

'n Korpusgebaseerde ondersoek na kohesieskakels as kenmerkende eienskap van outeurstyl van Dana Snyman en André le Roux

Anneen Church en Angeliqué van Niekerk

Anneen Church, Akademiese taal- en geletterdheidsontwikkeling, Onderrig-en-Leersentrum en
Angeliqué van Niekerk, Departement Afrikaans en Nederlands, Duits en Frans,
Universiteit van die Vrystaat

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). Die identifisering van die skrywer van 'n spesifieke teks val onder die studieveld van outeurskapidentifikasie. Joula (2008:249–51) 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 hierdie verkennende ondersoek word hoofsaaklik nagevors tot watter mate die gebruik van kohesieskakels in 'n teks (of tekste) 'n aanduiding van outeurskap kan wees en hoe dit gemeet kan word. Die vyf tipes kohesieskakels soos deur onder andere Halliday en Hasan (1976) uiteengesit, is as moontlike tekens van outeurstyl ondersoek. Hierdie vyf tipes is verwysing, plaasvervanging (ook bekend as substitusie), ellips, verbandsmerkers (ook bekend as konjunksie) en leksikale kohesie. As addisionele aspek van die navorsing is die voorkoms van kohesieskakels in twee tekstipes (en nie net die tekste van die onderskeie skrywers nie) ook beskryf.

'n Korpusgebaseerde benadering is gevolg in 'n ondersoek na die voorkoms van kohesieskakels in gekose rubrieke en kortverhale van twee populêre Afrikaanse skrywers, naamlik André le Roux en Dana Snyman. Die data-ontledingsproses bestaan uit verskeie fases, naamlik etikettering van kohesief-verwante items, sortering van etikette met die hulp van rekenaar-sagteware (Oxford WordSmith Tools), klassifikasie van etikette (vir tipes kohesieskakels) en statistiese ontleding van die voorkoms van die kohesieskakels.

Op grond van die literatuurstudie, data-ontleding en statistiese ontledingresultate is die aard van kohesieskakels as element van variasie tussen skrywers se skryfstyl bespreek. Die ontleding

ondersteun die vermoede dat daar wel variasie is in die manier waarop die onderskeie skrywers kohesieskakels gebruik, maar dat die mees betroubare manier om 'n persoonlike styl te identifiseer, steeds onduidelik is.

Trefwoorde: idiolek; kohesie; kortverhaal; outeurstyl; rekenaargesteuende outeurskap-identifikasie; rubriek

Abstract

A corpus-based study of cohesion links as characteristic of authorship of André le Roux and Dana Snyman

Individuals have preferences for certain linguistic items when using language. This means that writers leave traces of their authorship behind when they write a text (Louwse 2004:307). In theory, texts can therefore be identified as the work of a specific author based on their linguistic choices. The identification of a specific author as the producer of a certain text is a part of the academic field of authorship attribution studies. Joula (2008:249–51) suggests that there are certain aspects of authorship attribution studies which remain uncertain. These relate mainly to making the most appropriate decisions about methodology when conducting authorship attribution studies. Another aspect Joula mentions (2008:249–51) relates to the choice of linguistic item to be analysed as possible indication of authorship. Both the question about appropriate methodological considerations and that of a linguistic indication of authorship are explored in this study.

This hypothesis testing study mainly aims to explore a multistage process of isolating and analysing cohesion links within a text. The methodological contribution to the field of authorship attribution is made through its research design as the authors of all the texts in the corpus were known, but cohesion links as indicators of authorship and the best way to show possible patterns in cohesion link usage were unclear. This meant that predetermined linguistic variables (cohesion links) were chosen to be analysed in texts of known authors in order to identify possible patterns in the use of cohesion links as possible signs of authorship. 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 are reference, substitution, ellipsis, conjunction and lexical cohesion. Additionally, the way in which cohesion links appear in two text types (and not only 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. Each sub-corpus consisted of ten texts. Both the text analysis and the statistical analysis were made up of different stages to simplify the analysis process and improve reliability of the findings.

The process of text analysis was made up of three phases. In the first phase, texts (in .txt format) were tagged to create metadata to indicate cohesion. In this phase, linguistic items that were cohesively related to other items (regardless of the type of cohesion link) were tagged as being related to the antecedent. In the second phase, these .txt files were processed using computer

software (Oxford WordSmith Tools). The goal of this phase was to isolate parts of the texts that contained linguistic items that all related to a single antecedent. At this point they could also be sorted, so all tags that related to a single antecedent could be viewed together. The third phase in the text analysis process was classifying all the tags as examples of a specific cohesion link. At this point, the cohesion links were mapped and data relating to its frequency and position within the text could be used. This was done through statistical analysis.

The above-mentioned statistical analysis consisted of multiple stages (just like the process of text analysis) with the goal of systematically and reliably identifying and delineating relevant data. In each stage a data set was analysed by means of a specific statistical test to identify parts of the data set which could be unique to the author or text type in question. Data identified because of the statistical analysis became the data set for the subsequent stage of statistical analysis. First, descriptive statistics and relative frequencies were used to identify anomalies in the use of cohesion links in either of the sub-corpora. Cohesion links that appeared more or less frequently in certain sub-corpora were flagged for further analysis in the second stage. In the second stage, chi-square analyses were done using the frequencies of occurrence of the cohesion link types that were flagged in the previous stage. Cohesion link patterns in each sub-corpus that showed a statistically significant difference from other frequencies in the other sub-corpora were flagged as possible indications of authorship for the specific texts in the sub-corpora. Again, these cohesion links were flagged to be used as possible indications of authorship in the third stage of statistical analysis. In the third stage a binary regression model was generated using the cohesion link categories identified in the previous two stages. This model was used to test predictability of authorship of selected sub-corpora.

The statistical analysis supported various findings regarding authorship and the use of cohesive devices mapping as an authorship attribution strategy. Generally it was found that Dana Snyman's use of cohesion links is more consistent between the different genres than André le Roux's. Additionally, it was found that the use of lexical cohesion, reference and conjunction between the two authors was more distinct for each author than substitution and ellipsis. Lastly, it was found that the position in which the cohesion links appear in the text did not vary significantly between the two authors. Relative frequency of cohesion links in the whole text was a better indicator of authorship.

While this study provides useful insights into the patterns of cohesion link use of these two authors, there are still many questions about the practicalities of using cohesion link use as an identifier of authorship. The process of categorising linguistic items into types of cohesion links can be subjective and the best way to show variation statistically is still not clear. Nonetheless, the multistage process of text analysis and statistical analysis proved useful in promoting evidence-based methodology decisions within the research process. Closer collaboration with statisticians and more empirical research using this method may help to refine the process and provide more clarity on improved methodological practices within the field of authorship attribution.

Keywords: authorship style; cohesion; column; computer-based authorship attribution; idiolect; short story

1. Inleiding

Telesieprogramme soos die *CSI*-reekse prikkel dikwels kykers se nuuskierigheid en neem hulle mee op die reis van probleemoplossing. Karakters versamel bewyse van oortreders se optrede om skuldiges vas te trek. In die reeks *Manhunt: Unabomber*, wat op werklike gebeure gebaseer is, ondersoek die hoofkarakter uniekhede in die vermeende bommaker se taalgebruik om uiteindelik eienskappe van die onbekende verdagte te bepaal en hom uiteindelik op te spoor en vas te trek. Hoewel dié tipe uitbeeldings modern is, noem Kotzé (2007) dat ondersoeke om outeurskap te bepaal sedert die ontstaan van die vroegste geskrewe dokumente onderneem is.

Louwerse (2004:207) noem dat skrywers hulle sogenaamde merk op dokumente wat hul skryf, laat en dat so 'n dokument uitkenbaar is op grond van die voorkeur vir spesifieke taalkundige kenmerke in 'n teks, amper soos 'n vingerafdruk. Die beeld van 'n taalkundige vingerafdruk steun sterk op die idee van idiolek. *Idiolek* word deur Crystal (2008:235) beskryf as 'n term wat in die taalkunde gebruik word om na "die linguistiese sisteem van 'n individuele spreker [te verwys]" (eie vertaling; ook elders). Binne die Sitemies-Funksionele Grammatika (SFG) 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 en Grant 2013:180). Nini en Grant verduidelik verder dat idiolek uit die kodevariasie en dialektvariasie bestaan wat 'n taalgebruiker in elke situasie gebruik (2013:180). Hierdie variasie kan gemeet word indien taalgebruik in dieselfde konteks ondersoek word. Deur die konteks dieselfde te hou, word inmenging van registervariasie verminder.

In die literatuur, en veral in populêre media, word daar soms verwys na outeurskapidentifikasie-studie as die ondersoek van 'n tipe linguistiese vingerafdruk. Hoewel hierdie vergelyking maklik die verbeelding kan aangryp, kan die verwagting wat dit van outeurskapidentifikasie-studie skep, problematies wees (Coulthard 2005:14). Outeurskapidentifikasie behels "die wetenskap (daarvan) om karaktereieskappe van 'n skrywer af te lei van karaktereieskappe van 'n dokument wat deur die betrokke skrywer geskryf is" (Joula 2008:233).

