

'N ONDERSOEK NA AFRIKAANSE VERTAALKENMERKE IN 'N  
KORPUS KOERANTBERIGTE

deur

**Deirdré Roos**

*Tesis ingelewer ter gedeeltelike voldoening aan die  
vereistes vir die graad*

*MPhil (Vertaling)*

aan die

**Universiteit van Stellenbosch**

Departement Afrikaans en Nederlands

Studieleier: Prof. AE Feinauer

Maart 2009

*Pectora roborant animus rectus*

## VERKLARING

Deur hierdie tesis elektronies in te lewer, verklaar ek dat die geheel van die werk hierin vervat, my eie, oorspronklike werk is, dat ek die outeursregeienaar daarvan is (behalwe tot die mate uitdruklik anders aangedui) en dat ek dit nie vantevore, in die geheel of gedeeltelik, ter verkryging van enige kwalifikasie aangebied het nie.

Datum: .....



Kopiereg © 2008 Universiteit van Stellenbosch

Alle regte voorbehou

## OPSOMMING

Hierdie korpusvertaalstudie behels die bestudering van 'n eentalige vergelykende korpus van vertaalde en nievertaalde Afrikaanse koerantberigte uit *Die Burger* met behulp van *WordSmith Tools 4.0*. *WordSmith Tools* genereer statistiek, woordelyste en konkordansies wat op verskillende maniere gesorteer kan word en waarvan die data van die vertaalde en nievertaalde subkorpora met mekaar vergelyk word.

Hierdie studie volg op RG Bam (2005) se korpusvertaalstudie van Afrikaanse rugbyberigte wat bevind het dat vertaalde taal van nievertaalde taal verskil en dat dit ook verskil van die bevindinge vir Engels in 'n soortgelyke studie. Die verskil in bevindinge vir Afrikaans en Engels word toegeskryf aan die gemeenskaplikheid van die rugbyberigte. Vir hierdie studie is die domeine uitgebrei om ook berigte oor aktuele sake, kuns en vermaak, sakenuus, buitelandse gebeure en sport (rugby, atletiek, sokker, krieket, fietsry, hokkie en gholf) in te sluit. Met die uitgebreide domeine, stem my bevindinge met die vorige Afrikaanse studie ooreen ten opsigte van tipe-teken-verhouding, gemiddelde woordlengte en leksikale digtheid, maar nie ten opsigte van gemiddelde sinlengte en konvergensie nie. My bevinding ten opsigte van gemiddelde sinlengte kom ooreen met wat vir Engelse koerantberigte bevind is. Ten spyte hiervan, is dit duidelik dat Afrikaanse vertaalde tekste op 'n ander manier van nievertaalde tekste verskil as wat Engelse vertaalde en nievertaalde tekste van mekaar verskil.

'n Verdere uitbreiding op Bam se studie is die gebruik van 'n outomatiese Afrikaanse woordsoortetiketteerder wat in 2005 deur CTeXt ontwikkel is. Die geëtiketteerde data word met groot vrug aangewend by die berekening van leksikale digtheid en die vasstelling van die aantal voornaamwoorde in die onderskeie subkorpora.

Omdat korpusvertaalstudies 'n relatief jong veld is, word daar ook gekyk of die metodologie vir Engelse korpusondersoeke wat deur Laviosa-Braithwaite (1995) voorgestel is, op Afrikaans van toepassing is. Die metodologie is in die vorm van hipoteses waarvan sekere aspekte maklik met behulp van *WordSmith Tools* se verskillende funksies ondersoek word, maar ander, soos die voorkoms van superordinate, nie so maklik in die korpusmetodologie inpas nie.

## ABSTRACT

In this translation corpus study a monolingual comparable corpus of translated and nontranslated Afrikaans newspaper articles from *Die Burger* are compared with the use of *WordSmith Tools 4*. *WordSmith Tools* generates statistics, word lists and concordances that can be sorted in a variety of ways. The data generated for the translated and nontranslated subcorpora are then compared.

This study follows on a translation corpus study of Afrikaans rugby articles by RG Bam (2005), which found that translated language differs from nontranslated language and that it also differs from the results for English in a similar study. The difference between the findings for English and Afrikaans is attributed to the commonality of the rugby articles. For this study the domains are extended to include topical articles, arts and entertainment, business news, foreign news and sport (rugby, athletics, soccer, cricket, bicycling, hockey and golf). With the extended domains, my results are similar to the previous Afrikaans study regarding type-token ratio, average word length and lexical density, but not with regard to average sentence length and convergence. My finding on sentence length agrees with the finding for English newspaper articles. However, it is clear that Afrikaans translated articles differ from Afrikaans nontranslated articles and that Afrikaans differ from the way in which English translated articles differ from English nontranslated articles.

A further extension on Bam's study is the use of an automatic Afrikaans part-of-speech tagger that was developed by CTeXT in 2005. The tagged data was applied with good results to the calculation of lexical density and in determining the number of pronouns in the distinct subcorpora.

Because corpus translation studies is a relatively young field, the methodology suggested by Laviosa-Braithwaite (1995) for corpus studies in English is tested to see whether it is applicable to Afrikaans. The methodology is in the form of hypotheses. Certain aspects are investigated easily by means of *WordSmith Tools*, but other aspects, such as die occurrence of superordinates, is not so readily applicable to the corpus methodology.

## ERKENNING

Ek wil graag dankie sê vir my studieleier, prof. Ilse Feinauer, vir haar aanmoediging en ondersteuning, nie net in woord nie, maar ook in daad: Dankie vir die departementele en NNS-beurs wat daartoe gelei het dat ek my aandag voltyds op die navorsing kon fokus.

Dankie Gerhard vir jou onvoorwaardelike en deurlopende ondersteuning en natuurlik vir die grafieke!

Dankie ook vir Reinhardt wat verstaan het wanneer sy ma eerder wou "eksperimente" doen en haar "boek" skryf.



# 'N ONDERSOEK NA AFRIKAANSE VERTAALKENMERKE IN 'N KORPUS KOERANTBERIGTE

## INHOUDSOPGAWE

VERKLARING.....	i
OPSOMMING .....	ii
ABSTRACT.....	iii
ERKENNING .....	iv
HOOFSTUK 1 INLEIDING.....	1
1.1    Agergrond .....	1
1.2    Probleemstelling.....	2
1.3    Hipoteses.....	3
1.4    Metodologie.....	4
1.5    Verdeling van hoofstukke.....	6
HOOFSTUK 2 LITERATUURSTUDIE .....	7
2.1    Inleiding .....	7
2.2    Polisisteemteorie.....	8
2.3    Beskrywende Vertaalstudies.....	10
2.3.1    Veronderstelde vertaling.....	12
2.3.2    Norme.....	13
2.3.3    Universele kenmerke en waarskynlikheidswette.....	17
2.3.4    Samevatting.....	20
2.4    Die invloed van korpuslinguistiek.....	20
2.4.1    Vertaalgedrewe korpora.....	21
2.4.2    Saamstel van 'n korpus.....	26
2.4.3    Teksontleding.....	29
2.4.4    Samevatting.....	35

<b>2.5</b>	<b>Korpusvertaalstudies</b> .....	<b>35</b>
2.5.1	<i>Vertaalkenmerke</i> .....	36
2.5.2	<i>Ondersoeke na vertaalkenmerke</i> .....	40
2.5.3	<i>Metodologie vir ondersoeke na vertaalkenmerke</i> .....	42
2.5.4	<i>Bevindinge van korpusondersoeke van koerantberigte</i> .....	43
2.5.5	<i>'n Kritiese beskouing oor korpusvertaalstudies</i> .....	46
<b>2.6</b>	<b>Samevatting</b> .....	<b>53</b>
<b>HOOFSTUK 3 NAVORSINGSRESULTATE</b> .....		<b>54</b>
<b>3.1</b>	<b>Inleiding</b> .....	<b>54</b>
<b>3.2</b>	<b>Modus operandi</b> .....	<b>54</b>
<b>3.3</b>	<b>Stap 1: WordSmith Tools-gegenereerde statistiek</b> .....	<b>58</b>
3.3.1	<i>Tipe-teken-verhouding, gemiddelde sinlengte en gemiddelde woordlengte</i> .....	58
3.3.2	<i>Standaardafwyking-data vir TTV en gemiddelde sinlengte</i> .....	59
<b>3.4</b>	<b>Stap 2: Frekwensielys</b> .....	<b>61</b>
<b>3.5</b>	<b>Stap 3: Lemmatisering</b> .....	<b>63</b>
<b>3.6</b>	<b>Stap 4: Etikettering van data</b> .....	<b>64</b>
3.6.1	<i>Voornaamwoorde</i> .....	65
3.6.2	<i>Leksikale digtheid</i> .....	66
3.6.3	<i>Superordinate</i> .....	67
<b>3.7</b>	<b>Stap 5: Konkordansie</b> .....	<b>68</b>
3.7.1	<i>Opsionele dat</i> .....	68
3.7.2	<i>Kollokasiebotsings</i> .....	82
3.7.3	<i>Oortollige grammatikale items</i> .....	88
<b>3.8</b>	<b>Samevatting</b> .....	<b>89</b>
<b>HOOFSTUK 4 GEVOLGTREKKINGS EN AANBEVELINGS</b> .....		<b>90</b>
<b>4.1</b>	<b>Inleiding</b> .....	<b>90</b>
<b>4.2</b>	<b>Bespreking van bevindinge</b> .....	<b>91</b>
<b>4.3</b>	<b>Bevindinge ten opsigte van hierdie studie se hipoteses</b> .....	<b>96</b>
<b>4.4</b>	<b>Verdere navorsing</b> .....	<b>96</b>
<b>4.5</b>	<b>Samevatting</b> .....	<b>98</b>
<b>BRONNELYS</b> .....		<b>99</b>

<b>BRONNE</b> .....	99
<b>ANDER BRONNE GERAADPLEEG</b> .....	106
<b>BYLAAG A</b>	
Grafieke vir standaardafwyking van TTV en gemiddelde woord- en sinlengte .....	108
<b>BYLAAG B</b>	
Lemmatisering van subkorpora.....	115
<b>BYLAAG C</b>	
Artikel wat deur outomatiese woordsoortetiketteerder deur CText bewerk is.....	118
<b>BYLAAG D</b>	
Artikel 8 in die vertaalde en nievertaalde subkorpus.....	123
<b>BYLAAG E</b>	
Kollokasies met <i>dat</i> in Vertaalde subkorpus .....	125
<b>BYLAAG F</b>	
Kollokasies met <i>dat</i> in Nievertaalde subkorpus .....	126
<b>BYLAAG G</b>	
Soekterme gebruik vir konkordansies .....	127
<b>Lys Tabele</b>	
Tabel 2.1: Laviosa-Braithwaite (1995) se metodologie vir ondersoek na Engelse vertaalkenmerke .....	42



Tabel 2.2:	
Vergelyking van bevindinge van korpusondersoek van koerantberigte ten opsigte van vertaalde subkorpus (VS) en nievertaalde subkorpus (NS) .....	45
Tabel 3.1:	
Praktiese bewerking van Laviosa-Braithwaite (1995) se hipotese-metodologie.....	55
Tabel 3.2:	
Praktiese bewerking van aspekte wat nie in Laviosa-Braithwaite (1995) se metodologie gedek word nie .....	56
Tabel 3.3:	
Praktiese bewerking stapsgewys .....	57
Tabel 3.4:	
Statistiek genereer deur <i>WordSmith Tools</i> .....	58
Tabel 3.5:	
Standaardafwyking vir TTV, gemiddelde woord- en sinlengte .....	60
Tabel 3.6:	
Standaardafwyking in verhouding met TTV, gemiddelde woord- en sinlengte	61
Tabel 3.7:	
Die voorkoms van woorde in tekste .....	62
Tabel 3.8:	
Voorkoms van 108 frekwentste woorde teenoor hapax legomena .....	63
Tabel 3.9:	
Voorbeelde van die woordklas residu .....	64

Tabel 3.10:	
Voorbeelde van die woordklas uniek .....	65
Tabel 3.11:	
Frekwensie van voornaamwoorde .....	66
Tabel 3.12:	
Leksikale digtheid .....	66
Tabel 3.13	
Superordinate in artikel 8 .....	67
Tabel 3.14:	
Voorkoms van opsionele dat .....	82
Tabel 3.15:	
Soekwoorde se frekwensie in onderskeie subkorpora .....	83
Tabel 3.16:	
Kollokasiefrekwensietabel vir <i>span</i> in vertaalde subkorpus .....	84
Tabel 3.17:	
Kollokasiefrekwensietabel vir <i>span</i> in vertaalde subkorpus .....	85
Tabel 3.18:	
Woordsoorte se kombinasie met <i>span</i> .....	86
Tabel 4.1:	
Bevindinge van hierdie ondersoek ten opsigte van vertaalde subkorpus (VS) teenoor nievertaalde subkorpus (NS) .....	90

Pectora robotant cultus recti

# HOOFSTUK 1

## INLEIDING

### 1.1 Agtergrond

In hierdie korpusvertaalstudie maak ek gebruik van 'n eentalige vergelykende korpus van vertaalde en nievertaalde koerantberigte in Afrikaans om te kyk hoe vertaalde taal van nievertaalde taal verskil. Die korpus is van niks af opgebou deur vertaalde berigte in *Die Burger* te identifiseer en daarna nievertaalde berigte uit dieselfde koerant uit dieselfde domein (byvoorbeeld sport), onderwerp (byvoorbeeld krieket) en tydsraamwerk (Augustus tot Oktober 2006). Ek bou voort op die werk wat deur Bam in 2005 gedoen is op 'n korpus Afrikaanse koerantberigte oor die Wêreldbekerrugby.

