resultate van so 'n studie sou toepassingsmoontlikhede kon hê in tweedetaalonderrig.

7.9 Samevatting en slot

In hierdie hoofstuk is 'n samevatting gegee van die studie deur die belangrikste elemente uit die literatuurstudie, metodologie, analise en data-interpretasie uit te lig. Die navorsingsvrae is beantwoord en beperkinge van die studie is bespreek. Hierna is aanbevelings gemaak vir moontlike verdere navorsing. Met hierdie studie het die navorser gepoog om empiriese data by te voeg ter ondersteuning van die soeke na 'n standaardmetodologie vir die oplos van outeurskapstudie-vraagstukke. Empiriese studies soos hierdie een kan help om die lens waarmee daar gekyk word na outeurskapkwessies te fokus ten einde uiteindelik tekste akkuraat en betroubaar as behorende tot 'n sekere outeur te kan identifiseer.

Bronnelys

- Baayen, H., Halteren, H. V., Neijt, H., & Tweedie, A. 2002. An Experiment in Authorship Attribution. *IEEE Intelligent Systems & their Applications - IEEE INTELL SYST APPL.* Beskikbaar by https://www.researchgate.net/publication/253253587_An_experiment_in_authorship_attribution [10 Januarie 2020]
- Babbie, F., & Mouton, J. 1998. *The Practice of Social Research*. Cape Town: Oxford University Press.
- Biber, D. 1988. *Variation across speech and writing*. Cambridge University Press: Cambridge.
- Biber, D. 2012. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In *The Oxford Handbook of Linguistic Analysis*. Oxford University press. <https://doi.org/10.1093/oxfordhb/9780199544004.013.0008>
- Biber, D., & Conrad, S. 2019. *Register, Genre and Style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Brainerd, B. 1973. On the Distinction Between a Novel and a Romance: A Discriminant Analysis. *Computers and the Humanities*, 7: 259-270.
- Brainerd, B. 1974. *Weighing Evidence in Language and Literature: A Statistical Approach*. Toronto: University of Toronto Press.

- Brown, J. 2004. Research Methods for Applied Linguistics, Scope, Characteristics and Standards. In A. Davies, & C. Elder (Eds.), *The Handbook of Applied Linguistics*. Oxford: Blackwell Publishing. 476-500.
- Bruno, A.M. 1974. *Toward a Qualitative Methodology for Stylistic Analysis*. Berkeley, University of California Press.
- Carstens, W.A.M. 1993. Die studieterrein van die tekslinguistiek. *Suid-Afrikaanse tydskrif vir Taalkunde*, Supplement 15:1-14. doi:10.1080/10118063.1993.9724330
- Carstens, W.A.M. 1997. *Afrikaanse Tekslinguistiek: 'n Inleiding*. Pretoria: Van Schaik Uitgewers.
- Carstens, W.A.M., & Van de Poel, K. 2010. *Teksredaksie*. Spellenbosch: Africa Sun.
- Carter, R. 1982. *Language and literature: An introductory reader in stylistics*. New York: Harper Collins Publishers Ltd.
- Coulthard, M. 2004. Author identification, idiolect and linguistic uniqueness. *Applied linguistics*. doi:10.1093/applin/25.4.431.
- Coulthard, M. 2005. Some forensic applications of descriptive linguistics. *VEREDAS - Revista de Estudos Linguísticos*, 9(1-2), 9-28.
- Craig, H. 2004. Stylistic analysis and Authorship Studies. In Schreibman, S., Siemens, R. & Unsworth, J. (Eds.), *Companion to Digital Humanities*. Oxford: Blackwell Publishing. 273-288.
- Crystal, D. 2008. *A Dictionary of Linguistics and Phonetics*. London: Blackwell Publishing.