Hierdie tipe ontledings word al vir honderde jare onderneem (Kotzé 2007). Sedert die begin van formele outeurskapstudie het die metodologiese en teoretiese basis waarop hierdie ondersoeke gegrond is, groot kritiek ontlok, deur navorsers én kritici buite die veld (vergelyk Rudman 1998; Joula 2012 en Tyo, Dhingra en Lipton 2022). Die meeste kritiek is gerig teen die oënskynlik lukrake en subjektiewe metodiek wat in outeurskapstudies gebruik word. Selfs al het navorsers tans meer outomatiese verwerkingstechnologie tot hulle beskikking as ooit tevore, is daar steeds twee groot struikelblokke ten opsigte van die gepaste ondersoekmetodes. Hierdie probleme sluit in watter taalkundige kenmerke 'n aanduiding van outeurskap is, hoe die navorser besluit watter kenmerke om te gebruik en hoe dit betroubaar getoets kan word (Rudman 1998:252–3; Joula 2012:275–6). Veral die skep van betroubare en verifieerbare bewyse van outeurskap, bly 'n struikelblok vir die standaardisering van navorsing in dié veld (Tyo e.a. 2022). Navorsers reageer op hierdie kritiek met deurlopende voorstelle oor verbeterde tegnieke om navorsingsresultate herhaalbaar en betroubaar te maak.

In hierdie ondersoekende navorsing word gepoog om empiriese data by te voeg ter ondersteuning van die soeke na 'n standaardmetodologie vir die oplos van outeurskapstudievraagstukke. Oorhoofs is die ondersoek ingestel op die mate waartoe skrywers soos André le Roux en Dana Snyman kohesieskakels in herkenbare of unieke hoeveelhede en/of patrone gebruik in

uitgesoekte kortverhale en rubrieke. Verder is die moontlikheid van kohesiepatrone as statistiese voorspeller van outeurskap ook ondersoek.

Hoewel skrywers 'n sekere persoonlike skryfstyl het, beïnvloed verskeie situasionele faktore, soos genrevereistes, uiteraard hulle taalkundige keuses. Taavitsainen (2001:140) noem dat genre op konvensies gebaseer is. Hyland (2002:114) verduidelik verder 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 taal-handeling rig. Dit impliseer dat elke suksesvolle teks tekens sal toon van die teksproduseerder se bewustheid van die sosiale konteks waarbinne die teks geskep is (Hyland 2002:114). Dit is dus toepaslik 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.

In hierdie ondersoek is die patrone van kohesieskakelgebruik in rubrieke en kortverhale van Dana Snyman en André le Roux ondersoek. Die taalkundige veranderlike is dus vooraf gekies (kohesieskakels) en die skrywers van die tekste is bekend, sodat die geldigheid van kohesiepatrone as moontlik tekenend van 'n persoon se styl en/of benadering tot 'n spesifieke genre ondersoek kan word.

Die teksontleding is in drie fases gedoen. Tydens die eerste fase is tekste met die hand geëtiketteer, waarna die data met Oxford Wordsmith Tools verwerk is. Die verwerkte data is daarna verder gekategoriseer om die voorkoms van die onderskeie kohesieskakels te karteer. Hierdie kohesievoorkomsdata is daarna deur middel van statistiese ontledingsmetodes (beskrywende statistiek en binêre logistiese regressie-ontleding) bepaal.

Voorts word sowel die aard en kenmerke van kohesieskakels as die moontlike stilistiese kenmerke van rubrieke en kortverhale bespreek. Daarna volg 'n oorsig oor die ontledingsproses, metodologiese oorwegings en statistiese ontleding wat in hierdie ondersoek gevolg is. Bevindinge en implikasies word aan die hand van die statistiese ontleding bespreek.

2. Literatuuroorsig

In hierdie afdeling word kohesieskakels en stilistiese oorwegings ten opsigte van kortverhale en rubrieke bespreek. Hierdie bespreking is gefokus op die toepassing van hierdie skakels binne die bepaalde dataset en ontleding. 'n Volledige oorsig van kohesie en stilistiek val buite die bestek van hierdie artikel.

2.1 Kohesie

Kohesie is 'n verskynsel wat te make het met die eenheidstruktuur van 'n teks. Dit word bestudeer in die studieveld van tekslinguïstiek, wat op sy beurt 'n deel vorm van die teoretiese benaderings van Funktionale Grammatika.

Kohesie hou verband met die manier 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 en Dressler 1981:3; vergelyk ook Carstens en Van de Poel

2010:106 en Fouché 2016:17–8). Janjua (2012:150) wys op die onderskeie betekenisaspekte wat na vore kom in definisies van *kohesie* in verskeie bronne. Hy (2012:150) noem eerstens dat die *Concise Oxford dictionary of linguistics* (Matthews 1997) kohesie as 'n sintaktiese eenheid sien, terwyl Crystal (2008:85) die klem elders plaas en kohesie eerstens as 'n grammatiese eenheid beskou. Janjua (2012:150) noem laastens dat Halliday en Hasan (1976) fokus op kohesie as 'n semantiese eenheid. Richards en Schmidt (2010:94) se definisie veronderstel 'n wyer interpretasie deur te noem dat kohesie 'n “grammatiese *en/of* leksikale verhouding” (eie klem, eie vertaling) is. Die volgende definisie van Malmkjaer (1991:463) sluit aspekte van oppervlak- en dieptestruktuur in (dus sintaktiese, grammatiese en semantiese aspekte): “Kohesie het te make met die manier waarop die linguïstiese 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).

Kohesieskakels is skakels (“ties”) wat verwys na 'n spesifieke “instance of cohesion” van taalkundige items wat kohesief aan mekaar verwant is (Halliday en Hasan 1976:3). Daar is verskillende strategieë wat gebruik word om verwante items taalkundig aan mekaar te skakel om 'n samehangende geheel te vorm. Hierdie verskillende strategieë is in hierdie ondersoek gekategoriseer om die volgende kohesieskakels te verteenwoordig: verwysing, plaasvervanging (substitusie), ellips, verbandswoorde en leksikale kohesie. In hierdie studie word ondersoek of skrywers hierdie skakels op 'n unieke en/of uitkenbare manier gebruik al dan nie.

2.1.1 Verwysing

Soos vroeër vermeld is, word *verwysing* deur Crystal (2008:407) gedefinieer as 'n kohesieskakel wat gebruik word om 'n identiteitsverhouding tussen leksikale items uit te druk. '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). Carstens (1997:140) bespreek 'n verdeling wat aanvanklik deur Halliday en Hasan voorgestel is. Hierdie verdeling illustreer die tipes strukturele oorwegings wat met verwysing verband hou. Dit klassifiseer die tipe verwysing volgens die posisie van die verwysende element na die verwysde element (referent).

Benewens die bogenoemde verdeling, wat hoofsaaklik op die struktuur van die voorkoms van verwysing betrekking het, is daar ook ander klassifikasiestrukture. Fouché (2016:30–2) verduidelik byvoorbeeld Halliday en Hasan (1976:37–87) se indeling en noem die kategorieë van persoonlike verwysing (“personal reference”), aanwysende verwysing (“demonstrative reference”) en vergelykende verwysing (“comparative reference”). Hierdie kategorieë het meer 'n semantiese as 'n strukturele ordeningsbeginsel.

Nog 'n verdeling van tipes verwysing word deur Carstens (1997:136–7) gegee wanneer hy De Stadler (1989:18) se verdeling van hoofipes verwysing uiteensit, naamlik 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.

Vir die doel van hierdie ondersoek is twee tipes verwysing gemerk. Die eerste is eiename wat na 'n algemene entiteit buite die teks verwys, byvoorbeeld Kaapstad, Pieter-Dirk Uys, die Toring van Babel, ens. Die hoofrede waarom hierdie kategorie gekies is, is die betreklik

gereelde voorkoms van hierdie tipe verwysing in die tekste. Die tweede kategorie wat ingesluit is, kom veral deur die voorkoms van persoonlike, betreklike en aanwysende woorde tot stand (bv. onderskeidelik *hy, sy, wat, wie, hierdie, daardie*, ens.). Albei hierdie gekose subkategorieë val onder die oorhoofse kategorie van bepaalde verwysing.

2.1.2 Plaasvervanging

Plaasvervanging 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 plaasvervanging as 'n "leksikogrammatiese" een. Dit beteken dat plaasvervanging 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, normaalweg dieselfde grammatiese funksie te hê as die een wat vervang is. Dus moet 'n werkwoord gewoonlik met 'n werkwoord vervang word, 'n naamwoord met 'n naamwoord, ens.

Carstens (1997:237) som die Afrikaanse woorde wat gebruik word om woorde/sinsnedes te vervang soos volg op:

Naamwoordelike plaasvervanging:

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

Werkwoordelike plaasvervanging:

- *doen/maak*, ens. (+ dit/so/dieselfde)

Klousale plaasvervanging:

- *so*
- *nie*
- + *nie*.

Die aantal taalkundige items in 'n teks wat as plaasvervangings-elemente gekategoriseer word, is betreklik klein. Dit bied daarom nie groot uitdagings tydens die kategoriseringsproses nie.