'n Eentalige vergelykende korpus bestaan uit 'n stel vertalings en 'n stel sogenaamde "oorspronklike" dokumente - dokumente wat nie vertaal is nie. Die doel is om die twee stelle data te vergelyk op grond daarvan dat hulle in dieselfde omgewing geskep is om soortgelyke rolle te vervul (Kenny, 2001:59). Laviosa (2002:39-40) beweer dit is makliker om koerantberigte te gebruik vir vergelykbare data omdat mens met veiligheid kan aanneem dat die vertaalde en nievertaalde koerantberigte vir dieselfde leserskap bedoel is.

Koerantberigte is ook belangrik omdat dit *a series of snapshots of our life and our culture, often from a very specific viewpoint* (Reah, 1998:1) gee. Nuus - inligting oor onlangse gebeure wat van belang is vir 'n groot genoeg groep of wat die lewens van sodanige groep kan beïnvloed - word deur die redakteur vir die leser bepaal (Reah, 1998:3-4). Die koerantleser is nie die ontvanger van nûwe inligting oor onlangse gebeure nie, maar van geselekteerde inligting oor onlangse gebeure - wat ideologies weergegee kan word sodat dit vir die leser moeilik is om 'n onafhanklike besluit te neem oor wat sy/haar eie siening van die gebeure werklik is (Reah, 1998:9). Koerante is kulturele artefakte, want taal dra sy eie emosionele en kulturele lading en wanneer 'n boodskap

deur taal oorgedra word, is daar noodwendig waardes in die boodskap vasgelê (Reah, 1998:54-5).

Vertaling kan 'n rol speel tydens die insameling, skep en/of verspreiding van internasionale nuus (Kang, 2007:219). So kan inligting oor geleenthede wat in ander tale geskied, toeganklik gemaak word deur tolke (mondelling vertaling); 'n nuusagentskap soos Agence France-Presse (AFP) versprei byvoorbeeld inligting oor die Grand Prix en Parys se modenuus in Frans, Engels, Duits, Spaans, Portugees en Arabies ([www.afp.com](http://www.afp.com)); en die Afrikaanse koerant vertaal die nuus van die internasionale nuusagentskap. 'n Plaaslike voorbeeld is ook van toepassing: 'n Politieke vergadering in KwaZulu-Natal word waarskynlik in isiZulu behartig, 'n joernalis wat vir SAPA (South African Press Association) werk en nie isiZulu verstaan nie, maak dalk van 'n tolk gebruik om die inligting in te samel, skryf sy/haar artikel in Engels, SAPA versprei die Engelse artikel en *Die Burger* vertaal die Engelse artikel in Afrikaans vir gebruik in die koerant.

## **1.2 Probleemstelling**

Na wete is daar o.a. twee korpusvertaalstudies wat eentalige vergelykende korpora saamgestel het deur van koerantberigte gebruik te maak (sien 2.5.4): een vir Engels (Laviosa, 1997:315; 2002:62-3) en een vir Afrikaans (Bam, 2005). Bam (2005) bevind dat Afrikaanse vertaalde koerantberigte wel van Afrikaanse nievertaalde koerantberigte verskil. Hy bevind ook dat die resultate vir Afrikaanse koerantberigte verskil van die resultate vir Engelse koerantberigte. Hy skryf dit daaraan toe dat sy korpus beperk was tot slegs berigte oor die Wêreldbekerrugby.

Korpusvertaalstudies is 'n relatief nuwe veld wat in die laat jare 90 ontstaan het as 'n nuwe navorsingsgebied waarin korpuslinguistiek-hulpmiddels op vertaalstudies toegepas word (Kruger, 2002:70). Omdat die hulpmiddels nie primêr vir vertaalstudies ontwikkel is nie, word metodologieë nog verfyn en verbeter. Laviosa-Braithwaite (1995) stel 'n metodologie vir ondersoeke na vertaalkenmerke in Engels met gebruik van eentalige vergelykende korpora

voor. Die metodologie vir vertaalkenmerke is in die vorm van hipoteses in drie kategorieë (sien 2.5.3), naamlik eksplisitering<sup>1</sup> (sien 2.3.3), vereenvoudiging (sien 2.3.3) en normalisering (sien 2.3.3).

Met hierdie studie beoog ek om die volgende vrae te beantwoord:

- Sal 'n uitbreiding van domein lei tot ander resultate as dié van Bam (2005) vir Afrikaans? en
- Is die voorgestelde metodologie vir Engels ook van toepassing op Afrikaans?

### 1.3 Hipoteses

1. Vertaalde taal verskil van nievertaalde taal op grond van die
  - aantal funksiewoorde, soos voorsetsels en lidwoorde, teenoor inhoudswoorde, naamwoorde en byvoeglike naamwoorde - (dus die leksikale digtheid);
  - aantal woorde wat gereeld gebruik word teenoor dié wat nie gereeld gebruik word nie;
  - aantal kere wat woorde wat gereeld gebruik word, herhaal word; en
  - onderskeie vorme van die verskeidenheid woorde wat gereeld gebruik word (lemmas).
2. Laviosa-Braithwaite (1995) se metodologie vir ondersoek na vertaalkenmerke deur middel van die eentalige vergelykende korpus is van toepassing op Afrikaans ten opsigte van
  - die voorkoms van *dat* in indirekte rede;
  - oortollige grammatikale strukture (soos toutologie en pleonasme);

---

<sup>1</sup> Ek gebruik *eksplisitering* na aanleiding van 'n Afrikaanse artikel oor korpusvertaalstudies: Naudé, JA. 2003. Psalm 1-50 in Afrikaans. 'n Korpusgebaseerde vertaalkundige analise. *Acta Theologica* 1: 83-114.

- die aantal voornaamwoorde;
- die verhouding woordtipe teenoor woordtekens (as mens die woorde *die*, *die* en *'n* as voorbeeld gebruik, is daar drie woordtekens, naamlik *die*, *die* en *'n*, dus drie in totaal, maar net as twee woordtipes, naamlik *die* en *'n*) (die tipe-teken-verhouding dus);
- die voorkoms van superordinate;
- die gemiddelde sinlengte; en
- die voorkoms van kollokasiebotsings.

#### 1.4 Metodologie

Korpusvertaalstudies maak gebruik van die vergelykende model en in hierdie studie spesifiek word daar gebruik gemaak van 'n eentalige vergelyking. Die doelwit van die vergelykende model is om korrelasies tussen die twee kante van die verhouding te ontdek en dit kan nie gebruik word om oorsake te verduidelik en gevolge te voorspel nie (Chesterman, 2000). Daar kan wel stellings gemaak word en die studie is dus beskrywend van aard.

Korpusnavorsing word oor die algemeen gesien as empiriese beskrywing van taal (Kenny, 1998a:50), kwantitatief en/of kwalitatief (Tymoczko, 1998:652) en datagebaseer of datagedrewe (Tognini-Bonelli, 2001:81;84). Tognini-Bonelli (2001) tref 'n onderskeid tussen datagebaseerde en datagedrewe korpusstudies. Waar datagebaseerde studies korpora gebruik om 'n bestaande teorie te bewys, is die datagedrewe studies verbind tot die integriteit van die data as geheel. Kenny (1998) beskou datagedrewe studies as 'n *bottom-up* benadering teenoor datagebaseerd as *top-down*. Hierdie studie beoog om te toets of bestaande bevindinge en teorieë ook vir Afrikaans geld en kan dus as datagebaseerd beskou word.

Vir hierdie studie is 'n literatuuroorsig gedoen oor die teoretiese raamwerk van korpusvertaalstudies, die saamstel van 'n korpus, korpuslinguistiek-hulpmiddels waarmee data ontleed word en ondersoek wat in



korpusvertaalstudies gedoen is. 'n Korpus van vertaalde en nievertaalde koerantberigte is saamgestel en met behulp van *WordSmith Tools* ontleed.

As uitbreiding op Bam (2005) se studie oor Afrikaanse rugbyberigte, is 'n eentalige vergelykende korpus saamgestel uit 45 vertaalde en 45 nievertaalde koerantberigte wat in *Die Burger* geïdentifiseer is op grond van die bronaanduiding, soos SAPA, onderaan die berig. Die korpus se totale grootte is 36 733 woorde (of "lopende woorde" omdat dit al die woorde in die korpus is soos wat dit op mekaar volg - 17 450 vertaal en 19 283 nievertaal). Die volgende domeine is ingesluit: aktueel (3 berigte), kuns en vermaak (2 berigte), sake (12), buitelandse (3) en sport (25 berigte in totaal), bestaande uit rugby (5), atletiek (2), sokker (6), krieket (8), fietsry (1), hokkie (2) en gholf (1). Die indeling is soos dit in die koerant aangedui is. Die berigte val almal in die tydgleuf Augustus tot Oktober 2006.

Die vertaalde koerantartikels is eerstens in gedrukte koerante geïdentifiseer op grond van die aangeduide bron van die berigte, byvoorbeeld SAPA en SAPA-AFB of 'n naam onderaan die berig. Daarna is 'n nievertaalde berig uit dieselfde domein geselekteer met 'n soortgelyke onderwerp. Om artikels op grond van 'n vergelykbare lengte te kies, was nie haalbaar nie, omdat vertaalde berigte oor die algemeen korter is as nievertaalde berigte. Daarna is artikels van die elektroniese argief van *Die Burger* ([www.dieburger.com](http://www.dieburger.com)) afgetrek. Afgelaaste artikels is as tekslêers apart gestoor en die lêername sluit die nommer, datum, domein en soms onderwerp in, byvoorbeeld 14-2006-08-30-sake-vert-verbruikersvertroue. Hier word dus aangedui dat dit die 14de vertaalde artikel is en dat dit op 30 Augustus 2006 verskyn het in die domein "sake", met die onderwerp "verbruikersvertroue". Soms kon daar nie 'n vergelykbare nievertaalde artikel op dieselfde dag geïdentifiseer word nie, maar wel op 'n ander dag tussen Augustus en Oktober 2006. Dan verskil die datum van die vergelykbare nievertaalde berig, byvoorbeeld 14-2006-08-31-sake-nievert-verbruikersvertroue.

*Wordsmith Tools* is gebruik om die data te ontleed en kwalitatiewe ondersoeke na patrone en reëlmatighede in die vertaalde en nievertaalde berigte te doen.

## 1.5 Verdeling van hoofstukke

In hoofstuk 2 word die bevindinge uit die literatuurstudie uiteengesit.

Beskrywende Vertaalstudies word as die teoretiese raamwerk vir die studie bespreek. Daar word gekyk na die invloed van die polisisteamteorie op die ontwikkeling van Beskrywende Vertaalstudies, waarom vertaling as normbeheerde aktiwiteit beskou word, en hoe Toury se idee van veronderstelde vertaling die afbakening van die studie-objek vergemaklik. Daar word ook gekyk na Toury se waarskynlikheidswette, die wet op interferensie en die wet op groeiende standaardisering wat die soeke na universele vertaalkenmerke in korpusvertaalstudies inspireer. Die invloed van korpuslinguistiek en die teorie agter die saamstel van 'n korpus word ook bespreek, asook hoe die data met behulp van *WordSmith Tools* ontleed word. Ondersoeke wat deur middel van eentalige vergelykende korpora gedoen is en Laviosa se voorgestelde metodologie word bespreek waarna 'n kort kritiese beskouing oor korpusvertaalstudies gegee word.

In hoofstuk 3 word verslag gelewer oor die resultate van die ontleding van die vertaalde en nievertaalde subkorpora.

Gevolgtrekkings en aanbevelings word in hoofstuk 4 aangebied.





## HOOFSTUK 2 LITERATUURSTUDIE

### 2.1 Inleiding

Wehrmeyer (2004:214) glo dat korpusvertaalstudies kenmerkend is van die multidissiplinêre benadering van die afgelope tyd en dat dit 'n drievoudige teoretiese basis het, naamlik linguistiek, wiskunde en rekenaarprogrammering. Linguisties, omdat dit ontleen is aan die beginsels van Beskrywende Vertaalstudies en die polisisteemteorie; wiskundig, omdat dit staatmaak op 'n aantal statistiese tegnieke, en rekenaarprogrammering, omdat dit, uit die aard van die saak, 'n rekenaargesteunde dissipline is (Wehrmeyer, 2004:215-216). Ek verkies egter Laviosa (2002:5) se siening dat die twee hoofbronne van invloed en inspirasie vir korpusgebaseerde vertaalstudies korpuslinguistiek (wat gebruik maak van statistiek en rekenaarlinguistiek) en Beskrywende Vertaalstudies is.

In hierdie literatuurstudie word gekyk na die invloed van die polisisteemteorie op vertaalstudies, wat Beskrywende Vertaalstudies behels en hoe korpuslinguistiek se metodologie op vertaalstudies toegepas word. Toury se werk word as vertrekpunt vir die literatuurstudie gebruik, maar word aangevul deur sieninge van ander skrywers. Sy werk oor norme, die klem op die vertaling as behorende tot die doelsisteem - met die bronteks bloot 'n bron van inligting (Baker, 1993:238) - en die wegbeweeg van die bronteks, het die pad voorberei vir korpuswerk omdat dit die dissipline in staat gestel het om die obsessie met die bestudering van individuele instansies in isolasie (een vertaling teenoor een bronteks) af te gooi en die noodigheid geskep het vir die bestudering van groot hoeveelhede teks (Baker, 1993:237). Sy idees van 'veronderstelde vertaling' en waarskynlikheidswette is ook van toepassing.