- De Beaugrande, R. s.j.. Text Linguistics. In J. Zienkowski, J. Östman, J. Verschueren, J. Zienkowski, J. Ostman, & J. Verscheueren Reds., *Discursive Pragmatics*. Amsterdam: John Benjamins Publishing Company. 286-295.
- De Beaugrande, R., & Dressler, W. 1981. *Introduction to Text Linguistics*. London: Longman.
- Oxford Dictionaries. 2018. What can the Oxford English Corpus tell us about the English language? <https://en.oxforddictionaries.com/explore/what-can-corpus-tell-us-about-language/> [5 Julie 2018]
- Eggs, S. 2004. *An introduction to systemic functional linguistics*. New York: Continuum International Publishing Group.
- Engelbrecht, H. 1981. *Saterdagand Tuis*. Uitgewery Oranje.
- Flowerdew, L. 2005. An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24(3), 321-332. DOI: <https://doi.org/10.1016/j.esp.2004.09.002>.
- Fouché, N. 2016. *Kohesiemerkegebruik in die skryfwerk van Afrikaanssprekende graad 6- en graad 9-leerders*. Ongepubliseerde MA. Noordwes-Universiteit
- Fucks, W. 1952. On the mathematical analysis of style. *Biometrika*, 39:122-129.
- Giltrow, J., & Stein, D. 2009. Genres in the Internet: Innovation, evolution, and genre theory. In J. Giltrow, & D. Stein, *Genres in the internet: issues in the theory of genre*. Amsterdam: John Benjamins Publishing Co.

- Gouws, T. 2013. Kortkortverhaal. *Literêre terme en teorieë*. Beskikbaar by <http://www.literaryterminology.com/index.php/lemmas/18-k/95-kortkortverhaal> [24 Julie 2015].
- Graesser, A., McNamara, D., Louwse, M., & Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behaviour research methods, instruments & computers: A journal of the Psychonomic Society, Inc*, 36(2), 193-202. doi:10.3758/BF03195564
- Grieve, J. 2007. Quantitative Authorship Attribution: An Evaluation of Techniques. *Literary and Linguistic Computing*, 223, 251-270.
- Halliday, M.A.K. 1975. Learning how to mean: Explorations in the development of language. London: Edward Arnold.
- Halliday, M.A.K., & Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Halliday, M.A.K., & Matthiessen, C. M. 2014. *Halliday's Introduction to Functional Grammar*. Oxford: Routledge.
- Hasan, R., & Cloran, C. 1990. A sociolinguistic interpretation of everyday talk between mothers and children. In M. G. Halliday Red., *Learning, keeping and using language: Selected papers from the 8th world congress of applied linguistics*, Sydney, 16-21 Augustus 1987. Philadelphia: John Benjamins Publishing Company. 67-100.
- Holmes, D.I. 1994. Authorship Attribution . *Computers and Humanities*, 28, 87-101.
- Holmes, D.I., & Forsyth, R. S. 1994. The Federalist Revisited: New directions in authorship attribution. *Literary and Linguistic Computing*, 102, 111-127. Beskikbaar by <http://llc.oxfordjournals.org/> herwin [29 Januarie 2013].

- Hyland, D. 2002. Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics*, 34, 1091-1112.
- Hyon, S. 1996. Genre in three traditions: Implications for ESL. *TESOL Quarterly*, 304. Beskikbaar by <http://www.jstor.org/stable/3587930> [2 Julie 2015].
- Janjua, F. 2012. Cohesion and Meanings. *Canadian Social Science*, 82, 149-155.
- Johnson, A., & Wright, D. 2014. Identifying idiolect in forensic authorship attribution. *Language and the Law*, 11, 37-69.
- Jordaan, A. 2014. *Die gebruik van metadiskoers in Afrikaans T1-skryfwerk van eerstejaar-universiteitstudente*. Ongepubliseerde MA. Noordwes-Universiteit.
- Joula, P. 2008. Authorship Attribution. *Foundations and Trends in Information Retrieval*, 13, 233-334.
- Joula, P. 2012. Large scale experiments in authorship attribution. *English studies*, 933, 275-283. Beskikbaar by <http://dx.doi.org/10.1050/0013838X.2012.668792> [16 Desember 2014].
- Joula, P., Sofko, J., & Brennan, P. 2006. A Prototype for Authorship Attribution Studies. *Literary and Linguistic Computing*, 212, 169-178. Beskikbaar by van <http://llc.oxfordjournals.org/> herwin [8 Desember 2014].
- Junghaenel, D. U., Smyth, J. M., & Santner, L. 2008. Linguistic dimensions of Psychopathology. *Journal of Social and Clinical Psychology*, 271, 36-55. Beskikbaar by <https://doi.org/10.1521/jscp.2008.27.1.36> [20 Augustus 2018].