2.1.3 Ellips

Ellips word ook *weglating* genoem en word gekenmerk deur 'n proses wat soortgelyk is aan plaasvervanging, behalwe dat die vervangde element heeltemal weggelaat word. Hierdie elemente kan weggelaat word omdat die betekenis van die teks binne die betrokke konteks steeds verstaan kan word (Carstens e.a. 2022:644). Carstens en Van de Poel (2010:109) merk

op dat ellips, tesame met plaasvervanging, 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 en Linnegar 2012:76). Hierdie verskynsel is ook in hierdie ondersoek opgemerk. Omdat die gebruik van ellips die weglating van 'n element of elemente behels, soos vervolgens verduidelik word, is dit ook nie so duidelik identifiseerbaar tydens teksontleding soos sommige van die ander kohesiemiddele nie.

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

Buiten hierdie 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.

In hierdie ondersoek is slegs voorkomste van ellips in 'n neweskikkende konstruksie, ondersoek. 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.

2.1.4 Leksikale kohesie

Leksikale kohesie is nie 'n grammatiese bindmiddel nie, maar 'n semantiese een. Leksikale kohesie word deur Halliday en Hasan (1976:274) beskryf as “die kohesiewe effek wat deur woordkeuse bewerkstellig word”. Hulle sê voorts dat leksikale kohesie in 'n groot mate verskil van die ander soorte kohesie. Die grootste verskil is dat leksikale kohesie deel is van 'n sogenaamde “open set”, wat beteken dat daar nie 'n eindige versameling leksikale items is wat leksikale kohesie bewerkstellig nie.

Halliday en Hasan wys aanvanklik op vier kategorieë 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, sinonieme, superordinate en 'n epiteton of “algemene woord” (Carstens 1997:318–20; Carstens en Van de Poel 2010:111).

Die ander kategorie van leksikale kohesie wat deur Halliday en Hasan bespreek word, is kollokasie. Hulle beskryf kollokasie as kohesie wat bewerkstellig 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. Stotsky (1983) het in haar navorsing oor kohesie in studente se akademiese skryfwerk ook van 'n ruimer interpretasie van kollokasie gebruik gemaak, deur stelle of groepe woorde in te sluit, wat verwant is omdat dit gereeld in dieselfde kontekste gebruik word.

Stotsky (1983) én Halliday en Hasan erken dat die kohesieskakel kollokasie nie 'n duidelike een is om te omskryf nie en dat die vertolking van die skakel tot 'n mate afhanklik is van die leser, soortgelyk aan die manier waarop 'n leser koherensie in 'n teks sal gebruik om die teks te verstaan. Yonos (2019:4) bevestig ook die noue verband wat kollokasie met koherensie het en verduidelik dat die verskynsels soms verstrengel is en nie van mekaar geskei hoef te word

nie. In Fouché en Olivier (2020) se studie oor die realisering van leksikale kohesie in graad 6- en graad 9-leerders se skryfwerk laat hulle hierdie tipe kollokasie heeltemal uit hul ondersoek-kategorieë, aangesien die interpretasie daarvan, volgens hulle, te sterk op leserervaring steun en dus eerder onder koherensie tuishoort (Fouché en Olivier 2020:417). In hierdie ondersoek 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. Aangesien die studie hipotesetoetsend en daarom ondersoekend van aard is, is dit sinvol om, waar prakties moontlik, kategorieë van ondersoek in te sluit, eerder as uit te sluit. ’n Opsomming van hierdie twee verdelings (Halliday en Hasan 1976 en Stotsky 1983) kan in Tabel 1 gesien word.

Tabel 1. Opsomming van Halliday en Hasan (1976) en Stotsky (1983) se interpretasies van leksikale kohesie

Halliday en Hasan (1976)	Stotsky (1983)
1. Leksikale herhaling	1. Semanties verwante woorde
<ul style="list-style-type: none"> • Herhaling • Sinoniem • Superordinaat • Epiteton / Algemene woord 	<ul style="list-style-type: none"> • Herhaling • Sinoniem • Kontrasterende element of antoniem • Hiponiem • 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 • Kontrasterende element 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.

2.1.5 Verbandsmerkers

Verbandsmerkers (of konjunksiemerkers) 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) stel verbandsmerkers ’n ander tipe semantiese verhouding daar 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 (1976:238–9) stel moontlike soorte kategorieë voor waarin verbandsmerkers geplaas kan word, naamlik aaneenskakelend (“additive”), teenstellend (“adversative”), rede-gewend (“causal”) en tydsaanduidend (“temporal”). Carstens (1997:292) gebruik hierdie kategorieë om ’n lys Afrikaanse verbandsmerkers wat binne hierdie kategorieë pas, saam te stel.

Jordaan (2014:39–79 en 2016:157–8) herkategoriseer in haar korpusgebaseerde ondersoek oor die gebruik van verbandsmerkers in akademiese skryfwerk die bogenoemde indeling van

Carstens. Haar ondersoek toon dat die verbandsmerkers nie altyd in die verband gebruik word wat Carstens (1997:292) aangetoon het nie. Fouché, Van den Berg en Olivier (2017:822) beskryf Jordaan (2014 en 2016) se indeling van Afrikaanse verbandsmerkers as “besonder omvattend” en bou daarop voort deur verdere toepassings van verbandsmerkers in hul korpus te identifiseer en te kategoriseer. Ten tye van die ontleding wat in hierdie ondersoek beskryf word, was Jordaan se 2014-raamwerk die mees onlangse een en dus die een wat as raamwerk gebruik is om die voorkoms van verbandsmerkers in die tekste van André le Roux en Dana Snyman te ondersoek.

2.2 Genre en tekstipe

Biber en Conrad (2019:4) verduidelik dat taalkundige variasie stelselmatig plaasvind. Hulle verduidelik dat taalgebruikers ’n groot aantal taalkundige karaktereienskappe (bv. uitspraak, woordkeuse, grammatika, ens.) na gelang van nielinguistiese faktore aanpas. Hulle noem dat hierdie faktore onder meer die doelwit vir kommunikasie, verhouding tussen gespreksgenote, en sosiale eienskappe 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 hulle publikasies oor hierdie onderwerp variasie in beide gesproke en geskrewe taalgebruik uiteensit (Biber 1988; Biber, Conrad en Reppen 1998).

Die doel van hierdie afdeling is om te illustreer dat stilistiese aanpassings waarskynlik plaasvind wanneer ’n skrywer tekste in die onderskeie genres skryf. Literatuur oor die styl van die onderskeie genres verwys nie spesifiek na kohesieskakelgebruik nie en bied dus nie ’n teoretiese basis waarmee hierdie ondersoek se bevindinge vergelyk kan word nie.

2.2.1 Kortverhale

2.2.1.1 Werksdefinisie

Aangesien die kortverhale wat in hierdie ondersoek gebruik word, almal minder as 1 300 woorde lank is, word hulle as kortkortverhale 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 kortkortverhaal oor ’n duidelike aantoonbare verhaalmatigheid (beskik)”.

2.2.1.2 Stylkenmerke

Gouws (2013) verduidelik verder wat die styl van die kortkortverhaal ten opsigte van letterkundige én taalkundige eienskappe is. Ten opsigte 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. Gouws sê dat daar dikwels met tyd en vertellersperspektief geëksperimenteer word en dat die “vertellersperspektiefwisseling nou met die tydshantering (saamhang)”. Bogenoemde letterkundige elemente behoort ’n aantal taalkundige verskynsels wat te make het met die kohesiemiddele wat hier ondersoek word, tot gevolg te hê. So sou die feit dat daar met tyd en vertellersperspektief geëksperimenteer word, ’n invloed kon hê op onderskeidelik tydsaan-duidende konjunksie en referensiële verwysing.

In terme van taalkundige elemente verduidelik Gouws dat die woordgebruik in die kortkortverhaal dikwels 'n poëtiese aard het. Verder sê hy dat verskynsels soos herhaling, parallelismes en anaforiek dikwels in kortkortverhale voorkom.

Waldell (2014) gee in sy inligtingstuk “Basic prose style and mechanics” ’n oorsig oor stylkenmerke van goeie prosaskryfwerk. Anders as Gouws se oorsig oor die kortkortverhaal, 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 skrywer en die leser (vgl. afdeling 2.2.2.2). Hy gee 13 voorstelle vir taalkundige aspekte wat nagestreef of vermy moet word wanneer prosatekste geskep word. Hierdie voorstelle sluit onder meer die volgende in:

- Skryf in die bedrywende vorm
- Vermy nominalisering
- Druk stelsinne in die positiewe vorm uit (dus nie negativering nie)
- Varieer sinspatrone
- Vermy oortolligheid
- Vermy werkwoordfrases.

2.2.2 Rubrieke

2.2.2.1 Werksdefinisie

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”. Albei beskrywings plaas die manier waarop rubrieke aan die leser gelewer word, voorop.

2.2.2.2 Stylkenmerke

Rabe (2015:183–4) 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 2012:174–7 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.