Korpusvertaalstudies gebruik metodes wat aan korpuslinguistiek ontleen is, maar nie alles is direk van toepassing op vertaalstudies nie. Daar word gekyk na die definisie van 'n 'korpus', hoe 'n korpus saamgestel word, watter tipe

korpora in vertaalstudies gebruik word en die belangrikheid van vergelykbaarheid vir 'n eentalige vergelykbare korpus. Die teksontledingstegnieke van *WordSmith Tools* word daarna kortliks bespreek.

Die hoofstuk sluit af met 'n bespreking van korpusvertaalstudies veral ten opsigte van die sogenaamde universele vertaalkenmerke. Die omstredenheid wat betref die universele aard van die kenmerke word bespreek, waarna daar gekyk word na Laviosa-Braithwaite (1995) se voorgestelde metodologie vir die bestudering van vertaalkenmerke. Bam (2005) se bevindinge in sy studie oor Afrikaanse vertaalde en nievertaalde koerantberigte word ook kortliks aangebied. Die hoofstuk sluit af met 'n kritiese beskouing oor korpusvertaalstudies.

## **2.2 Polisistemeorie**

Die term 'polisisteam' is geskep deur die Israeliese teoretikus Itamar Even-Zohar in die jare 70 (Øverås, 1998:1) en verwys na die idee van 'n sisteem binne 'n sisteem binne sisteme (Hodges, s.d.:5).

Even-Zohar is sterk beïnvloed deur die Russiese Formaliste wat nie net 'letterkunde' op grond van estetiese waardes of enige ander tipe kulturele reputasie aanvaar het as wettige onderwerp vir literatuurwetenskap nie, maar ook verskynsels soos "massaliteratuur" - *popular literature of all sorts, marginal and peripheral literature (such as "journalese" semi-literary texts, folktales, etc.* (Even-Zohar, 1973:11). Even-Zohar gebruik die idee van literatuur as 'n sisteem met 'n effense terminologiese aanpassing en noem dit 'n polisisteam sodat dit moontlik is om van literêre sisteme as deel van hierdie polisisteam te praat (Even-Zohar, 1973:12).

Die polisisteamteorie gebruik 'n tipologie waarin literatuur in 'n hiërgargiese digotomie teenoor mekaar geplaas word as gekanoniseerde en niegekanoniseerde sisteme, elkeen met sy eie subsisteme. Gekanoniseerde literatuur verwys na die werk wat as letterkunde beskou word en gewoonlik deur die gemeenskap as deel van die kulturele erfenis bewaar word.

Niegekanoniseerde literatuur verwys na die werk wat dikwels afgemaak word as nie esteties genoeg nie, soos sentimentele romans. Albei sisteme kan verder in subsisteme (genres) geklassifiseer word. Party sisteme is meer sentraal as ander op 'n spesifieke tyd, maar verskuiwings vind gedurig plaas sodat die sisteme wat vandag in 'n primêre (sentrale) posisie is, môre na 'n sekondêre posisie verskuif en dié in 'n sekondêre posisie verskuif na die sentrale posisie. (Even-Zohar, 1973:12-13). Die grondbeginsel wat hierdie verskuiwing beheer, is die opposisie tussen die primêre en sekondêre sisteme en word as universeel aan enige kulturele sisteem beskou (Even-Zohar, 1978:7).

Die polisisteamteorie is belangrik vir vertaalstudies omdat daar erkenning gegee word aan vertaalde literatuur. Even-Zohar (1973:12) sê *no observer of history of any literature can avoid recognizing as an important fact the impact of translations and their role in the synchrony and diachrony of a certain literature*. Vertaalde literatuur word as 'n versameling tekste gesien wat gestruktureer is en as 'n sisteem funksioneer omdat die verhouding daarvan met ander sisteme die besluit beïnvloed van wát vertaal moet word en watter spesifieke norme, gedrag en beleid aangeneem word. Dít korreleer met wat in ander sisteme plaasvind. (Even-Zohar, 1976:18-19).

Wanneer vertaalde werke 'n sentrale posisie in die literêre polisisteam van die doelkultuur inneem, is dit in 'n sterk posisie en hoef daar nie aangepas te word by doelkultuurkonvensies nie. Die vertaler probeer dus nie aan te pas by die doeltaal nie en bly dus naby aan die oorspronklike brontaal. Wanneer die posisie van vertaalde literatuur swak is, neig vertalers om meer doelkultuurgerigte werk te lewer. (Hodges, s.d.:6).

Munday (2001:111) is van mening dat die polisisteam vertaalstudies vorentoe laat beweeg het *into a less prescriptive observation of translation within its different contexts*. Die belangrike rol van die teorie in die onlangse ontwikkeling van vertaalstudies en korpusvertaalstudies is drieledig. Eerstens bevestig die polisisteamteorie dat vertaalde literatuur 'n sisteem is wat die moeite werd is om bestudeer te word. Tweedens word vertaalde tekste as iets

eiesoortigs beskou wat as groep (as 'n korpus) ondersoek kan word. Dertens word geldigheid verleen aan die bestudering van vertaalde literatuur teen die agtergrond van nievertaalde literatuur in dieselfde doelkultuur aangesien vertaalde literatuur as 'n sisteem in die doelkultuurpolisisteem funksioneer. (Kenny, 2001:49-50). Die polisisteemteorie begin ook die soeke na die universele wette vir vertaling (Hodges, s.d.:9) wat in korpusvertaalstudies groot aanhang geniet. Lind (2007:3) verwys in dié verband na die *wolf-pack pursuit of translation universals*.

Kritiek teen die polisisteem hou verband met die oorveralgemening van universele wette, die belangrikheid en relevansie van die Russiese Formalisme en die abstraktheid en subjektiwiteit daarvan (Gentzler in Hodges, s.d.:7 en ook in Munday, 2001:111). Nietemin beskou Even-Zohar (2005:1) sisteme as een van die vernaamste idees van ons tyd, wat in die meeste menswetenskappe 'n invloed gehad het. Dit het daartoe gelei dat menslike kommunikasiepatrone wat deur simbole soos kultuur, taal en literatuur beheer word, beter verstaan en bestudeer kon word. Die polisisteemteorie bevorder nie net ons kennis omdat dit ons in staat stel om verhoudings raak te sien op plekke waar dit nog nie voorheen opgemerk is nie, maar dit help ook om die meganismes van hierdie verhoudings en, daaropvolgend, die spesifieke posisie en rol van literêre tipes in die historiese bestaan van die literatuur, te verduidelik (Even-Zohar, 1976:19).

In hierdie afdeling is gesien hoe die polisisteemteorie die weg baan vir die bestudering van vertaling as 'n sisteem binne 'n literêre sisteem binne die historiese, sosiale en kulturele sisteme van die doelkultuur. Die polisisteemraamwerk is deur Gideon Toury opgeneem en uitgebrei in die ontwikkeling van Beskrywende Vertaalstudies (Øverås, 1998:1), wat vervolgens bespreek word.

### **2.3 Beskrywende Vertaalstudies**

Beskrywende Vertaalstudies (BVS) word nie as 'n verenigde benadering beskou nie en verskeie geleerdes het die benadering in verskeie rigtings

ontwikkel (Ulrych en Bollettieri Bosinelli, 1999:222-3), maar die onderliggende basis van BVS is die doelbewuste opposisie teen preskriptiewe vertaalstudies. Die idees volg uit die polisisteamteorie se siening van vertaalde literatuur as 'n sisteem op sy eie wat spesifieke norme, gedrag en beleid aanneem na aanleiding van die wisselwerking met ander sisteme. Voorstanders van die BVS poog om met werklike vertaalverskynsels te werk, vertalings te beskryf soos dit werklik voorkom en rekenskap te gee van waargenome kenmerke van vertalings met verwysing na die literêre, kulturele en historiese konteks waarin die vertalings geskep is (Kenny, 2001:49).

Volgens Hermans (1999b:32) bestaan die leerstellings van BVS daaruit dat a) literatuur as 'n komplekse en dinamiese sisteem gesien word, b) daar 'n deurlopende wisselwerking behoort te wees tussen die teorie en praktiese gevallestudies, c) dat die benadering tot literêre vertaling beskrywend, doelgeoriënteerd, funksioneel en sistemies is en d) dat daar 'n belangstelling is in die norme en beperkinge wat die produksie en ontvangs van vertalings beheer in die verhouding tussen vertaling en ander soorte teksprosesse en in die plek en rol van vertalings in 'n gegewe literatuur en in die interaksie tussen literatuur.

Kenny (2001:50) beskou nie net doelgeoriënteerdheid as 'n belangrike kenmerk van BVS nie, maar ook 'n klem op empiriese data. Van al die voorstanders van BVS, is dit Gideon Toury wat die meeste met empirisisme geassosieer word. Sentraal tot Toury se benadering is die oortuiging dat vertaalstudies 'n empiriese wetenskap is en as sodanig beoefen moet word met werklike vertaalverskynsels en nie 'n geïdealiseerde weergawe daarvan nie. As empiriese dissipline probeer BVS om verskynsels te beskryf, te verduidelik en te voorspel. Reëlmatighede in waargenome gedrag word verduidelik deur die werking van norme. (Kenny, 2001:50).

Vervolgens word Toury se benadering bespreek aan die hand van sy idees oor veronderstelde vertalings, norme en waarskynlikheidswette.



### 2.3.1 Veronderstelde vertaling

Toury (1995:35) beskryf sy idee van veronderstelde vertaling (*assumed translation*) soos volg:

*...an assumed translation would be regarded as any target-culture text for which there are reasons to tentatively posit the existence of another text, in another culture and language, from which it was presumedly derived by transfer operations and to which it is now tied by certain relationships some of which may be regarded - within that culture - as necessary or sufficient.*

Hierdie konsep is dus gebaseer op drie aannames. Eerstens word daar aangeneem dat daar in 'n ander taal of kultuur 'n ander, of selfs meer as een teks bestaan wat kronologies en logies die vertaling voorafgaan en wat as uitgangspunt daarvoor gedien het. Meer as een bronteks kan in vertalings in koerantartikels voorkom: teks A word vertaal, maar aangesien die inligting nog nie voorheen in die spesifieke koerant verskyn het nie, word inligting uit verskillende tekste in die vertaling geïnkorporeer. Tweedens behels die proses waardeur die vertaling tot stand kom, oordrag van sekere kenmerke wat die twee tekste nou gemeen het. Derdens kan daar rekenskap gegee word van 'n verhouding tussen die twee tekste ten opsigte van wat die twee gemeen het en wat vermoedelik oor die kultureel-semiotiese en linguïstiese grens oorgedra is. (Toury, 1995:34-35).

Snell-Hornby (2006:157) is egter van mening dat die konsep van 'veronderstelde vertaling' nie voldoende is vir objektiewe wetenskaplike data nie omdat daar in vandag se globalisering en hibridiese samelewing moeilik bepaal kan word wat in werklikheid *nievertaalde* teks is. Wat sou mens as relevant beskou: *the conscious intention to translate with a specific translation brief, or the mere fact of linguistic interference?* (Snell-Hornby, 2006:157). In die meertalige Suid-Afrikaanse opset en met die hegemones invloed van Engels (Heugh, 1995:329; Du Plessis, 2000:104; Prabhakaran, 2001:8, 27) is selfs nievertaalde tekste blootgestel aan interferensie van Engels (sien 2.3.3).

Volgens Hermans skep Toury se definisie van vertaling meer probleme as oplossings, veral omdat hy aan die een kant die idee van ekwivalensie weer probeer instel, maar dit aan die ander kant probeer beperk met *a reliance on norms and conventions* (Hermans, 1999a:59).

Ten spyte van Snell-Hornby se toepaslike opmerking en Hermans se kritiek, het Toury se konsep van 'veronderstelde vertaling' die voordeel dat die navorsingsobjek uitgebrei word in ooreenstemming met *real-life situations* (Toury, 1995:32-33) en is die navorser dus geregverdig om koerantberigte wat van 'n nuusagentskap af kom (en as sodanig gemerk is) as vertaalde tekste te aanvaar. Die joernalis verantwoordelik vir die nievertaalde teks kon egter ook inligting by 'n anderstalige bron gekry het en die sogenaamde nievertaalde teks kan dus ook onder invloed van 'n brontaal wees. Dit sal in gedagte gehou moet word tydens die interpretasie van die resultate.

Waar die 'ideale' vertaling voorheen deur middel van die bronteks bepaal is, verskuif Toury se idee van veronderstelde vertaling die aandag na die doelkultuur (Kenny, 2001:49). Hoewel die invloed van die bronteks erken word, gaan BVS onder invloed van die polisisteamteorie ook oor die sistemiese verskille tussen die tale en tekstuele tradisies wat in die proses betrokke is. Vervolgens word gekyk na die rol van norme in vertaling.

### 2.3.2 Norme

Hoewel Toury nie die eerste persoon was wat norme met vertaling geassosieer het nie, was hy wel verantwoordelik vir 'n *heavy dose of norms into the veins of Translation Studies in the 1970s and early 1980s* (Toury, 1998:13). Norme is die oorsetting van algemene waardes of idees wat deur 'n groep gedeel word (ten opsigte van wat reg of verkeerd en voldoende of onvoldoende is) na gedragsinstruksies wat gepas is vir en toepaslik is op spesifieke situasies. Norme skryf voor en verbied, maar gee ook 'n aanduiding van wat geduld en toegelaat sal word in sekere gedragsituasies. (Toury, 1998:15). Norme is nie bevels of voorskrifte wat deur een of ander

gesagstruktuur aan ondergeskiktes gegee word nie, maar eerder gedeelde idees oor die korrektheid van gedrag in 'n spesifieke gemeenskap (Chesterman, 1997:54). Norme is dus 'n belangrike begrip en fokuspunt in enige wetenskaplike benadering tot die bestudering en beskrywing van sosiale verskynsels, en veral gedragsaktiwiteite, omdat die blote bestaan daarvan asook die wye omvang van situasies waarop dit van toepassing is, die hoofkategorie is waarmee sosiale orde geskep en behou word (Toury, 1978:84; 1995:55).