- Kirsten, J. 2012. Onthou: Dis jou opinie! In R. Scheepers, & L. Kleyn, *Die Afrikaanse Skryfgids*. Johannesburg: Penguin.
- Kress, G. 1989. *Linguistic processes in sociocultural practice*. Oxford: Oxford University Press.
- Le Roux, A. 2008. *Kom ons sit sommer by die tafel*. Kaapstad: Human & Rousseau.
- Lee, D. 2001. Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language learning & Technology*, 53, 37-72. Beskikbaar by <http://lt.msu.edu/vol5num3/lee/> [21 Junie 2016].
- Louwerse, M. 2004. Semantic Variation in Idiolect and Sociolect: Corpus Linguistic Evidence from Literary Texts. *Computers and Humanities*, 38, 207-221.
- Louwerse, M., McCarthy, P., McNamara, D., & Graesser, A. 2004. *Variation in Language and Cohesion across Written and Spoken Registers*. Proceedings of the Annual Meeting of the Cognitive Science Society, 2626. Beskikbaar by <http://escholarship.org/uc/item/7d8631cr> [14 Mei 2018]
- Malmkær, K. 1991. *The Linguistics Encyclopedia*. New York: Routledge.
- Martin, J.R. 1983. Conjunction: the logic of English text. In Petöfi, J.S. en Sözer, E. (Red.) *Micro and Macro Connexity of Texts*. Hamburg: Helmut Buske Verlag, 1-72.
- Martin, J.R. 1984. Language, register and Genre. In Christie, F. (Red) *Children Writing: A reader*. Geelong, Vic.: Deakin University Press, 21-29.
- Matthiessen, C. M. 2007. The 'architecture' of language according to systemic functional theory: developments since the 1970s. In R. Hasan, C. Matthiessen, & J. J. Webster

- Reds., *Continuing discourse on language: A functional perspective*. London: Equinox. 505-561.
- McEnery, T., Xiao, R., & Tono, Y. 2006. *Compus-based language studies: An advanced resource book*. New York: Routledge.
- McMenamin, G. R. 1993. *Forensic stylistics*. Amsterdam: Elsevier.
- Meintjes, Z. 2014. Koherensie in die argumentatiewe skryfwerk van eerstejaarstudente: 'n teksgebaseerde ondersoek. Ongepubliseerde PhD: Noordwes-Universiteit.
- Mendenhall, T.C. 1887. The characteristic curves of composition. *Science*, IX, 237-249.
- Mosteller, F. & Wallace, D.I. 1964. *Interference and Disputed Authorship: The Federalist*. Lesing, MA: Addison-Wesley.
- Mouton, J., & Marais, H. 1990. *Basiese begrippe: metodologie van die geesteswetenskappe*. Pretoria: RGN-Uitgewers.
- Nascione, A. 2010. *Stylistic use of phraseological units in discourse*. Amsterdam: John Benjamins.
- Nel, M. 2014. *Joan Hambidge se idiolek oor die grense van genres: 'n korpuslinguistiese ondersoek*. Ongepubliseerde MA. Noordwes-Universiteit.
- Nini, A., & Grant, T. 2013. Bridging the between Stylistic and Cognitive approaches to authorship analysis using Systemic Functional Linguistics and Multidimensional analysis. *The International Journal of Speech, Language and the Law*, 202, 173-202.
- Oakes, M. P. 1998. *Statistics for Corpus Linguistics*. Oxford: Edinburg University Press.
- Oakshott, L. 2006. *Essential Quantitative Methods for Business, Management and Finance*. New York: Palgrave Macmillan.