2.2.2.3 Inhoud

Rabe (2015:183–4) 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 'n spesifieke rubriek deur 'n spesifieke skrywer 'n baie spesifieke identiteit het. 'n Tweede vereiste wat die inhoud betref, is dat die inhoud die skrywer se eie opinie moet weerspieël. Kirsten (2012:275) en Rabe (2015:183–4) maak albei van hierdie vereiste melding. Kirsten se afdeling oor rubrieke in *Die Afrikaanse skryfgids* (Scheepers en Kleyn 2012) se titel is selfs: “Rubriek: Onthou dis jou opinie”.

2.2.2.4 Skrywer se persoonlikheid

Omdat 'n rubriek gereeld deur dieselfde skrywer in dieselfde publikasie verskyn, bestaan daar by lesers 'n verwagting om 'n sekere produk op 'n sekere tyd te kry. Dit beteken 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 2015:183–4).

Daar is verskeie eienskappe wat van 'n rubriekskrywer 'n goeie vriend vir sy leserspubliek maak. Engelbrecht (1981) sê dat dit belangrik vir 'n rubriekskrywer is om voortdurend aan sy/haar skrywerspersoonlikheid te bou. Rabe (2015:12–3) verwys daarna dat die rubriekskrywer 'n definitiewe stem of persoonlikheid moet hê. Die skrywer se skrywerspersoonlikheid moet dus duidelik wees en die skrywer moet sy/haar opinie duidelik kan artikuleer (Kirsten 2012:275). Verder noem Rabe (2015:12–3) dat die rubriekskrywer 'n hele aantal belangstellings en opinies moet hê. Dit impliseer 'n belese rubriekskrywer met 'n wye verwysingsraamwerk wat 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 eerste persoon geskryf kan word, aangesien die skrywers hulle eie mening uitdruk. Wat wel duidelik is uit bogenoemde stylkenmerke van 'n rubriek, is dat dit 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.

3. Metodologie

'n Korpusgebaseerde benadering is in hierdie ondersoek gevolg ten einde die navorsingsdoelwitte te bereik. Om mee te begin, word 'n kort oorsig oor korpuslinguistiek as metodologiese benadering gegee, waarna die korpuskeuse en die ontledingsproses asook die statistiese verwerkingsmetodes bespreek word.

'n *Korpus* word vir ons doeleindes beskou as 'n groep tekste wat in elektroniese formaat beskikbaar is en wat navorsers met behulp van doelgemaakte sagteware, outomaties kan deursoek vir 'n verskeidenheid van kenmerke (Szudarski 2017). Benewens 'n verskeidenheid kenmerke wat ondersoek kan word, noem Botha en Partridge (2022:581) ook dat korpusstudies gebruik kan word om sekere taalpatrone te ondersoek.

Biber e.a. (1998:4) noem 'n aantal eienskappe van 'n korpusgebaseerde benadering. Die korpusgebaseerde benadering is eerstens empiries van aard en ondersoek patrone in natuurlike tekste. Tweedens word die korpusgebaseerde benadering gebruik wanneer tekste wat op 'n

natuurlike wyse geskep is, ontleed word (Biber e.a. 1998:4). Derdens maak die korpusgebaseerde benadering grootliks van rekenaars en doelontwerpte sagteware gebruik om spesifieke verskynsels te soek, te groepeer en te verwerk (Biber e.a. 1998:4–5). Laastens steun ’n korpusgebaseerde benadering op sowel kwalitatiewe as kwantitatiewe metodes om data te ontleed en gevolgtrekkings of afleidings te maak (Biber e.a. 1998:4). Kwalitatiewe dataontleding 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.

Botha en Partridge (2022:581) wys ook sekere praktiese oorwegings van korpusgebaseerde ondersoeke uit. Hierdie oorwegings hou onder meer in dat ’n korpus in só ’n formaat moet wees dat dit deur ’n rekenaar verwerk moet kan word. Dit moet ook “omvattend genoeg [...] wees om gesien te kan word as verteenwoordigend van ’n spesifieke taalvariëteit of tekstipe” (Botha en Partridge 2022:581). Verder beklemtoon hulle, soos Biber e.a. (1998), dat taalgebruik binne ’n sekere konteks die ondersoekfokus moet wees. Hulle (2022:581) verduidelik ook dat frekwensie en konkordansie ’n sentrale rol speel in korpuslinguistiese ondersoeke. Dit wil sê, die reëlmaat waarmee sekere items verskyn, en die kombinasies waarbinne dit in die konteks verskyn, is ondersoekmoontlikhede binne hierdie studies. ’n Korpusgebaseerde benadering is dus ’n gepaste benadering vir hierdie studie omdat die fokus val op frekwensie en kombinasies van die voorkoms van kohesiemerkers in natuurlike tekste.

3.1 Tipe data

Die data wat in hierdie ondersoek gebruik word, is kwalitatiewe data omdat dit uit tekste, meer spesifiek rubrieke en kortverhale, bestaan. Omdat dit ’n korpusgebaseerde ondersoek is, bestaan die korpus uit ’n versameling tekste wat tydens die navorsing ondersoek is. McEnery, Xiao en Tono (2006:13) verduidelik dat ’n korpus ontwerp moet word om ’n spesifieke taal of taalvariëteit te verteenwoordig.

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 ontledingsproses word volledig in afdeling 3.3 beskryf. Dit is egter sinvol om op hierdie stadium te noem dat die rou data van die korpus gemerk is om ’n stel kwantitatiewe metadata te skep, wat uiteindelik statisties ontleed is.

3.2 Keuse van datastel

Vir hierdie ondersoek is daar besluit om rubrieke en kortverhale van die populêre skrywers André le Roux en Dana Snyman te kies. Kortverhaal- en rubriektekste van hierdie skrywers is algemeen beskikbaar en daar is ’n voldoende aantal tekste van elke tipe om vergelykbare subkorpora saam te stel.

Nadat die kortverhaalttekste gekies is, is daar gepoog om rubriektekste te vind wat met die kortverhaalttekste vergelyk kon word. Ten einde eksterne bronne van taalvariasie soveel as moontlik te beperk, is tekste wat in vergelykbare publikasies verskyn het (Afrikaanse koerante) en tydens ’n soortgelyke periode gepubliseer is, gekies. Daar is dus twee tipes tekste gebruik: rubrieke uit koerante en kortverhale uit verskillende kortverhaalbundels. Aangesien minder rubrieke van Le Roux in Afrikaanse koerante verskyn het, was dit nodig om rubrieke wat oor ’n langer tydperk verskyn het, te kies. Uiteindelik is rubrieke van Le Roux wat tussen 2000 en 2014 in *Beeld*, *Volksblad* en *Die Burger* verskyn het en rubrieke van Dana Snyman wat tussen

2006 en 2012 in dieselfde koerante verskyn het, gekies. Die kortverhaalttekste is uit die bundels ingeskandeer en die rubriektekste is van die argiewe van die betrokke koerante in txt-formaat verkry.

3.2.1 Beskrywing van die korpus

Die korpus het bestaan uit 'n totaal van 40 tekste: 20 tekste per skrywer, 10 tekste per skrywer per genre. Die totale getal lopende woorde in die korpus is 31 658. 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.

3.3 Ontledingsproses

Die doel van die aanvanklike ontledingsproses was om al die gevalle van kohesie te dokumenteer sodat die voorkoms van hierdie skakels uiteindelik ondersoek kon word en patrone wat duidelik mag wees, statisties weergegee kon word. Om egter die voorkoms van kohesieskakels te dokumenteer en uiteindelik te ontleed, moes die skakels geïsoleer word, omdat die voorkoms van die skakel en nie die voorkoms van 'n spesifieke leksikale item ondersoek word nie. Die ontledingsproses was derhalwe daarop gemik om hierdie kohesieskakels te identifiseer en dan te isoleer ten einde die voorkoms daarvan te kan ontleed.

Die ondersoek is in drie fases gedoen, waarvan twee interpretasie van die navorsers geverg het en een rekenaarmatig was.

3.3.1 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 met mekaar verbind is, ongeag die kategorie van kohesieskakel wat gebruik is. Die antesedent is telkens gemerk en alle voorafgaande of daaropvolgende items wat na die bepaalde antesedent verwys, is met dieselfde etiket gemerk.

3.3.2 Fase 2: WordSmith-verwerking

Die rekenaarprogrammatuur Oxford WordSmith Tools, ontwikkel deur Oxford University Press, is as navorsingsinstrument gebruik. WordSmith Tools maak dit moontlik 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.

As gevolg van die aard van die linguistiese 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 plaasvervanging verteenwoordig, vorm 'n enorme groep woorde en die voorkoms daarvan kan nie opgespoor word bloot deur na die voorkoms van spesifieke leksikale items te kyk nie. Daarom is dit nie moontlik of sinvol om vir die opsporing van hierdie skakels 'n voorafbepaalde lys woorde op te soek nie. Om hierdie kohesieskakels (verwysing, plaasvervanging en leksikale kohesie) te isoleer, is daar op grond van die etikette 'n konkordansielys opgestel. Nadat die voorkoms van al die etikette gesoek is, is daar 'n lys saamgestel wat die etikette of dan gevalle van kohesie lys. Kategorisering is in die volgende fase gedoen.