Vertaling word as 'n normbeheerde aktiwiteit beskou wat werkzaam is tussen twee stelsels (twee tale en kulturele tradisies) (Toury, 1995:56). Vertaling is kommunikasie: 'n vorm van sosiale gedrag (Hermans, 1999b:80). Vertaling vervul 'n sosiale funksie in 'n sosiokulturele konteks en is dus onderhewig aan die beperkinge van wat deur 'n gemeenskap as gepas beskou word (Toury, 1995:53; 1998:19). Hierdie beperkinge lê op 'n skaal tussen twee uiterstes: objektiewe, relatief absolute reëls aan die een kant en heeltemal subjektiewe idiosinkrasieë aan die ander kant (Toury, 1978:83). In die praktyk is norme dus beperkinge op die vertaler se gedrag (Hermans, 1996:1). Omdat norme 'n mate van sosiale en psigologiese druk impliseer, tree dit op as praktiese beperkinge op die individu se gedrag deur sekere keuses en opsies uit te sluit, hoewel dit in beginsel steeds beskikbaar bly. Terselfdertyd maak dit voorsiening vir 'n spesifieke keuse uit 'n reeks moontlike keuses. (Hermans, 1996:5-6; 1999b:82).

Toury (1978:86-88; 1980:53-4; 1995:56-61) onderskei tussen preliminêre, inisiële en operasionele norme. Preliminêre norme het te make met die keuse van wat vertaal moet word en of dit direk uit die brontaal vertaal word of deur middel van 'n ander taal (soos om die Bybel in Afrikaans uit 'n Engelse bronteks te vertaal). Inisiële norme lei die vertaler se keuse tussen onderhewigheid aan die brontaal (wat lei tot 'n adekwate vertaling) en die doeltaal (wat lei tot 'n aanvaarbare vertaling). Operasionele norme reguleer die besluite wat tydens die vertaalhandeling geneem word ten opsigte van die makro- en mikrostruktuur van die teks. Hieronder ressorteer matriksnorme - wat die makrostruktuur van die teks help bepaal: wat word bygevoeg, wat



word weggelaat - en tekslinguistiese norme - wat die mikrostruktuur help bepaal: sinskonstruksie, woordkeuse, gebruik van skuins- of vetdruk (Hermans, 1999b:76).

Hermans (1999b:76-77) voer aan dat die enigste adekwate teks die oorspronklike teks is en redeneer dat dit die leser is wat die tekstuele verhouding bepaal. Hy voel dat mens eerder die terme 'doelgeoriënteerd' en 'brongeoriënteerd' of 'prospektief' en 'retrospektief' moet gebruik eerder as 'adekwaat' en 'aanvaarbaar'. Hy glo ook nie dat die inisiële norm 'n gedwonge keuse tussen slegs twee pole is nie, maar dat dit veelvuldige faktore behels, soos: Hoe word die bronteks gesien? Is die teks of soortgelyke tekste al voorheen vertaal? Word die vertaling ingevoer of uitgevoer? Deur 'n spreker van watter taal is dit vertaal? en Vir watter gehoor en funksie? Hy voer verder aan dat as vertaling 'n sosiokulturele aktiwiteit is soos wat met die normkonsep voorgestel word, dit onnodig is om dit ingevolge 'n enkele as te konseptualiseer.

Volgens Hermans (1999b:70) brei Chesterman met sy norme Toury se norme uit en word ander perspektiewe buiten dié van die vertaler ingesluit. Chesterman (1997:64-69) sien Toury se operasionele en inisiële norme uit 'n ander hoek. Hy onderskei twee soorte vertaalnorme, naamlik produk- en prosesnorme. Produknorme word ook beskou as verwagtingsnorme, omdat die produk uiteindelik bepaal word deur die verwagtinge van die doeltaallesers. Hierdie verwagtinge is egter nie permanent en staties nie en is hoogs sensitief vir tekstipe. Prosesnorme word beskou as professionele norme omdat die prosesnorm bepaal word deur die aard van die eindproduk waarna daar gemik word. Onder prosesnorme word verantwoordingsnorme ('n etiese norm wat verwys na professionele standarde van integriteit en deeglikheid), kommunikasienorme ('n sosiale norm waar die vertaler werk om maksimale kommunikasie tussen die partye te verseker) en verhoudingsnorme ('n linguistiese norm wat verband hou met die verhouding tussen bron- en doeltaal) onderskei (Van Rooyen, 2005:18).

Ten opsigte van die koerantberigte in hierdie studie, speel preliminêre norme 'n rol in die keuse van watter berigte vertaal word. In *Die Burger* is nuusberigte oor die Midde-Ooste meestal vertaalde artikels. Nievertaalde berigte oor gebeure in die VSA word dikwels aangevul deur vertaalde berigte. In sulke gevalle sal die vertaalde en nievertaalde berigte byvoorbeeld uit verskillende perspektiewe aangebied word. Die invloed van die inisiële en produknorme by die vertaling van koerantberigte in *Die Burger* is dat berigte maklik lees in die doeltaal en dikwels bloot onderskei kan word weens ekstralinguistiese aanduidings soos SAPA of AP as die bron (O'Connor, 2008). Die operasionele en produknorme hang af van die doelgehoor van die koerant, wát die redakteurs as nuus beskou, en ook dít wat reeds in die koerant verskyn het. Aangesien die doelgehoor Wes-Kaapse Afrikaanssprekendes is, is die nievertaalde berigte hoofsaaklik plaaslike nuus en is die internasionale berigte meestal vertaalde berigte. Watter inligting aangebied word en hoeveel ekstra inligting nodig is, word beïnvloed deur daar reeds oor die onderwerp berig is en of die perspektief verskil.

Een van die redes waarom norme in die studie van vertaling van strategiese belang is, is die verskuiwing weg van die idee van ekwivalensie na 'n fokus op die faktore wat die verhouding tussen bron- en doelt tekste bepaal (Hermans, 1999a:59). Toury (1995:61) redeneer dat norme die tipe en vlak van ekwivalensie bepaal wat in werklike vertalings ontbloom word. Hy glo dit is belangrik om norme te bestudeer omdat dit die eerste stap is om te bepaal hoe die funksie en betrekkinge van ekwivalensie tussen die twee tekste verwesenlik word. So behou Toury dus die begrip 'ekwivalensie' maar verander dit van 'n ahistoriese, grootliks preskriptiewe konsep na 'n historiese konsep.(Toury, 1995:61).

Hermans (1999a:59) glo dat hierdie siening van Toury - *it is norms that determine the (type and extent of) equivalence manifested by actual translations* (Toury, 1995:61) - ekwivalensie effektief opsy skuif en van sy spreekwoordelike troontjie afstoot. In preskriptiewe benaderings is ekwivalensie gesien as die doelwit en voorvereiste van vertaling. Toury

redukeer ekwivalensie tot 'n blote benaming van die uitkoms van 'n handeling wat om een of ander rede as vertaling beskou word. (Hermans, 1999a:58-9).

In die bestudering van vertalings is dit belangrik om in gedagte te hou dat die proses van besluitneming, en dus die werking van norme in die vertaling, hoofsaaklik in die vertaler se kop plaasvind en dus onsigbaar is (Hermans, 1996:4). Norme is nie direk waarneembaar nie. Wat mens wel raaksien, is herhaalde patrone van byvoorbeeld die keuse van tekste wat in 'n spesifieke taal vertaal word of linguistiese vorme wat in vertaalde tekste gebruik word. Hierdie patrone word dan geïnterpreteer as bewyse vir die werking van norme in die teks (Kenny, 2001:52). Toury waarsku egter dat verskynsels met 'n hoë frekwensie nie outomaties norme is nie, maar dat dit universele eienskappe van vertaalgedrag kan wees (Toury, 1978:95). Waar norme sosiaal en kultureel bepaal word en oor tyd verander, verteenwoordig universele kenmerke algemene tendense wat ongeag die vertaler, taal, genre of tydperk voorkom (Kenny, 2001:52-3). Vervolgens word daar verder gekyk na universele kenmerke en waarskynlikheidswette.

### *2.3.3 Universele kenmerke en waarskynlikheidswette*

Vertalings word beskou as *facts of the culture that hosts them* met die gepaardgaande aanname dat wat ook al die funksie en identiteit, dit binne daardie kultuur is en dus 'n eie samestelling weergee (Toury, 1995:24). Hoewel vertalings bedoel is om vir die behoeftes van die doelkultuur voorsiening te maak, wyk dit ook (op een of ander vlak) af van die goedgekeurde patrone van die doelkultuur, omdat daar altyd 'n paar kenmerke van die bronteks behou word. Om hierdie rede is vertalings dikwels duidelik onderskeibaar van nievertaalde teks en nie noodwendig omdat dit slegte vertalings met foute is nie, maar omdat 'n sekere mate van afwyking nie net regverdigbaar of aanvaarbaar nie, maar selfs verkieslik is (Toury, 1995:28). Omdat vertalings onderhewig is aan ander beperkinge as nievertaalde teks, beskou Toury dit, in ooreenstemming met die polisisteamteorie, as 'n goeie rede om vertalings as 'n spesiale sisteem of genre binne hul eie kultuur te beskou (Toury, 1995:28). Hy beveel ook aan dat

die ontleding van intertalige vorme wat in vertalings voorkom, 'n integrale deel van enige sistematiese Beskrywende Vertaalstudie as empiriese verskynsel behoort te wees (Touy, 1980:73).

Hierdie intertalige vorm het William Frawley as 'n derde kode gesien wat ontstaan uit die wedersydse oorweging van die bron- en doelkodes en dat dit dus 'n soort subkode is van elkeen van die betrokke kodes (Frawley aangehaal in Øverås, 1998:3). Die idee van vertaling as 'n soort aparte subtaal is nie nuut nie. Wat wel nuut is, is die nie-evaluerende siening binne Beskrywende Vertaalstudies van 'intertaal' as 'n noodwendige aspek van vertaling. (Øverås, 1998:3). Øverås (1998:3) onderskei tussen (aanvaarbare) vervreemding (*foreignization*) en (afgekeurde) vertaaltaal (*translationese*). Sy sien *translationese* soos Baker (1993:248) dit beskryf as die ongewone verspreiding van kenmerke wat duidelik die resultaat is van 'n vertaler se onbedrewenheid of onbevoegdheid in die doeltaal. Tirkkonen-Condit (2002:207) daarenteen, onderskei tussen *translationese* (dieselfde as Baker se beskrywing) en vertaaluniversele kenmerke (*translation universals*). Hierdie vertaaluniversele kenmerke is nie vertaalfoute nie en ook nie dieselfde as Øverås se vervreemding nie, maar dui op *frequencies of lexical items, syntactic patterns, etc. which deviate from those in originally produced texts* (Tirkkonen-Condit, 2002:208).

Baker het in 1996 vier kategorieë moontlik universele kenmerke geïdentifiseer naamlik eksplisitering, die geneigdheid om inligting in vertaling uit te spel eerder as om dit implisiet te hou; vereenvoudiging, die geneigdheid om die taal of die boodskap in die doelteks te vereenvoudig; normalisering of konserwatisme, die geneigdheid om aan te pas of selfs die patrone en praktyke wat tipies in die doeltaal voorkom, te oordryf en gelykmaking (*levelling out*), die geneigdheid van die vertaalde teks om gemiddeld te wees en weg te bly van uiterstes (Ulrych en Bollettieri Bosinelli, 1999:235). (Sien 2.5.) Norme en universele kenmerke verteenwoordig verduidelikings vir patrone wat herhaaldelik in vertaalde tekste waargeneem word. Maar dit kan ook gebruik word vir voorspelling: as ons aanvaar dat 'n tipe gedrag 'n

universele kenmerk is, dan voorspel ons dat daardie kenmerk sal voorkom in vertalings wat nog nie ondersoek is nie. (Kenny, 2001:54).

Toury (1995:259) glo alle wetenskappe word gekenmerk deur 'n soeke na wette. Hy beskou die formulering van sodanige wette as die uiteindelijke doelwit van vertaalstudies. Die wette is egter nie absoluut nie, maar dui op die waarskynlikheid dat 'n sekere gedrag in gegewe toestande sal plaasvind (Toury, 1995:16). Toury (1995:259) gee twee voorbeelde van waarskynlikheidswette vir vertaling, naamlik die wet van groeiende standaardisering en die wet van interferensie.

Die wet van groeiende standaardisering maak die aanspraak dat tekstuele verhoudinge in die oorspronklike teks dikwels aangepas word, soms selfs heeltemal geïgnoreer word, ten gunste van gebruikelike opsies in die doeltaal (Toury, 1995:268). As gevolg hiervan manifesteer vertalings groter standaardisering as in die brontekste, word dubbelsinnigheid uitgeskakel en eenvoudiger weergegee (Toury, 1995:268-270). Toury gebruik 'n aanhaling van Even-Zohar (Toury, 1995:272) om te impliseer dat vertaling geneig is om konserwatief te wees: *translation tends to assume a peripheral position in the target system, generally employing secondary models and serving as a major factor of conservatism.*

Die wet van interferensie voorspel dat verskynsels (positief of negatief) wat aan die bronteks behoort, aan die doelteks oorgedra word (Toury, 1995:275). Positiewe oordrag hou verband met die waarskynlikheid dat eienskappe wat in die doeltaal bestaan en wel gebruik word, gekies word (Toury, 1995:275), soos dieselfde sin- en paragraafstruktuur (Pym, 2007:7). Negatiewe oordrag is waar daar afgewyk word van die normale praktyk in die doelsisteem (Toury, 1995:275). Omdat interferensie negatiewe sowel as positiewe oordrag impliseer, beteken dit dus dat interferensie outomaties deel sal wees van vertalings (Toury, 1995:275).