- Olsson, J. 2008. *Forensic linguistics*. London: Continuum International Publishing.
- Oxford Dictionaries. 2018. What can the Oxford English Corpus tell us about the English language? Beskikbaar by <https://en.oxforddictionaries.com/explore/what-can-corpus-tell-us-about-language/> [12 Augustus 2018].
- Paltridge, B. 1996. Genre, text type, and the language learning classroom. *ELT Journal*, 50/3. 237-243.
- Ponelis, F. 1979. *Afrikaanse sintaksis*. Pretoria: JL van Schaik.
- Rabe, L. 2015. “Rubriekmaak” à la ryk, ryker, Rykie – Rykie van Reenen en enkele merkers van hedendaagse blog-rubriekskrywers. *LitNet Akademies*, 12(2):172-195. Beskikbaar by <https://www.litnet.co.za/rubriekmaak-a-la-ryk-ryker-rykie-rykie-van-reenen-en-enkele-merkers-vir-hedendaagse-blog-rubriekskrywers/> [23 Junie 2016]
- Rudman, J. 1998. The State of Authorship Attribution Studies: Some Poblems and Solutions. *Computers and Humanities*, 31, 351-365.
- Schlepperegrell, M. 2013. Systemic functional linguistics. In J. P. Gee, & M. Handford Reds., *The Routledge handbook of discourse analysis*. London: Taylor & Francis Ltd. 21-34.
- Schlepperegrell, M. 2012. Academic Language in Teaching and Learning: Introduction to the special issue. *The Elementary School Journal*, 1123, 409-418. doi:<https://doi.org/10.1086/663297>
- Sieben, I., & Linssen, L. 2009. Logistiese regressie analyse: Een handleiding . Beskikbaar by <https://www.ru.nl/socialewetenschappen/rtog/> [21 Paril 2019]
- Simpson, E.H. Measurement of diversity. *Nature*, 163:688.

- Simpson, P. 2004. *Stylistics: A Resource Book for Students*. New York: Routledge.
- Smith, M.W.A. 1983. Recent experience and new developments of methods for the determination of authorship. *Association for Literary and Linguistic Computing Bulletin*, 11:73-82.
- Snyman, D. 2008. *Op die agterpaaie*. Tafelberg Uitgewers Ltd.
- Solan, L. M. 2012. Intuition versus Algorithm: The Case for Forensic Authorship Attribution. *Journal of Law and Policy*, XXI, 551-576.
- SPSS. 2018. Chi-square independence test - What and why? Beskikbaar by SPSS Tutorials: <https://www.spss-tutorials.com/chi-square-independence-test/> herwin [21 November 2018].
- Stamatatos, E. 2008. A Survey of Modern Authorship Attribution Methods. *Journal of the American Society for Information Science and Technology*, 603, 538-556.
- Stamatatos, E., Daelemans, W., Verhoeven, B., Pothast, M., Stein, M., Joula, P., Sanchez-Perez, M.A. & Barron-Cedeno, A. 2014. Overview of the Author Identification Task at PAN 2014. In L. Cappellato, N. Ferro, M. Halvey, & W. Kraaij, CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers. CEUR Workshop Proceedings. CLEF and CEUR-WS.org. Beskikbaar by <http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-StamatosEt2014.pdf> [10 November 2019]
- Stotsky, S. 1983. Types of Lexical Cohesion in Expository Writing: Implications for Developing the Vocabulary of Academic Discourse. *College Composition and Communication*, 430-446. doi:10.2307/357899
- Scott, M. 2008. Wordsmith tools version 6. Liverpool: Lexical analysis software.