Die tweede wyse waarop kohesieskakels geïsoleer kan word, hou verband met die feit dat verbandsmerkers en ellips altyd met sekere linguistiese items geassosieer 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 as soekfrase vir 'n konkordansielys ingevoer word, word 'n lys geskep met die voorkoms van al die betrokke woorde en moet dit bloot gekategoriseer word. Deur die konkordansielyste van bogenoemde twee prosesse bymekaar te voeg, is die gevalle waar kohesie voorgekom het, almal geïsoleer en kon hulle gekategoriseer word.

3.3.3 Fase 3: Kategorisering van kohesief verwante items

In die laaste fase van die teksontleding is die konkordansielys soos in afdeling 3.3.2 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.

Die konkordansielyste waarop voorkomste van verwysing, leksikale kohesie en plaasvervanging geëtiketteer is, is met die hand gekategoriseer. Die lys met moontlike ellips in 'n neweskikkende sinskonstruksie is gebruik om ellips te kategoriseer en die lys met die onderskeie tipes verbandsmerkers is deurgegaan om te bevestig dat die geïdentifiseerde leksikale items wel in die betrokke verband gebruik is, voordat dit as sodanig gekategoriseer is.

3.4 Statistiese ontledingsmetode

Soos reeds vermeld is, is daar van statistiese metodes gebruik gemaak om die data te ontleed. Aangesien die ondersoek tot 'n mate hipotesetoetsend is, is daar nie 'n standaard statistiese ontledingsmetode wat normaalweg vir hierdie tipe ondersoek gebruik word nie. Beskrywende data soos rou hoeveelhede, persentasies en die chi-kwadraattoets is gebruik om modelle te ontwikkel vir die binêre logistiese regressie-ontleding. Dit is 'n ontledingsmetode wat 'n aanduiding gee van die mate waartoe sekere veranderlikes voorspellingswaarde het ten opsigte van 'n sekere uitkoms. In hierdie geval is die moontlike veranderlikes die kohesieskakels, en die onderskeie skrywerstyle is moontlike uitkomst.

4. Data-ontleding en interpretasie

4.1 Beskrywende data: relatiewe hoeveelhede

4.1.1 Algemene voorkoms van kohesieskakels

In die hele korpus, is daar altesaam 9 098 kohesieskakels gemerk en geëtiketteer. Hiervan het 4 654 in die tekste van André le Roux en 4 444 in die tekste van Dana Snyman voorgekom. Die aantal voorkomste van elke kohesieskakel word ook in Tabel 3 aangedui.

Ter wille van vergelykbaarheid van die aantal voorkomste in elke korpus word voorkomste as genormaliseerde getalle getoon. Dit is getalle wat verwerk is om 'n spesifieke aantal voorkomste van 'n verskynsel per x aantal woorde in daardie spesifieke korpus uit te druk.

Tabel 2 wys die genormaliseerde getal kohesieskakels per 100 woorde in elke korpus (elke genre van elke skrywer). Die laaste ry in die tabel verwys na die persentasie kohesieskakels wat in elk van die onderskeie korpusse gemerk is. Ongeveer 30% van alle woorde wat in hierdie korpus gebruik is, is op een of ander wyse aan 'n ander woorde of sinsdele verbind.

Tabel 2. Genormaliseerde getal kohesieskakels per 100 woorde in die korpus

Kohesie-skakel	André le Roux			Dana Snyman			Totale korpus
	Kort-verhale	Rubrieke	Gemiddeld ¹	Kort-verhale	Rubrieke	Gemiddeld ²	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
Plaasvervanging	0,10	0,22	0,15	0,11	0,15	0,13	0,14
Verbandsmerker	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
Getal woorde per korpus³	8 977 (28%)	7 249 (23%)	16 226 (51%)	7 991 (25%)	7 441 (24%)	15 432 (49%)	31 658 (100%)

In Tabel 2 kan gesien word dat die subkorpora van die twee skrywers soortgelyk in grootte is (51% en 49%) in terme van die totale getal woorde van die algemene korpus. Die gemiddelde aantal kohesieskakels wat skrywers per 100 woorde gebruik het (28,68 en 28,8), is ook soortgelyk. Die genres toon egter afwykings vir die onderskeie skrywers. Terwyl Le Roux 'n soortgelyke getal kohesieskakels per 100 woorde in sy kortverhale en rubrieke gebruik (28,86 en 28,46), gebruik Snyman minder kohesieskakels in sy kortverhale as in sy rubrieke (27,94 teenoor 29,71).

Tussen Snyman se kortverhale en rubrieke word onderskeidelik die meeste variasie veroorsaak deur die voorkoms van verwysing (9,93 teenoor 8,45). Die voorkoms van die ander kohesieskakels in die tekste van Snyman is soortgelyk ten opsigte van die verskeie genres. Le Roux se verdeling van kohesieskakels tussen genres is egter meer dinamies, ten spyte van die feit dat die algehele getal per genre basies dieselfde is. Die twee kohesieskakels wat grootliks hierdie verskynsel illustreer, is leksikale kohesie en verbandsmerkers. Le Roux maak meer gereeld van leksikale kohesie in sy kortverhale as in sy rubrieke gebruik (10,39 teenoor 9,12), maar minder gereeld van verbandsmerkers (onderskeidelik 5,96 teenoor 7,12). 'n Soortgelyke tendens as wat hier bo beskryf is (groter verskille in Le Roux se toepassing van kohesieskakels as in Snyman s'n tussen die genres), is ook vir ellips en plaasvervanging waargeneem, maar as gevolg

van die relatief klein aantal voorkomste van hierdie kohesieskakels blyk die invloed daarvan nie groot te wees nie.

4.1.2 Algemene voorkoms van subkategorieë

Soos daar in afdeling 2.2 uiteengesit is, is daar vir leksikale kohesie, verbandsmerkers en verwysing, data ten opsigte van subkategorieë beskikbaar.

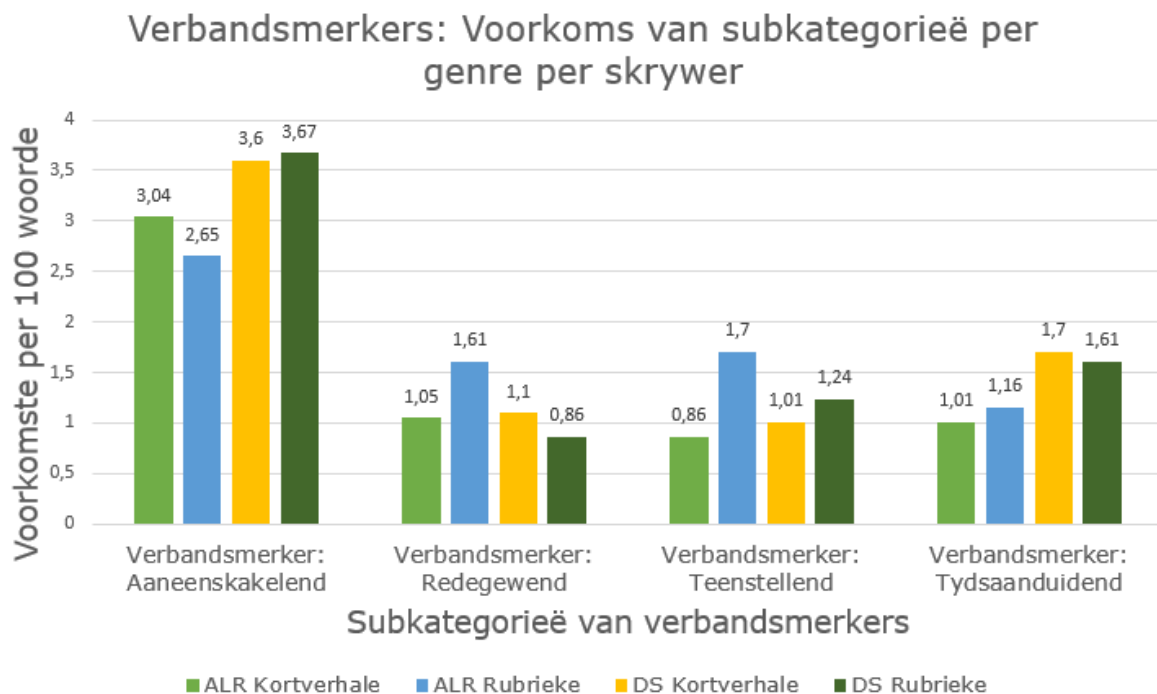
Tabel 3 wys die genormaliseerde getal voorkomste van die onderskeie subkategorieë van kohesieskakels in die onderskeie korpusse. Slegs tersaaklike tendense word vervolgens uitgewys.