Volgens Toury (1995:265) moet die soeke na wette reëlmatighede van werklike gedrag in ag neem deur *an evergrowing (and ever more variegated)*



*series of studies into well-defined corpuses.* Die bevindinge van goed deurdagte studies beïnvloed die onderliggende teorieë en dra dus so by tot die staving of weerlegging daarvan (Toury, 1995:15). Bevindinge van studies wat op mekaar voortbou, maak dit moontlik om 'n reeks koherente wette te formuleer wat die inherente verhoudings tussen al die veranderlikes van toepassing op vertaling weergee (Toury, 1995:16).

In hierdie onderafdeling is daar gekyk na hoe die bestudering van reëlmatighede kan lei tot die identifisering van moontlike universele kenmerke van vertaling. Toury se wet van groeiende standaardisering en die wet van interferensie is waarskynlikheidswette omdat hulle gebruik kan word om voorspellings oor die vertaalproduk te maak.

#### 2.3.4 Samevatting

Toury se Beskrywende Vertaalstudies is hierbo beskryf na aanleiding van sy idees oor veronderstelde vertalings, norme en waarskynlikheidswette. Dat enige teks wat as vertaling aanvaar word in die doelkultuur as vertaling geklassifiseer word, beteken dat vertaalde koerantberigte bestudeer kan word as vertalings omdat dit as sodanig in koerante aangedui word deur die aanduiding van bron onderaan die berig.

Hoewel norme nie waargeneem kan word nie, kan reëlmatighede in tekste wel geïdentifiseer word en dit kan lei tot die identifisering van moontlike vertaalkenmerke of waarskynlikheidswette. Vervolgens word daar gekyk hoe metodologie uit die korpuslinguistiek gebruik kan word om reëlmatighede te ondersoek.

## 2.4 Die invloed van korpuslinguistiek

Korpusgebaseerde vertaalnavorsing het in die laat jare 90 uit korpuslinguistiek ontstaan (Kruger, 2002:70). Korpuslinguistiek is die tak van linguistiek wat taal met behulp van korpora, 'n versameling tekste wat volgens eksplisiete ontwerp kriteria vir 'n spesifieke doel bymekaar gemaak is,

bestudeer en is 'n empiriese benadering tot die beskrywing van taal (Kenny, 1998a:50). Volgens Stubbs (in Olohan, 2004:16) is korpuslinguistiek gemoed met *what frequently and typically occurs as opposed to isolated, unique instances of language*. Maar dit is nie 'n breinlose proses van outomatiese taalbeskrywing nie (Kennedy, 1998:2).

Hoewel daar verskeie definisies vir 'korpora' is, is daar toenemende konsensus in korpuslinguistiek dat 'n korpus 'n versameling masjienleesbare, outentieke tekste is (geskrewe en gesproke data) waarvan monsters geneem is sodat dit 'n spesifieke taal of taalvariëteit sal verteenwoordig (McEnery, Xiao en Tono, 2006:5). In vertaalstudies was 'n korpus oorspronklik enige versameling tekste, in geprosesseerde of ongeprosesseerde vorm, gewoonlik van 'n spesifieke skrywer. Met korpuslinguistiek se invloed, het die definisie op drie maniere verander: i) 'n korpus beteken nou hoofsaaklik 'n versameling masjienleesbare tekste wat outomaties of semi-outomaties op verskeie maniere ontleed kan word; ii) 'n korpus is nie beperk tot geskrewe werk nie, maar sluit ook gesproke taal in; en iii) 'n korpus kan 'n groot aantal tekste uit verskeie bronne deur verskillende skrywers en sprekers en oor verskeie onderwerpe insluit. Wat ook belangrik is, is dat dit vir 'n spesifieke doel en volgens eksplisiete ontwerp kriteria saamgestel is om seker te maak dit is verteenwoordigend van 'n gegewe area of monster van taal waarvoor daar verslag gelewer sal word. (Baker, 1995:225).

Enige korpuslinguistiekontleding is afhanklik van die skep van die korpus en programmatuur vir die waarneming, ontleding en prosessering daarvan (Laviosa, 2002:6). Daar word nou eers gekyk na die tipes korpora wat in vertaalstudies gebruik word en dan na hoe 'n korpus saamgestel word en die hulpmiddels wat vir die ontleding gebruik word.

#### 2.4.1 Vertaalgedrewe korpora

Vertaalgedrewe korpora is volgens Zanettin (2000:105-6) 'n eentalige vergelykende korpus, 'n tweetalige vergelykende korpus en 'n parallelkorpus. Die eentalige vergelykende korpus bestaan uit 'n stel vertalings en 'n

vergelykbare stel tekste wat spontaan in dieselfde taal geskep is en volgens soortgelyke ontwerp kriteria geselekteer is (Zanettin, 2000:105). Dit is 'n vertaalgedrewe korpus omdat die nievertaalde komponent saamgestel is na aanleiding van die vertaalde komponent (Laviosa, 1997:293). In hierdie studie word van 'n eentalige vergelykende korpus gebruik gemaak. Die tweetalige vergelykende korpus bestaan uit tekste wat spontaan geproduseer is in twee tale, onder soortgelyke omstandighede en binne dieselfde domeine. Die doel hiervan is die onttrekking van tweetalige terminologie en kan vir die opleiding van vertalers aangewend word. Die parallelkorpus bestaan uit vertalings en die onderskeie brontekste daarvan. (Zanettin, 2000:106). Dit kan ook sin vir sin op mekaar afgestem wees (*aligned*) soos by vertaalgeheue-programmatuur. Baker (1995:232) onderskei ook veeltalige korpora, wat sy beskryf as *sets of two or more monolingual corpora in different languages, built up either in the same or different institutions on the basis of similar design criteria*. As dit dus 'n stel van twee eentalige korpora in twee verskillende tale is, is dit dieselfde as Zanettin se tweetalige vergelykende korpus.

Baker (1995:235) stel voor dat daar wegbeweeg word van die vergelyking tussen bron- en doeltaal en dat die verskille tussen nievertaalde teks en vertaalde teks verken word. Die verandering in fokus is na bewering noodsaaklik om die linguistiese patrone wat spesifiek van toepassing is op vertaalde tekste, onafhanklik van die betrokke bron- of doeltaal, te ondersoek (Laviosa-Braithwaite, 1995:157). Om dit te doen, word daar gebruik gemaak van 'n eentalige vergelykende korpus. Die twee stelde data moet dan 'n soortgelyke domein, taalvariëteit en tydperk voorstel en ook vergelykbaar wees ten opsigte van lengte (Baker, 1995:234). Maar Baker (2004b:185) is ook van mening dat vergelykbaarheid altyd 'n relatiewe begrip sal bly en dat mens dit in gedagte moet hou wanneer bevindinge bespreek word. Laviosa (1997:290) beskou een van die hoofeienskappe van hierdie tipe korpus die vergelykbaarheid tussen die vertaalde en nievertaalde komponent. Dit moet dus soortgelyk wees en in soveel aspekte moontlik ooreenstem sodat enige linguistiese verskille wat gevind word, betroubaar toegeskryf kan word aan die



verskil in status (vertaal teenoor nievertaal) eerder as aan verwarrende veranderlikes.

Volgens Laviosa se korpustipologie vir vertaalstudies (1997; 2002:34-38) is die korpus wat in hierdie studie gebruik word 'n volteks, sinkroniese, algemene, eentalige vergelykbare vertaalkorpus van geskrewe Afrikaans. Sy onderskei ook 'n vierde vlak waarvolgens 'n vertaalkorpus onderskeid kan maak tussen een brontaal, twee brontale of meer brontale. Selfs indien tekste gemerk is dat dit byvoorbeeld van DPA (die Duitse nuusagentskap) afkomstig is, kon dit in Engels verskaf gewees het, omdat DPA ook nuus in Engels verskaf. Daar word dus aangeneem dat die brontaal van vertaalde artikels in hierdie studie hoofsaaklik Engels is.

Die subkorpora van hierdie navorsing is dus vergelykbaar ten opsigte van die volgende aspekte: volledige artikels word gebruik vir albei stalle (volteks); teks word gekies uit 'n beperkte tydperk, naamlik Augustus, September en Oktober van 2006 (sinkronies); die berigte word beskou as verteenwoordigend van alledaagse, niegespesialiseerde taalgebruik (algemeen); die fokus is op Afrikaans vir albei stalle data (eentalig). Verder is die teks van albei ook geskryf om gelees te word (geskrewe). Sy onderskei verder vertaalmetode en kenmerke van vertalers as kategorieë. Ek neem aan dat die vertaalmetode vir die koerantberigte menslike vertaling is, maar aangesien die ouderdom, geslag en ervaring van die vertalers nie aangedui word nie, kan ek nie 'n onderskeid daarvolgens maak nie.

Om vergelykbare koerantberigte te kry, is na bewering (Laviosa, 2002:39-40) makliker omdat beide die vertaalde en nievertaalde tekste in dieselfde koerant geïdentifiseer kan word en daar om daardie rede aangeneem kan word dat die leserskap dieselfde is. Laviosa (2002:40) beweer dat die dimensies van vergelykbaarheid groter en ook minder problematies is as vir ander genres omdat daar vertaalde en nievertaalde tekste in koerante beskikbaar is en dit maklik is om elkeen se onderwerp te identifiseer uit die titel en subtitel.

In hierdie studie is die stelde data op verskeie aspekte vergelykbaar. Om te verseker dat die data ook vergelykbaar is ten opsigte van domein, is vertaalde berigte eerste geïdentifiseer en daarna nievertaalde berigte uit dieselfde domein en oor dieselfde onderwerp. Prakties het die beskikbaarheid van elektroniese weergawes van die artikels op *Die Burger* se webargief die seleksie beïnvloed.

Die vertaalde artikels wat in die korpus ingesluit is, was hoofsaaklik afkomstig van die volgende nuusagentskappe: die South African Press Association (SAPA), Agence France-Presse (AFP), Associated Press (AP), Deutsche Presse-Agentur (DPA), New Zealand Press Association (NZPA) en Reuters.

- SAPA is 'n nieregeringsnuusagentskap wat in 1938 gestig is. Dit is die hoofverskaffer van internasionale nuus aan SA nuusmedia. Die nuusdiens dek die volgende kategorieë: binnelands, internasionaal, finansies, sportstories, sportuitslae, perdewendrenne, Afrika, dagboeke (inligting oor die dag se verwagte gebeure), vermaak, parlementêr. (Van Rooyen, 2005:68; [www.sapa.co.za](http://www.sapa.co.za)).
- AFP is die oudste nuusagentskap en is in 1835 gestig. Hulle verdeel die nuus in algemene, sake- en sportnuus, en dek ook die Formule 1 Grand Prix en veral Parys se modenuus. Nuus word in Frans, Engels, Duits, Spaans, Portugees en Arabies verskaf. ([www.afp.com](http://www.afp.com)).
- AP is 'n globale nuusnetwerk wat in 1846 gestig is en is die grootste en betroubaarste bron van onafhanklike nuus en inligting, volgens hulle webwerf ([www.ap.org](http://www.ap.org)). Hulle lewer internasionale, nasionale (VSA), staats-, sport-, sake- en vermaaknuus.
- DPA is 'n nuusagentskap wat in 1949 in Duitsland gestig is. Hulle verskaf plaaslike nuus (Duitsland), verbruikersnuus (dit sluit in reis, toerisme, motors en vervoer, rekenaar, internet, multimedia, tegnologie en telekommunikasie, onderwys en beroepe, geboue en huise, tuinmaak, gesondheid, geld en regsake, gesin, kinders en aftrede, en ook mode en lewenstyl) asook nuus wat van internasionale belang is. Dit word in Duits, Engels, Spaans en Arabies gelewer ([www.dpa.de](http://www.dpa.de)).

- NZPA is in 1879 gestig en verskaf plaaslike nuus (Nieu-Seeland) oor algemene, politieke, sake-, sport- en vermaaknuus en ook internasionale nuus ([www.nzpa.co.nz](http://www.nzpa.co.nz)).
- Reuters is in 1851 in Londen gestig en is die wêreld se grootste internasionale multimedia-nuusagentskap, volgens hulle webwerf ([www.reuters.com](http://www.reuters.com)). Hulle verskaf inligting oor finansiële dienste, media en korporatiewe markte.

Hoewel daar artikels van Reuters (vertaalde artikels) in die korpus ingesluit is, was daar dikwels Reuters-artikels wat nie beskikbaar was nie. Omdat dit nie elektronies beskikbaar was nie, is daardie artikels dan nie geselekteer nie. Dit is ook so dat mens vertaalde berigte identifiseer en dan geen nievertaalde berigte kan kry wat daarmee vergelyk kan word nie.