- Tallentire, D.R. 1972. *An Appraisal of Methods and Models in Computational Stylistics, with Particular Reference to Author Attribution*. Ph.D. Cambridge Universiteit.
- Taavitsainen, I. 2001. Changing Conventions of Writing: The Dynamics of Genres, Text Types, and Text Traditions. *European Journal of English Studies*, 52, 139-150.
doi:10.1076/ejes.5.2.139.7309
- Trask, R., & Stockwell, P. 2007. *Language and Linguistics: The key Concepts*. New York: Routledge.
- Van de Poel, K., Carstens, W.A.M., & Linnegar, J. 2012. *Text editing: a handbook for students and practitioners*. Antwerp: UPA University.
- Vandermay, J., Houlihan, D., Klein, L. A., Lewinski, W., & Buchanan, J. 2008. Command sequence in police encounters: Search for a linguistic fingerprint. *Law Enforcement Executive Forum*, 83. Beschikbaar by <https://guilfordjournals.com/doi/abs/10.1521/jscp.2008.27.1.36> [20 Augustus 2018].
- Waddell, C. 2014. Basic prose style and mechanics. Michigan, Houghton. Beschikbaar by http://www.mtu.edu/humanities/department/faculty-staff/faculty/waddell/Basic_Prose_Style_and_Mechanics.pdf [24 Junie 2015].
- Yule, G.U. 1938. On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship. *Biometrika*, vol.30:363-390.
- Zipf, G.K. 1932. *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press.

Bylaag A

Geobserveerde hoeveelheid voorkomste van kohesieskakels in vier posisies op sinsvlak

| | André le Roux | | | Dana Snyman | | | Totaal |
|--------------------------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Kortverhale | Rubrieke | Totaal | Kortverhale | Rubrieke | Totaal | |
| Ellips | 158 | 77 | 235 | 124 | 101 | 225 | 460 |
| Posisie 1 | 14 | 9 | 23 | 16 | 9 | 25 | 48 |
| Posisie 2 | 36 | 19 | 55 | 39 | 34 | 73 | 128 |
| Posisie 3 | 67 | 22 | 89 | 35 | 32 | 67 | 156 |
| Posisie 4 | 41 | 27 | 68 | 34 | 26 | 60 | 128 |
| Leksikale Kohesie | 933 | 661 | 1594 | 832 | 809 | 1641 | 3235 |
| Posisie 1 | 91 | 72 | 163 | 121 | 84 | 205 | 368 |
| Posisie 2 | 239 | 147 | 386 | 209 | 199 | 408 | 794 |
| Posisie 3 | 224 | 168 | 392 | 193 | 195 | 388 | 780 |
| Posisie 4 | 379 | 274 | 653 | 309 | 331 | 640 | 1293 |
| substitusie | 9 | 16 | 25 | 9 | 11 | 20 | 45 |
| Posisie 1 | 1 | 2 | 3 | 1 | 3 | 4 | 7 |
| Posisie 2 | 3 | 4 | 7 | 1 | 2 | 3 | 10 |
| Posisie 3 | 5 | 5 | 10 | 4 | 3 | 7 | 17 |
| Posisie 4 | | 5 | 5 | 3 | 3 | 6 | 11 |
| Verbandsmerker | 535 | 516 | 1051 | 593 | 551 | 1144 | 2195 |
| Posisie 1 | 111 | 148 | 259 | 150 | 137 | 287 | 546 |
| Posisie 2 | 138 | 129 | 267 | 187 | 169 | 356 | 623 |
| Posisie 3 | 181 | 126 | 307 | 152 | 154 | 306 | 613 |
| Posisie 4 | 105 | 113 | 218 | 104 | 91 | 195 | 413 |
| Verwysing | 956 | 793 | 1749 | 675 | 739 | 1414 | 3163 |
| Posisie 1 | 213 | 152 | 365 | 179 | 171 | 350 | 715 |
| Posisie 2 | 290 | 244 | 534 | 171 | 212 | 383 | 917 |
| Posisie 3 | 224 | 193 | 417 | 148 | 153 | 301 | 718 |
| Posisie 4 | 229 | 204 | 433 | 177 | 203 | 380 | 813 |
| Groottotaal | 2591 | 2063 | 4654 | 2233 | 2211 | 4444 | 9098 |