Tabel 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
	Kort-verhale	Rubrieke	Totaal	Kort-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
Getal woorde per korpus	8 977	7 249	16 226	7 991	7 441	15 432	31 658
	28%	23%	51%	25%	24%	49%	100%

In terme van leksikale kohesie wyk die aantal kere wat Le Roux beide tipes leksikale kohesie in rubrieke gebruik, af van die patroon waarvolgens dit in die ander subkorpora voorkom. Die voorkoms van “Leksikale kohesie: Kollokasie” is die hoogste in die rubrieke van 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 Le Roux en onderskeidelik 2,16 en 2,51 in die kortverhale en rubrieke van Snyman). Die voorkoms van “Leksikale kohesie: Semanties” in die rubrieke van 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 subkorpusse 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 vanuit die tabel te interpreteer 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 1 stel hierdie gegewens voor.



Figuur 1. Uiteensetting van voorkoms van subkategorieë van verbandsmerkers in die hele datastel

In Figuur 1 is dit duidelik dat die rubrieke van 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 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 Le Roux se rubrieke soortgelyk. In die geval van tydsaanduidende verbandsmerkers is die voorkomste in beide genres van Le Roux laer as in die tekste van Snyman.

Wanneer daar na die onderafdeling van die voorkomste van tipes verwysing gekyk word, kan waargeneem word dat Le Roux in sowel kortverhale as rubrieke meer van voornaamwoord-verwysing gebruik maak as Snyman (10,07 en 9,02 teenoor 7,56 en 8,74 voorkomste per 100 woorde). In die afdeling eienaamverwysing is dit opmerklik dat die kortverhale van Le Roux die minste voorkomste per 100 woorde bevat (0,53), maar sy rubrieke die meeste (1,89). Hierdie verskynsel strook met die voorkomste van sommige van die ander skakels tussen die genres van die skrywers waar die tekste van Snyman meer soortgelyk is en dié van Le Roux meer verskillend.

4.2 Beskrywende data: chi-kwadraattoets

4.2.1 Algemene voorkoms van kohesieskakels

'n Chi-kwadraattoets is gedoen met die data wat op die oog af lyk of dit tekens van 'n patroonmatige voorkoms toon. 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.

Die webplatform Social Science Statistics stel navorsers in staat om 'n chi-kwadraattoets uit te voer wat ook 'n chi-kwadraatstatistiek aan elke veranderlike toeken en nie net aan die reeks veranderlikes nie. Dit beteken dat die veranderlike(s) wat binne die reeks statisties beduidend van mekaar verskil, ook aangedui kan word. Getalle wat tussen blokhakies aangedui is, is die sogenaamde kritieke waarde wat die chi-kwadraat van daardie spesifieke skakel aandui. Dit dui aan of die voorkoms van daardie spesifieke kohesieskakel as 'n enkele veranderlike binne die reeks statisties beduidend verskil na gelang van die toepaslike veranderlike (skrywer of kategorie van verwysing).

Chi-kwadraattoets is op die volgende stelle data uitgevoer:

1. Algemene voorkoms van kohesieskakels in die tekste van die twee skrywers
2. Algemene voorkoms van subkategorieë van leksikale kohesie in die tekste van die onderskeie skrywers.
3. Algemene voorkoms van subkategorieë van verbandswoorde in die tekste van die onderskeie skrywers.
4. Algemene voorkoms van subkategorieë van verwysing in die tekste van die onderskeie genres vir die onderskeie skrywers.

Tabelle 4 tot 7 stel die resultate van die chi-kwadraattoets voor. Datareekse wat p-waardes laer as 0,05 oplewer, dui op 'n statisties beduidende verskil tussen die twee reekse data.

Tabel 4. Chi-kwadraattoets: Algemene voorkoms van die kohesieskakels in die twee skrywers se tekste

	Ellips	Leksikale kohesie	Verwysing	Verbandsmerkers	Verwysing	Totale	Chi-kwadraat
André le Roux	235 [0,00]	1 594 [2,24]	25 [0,17]	1 051 [4,60]	1 749 [10,61]	4 654	Chi- kwadraatstatistiek: 36,0487 p-waarde: 0,00001
Dana Snyman	225 [0,00]	1 641 [2,34]	20 [0,18]	1 144 [4,81]	1 414 [11,11]	4 444	
Totaal	460	3 235	45	2 195	3 163	9 098 (groot- totaal)	
Kritieke waarde vir $p < 0,05$ is 3,841							

Tabel 5. Chi-kwadraattoets: Algemene voorkoms van subkategorie van leksikale kohesie in die tekste van die skrywers

	Leksikale kohesie: Kollokasie	Leksikale kohesie: Semanties	Totaal	Chi-kwadraat
André le Roux	435 [4,73]	1 157 [1,54]	1 592	Chi- kwadraatstatistiek: 12,3669 p-waarde: 0,000437
Dana Snyman	360 [4,60]	1 277 [1,50]	1 637	
Totaal	795	2 434	3 229 (groot- totaal)	
Kritieke waarde vir $p < 0,05$ is 3,89				

Tabel 6: Chi-kwadraattoets: Algemene voorkoms van subkategorieë van verbandswoorde in die tekste van die skrywers

	Aaneenskakelend	Redegewend	Teenstellend	Totaal	Chi-kwadraat
André le Roux	465 [2,90]	211 [3,84]	200 [0,71]	876	Chi-kwadraatstatistiek: 15,0441 p-waarde: 0,000541
Dana Snyman	531 [2,96]	154 [3,92]	173 [0,72]	858	
Totaal	996	365	373	1 734 (groot-totaal)	
Kritieke waarde vir $p < 0,05$ is 5,99					

Tabel 7: 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: Voornaamwoord	904 [2,26]	854 [1,89]	604 [0,04]	850 [0,00]	3 212	Chi-kwadraatstatistiek: 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 (groot-totaal)	
Kritieke waarde vir $p < 0,05$ is 7,815						

Soos gesien kan word in tabelle 4 tot 7, verskil al die reekse data wat ondersoek is, statisties beduidend van mekaar. Dit beteken dat:

- Le Roux en Snyman die vyf kohesieskakelkategorieë in statisties beduidend verskillende ratio's gebruik (Tabel 4; chi-kwadraatstatistiek: 36,0487; p-waarde: 0,00001)
- Le Roux en Snyman die subkategorieë van kollokasie in statisties beduidend verskillende ratio's gebruik (Tabel 5; chi-kwadraatstatistiek: 12,3669; p-waarde: 0,000437);
- Le Roux en Snyman die subkategorieë van verbandswoorde in statisties beduidend verskillende ratio's gebruik (Tabel 6; chi-kwadraatstatistiek: 15,0441; p-waarde: 0,000541);

- Die subkategorieë van verwysing in statisties beduidend verskillende ratio's voorkom ten opsigte van die vier subkorpusse (Tabel 7; chi-kwadraatstatistiek: 43,344; p-waarde: 0,00001). Benewens die reekse data is daar ook individuele veranderlikes wat statisties beduidend in die onderskeie datastelle verskil. In Tabelle 5 tot 8 is dit veranderlikes waarvan die individuele chi-kwadraatstatistiek (aangedui tussen blokhakies) hoër is as die kritieke waarde vir daardie reeks. Hierdie individuele veranderlikes sluit die volgende in:
- **Verwysing** vir albei skrywers in Tabel 4: Die chi-kwadraat vir Le Roux en Snyman is onderskeidelik 10,61 en 11,11, wat albei hoër is as die kritieke waarde van 9,4.
- **Leksikale kohesie: kollokasie** vir albei skrywers in Tabel 5: Die chi-kwadraatstatistiek vir Le Roux en Snyman is onderskeidelik 4,73 en 6,6, wat albei hoër is as die kritieke waarde van 3,84.
- **Verwysing: Eienaam** in albei genres van Le Roux in Tabel 7: Die chi-kwadraatstatistiek vir Le Roux se kortverhale en rubrieke is onderskeidelik 21,11 en 17,65, wat albei hoër is as die kritieke waarde van 7,815.

4.3 Binêre logistiese regressie-ontleding

Die derde tipe statistiese ontleding wat gebruik is, is 'n binêre logistiese regressie-ontleding. Hierdie tipe ontleding word gebruik wanneer een van twee moontlike uitkomstes op grond van 'n veranderlike of veranderlikes voorspel wil word. In die geval van hierdie ondersoek wil ondersoek word tot watter mate 'n sekere stel veranderlikes kan voorspel of 'n teks deur Le Roux of deur Snyman geskryf is.

Op grond van die ontledings wat in afdelings 4.1 en 4.2 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 (vervanging uitgesluit)
2. die algemene voorkoms van die subkategorieë van die kohesieskakels

Die resultate van hierdie twee ontledings word vervolgens bespreek.

4.3.1 Binêre logistiese regressie-ontleding: voorstelling van nulmodelle⁴

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 ondersoek.