Volgens Atkins (in Laviosa, 1997:302) kan afsonderlike tekste geïdentifiseer word op grond daarvan dat dit 'n beredeneerde eenheid is met koherente sinne en paragrawe en dat dit geïntegreerd is omdat dit 'n doelbewuste produksie van 'n skrywer is wat stilisties homogeen is. Om daardie rede besluit ek om volledige tekste in die korpus in te sluit. Dit het egter tot gevolg dat die vertaalde subkorpus minder lopende woorde (17 450) bevat as die nievertaalde subkorpus (19 283) omdat vertaalde berigte in *Die Burger* oor die algemeen korter as nievertaalde berigte is. Dit is dus moeilik om te verseker dat die twee stelle data vergelykbaar in lengte is. Wanneer mens data van verskillende groottes vergelyk, word die rou frekwensie (die werklike telling) genormaliseer na 'n gemene basis (McEnery, Xiao en Tono, 2006:52-3) (sien 2.4.3). Die lengte-aspek word dus statisties geneutraliseer en is in hierdie geval nie 'n absolute noodsaaklikheid nie.

Praktiese implikasies, en ook die afgebakende tyd waarin die studie voltooi moes word, het noodgedwonge daartoe gelei dat ek moes "vat wat ek kon kry". Teubert en Cermáková (2007:70) noem so 'n korpus wat die gevolg is van *collecting all the corpora one can lay hands upon*, 'n opportunistiese korpus wat dan ook nie aansprake daarop maak dat dit die taal verteenwoordig nie en werk op die aanname dat alle korpora in wese

ongebalanseerd is. 'n Opportunistiese korpus moet baie omvattend gedokumenteerd word om van nut te wees vir navorsers wat die resultate van die studie verder sou wou bestudeer (Teubert en Cermáková, 2007:70).

Die nadeel van so 'n redelik homogene korpus is dat die bevindinge waarskynlik net waar sal wees vir die spesifieke genre of domein wat bestudeer word (Baroni en Bernardini, 2005:6). Dit is egter belangrik om eksterne veranderlikes te beheer sodat daar op die algemene toepasbaarheid van die resultate deur middel van akkumulering van verskeie eksperimente voortgebou kan word: elke veralgemening is gebaseer op 'n klein homogene korpus, eerder as 'n enkele eksperiment met 'n groot gevarieerde korpus, waar verwarrende faktore moeilik is om te beheer (Baroni en Bernardini, 2005:6).

In hierdie onderafdeling is die eentalige vergelykende korpus bespreek waar twee subkorpora (vertaalde dokumente teenoor nievertaalde dokumente) in dieselfde taal met mekaar vergelyk word. Daar is veral aandag gegee aan die aspek "vergelykbaarheid". Vervolgens word daar gekyk na belangrike aspekte by die samestelling van 'n korpus.

#### 2.4.2 Saamstel van 'n korpus

Daar is drie oorkoepelende ontwerpkwessies volgens McEnery *et al.* (2006:8) en Olohan (2004:45-46): permanensie, verteenwoordigendheid en korpusgrootte.

Permanensie het te make met die vraag of die korpus staties of dinamies is: Word die taal van 'n spesifieke tyd weergegee of word deurlopende verandering weerspieël? Indien die inhoud beperk is tot 'n spesifieke tydperk, word tekste uit daardie tydperk geïdentifiseer en as dit gedoen is, is die samestelling van die korpus afgehandel en is die inhoud staties. Indien veranderinge oor die loop van tyd weerspieël moet word, word tekste deurlopend bygevoeg sodat die mees onlangse tydperk ook in die korpus verteenwoordig word; dus is die inhoud dinamies. Vir hierdie ondersoek word

'n statiese korpus wat beperk is tot artikels in die maande Augustus, September en Oktober van 2006, gebruik.

As 'n korpus voldoende verteenwoordigend is, is dit moontlik om veralgemenings te maak op grond van die ontleding van die data uit daardie korpus (Van Doorslaer, 1995:248). Dit hou dan verband met die geldigheid en betroubaarheid van die navorsing wat op daardie spesifieke korpus gebaseer is (Kennedy, 1998:60). Een van die groot uitdagings in korpusontwerp is of die korpus verteenwoordigend is van 'n groter tekstopulasie. Dit is ook die basis waarvolgens 'n korpus van bloot 'n versameling elektroniese tekste onderskei kan word. (Zanettin, 2000:107).

In korpuslinguistiek word die verteenwoordigendheid van 'n korpus in 'n groot mate deur twee faktore bepaal, naamlik die omvang van genres of domeine wat ingesluit word (met ander woorde, die balans) en hoe die stukke teks vir elke genre of domein geselekteer word (met ander woorde, wát van elke genre word ingesluit) (McEnery, Xiao en Tono, 2006:13). *Die Burger* merk die bladsye bo-aan om aan te dui tot watter domein die nuus op die bladsy behoort, byvoorbeeld aktueel, buiteland, kuns en vermaak, kommentaar, forum, sport en sake-nuus. In hierdie studie sou mens die domeine wat in die koerant gedek word, kon gebruik om 'n idee van die balans van die korpus te kry. Eers word vertaalde berigte geïdentifiseer (op grond van ekstralinguistiese aanwysings, soos SAPA en AP onderaan die berig). Daarna word 'n nievertaalde berig uit dieselfde domein geïdentifiseer.

Halverson (1998:2-3) se gevolgtrekking oor verteenwoordigendheid is dat 'n mens vooraf moet besluit wat die teikenpopulasie is (wat ingesluit moet word en wat uitgesluit moet word) en dit hiërargies organiseer. Indien dit nie gedoen word nie, is dit nie moontlik om te evalueer of die korpus geskik of verteenwoordigend is nie, want daar is nie 'n goed gedefinieerde begrip van wat die monster moet verteenwoordig nie. 'n Monsterraamwerk (*sampling frame*) is die fisiese lys van die moontlike populasie waaruit 'n verteenwoordigende monster gekies kan word (Halverson, 1998:4).



Korpusgrootte word bepaal deur die besluit of volledige tekste of slegs dele van tekste gebruik word. Olohan (2004:46) waarsku dat groter korpora nie noodwendig meer bruikbaar is nie. Biber (in Kennedy, 1998:69) beskou 'n korpus van 2 000 tot 5 000 woorde as groot genoeg om 'n tekskategorie te verteenwoordig. Hoewel groot verwysingskorpora belangrik is om bewerings teen te toets (Kenny, 1998b:3) kan kleiner, versigtig saamgestelde korpora vollediger ondersoek word (Malmkjær, 1998:6).

Omdat hierdie studie uit Bam (2005) se studie voortvloei, is die ontwerp kriteria opgestel as voortsetting van en uitbreiding op sy werk. Hy gebruik die onderstaande seleksiekriteria vir sy korpus (Bam, 2005:43-44):

- Tekste in die korpus is vergelykbaar volgens onderwerp, doel, tydperk en styl;
- Tekste is ongeveer dieselfde lengte;
- Tekste is willekeurig gekies binne dieselfde onderwerp, doel, tydperk en styl;
- Die korpus is groot genoeg om weg te doen met enige vooroordeel en die geldigheid van die data te verseker;
- Die korpus is klein genoeg vir genoegsame diepte en detail;
- Die korpus het dieselfde medium van verspreiding;
- Volledige tekste is ingesluit omdat linguïstiese kenmerke ewerediger daarin versprei is ; en
- Die korpus is genoegsaam verteenwoordigend van die hele genre.

Die teikenpopulasie vir hierdie ondersoek is vertaalde en nievertaalde berigte uit *Die Burger*. Bam het 60 berigte (30 vertaal en 30 nievertaald) gebruik, geselekteer oor twee maande (Junie en Julie 2004) en beperk tot sportberigte (spesifiek oor die Wêreldbekerrugby). Vir hierdie studie word 90 berigte (45 vertaal en 45 nievertaal) oor 'n tydperk van drie maande geselekteer oor al die domeine wat in die koerant gedek word. Die monsterraamwerk is dus die domeine (soos in die koerant gemerk): aktueel, kuns en vermaak, kollig, sake, buitelandse nuus en sport.

In hierdie onderafdeling is daar gekyk na die aspekte permanensie, verteenwoordigendheid en korpusgrootte. Die korpus wat vir hierdie ondersoek gebruik word, is staties omdat 'n beperkte tydperk ingesluit is. Verteenwoordigendheid is verkry deur artikels uit verskeie domeine in die koerant in te sluit. Die korpus is groot genoeg vir verteenwoordigendheid van koerantberigte as 'n tekskategorie en is 'n uitbreiding op die vorige Afrikaanse korpusvertaalstudie.

Vervolgens word gekyk na die elektroniese hulpmiddels vir die ontleding van tekste.

### *2.4.3 Teksontleding*

Korpora is waardevolle hulpbronne in hedendaagse linguistiek, maar sonder tegnieke om die groot hoeveelheid data wat ingesluit is te soek, sorteer, tel en vertoon, sou dit geen praktiese waarde gehad het nie (Kenny, 2001:33). Die basiese twee stukke gereedskap van korpusontleding is frekwensielyste en konkordansies (Partington, 2001:46). Op sy eenvoudigste is 'n frekwensielys bloot 'n lys items ('n woordelys) van gewoonlik enkelwoorde wat wys hoeveel keer daardie woord in 'n versameling tekste voorkom (Partington, 2001:46). Dit kan ook alfabeties gerangskik word. 'n Konkordansie is 'n lys teksreëls wat 'n spesifieke nodus (sleutelwoord/soekwoord) bevat wat die woord in sy konteks voorstel (Teubert en Cermáková, 2007:140). Konkordansies word gebruik om onderskeidinge tussen verskillende gebruike, patrone en betekenisse te bestudeer (Olohan, 2004:73). Die algemeenste tipe konkordansie is die sleutelwoord-in-konteks-konkordansie, wat net een reël van die uittreksel wys met die sleutelwoord in die middel van die reël (Partington, 2001:47). So kry mens 'n minder ontwrigtende perspektief op die korpus deurdat daar 'n bietjie konteks bygevoeg word. Die sleutel-in-konteks-konkordansie kan alfabeties gelys word (volgens óf die voorafgaande items óf die daaropvolgende woorde). (Barlow, 2004:209).

Die beskrywing van hierdie hulpmiddels klink heel eenvoudig, maar dit is waarskynlik die radikaalste verandering van 'n teks wat in linguïstiese ontledings gebruik word: die teks word letterlik uitmekaar geskeur om 'n woordelys te kan produseer (Barlow, 2004:207). Hoewel die rekenaar so 'n lys outomaties kan voorberei, moet dit ook in gedagte gehou word dat daar maklik foute kan insluip. Woorde word algemeen afgegrens deur spasies en verskillende soorte leestekens (Barlow, 2004:207). Koppeltekenwoorde word in *WordSmithTools* opgebreek (Munday, 1998:3), tensy mens dit spesifiseer (*Settings, Language, Hyphens separate words* moet nie gemerk wees nie). Net so word plekname soos New York en Sri Lanka en vanne soos Du Toit en Van der Merwe dan as meer as een woord getel. Spelvariasies en variasies ten opsigte van hoof- en kleinletters word nie in ag geneem nie (Kenny, 2001:34-5). 'n Stoplyst kan gebruik word om sekere woorde uit te filtreer (Barlow, 2004:208).

Woordelyste en konkordansies word tipies ontleed deur gebruik te maak van die tipe-teken-verhouding, leksikale digtheid en gemiddelde woord-, sin- en paragraaflengte. Hierdie basiese statistiek gee 'n idee van die algemene tekstuur van die data en is waardevol omdat mens hieruit moontlike vrugbare ondersoekareas kan identifiseer (Munday, 1998:7).

Tekens verwys na enige volgorde van letters met 'n ortografiese spasie aan albei kante (Baker, 1995:236). Tipes verwys na die aantal verskillende woorde in die teks (Munday, 1998:4). Die woorde *die*, *die* en *'n* word dus as drie tekens beskou, maar aangesien *die* twee keer voorkom, is daar net twee verskillende tipes (*die* en *'n*). Die tipe-teken-verhouding (TTV) word bereken deur die aantal tekens deur die aantal tipes te deel en is 'n eenvoudige aanduiding van die oppervlakkige leksikale kompleksiteit van 'n teks (Munday, 1998:4). Maar die TTV is nou verwant aan die tekslengte (Chipere, Malvern, Richards en Duran, 2004:127). *WordSmith Tools* gebruik 'n gestandaardiseerde TTV soortgelyk aan Chipere *et al.* (2004:128) se MSTTR (*mean segmental type-token ratio*) waar die TTV vir aaneenlopende segmente van dieselfde lengte bereken word (Scott, 2008a). Die voordeel hiervan is dat die aantal woorde van die kortste artikel dus gebruik kan word



as die basisgrootte van die segment (die gestandaardiseerde TTV-basis) (Chipere *et al.*: 128). Aangesien die vertaalde artikels oor die algemeen korter is, stel ek die n-waarde gelyk aan die aantal woorde in die kortste artikel, naamlik 179 (*Settings, Specific to wordlist, Standardised Type/Token*) vir albei subkorpora sodat die data met mekaar vergelyk kan word.

Die tipe-teken-verhouding gee ons 'n aanduiding van die verskeidenheid woordvorms wat in die korpus gebruik word (Kenny, 2001:34). Hoe hoër die persentasie, hoe meer gevarieerd die woordeskat (Baker, 1995:236). Hier moet mens in gedagte hou dat die rekenaar nie 'n onderskeid kan tref tussen polisemiese woorde (byvoorbeeld bank, die meubelstuk, teenoor bank, die finansiële instelling) en meerdelige eenhede (byvoorbeeld Sri Lanka) nie. Zanettin (2000:113) beskou TTV nie as voldoende bewys van leksikale variasie nie en is van mening dat dit saam met ander aanwysers soos leksikale digtheid en die verhouding hapax legomena (woorde wat net een keer in die korpus voorkom) tot die totale aantal woorde, oorweeg moet word.