Bylaag B

Geobserveerde hoeveelheid kohesieskakels in onderskeie posisies op paragraafvlak

| | André le Roux | | | Dana Snyman | | DS Totaal | Groot totaal |
|--------------------------|---------------|------------|-------------|-------------|------------|-------------|-----------------|
| | Kortverhale | Rubrieke | Totaal | Kortverhale | Rubrieke | Totaal | |
| Ellips | 158 | 77 | 235 | 124 | 101 | 225 | 460 |
| 1 | 20 | 5 | 25 | 13 | 9 | 22 | 47 |
| 2 | 15 | 8 | 23 | 12 | 9 | 21 | 44 |
| 3 | 22 | 4 | 26 | 13 | 10 | 23 | 49 |
| 4 | 13 | 5 | 18 | 14 | 10 | 24 | 42 |
| 5 | 10 | 9 | 19 | 11 | 9 | 20 | 39 |
| 6 | 15 | 9 | 24 | 13 | 16 | 29 | 53 |
| 7 | 15 | 10 | 25 | 9 | 4 | 13 | 38 |
| 8 | 11 | 11 | 22 | 15 | 11 | 26 | 48 |
| 9 | 24 | 7 | 31 | 10 | 14 | 24 | 55 |
| 10 | 13 | 9 | 22 | 14 | 9 | 23 | 45 |
| Leksikale Kohesie | 933 | 661 | 1594 | 832 | 809 | 1641 | 3235 |
| 1 | 82 | 54 | 136 | 66 | 51 | 117 | 253 |
| 2 | 137 | 63 | 200 | 81 | 63 | 144 | 344 |
| 3 | 111 | 75 | 186 | 75 | 86 | 161 | 347 |
| 4 | 94 | 57 | 151 | 91 | 89 | 180 | 331 |
| 5 | 108 | 55 | 163 | 91 | 82 | 173 | 336 |
| 6 | 87 | 68 | 155 | 79 | 96 | 175 | 330 |
| 7 | 99 | 69 | 168 | 88 | 82 | 170 | 338 |
| 8 | 79 | 69 | 148 | 89 | 90 | 179 | 327 |
| 9 | 69 | 86 | 155 | 79 | 99 | 178 | 333 |
| 10 | 67 | 65 | 132 | 93 | 71 | 164 | 296 |
| Substitusie | 9 | 16 | 25 | 9 | 11 | 20 | 45 |
| 1 | | 3 | 3 | | 1 | 1 | 4 |
| 2 | | 2 | 2 | 2 | 1 | 3 | 5 |
| 3 | 1 | 2 | 3 | 1 | 2 | 3 | 6 |
| 4 | | | | | 1 | 1 | 1 |
| 5 | 1 | 1 | 2 | | | | 2 |
| 6 | 1 | | 1 | 3 | 2 | 5 | 6 |
| 7 | 1 | | 1 | | 2 | 2 | 3 |
| 8 | 4 | 1 | 5 | 1 | 1 | 2 | 7 |
| 9 | | 3 | 3 | 1 | 1 | 2 | 5 |
| 10 | 1 | 4 | 5 | 1 | | 1 | 6 |
| Verbandsmerker | 535 | 516 | 1051 | 593 | 551 | 1144 | 2195 |
| 1 | 54 | 43 | 97 | 47 | 40 | 87 | 184 |
| 2 | 45 | 53 | 98 | 56 | 43 | 99 | 197 |