In Tabel 8 word die akkuraatheid van die onderskeie nulmodelle, die statistiese beduidendheid van die model (d.w.s. of die model 'n statisties beduidende voorspeller van 'n sekere uitkomstes is) en die aantal 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 'n 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 getal algemene voorkoms van kohesieskakels 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 ten opsigte van die getal kere dat 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 ontleding kom egter voor by die statistiese beduidendheid van die nulmodelle. Soos gesien kan word, bied elkeen van die nulmodelle 'n uitkoms wat as statisties beduidend ("statistically significant") geklassifiseer kan word (dit is statisties beduidend wanneer die sig.-waarde kleiner is as 0,05, dus 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-kwadraattoets die geval was. Omdat die datastelle so groot is (n=8 593, n=8 576 en n=5 351), 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 8: Voorstelling van nulmodelle van onderskeie logistiese regressie-ontledings

Model	Persentasie akkuraatheid	Statistiese beduidendheid van nulmodel (p<0,05)	N
1. Effek van algemene voorkoms van kohesiemerkers as voorspeller van skrywer	51,2 %	0,028	9 098
3. Effek van die voorkoms van die subkategorieë van leksikale kohesie, verbandswoorde en verwysing as voorspeller van skrywer	51,2%	0,028	9 098

4.3.2 Binêre logistiese regressie-ontleding: Model 1

Tabel 9 stel die resultate voor van die logistiese regressie-ontleding van die algemene voorkoms van leksikale kohesie, verbandsmerkers en verwysing. Die tabel is in die formaat waarin SPSS Statistics resultate vir hierdie statistiese toets gee. Daar is bykomende inligting onder die tabel ingevoeg wat ook deur SPSS Statistics gegee word. Vir verklarings ten opsigte van die statistiese elemente wat in die tabel ingesluit is, kan 'n SPSS-handleiding, soos die een deur Pallant (2016) geraadpleeg word.

Indien navorsers 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 statistieke in hierdie kolom gee 'n aanduiding van die waarskynlikheid dat die voorkoms van hierdie betrokke veranderlike een uitkoms bo 'n ander kan verseker. 'n Voorbeeld hiervan kan 'n ondersoek wees waarin daar gepoog word om studentesukses te voorspel deur 'n aantal veranderlikes in ag te neem. Daar kan dan byvoorbeeld twee moontlikhede ten opsigte van die uitkoms wees, naamlik, suksesvol of onsuksesvol. Dan kan daar ook 'n aantal moontlikhede ten opsigte van die veranderlikes wees, byvoorbeeld demografiese oorwegings, vorige akademiese prestasie, sosio-ekonomiese agtergrond, ens. Die hoë of lae punte wat 'n spesifieke student voorheen behaal het, kan beskryf word as iets wat sy/haar kans op studentesukses byvoorbeeld een of twee of drie keer verhoog. In die huidige navorsing kan dieselfde egter nie van 'n enkele voorkoms van 'n kohesiemerker gesê word nie. 'n Enkele voorkoms

van byvoorbeeld verwysing is nie genoeg om 'n teks as 'n produk van 'n sekere skrywer te eien nie. Die kollektiewe voorkoms van hierdie merkers sou egter 'n gebruikspatroon eie aan 'n spesifieke skrywer kon aandui. Daarom sal daar nie in hierdie bespreking van die binêre logistiese regressie-ontleding op hierdie $\text{Exp}(B)$ gefokus word nie, maar op aanduidings wat vir die datastel in sy geheel geld.

Wanneer hierdie binêre logistiese regressie-ontleding bespreek word, val die fokus eerstens op modelakkuraatheid. In Tabel 8 is aangedui dat die nulmodel 51,2% akkuraat is. In Tabel 9 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 Snyman as die teksproduseerder as vir Le Roux. Volgens hierdie model kan Snyman se tekste 62,7% van die tyd van Le Roux s'n onderskei word, terwyl die persentasie vir Le Roux se tekste slegs 43,2% is.

Ondanks die betreklike klein verbetering op die modelakkuraatheid van die nulmodel, word hierdie stel veranderlikes as voorspellingsmodel steeds as 'n statisties beduidende verbetering op die nulmodel gesien (chi-kwadraatstatistiek 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 10. Binêre logistiese regressie-ontleding van die algemene voorkoms van leksikale kohesie, verbandsmerkers en verwysing: SPSS-ontleding

Veranderlikes in die vergelyking		B (konsekwenheids- koeffisiënt)	S.E. (standaardfout)	Wald (wald-statistiek)	Df (vryheidsgrade)	Sig. (beduidendheid)	Exp(B) (eksponent van konsekwenheids- koeffisiënt)
Stap 1 ^a	Leksikale kohesie	0,088	0,096	0,854	1	0,355	1,093
	Verbandsmerkers	0,144	0,099	2,132	1	0,144	1,155
	Verwysing	-0,153	0,096	2,549	1	0,110	0,858
	Konstante	-0,020	0,190	0,011	1	0,916	0,980
Modelakkuraatheid = 52,7% Akkuraatheid per skrywer: Dana Snyman: 62,7% André le Roux: 43,2% Chi-kwadraatstatistiek = 53,770 ; $p = 0,000$ Pseudo- R^2 (Nagelkerke) = 0,005 N = 9 098							

4.3.3 Binêre logistiese regressie-ontleding: Model 2

Tabel 10 stel die resultate voor van die regressie-ontleding op die onderafdelings wat binne leksikale kohesie, verbandsmerkers en verwysing geklassifiseer is. Die tabel is weer eens in die formaat waarin die statistiese verwerkingsagteware SPSS Statistics die resultate na afloop van die ontleding verskaf.

Eerstens val die fokus op modelakkuraatheid. In Tabel 8 is aangedui dat die nulmodel vir hierdie stel veranderlikes 51,2% akkuraat is. In Tabel 10 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 4.3.2). 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 Snyman se tekste meer akkuraat kon voorspel, voorspel hierdie model Le Roux se tekste meer akkuraat. Volgens hierdie model kan Le Roux se tekste 61,3% van die tyd van Snyman s'n onderskei word, terwyl die persentasie vir Snyman se tekste 47,1% is.

Soos in die geval van die vorige model kan hierdie model ten spyte van die oënskynlik klein verbetering op die akkuraatheid van die nulmodel ook as 'n statisties beduidende verbetering daarvan gesien word (chi-kwadraatstatistiek: 79,770; $p < 0,05$). Die Nagelkerke-pseudo-R²-waarde dui aan dat 1,1% van die verbetering van modelakkuraatheid waarskynlik aan die veranderlikes in die model te danke is.

Tabel 11. Logistiese regressie-ontleding van die algemene voorkoms van die subkategorieë binne leksikale kohesie, verbandsmerkers en verwysing

Veranderlikes in die vergelyking		B (konsekwentheids-koëffisiënt)	S.E. (standaardfout)	Wald (wald-statistiek)	Df (vryheidsgrade)	Sig. (beduidendheid)	Exp(B) (eksponent van konsekwentheids-koëffisiënt)
Stap 1 ^a	Verwysing: Eienaam	-0,597	0,105	32,275	1	0,000	0,550
	Verbandswoorde: Teenstellend	-0,525	0,143	13,532	1	0,000	0,591
	Verbandswoorde: Aaneenskakelend	-0,193	0,116	2,740	1	0,098	0,825
	Verbandswoorde: Redegewend	-0,695	0,144	23,183	1	0,000	0,499
	Leksikale kohesie: Semanties	-0,282	0,106	7,043	1	0,008	0,754

Veranderlikes in die vergelyking						
	B (konsekwenheids- koëffisiënt)	S.E. (standaardfout)	Wald (wald-statistiek)	Df (vryheidsgrade)	Sig. (beduidendheid)	Exp(B) (eksponent van konsekwenheids- koëffisiënt)
Leksikale kohesie: Kollokasie	-0,570	0,121	22,078	1	0,000	0,566
Verwysing: Eienaam	-0,557	0,147	14,462	1	0,000	0,573
Konstante	3,039	0,626	23,544	1	0,000	20,887

Modelakkuraatheid = 54,7%
 Dana Snyman: 47,1%
 André le Roux: 61,3%
 Chi-kwadraatstatistiek = 79,308; p = 0,000
 Pseudo-R² (Nagelkerke) = 0,011
 N = 9 098

5. Bevindinge

Die gebruik van kohesieskakels in die werk van Dana Snyman en André le Roux dui tot 'n mate op 'n voorkeur in tekste van die onderskeie skrywers. Die skrywer- of genrespesifieke patrone wat gevind is, sluit in 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 getal 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
- Snyman se kortverhale en rubrieke meer soortgelyk in terme van kohesieskakelgebruik is as Le Roux s'n
- Le Roux ten opsigte van sekere kohesieskakels sy styl waarneembaar aanpas na gelang van die genre. Die duidelikste voorbeeld hiervan is eienaamverwysing.

Verder is hierdie verskynsels as 'n moontlike voorspeller van outeurskap, deur middel van 'n binêre logistiese regressie-ontleding, ook ondersoek. Hoewel die regressie-ontleding tog daarop gedui het dat so 'n voorspellingsmodel meer akkuraat is om outeurskap van 'n teks te raai, is die modelakkuraatheid nie hoog genoeg om onomwonde 'n akkurate voorspeller in die praktyk te wees nie. Dit is wel 'n hulpbron.

6. Ten slotte

6.1 Beperkings

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

Die statistiese ontledingsmetodes is waarskynlik die grootste beperking van die ondersoek. Die belangrikste kwessie is dat die gepaste statistiese ontledingsmetode 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) voorspel. Nouer samewerking met statistici en uitgebreide kennis van die statistiese ontleding deur die navorser sou in die toekoms die effek van hierdie beperking kon minimaliseer.