Waardes vir gemiddelde lengte van woorde, sinne en paragrawe word outomaties deur *WordSmith Tools* gegee. Die program gebruik konvensionele slotaanduiders soos punte, uitroeptekens en vraagtekens gevolg deur 'n spasie en 'n aanvangshoofletter om sinne te identifiseer. Paragrawe word herken deur etikette (soos die paragraafmerker) in die teks. (Olohan, 2004:81). Statistiek oor hoeveel woorde van watter lengte (een letter/twee letters, ens.) word ook gegee (Kenny, 2001:35).

Meestal maak navorsers slegs gebruik van die statistiese berekenings wat in *WordSmith Tools* ingebou is. Wanneer statistiek wel toegepas word, word daar gefokus op beduidendheid. Daar is egter verskillende sieninge onder linguïste en vertaalgeleerders oor hoe bruikbaar hierdie toetse is en baie voel dat dit nie nodig is nie. Die data gee immers 'n aanduiding van 'n algemene geneigdheid en die statistiese bewerking gee dikwels nie meer inligting as wat die rou data gee nie. (Olohan, 2004:86). Syfers en frekwensie is slegs 'n beginpunt en fokus volgens Baker (2004b:183) ons aandag op sommige kenmerke wat waarskynlik die moeite werd sal wees om in meer

besonderhede te ondersoek. 'n Mens moet noukeurig na die data kyk en 'n gevoel vir die tekste kry sodat mens verby *low-level description* na *situated explanation* kan beweeg (Baker, 2004b:183). Mauranen (2000:120) het bevind dat noukeuriger linguistiese ontleding nuwe aspekte van frekwente items ten toon stel, naamlik dat dit ander kombinasies vorm en dikwels met ander frekwente items. Onder die frekwensie is dus 'n ander tipe variasie wat mens deur middel van uitgebreide stringe eerder agterkom as met enkele leksikale items. Mens sou dus kollokasiefrekwensietabelle en wyer konteks op hierdie studie kon toepas.

'n Kollokasiefrekwensietabel is 'n alternatief tot die deurgaang van groot hoeveelhede konkordansiereëls. Dit toon die frekwensie van woorde binne 'n spesifieke omvang van die soekwoord, soos tweede woord links (L2) of eerste woord links (L1) van soekwoord en eerste woord regs (R1) of tweede woord regs (R2) van soekwoord. (Barlow, 2004:212). Kollokasies is die boublokke van taal en kan dus gesien word as fundamentele eenhede van taal in gebruik (Barlow, 2004:213). Daar is twee benaderings om kollokasies uit 'n korpus te onttrek. Die een word veral deur leksikograwe gebruik en die doelwit is om so veel kollokasie-inligting oor soveel woorde moontlik te onttrek. In die tweede benadering, wat voorkeur geniet in korpusvertaalstudies, word 'n woord gekies en word alle, of verskynsels wat willekeurig gekies word, sistematies ondersoek. (Kenny, 2000:98-9). Kenny (1998b:4) beweer dat 'n noukeurige bestudering van kollokasiepatrone in vertaalde teks lig kan werp op die kulturele magte wat 'n rol speel, omdat kultuur en taal onontkombaar aan mekaar gekoppel is en dit dus nie sin maak om te beweer dat kulturele en linguistiese benaderings tot die bestudering van vertaling mekaar uitsluit nie.

Om die grootte en aard van die konteks te verander, lei tot ander resultate (Barlow, 2004:212). 'n Goed ontwerpte program openbaar maklik patrone in die data aan die gebruiker, maar mens moet in gedagte hou dat die programme nie net die tipe patrone wat uit die teks onttrek kan word, uitlig nie, maar dit ook beperk (Barlow, 2004:212). Om na opeenvolgings van twee woorde te kyk, openbaar dalk 'n aantal sterk kollokasies, maar 'n wyer perspektief soos opeenvolgings van vyf woorde en selfs meer, kan dalk

ander, moontlik meer subtiële assosiasiepatrone openbaar (Barlow, 2004:217).

Rekenaarlinguistiek gebruik ook markering en annotasie om die tipe elektroniese bewerking wat op 'n korpus gedoen kan word, te verbreed. Markering is 'n stelsel standaardkodes wat in 'n elektroniese dokument gevoeg word om inligting oor die teks te verskaf en het te make met die tekstuele struktuur van elke teks. Markering voeg waarde toe tot 'n korpus en laat 'n groter verskeidenheid navorsingsvrae toe. (McEnery *et al.*, 2006:22-3; 29). Annotasie vermeerder ook die verskeidenheid navorsingsvrae wat beantwoord kan word en het te make met geïnterpreteerde taalinligting op fonologiese, morfologiese, leksikale, sintaktiese, diskoers-, pragmatiese en stilistiese vlak (McEnery *et al.*, 2006:33). Geannoteerde teks maak dit makliker om inligting te onttrek en verskaf 'n standaardverwysingsbron sodat daaropvolgende studies op 'n gemeenskaplike basis vergelyk kan word (McEnery *et al.*, 2006:30). Hierdie bewerkings is egter baie tyds- en arbeidsintensief (Kenny, 2000:95) en, as gevolg van die subjektiewe aard van kategorisering, ook nie 100% akkuraat nie (Olohan, 2004:71). Om funksionele etikette by linguistiese eenhede te voeg, is 'n hoogs subjektiewe aktiwiteit wat moeilik is om konsekwent toe te pas en dit moeilik maak om die navorsing te replikeer (Reppen, Fitzmaurice en Biber, 2002:115). Zanettin (2000:114) is van mening dat mens belangrike inligting oor die konteks en visuele inligting so kan verloor. Die annotasievlak wat die meeste gebruik word, is leksikaal en wel woordsoortetikettering (McEnery *et al.*, 2006:34).

'n Outomatiese Afrikaanse woordsoortetiketteerder is in 2005 deur Suléne Pilon ontwikkel as deel van 'n NNS-befondsde projek *Afrikaans Computational Morphology* (Pilon, 2005). Alhoewel die akkuraatheid nie so goed vaar soos soortgelyke etiketteerders vir Engels en Nederlands nie, is die akkuraatheid tussen 85,87% en 93,69% (Pilon, 2005:6; 118). Ek het besluit om die woordsoortetikettering te laat doen omdat dit nuwe tegnologie is wat beskikbaar is vir Afrikaans, wat na my wete nog nie op 'n Afrikaanse korpusvertaalstudie toegepas is nie en omdat dit replisering ten opsigte van die berekening van leksikale digtheid sal vergemaklik. Etikettering is 'n

subjektiewe proses. Ek sou byvoorbeeld afkortings soos *mnr.* as naamwoorde gemerk het en simbole soos % verkeerdelik as punktuasie, waar die etiketteerder geprogrammeer is om dit as residu (sien 3.6) te klassifiseer. Net so met die klassifisering van die partikels, wat ek waarskynlik as voegwoorde en voorsetsels sou beskou het, is die etiketteerder geprogrammeer om dit as uniek (sien 3.6) te merk (skakel *en* word as partikel beskou, maar *as* soos in *meer as* steeds as voegwoord).

Om leksikale digtheid te bereken, moet mens die inhoudswoorde (naamwoorde, byvoeglike naamwoorde en werkwoorde) en funksiewoorde (lidwoorde en voorsetsels) identifiseer. Leksikale digtheid is die persentasie inhoudswoorde teenoor funksiewoorde en word bereken deur die aantal inhoudswoorde deur die totale aantal woorde van die korpus te deel en te vermenigvuldig met 100 om 'n persentasie te kry (Baker, 1995:237). De Villiers (1983:55-62) se onderskeid tussen oop en geslote woordsoorte is hier van toepassing: Oop woordsoorte is naamwoorde, werkwoorde en byvoeglike naamwoorde omdat dit meer en minder kan word en ook kan verander. Geslote woordsoorte is voornaamwoorde, lidwoorde, telwoorde, voorsetsels, voegwoorde en sekere bywoorde omdat dit beperk is en selde verandering ondergaan. Ek beskou oop woordsoorte as inhoudswoorde en geslote woordsoorte as funksiewoorde. Leksikale digtheid hou verband met inligtingstempo en -lading wat geassosieer word met die gebruik van tegniese teenoor algemene woordeskat, persentasie bekende teenoor onbekende inligting, die algehele lengte van die teks en die hoeveelheid besonderhede in die beskrywing van 'n gebeurtenis (Baker, 1995:237). Moeiliker tekste het 'n hoër leksikale digtheid as makliker tekste (Baker, 1995:238). Die leksikale digtheid word nie outomaties deur *WordSmith Tools* verskaf nie en omdat die navorser self die inhoudswoorde moet identifiseer en die berekening doen, beskou ek die outomatiese woordsoortetiketteerder as 'n nuttige hulpmiddel.

In hierdie onderafdeling is daar gekyk na tegnieke wat gebruik word om die inhoud van korpora te ontleed. Alfabetiese en frekwensiewoordelyste en konkordansies en die gepaardgaande statistiek word gebruik om gebruike en patrone in korpora waar te neem en te bestudeer. Die inhoud van die korpus

kan ook van addisionele inligting, soos die aanduiding van woordsoorte, voorsien word om die bewerking te verbreed.

#### 2.4.4 Samevatting

In die voorafgaande afdeling is daar gekyk na wat onder die term *korpus* verstaan word en watter soorte korpora vir bestudering van vertaling gebruik word. Die eentalige vergelykende korpus word in hierdie studie gebruik en is vergelykbaar omdat albei subkorpora bestaan uit volteks-artikels uit dieselfde tydperk uit dieselfde koerant. Verder is die subkorpora vergelykbaar op grond van domein en onderwerp. Die korpus vir die ondersoek is staties en verteenwoordigend van domeine wat in *Die Burger* gedek word en dit is groot genoeg om die tekskategorie koerantberigte, te verteenwoordig.

Die hulpmiddels vir teksontleding is bespreek met verwysing na die twee basiese stukke gereedskap, woordelyste (alfabeties en volgens frekwensie) en konkordansies, asook verdere funksies soos sleutelwoord-in-konteks, kollokasies en statistiese bewerkings soos tipe-teken-verhouding, leksikale digtheid en gemiddelde woord-, sin- en paragraaflengte.

Daar is kortliks verwys na die annotasie van die data in 'n korpus op leksikale vlak, naamlik woordsoortetikettering en hoe dit die berekening van leksikale digtheid kan vergemaklik. In die volgende afdeling word korpusvertaalstudies bespreek.

### 2.5 Korpusvertaalstudies

Tot dusver is gekyk na die invloede wat gelei het tot die ontwikkeling van korpusvertaalstudies. Polisisteamteorie het gesorg vir die erkenning van vertaalde werke (as 'n korpus) as geldige navorsingsonderwerp, het die klemverskuiwing na die doelkultuur asook die soeke na universele wette vir vertaling begin, en het gelei tot Toury se ontwikkeling van Beskrywende Vertaalstudies (BVS). BVS beklemtoon die bestudering van werklike vertaalverskynsels eerder as 'n geïdealiseerde vorm, kyk na die invloed van



norme op die vertaalproduk, en bevestig die soeke na wette as 'n empiriese doelwit vir vertaalstudies.

Paloposki (2001:265-6) is van mening dat die belangstelling in universele vertaalkenmerke te danke is aan Toury se *Beskrywende Vertaalstudies* en die beskikbaarheid van die moderne metodes van korpuslinguistiek. In die volgende afdeling word die ondersoek na die onderskeie vertaalkenmerke bespreek, word daar kortliks gekyk na die universaliteit van die kenmerke, en word daar afgesluit met 'n kritiese beskouing van korpusvertaalstudies.

### 2.5.1 *Vertaalkenmerke*

Laviosa (2002:59) beskou Mona Baker as die 'moeder' van korpusvertaalstudies omdat sy voorgestel het dat korpusmetodes toegepas word om vertaalde tekste te bestudeer. Baker was ook die voorstander daarvan dat vertalings met nievertaalde tekste in dieselfde taal vergelyk word om die kenmerke van die vertaalde teks te ondersoek (Olohan, 2004:91). In 1993 het Baker voorspel dat die beskikbaarheid van groot korpora van oorspronklike en vertaalde tekste asook die ontwikkeling van korpusgedrewe metodologie, mens in staat sou stel om die aard van vertaalde teks as 'n *mediated communicative event* (Baker, 1993:243) deur die sogenaamde vertaaluniversele kenmerke (sien 2.3.3) te ondersoek (Laviosa, 2002:18).

Vertaaluniversele kenmerke is linguistiese kenmerke wat tipies voorkom in vertaalde eerder as nievertaalde teks en is vermoedelik onafhanklik van die spesifieke taalpare betrokke in die proses (Baker, 1993:243). Die term *translationese* is oorspronklik deur Gellerstam gebruik (in Baroni en Bernardini, 2005:3) om te dui op die vingerafdrukke wat een taal tydens vertaling op 'n ander laat. Deesdae het *translationese* egter 'n pejoratiewe konnotasie wat dui op die resultate van onbevoegde vertalers (Tirkkonen-Conditt, 2002:208) (sien 2.3.3).