| | | | | | | | |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 3 | 55 | 47 | 102 | 56 | 47 | 103 | 205 |
| 4 | 59 | 50 | 109 | 60 | 50 | 110 | 219 |
| 5 | 54 | 55 | 109 | 63 | 55 | 118 | 227 |
| 6 | 53 | 54 | 107 | 67 | 64 | 131 | 238 |
| 7 | 49 | 55 | 104 | 59 | 54 | 113 | 217 |
| 8 | 56 | 61 | 117 | 69 | 68 | 137 | 254 |
| 9 | 58 | 54 | 112 | 61 | 75 | 136 | 248 |
| 10 | 52 | 44 | 96 | 55 | 55 | 110 | 206 |
| Verwysing | 956 | 793 | 1749 | 675 | 739 | 1414 | 3163 |
| 1 | 109 | 91 | 200 | 73 | 67 | 140 | 340 |
| 2 | 94 | 88 | 182 | 67 | 72 | 139 | 321 |
| 3 | 113 | 76 | 189 | 72 | 71 | 143 | 332 |
| 4 | 109 | 83 | 192 | 71 | 80 | 151 | 343 |
| 5 | 90 | 88 | 178 | 62 | 73 | 135 | 313 |
| 6 | 102 | 76 | 178 | 62 | 78 | 140 | 318 |
| 7 | 85 | 76 | 161 | 69 | 68 | 137 | 298 |
| 8 | 87 | 90 | 177 | 61 | 76 | 137 | 314 |
| 9 | 74 | 60 | 134 | 67 | 70 | 137 | 271 |
| 10 | 93 | 65 | 158 | 71 | 84 | 155 | 313 |
| Groototaal | 2591 | 2063 | 4654 | 2233 | 2211 | 4444 | 9098 |

Bylaag C

Fyner verdeling van die voorkoms van verskillende tipes leksikale kohesie, verbandsmerkers en verwysing in die datastel

| Kohesieskakel | André le Roux | | | Dana Snyman | | | Totaal |
|-----------------------------------|---------------|--------------|--------------|--------------|--------------|--------------|---------------|
| | Kortverhale | Rubrieke | Totaal | Kortverhale | Rubrieke | Totaal | |
| Leksikale kohesie: | 194 | 241 | 435 | 173 | 187 | 360 | 795 |
| Kollokasie | 6,0% | 7,5% | 13,5% | 5,4% | 5,8% | 11,1% | 24,6% |
| Leksikale kohesie: | 738 | 421 | 1159 | 657 | 620 | 1277 | 2436 |
| Semanties | 22,8% | 13,0% | 35,8% | 20,3% | 19,2% | 39,5% | 75,4% |
| Totaal: Leksikale kohesie. | 932 | 660 | 1592 | 83 | 807 | 1637 | 3232 |
| | 28.8% | 20.4% | 49,3% | 25.7% | 25.0% | 50,7% | 100% |
| Verbandsmerker: | 273 | 192 | 465 | 288 | 273 | 561 | 1026 |
| Aaneenskakelend | 12.4% | 8.7% | 21.1% | 13.1% | 12.4% | 25.5% | 46.7% |
| Verbandsmerker: | 94 | 117 | 211 | 88 | 66 | 154 | 365 |
| Redegewend | 4.2% | 5.3% | 9.6% | 4.0% | 3.0% | 7.0% | 16.6% |
| Verbandsmerker: | 77 | 123 | 200 | 81 | 92 | 173 | 373 |
| Teenstellend: | 3.5% | 5.6% | 9.1% | 3.6% | 4.1% | 7.8% | 16.9% |
| Verbandsmerker: | 91 | 84 | 175 | 136 | 120 | 256 | 431 |
| Tydsaanduidend | 4.1% | 3.8% | 7.9% | 6.2% | 5.4% | 11.6% | 19.6% |
| Totaal: Verbandsmerker | 535 | 516 | 1051 | 593 | 551 | 1144 | 2195 |
| | 24.3% | 23.5% | 47.8% | 27.0% | 25.1% | 52.1% | 100% |
| Verwysing: | 904 | 654 | 1558 | 604 | 650 | 1254 | 2812 |
| Bepaald | 28.6% | 20.7% | 49.4% | 19.1% | 20.6% | 39.7% | 89.1% |
| Verwysing: | 48 | 137 | 185 | 70 | 89 | 159 | 344 |
| Eienaam | 1.5% | 4.3% | 5.9% | 2.2% | 2.8% | 5.0% | 10.9% |
| Totaal: Verwysing | 952 | 791 | 1743 | 674 | 739 | 1413 | 3156 |
| | 30.2% | 25.1% | 55.2% | 21.4% | 23.4% | 44.8% | 100.0% |