6.2 Aanbevelings

Deur die loop van die ondersoek en na afloop van die data-ontleding 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. In samewerking met statistici sou dieselfde data aan ander statistiese ontledingsmetodes onderwerp kon word ten einde te bepaal of daar metodes is wat meer bruikbaar is vir die doel van bepaalde navorsing.
- c. 'n Ander taalkundige element as kohesieskakels sou op dieselfde wyse ondersoek kon word ten einde te bepaal of dit 'n duideliker aanduiding van outeurskap is.
- d. 'n Soortgelyke ondersoek sou in 'n ander taal soos Engels onderneem kon 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 skrywerspesifieke of 'n taalspesifieke verskynsel is. Resultate van so 'n ondersoek sou toepassingsmoontlikhede kon hê in tweede- of vreemdetaalonderrig.

Met hierdie ondersoek het die navorsers gepoog om empiriese data by te voeg ter ondersteuning van die soeke na 'n standaardmetodologie vir die oplos van outeurskapstudievraagstukke.

Empiriese ondersoeke soos hierdie een kan help om die lens waarmee daar na outeurskap-kwessies gekyk word te fokus ten einde uiteindelik tekste akkuraat en betroubaar as behorende tot 'n sekere skrywer te kan identifiseer.

Bibliografie

- Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. en S. Conrad. 2019. *Register, genre and style*. Cambridge: Cambridge University Press.
- Biber, D., R. Reppen en S. Conrad. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Botha, Y. en M. Partridge. 2022. Korpuslinguistiek. In Carstens en Van Dyk (reds.) 2022.
- Carstens, W.A.M. 1997. *Afrikaanse tekslinguistiek: 'n inleiding*. Pretoria: Van Schaik Uitgewers.
- Carstens, W.A.M., N. Fouché, M. Pienaar en H. Beyer. 2022. Tekslinguistiek. In Carstens en Van Dyk (reds.) 2022.
- Carstens, W.A.M. en K. van de Poel. 2010. *Teksredaksie*. Stellenbosch: African Sun Media.
- Carstens, W.A.M. en T.J. van Dyk (reds.). 2022. *Toegepaste taalkunde in Afrikaans*. Pretoria: Van Schaik Uitgewers.
- Carter, R. 1982. *Language and literature: an introductory reader in stylistics*. New York: Harper Collins Publishers Ltd.
- Church, A. 2019. 'n Korpusgebaseerde ondersoek na kohesiepatrone as moontlike stilistiese kenmerk van outeurstyl. MA-verhandeling, Universiteit van die Vrystaat.
- Coulthard, M. 2005. Some forensic applications of descriptive linguistics. *VEREDAS – Revistade Estudios Linguisticos, Juiz de Fora*, 9(1/2):9–28.
- Crystal, D. 2008. *A dictionary of linguistics and phonetics*. Oxford: Blackwell Publishing.
- De Beaugrande, R. en W. Dressler. 1981. *Introduction to text linguistics*. Londen: Longman.
- De Stadler, L.G. 1989. *Afrikaanse semantiek*. Johannesburg: Southern Boekuitgewers.
- Engelbrecht, H. (red.). 1981. *Saterdagavond tuis*. Pretoria: Uitgewery Oranje.
- Fouché, N. 2016. Kohesiemerkergebruik in die skryfwerk van Afrikaanssprekende graad 6- en graad 9-leerders. MEd-verhandeling, Noordwes-Universiteit.

- Fouché, N. en J. Olivier. 2020. Die realisering van leksikale kohesie: 'n Korpuslinguistiese ondersoek van Afrikaanssprekende graad 6- en graad 9-leerders se skryfwerk, *LitNet Akademies*, 17(1):410–39. https://www.litnet.co.za/wp-content/uploads/2020/04/LitNet_Akademies_17-1_Fouche-Olivier_410-439.pdf.
- Fouché, N., R. van den Berg en J. Olivier. 2017. Carstens se raamwerk van Afrikaanse konjunksiemerkers: 'n akkurater en vollediger weergawe. *Tydskrif vir Geesteswetenskappe*, 57(3):814–36.
- Gouws, T. 2013. Kortkortverhaal. *Literêre terme en teorieë*. [https://www.literaryterminology.com/index.php/lemmas/18-k/95-kortkortverhaal%20\(17 Februarie 2023 geraadpleeg\)](https://www.literaryterminology.com/index.php/lemmas/18-k/95-kortkortverhaal%20(17%20Februarie%202023%20geraadpleeg)).
- Halliday, M. en R. Hasan. 1976. *Cohesion in English*. Londen: Longman.
- Hyland, D. 2002. Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics*, 34:1091–112.
- Janjua, F. 2012. Cohesion and meanings. *Canadian Social Science*, 8(2):149–55.
- Jordaan, A. 2014. Die gebruik van metadiskoers in Afrikaans T1-skryfwerk van eerstejaar-universiteitstudente. MA-verhandeling, Noordwes-Universiteit.
- . 2016. Afrikaanse verbandsmerkers: uitbreiding en herkategorisering van voorbeeldwoorde. *Southern African Linguistics and Applied Language Studies*, 34(2):147–67.
- Joula, P. 2008. Authorship attribution. *Foundations and trends in Information Retrieval*, 1(3):233–334.
- . 2012. Large scale experiments in authorship attribution. *English Studies*, 93(3):275–83.
- Kirsten, J. 2012. Onthou: Dis jou opinie! In Scheepers en Kleyn (reds.) 2012.
- Kotzé, E.F. 2007. Die vangnet van die woord. *South African Linguistics and Applied Language Studies*, 25(3):385–99.
- Lombard, J. 1979. Die kort-kortverhaal. MA-verhandeling, Universiteit Stellenbosch.
- Louwerse, M. 2004. Semantic variation in idiolect and sociolect: corpus linguistic evidence from literary texts. *Computers and Humanities*, 38:207–21.
- Malmkaer, K. 1991. *The linguistics encyclopedia*. New York: Routledge.
- Matthews, P.H. 1997. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.
- McEnery, A., R. Xiao en Y. Tono. 2006. *Corpus-based language studies: an advanced resource book*. New York: Routledge.

- Nini, A. en T. Grant. 2013. Bridging the gap between stylistic and cognitive approaches to authorship analysis using Systemic Functional Linguistics and multidimensional analysis. *The International Journal of Speech, Language and the Law*, 20(2):173–202.
- Pallant, J. 2016. *SPSS survival manual: a step by step guide to data analysis using the SPSS program*. 6de uitgawe. Londen: McGraw-Hill Education.
- Rabe, L. 2015. “Rubriekmaak” à la ryk, ryker, Rykie – Rykie van Reenen en enkele merkers van hedendaagse blog-rubriekskrywers. *LitNet Akademies*, 12(2):172–95. https://www.litnet.co.za/wp-content/uploads/2015/11/LitNet_Akademies_12-2_Rabe_172-195.pdf.
- Richards, J.C. en R.W. Schmidt. 2010. *Longman dictionary of language teaching and applied linguistics*. Londen: Pearson.
- Rudman, J. 1998. The state of authorship attribution studies: some problems and solutions. *Computers and Humanities*, 31:351–65.
- Scheepers, R. en L. Kleyn (reds.) 2012. *Die Afrikaanse skryfgids*. Johannesburg: Penguin.
- Stotsky, S. 1983. Types of lexical cohesion in expository writing: implications for developing the vocabulary of academic discourse. *College Composition and Communication*, 40(3):430–46. DOI:10.2307/357899.
- Szudarski, P. 2017. *Corpus linguistics for vocabulary: a guide for research*. New York: Routledge. DOI:10.4324/9781315107769.
- Taavitsainen, I. 2001. Changing conventions of writing: the dynamics of genres, text types, and text traditions. *European Journal of English Studies*, 5(2):139–50. DOI:10.1076/ejes.5.2.139.7309.
- Tyo, J., B. Dhingra en Z.C. Lipton. 2022. On the state of the art in authorship attribution and authorship verification. *arXiv*. DOI: 10.48550/ARXIV.2209.06869.
- Van de Poel, K., W.A.M. Carstens en J. Linnegar. 2012. *Text editing: a handbook for students and practitioners*. Antwerp: UPA University.
- Waldell, C. 2014. *Basic prose style and mechanics*. Michigan: Houghton. <https://www.mtu.edu/humanities/department/emeriti-faculty/waddell/basicprosestyleandmechanics.pdf>.
- Weiner, B. 1996. *Webster’s new world dictionary of media and communications*. New York: Macmillan.
- Yonos, E. 2019. A comparative analysis of lexical cohesion in native and non-native speaker writing: text linguistics and corpus perspectives. Doktorale proefskrif, Universiteit van Birmingham.

Eindnotas

¹ Dui die gemiddelde getal 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.

² Dui die gemiddelde getal 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.

³ 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 ontleding.

⁴ Wanneer 'n binêre logistiese regressie-ontleding 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-ontleding 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-ontleding 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 uitkomstes wat deur die afhanklike veranderlikes voorgestel word. Die model wat aan statistiese ontleding 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.