Na aanleiding van studies deur Blum-Kulka, Vanderauwera, Schlesinger en Toury wat voor die koms van elektroniese korpora gedoen is, onderskei Baker



aanvanklik ses eienskappe as moontlike universele eienskappe (Baker, 1993:234), naamlik:

- 'n hoër vlak van eksplisitering of verduideliking vergeleke met spesifieke brontekste en oorspronklike tekste in die algemeen;
- 'n geneigdheid tot ondubbelsinnigheid en vereenvoudiging;
- 'n sterk voorkeur vir konvensionele grammatikaliteit;
- 'n geneigdheid om herhalings te vermy, deur dit weg te laat of anders te verwoord;
- algemene geneigdheid om kenmerke van die doeltaal te oordryf; en
- 'n spesifieke soort verspreiding van spesifieke kenmerke van vertaalde tekste teenoor brontekste en oorspronklike tekste in die doeltaal (as voorbeeld 'n bevinding dat *say* en *day* meer voorkom in Engelse tekste wat uit Arabies vertaal is vergeleke met nievertaalde Engelse tekste. Maar al kom dit meer in die Engels voor, kom *say* en *day* nog steeds minder in die Engels voor as wat dit in die Arabiese bronteks (van die vertaalde Engelse teks) voorkom).

In 1996 vat Baker bogenoemde kenmerke saam in vier hoofkategorieë, naamlik *eksplisitering*, *vereenvoudiging*, *normalisering* en *gelykmaking* (sien 2.3.3 en 2.5.2) (Ulrych en Bollettieri Bosinelli, 1999:235; Olohan, 2004:91).

Laviosa (2002:75-78) erken dat die vertaalkenmerke mekaar soms weerspreek en sê die volgende daaroor:

*Providing one recognises the probabilistic nature of translational universals, it is possible to feel comfortable with the co-occurrence of trends and of counter-examples to such trends with no ill consequences for the validity of the notion itself.*

(Laviosa, 2002:77).

Sy (Laviosa, 2002:78) is van mening dat ondersoek na universele vertaalkenmerke daarop gemik is om neigings, tendense en reëlmatighede te ontdek in 'n ryk, ingewikkelde, dinamiese wêreld van diversiteit en kontraste.

Toury (in Lind, 2007:3) sê in 2001 by 'n konferensie oor universele vertaalkenmerke ("Translation Universals: Do they exist?") dat hy die term *universeel* met teensin aanvaar, maar *wet* verkies, nie net omdat die moontlikheid van uitsondering daarby ingebou is nie, maar hoofsaaklik omdat dit altyd moontlik moet wees om skynbare uitsonderings met behulp van ander wette te kan verduidelik. Reëlmatighede in vertaling is waarskynlik en nie absoluut nie, en daarom verkies hy die idee van voorwaardelike universele vertaalkenmerke (Toury in Lind, 2007:3).

Naas Toury se wette van standaardisering en interferensie, onderskei Chesterman (in Lind, 2007:4) *S-universals* en *T-universals*. Eersgenoemde verwys na die verskille tussen vertalings en die bronteks, ongeag die taal (byvoorbeeld, interferensie, standaardisering, eksplisitering) en laasgenoemde na die verskille tussen vertalings en vergelykbare tekste in die doeltaal (byvoorbeeld vereenvoudiging ten opsigte van minder leksikale variasie, laer leksikale digtheid en minder doeltaal-spesifieke items). Hierdie onderskeid gee erkenning daaraan dat bewerings oor vertaalde tekste gemaak word ten opsigte van die brontaal en ten opsigte van oorspronklike tekste in die doeltaal (Lind, 2007:4).

Pym (2007:10-11) is van mening dat die vier hoofkategorieë universele vertaalkenmerke verskillende aspekte van een onderliggende universele kenmerk en 'n uitbreiding van Toury se wet van groeiende standaardisering is. Eksplisitering en vereenvoudiging maak tekste makliker om te lees, net so die normalisering van puntuasie ('n voorbeeld van normalisering/konvensionalisering). Indien alle vertalings die kenmerke vereenvoudiging, eksplisitering en normalisering deel, volg dit waarskynlik dat vertalings na die middel van 'n kontinuum neig, want dit verwys na dieselfde linguistiese veranderlikes soos leksikale digtheid, tipe-teken-verhouding, en sinlengte (Pym, 2007:10). Hy redeneer verder dat indien elkeen van die kenmerke werklik universeel was, nie een van die ander kenmerke op universaliteit kon aanspraak maak nie (Pym, 2007:11). Hy verenig Toury se twee wette (interferensie en standaardisering) en die sogenaamde universele

vertaalkenmerke deur voor te stel dat dit strategieë is waardeur vertalers risiko vermy: Vertalers sal risiko probeer vermy deur taal te standaardiseer en/of interferensie positief of negatief te kanaliseer indien en wanneer hulle nie beloon word om die teenoorgestelde te doen nie (Pym, 2007:19-20). Met ander woorde, vertalers verkies om risiko te vermy - hulle sal aanpas by doelnorme (deur eksplisitering of vereenvoudiging byvoorbeeld) wanneer dit is waar die beloning is (duidelike kommunikasie) en hulle sal interferensie van die bronteks toelaat (deur byvoorbeeld letterlike vertaling) as dit is waar die beloning lê (byvoorbeeld by Bybelvertaling) (Lind, 2007:4).

Tymoczko (1998:653) ondersteun nie die soeke na universele vertaalkenmerke nie omdat die waarnemer of navorser se perspektief in elke ondersoek geënkodeer word en dus inbreuk maak op die voorwerp en resultate van die studie. Sy is van mening dat korpusvertaalstudies die soeke na universele vertaalkenmerke moet opgee omdat nie alle navorsingsbevindinge van toepassing is op alle vertaaltipes of vertaalkontekste nie (Tymoczko, 1998:656; 2005:10). Sy voel ook dat korpusvertaalstudies moet wegbeweeg van stelsels van verhoudings wat fokus op gelykheid en dit vermy om in 'n binêre apparaat vasgevang te wees wat net persepsies van absolute gelykheid of ongelykheid erken (Tymoczko, 1998:657).

Of die kenmerke dus wel universeel is of nie, die identifisering van reëlmatighede in vertaalde teks teenoor nievertaalde teks kan insig verskaf in vertaalprosesse (Lind, 2007:5). Laviosa (2002:77) glo dat die bestudering van reëlmatighede mens meer kan leer oor die aard van vertaalde teks en vertaling en so kan mens bewus word van die komplekse, wederkerige verhouding tussen taal en kultuur. Venuti (2005:808) is van mening dat navorsing oor vertaaluniversele eienskappe nie onproduktief is nie, maar waarsku dat dit versigtig gekontekstualiseer moet word om onderskeid te tref tussen hedendaagse en eertydse norme.

### 2.5.2 *Ondersoeke na vertaalkenmerke*

Daar is reeds verskeie ondersoeke na vertaalkenmerke gedoen. Laviosa (1998a; 2002; 2004) en Olohan (2004) gee elk 'n oorsig oor korpusondersoeke wat navorsers ná die koms van korpora oor kenmerke van vertaling gedoen het. Ek gee hier 'n kort samevatting van aspekte wat in tekste ondersoek kan word om te kyk vir eksplisitering, vereenvoudiging, normalisering en afplatting.

Tekstuele aanduidings van eksplisitering kan die vorm aanneem van leksikale, sintaktiese of semantiese byvoegings, uitbreidings of vervangings, byvoorbeeld 'n hoër verhouding funksiewoorde tot inhoudswoorde (laer leksikale digtheid) omdat funksiewoorde grammatiese verhoudings eksplisiet maak, spesifieke terme kan algemener terme vervang, dubbelsinnige voornaamwoorde kan vervang word, verhoogde grammatiese en leksikale kohesie, invoeg van materiaal wat uitgelaat is, byvoeg van bepalers, langer sinlengte of langer teks (met voorsiening vir morfosintaktiese verskille tussen die betrokke tale). Dit sluit nie die inligting in wat voorsien word om te vergoed vir kulturele gapings nie. (Lind, 2007:2).

Vereenvoudiging kan leksikaal, sintakties of semanties in tekste uitgedruk word. Sommige tekstuele eienskappe van vereenvoudiging kan die teenoorgestelde van eksplisitering wees - algemener terme vervang spesifieke terme, korter sinne vervang langer sinne en 'n laer gemiddelde sinlengte, uitlaat of wysiging van frases en woorde - of dit kan dieselfde wees byvoorbeeld laer leksikale digtheid wanneer die informasielading verminder word. Ander tipes vereenvoudiging is vermindering en uitlaat van herhaling, 'n woordeskat van nouer omvang en 'n laer tipe-teken-verhouding. (Lind, 2007:2). Zanettin (2000:113) wys daarop dat die tipe-teken-verhouding nie op sigself voldoende bewys is van leksikale variasie nie. Hy glo dat die tipe-teken-verhouding saam met leksikale digtheid, die verhouding inhouds- tot lopende woorde of die verhouding hapax legomena (woorde wat net een keer in 'n korpus voorkom) tot lopende woorde gebruik moet word om 'n aanduiding van vereenvoudiging te gee.

Pym (2007:11) twyfel oor die geskiktheid van sinlengte as aanduiders van vereenvoudiging (omdat langer sinne op eksplisitering dui). Laviosa (2002:61-2) gebruik sinlengte omdat sinlengte lank reeds in gebruik is as metode om outeurskap te bepaal en verskeie leesbaarheidsformules onder andere kort sinne as kriterium gebruik om leesbaarheid te bepaal.

Normalisering kan ondersoek word deur verskuiwings in puntuasie, woordeskat of sintaksis. Net soos sommige eienskappe as eksplisitering of vereenvoudiging verklaar kan word, kan baie normaliseringsverskuiwings as vereenvoudiging verklaar word en sommige navorsers beskou vereenvoudiging as 'n subkategorie van normalisering (Lind, 2007:2-3). Normalisering kan ook verband hou met lae leksikale variasie en 'n relatiewe hoër gebruik van superordinate teenoor hiponieme (Laviosa-Braithwaite, 1995:162).

Volgens McQuial (in Van Rooyen, 2005:58) is een van die kenmerke van die massakommunikasieproses dat die boodskap van massakommunikasie dikwels op standaardwyse geskep word, eerder as op 'n unieke, kreatiewe of onvoorspelbare manier. Na verwagting sal normalisering en vereenvoudiging dus vertaalkenmerke wees wat in 'n korpus vertaalde koerantberigte figureer.

Gelykmaking (*levelling out*) verwys na die geneigdheid van vertalers om 'n konvensionele teks te lewer wat die middelweg baan en wegbly van uiterstes. Die term (*levelling out*) is in 1989 deur Shlesinger (in Laviosa, 2002:71) geskep na aanleiding van haar waarneming van getolkte tekste en die vergelyking van brontekste en doeltekste. Ongrammatikale uitsprake word grammatikaal reg verwoord en korreksies wat deur sprekers self gemaak word, word byvoorbeeld uitgelaat.

In haar soeke na die verskynsel gelykmaking, bevind Laviosa (2002:71) dat vertaalde nuusberigte teenoor nievertaalde berigte meer uniform is ten opsigte van elke artikel se tipe-teken-verhouding, leksikale digtheid en gemiddelde sinlengte. Sy verkies egter om die verskynsel *konvergensie* te



noem eerder as *gelykmaking* en voer aan dat die verskynsels konseptueel en metodologies verskil (Laviosa, 2002:72-3).

Vervolgens word daar gekyk na Laviosa-Braithwaite se metodologie vir die ondersoek na eksplisitering, vereenvoudiging en normalisering in eentalige vergelykende korpora.

### 2.5.3 Metodologie vir ondersoeke na vertaalkenmerke

Laviosa-Braithwaite (1995:160-1) stel 'n metodologie in die vorm van hipoteses voor waardeur eksplisitering, vereenvoudiging en normalisering deur middel van eentalige vergelykende korpora ondersoek kan word. Die metodologie word in tabel 2.1 weergegee.

**Tabel 2.1: Laviosa-Braithwaite (1995) se metodologie vir ondersoeke na Engelse vertaalkenmerke**

<p><i>Eksplisitering</i></p>	<ul style="list-style-type: none"> <li>• Vertaalde tekste het 'n hoër frekwensie van die opsionele <i>that</i> in indirekte rede as nievertaalde tekste.</li> <li>• Vertaalde tekste het 'n hoër frekwensie van herhaalde, dus oortollige, grammatikale items as nievertaalde tekste, byvoorbeeld <i>on arrival but not on departure; the feature of relevance and of readability</i>.</li> <li>• Vertaalde tekste het 'n laer frekwensie voornaamwoorde as nievertaalde tekste.</li> </ul>
<p><i>Vereenvoudiging</i></p>	<ul style="list-style-type: none"> <li>• Vertaalde tekste het 'n laer tipe-teken-verhouding as nievertaalde tekste.</li> <li>• Vertaalde tekste het 'n hoër frekwensie van superordinate as nievertaalde tekste.</li> <li>• Vertaalde tekste het 'n laer leksikale digtheid as nievertaalde tekste.</li> <li>• Vertaalde tekste het 'n laer gemiddelde sinlengte as</li> </ul>



	<p>nievertaalde tekste.</p> <ul style="list-style-type: none"> <li>Die omvang (sien 2.5.4) van woordskat in vertaalde tekste is nouer as die omvang van woordeskat in nievertaalde tekste.</li> </ul>
<i>Normalisering</i>	<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n laer frekwensie kollokasiebotsings (<i>collocational clashes</i>) wat aanleiding gee tot doelbewuste ironie of ongewone metafore as nievertaalde tekste.</li> </ul>

Bogenoemde metodologie sal in hierdie ondersoek gebruik word om vas te stel of dit ook op 'n Afrikaanse korpus van toepassing sal wees.

#### *2.5.4 Bevindinge van korpusondersoeke van koerantberigte*

Laviosa (1998b:10) gebruik 'n subkorpus van 102 vertaalde koerantberigte uit *The Guardian* en 64 vertaalde berigte uit *The European* met 'n totaal van 74 791 lopende woorde. Hoewel dit nie duidelik in my beskikbare literatuur gestel word nie, neem ek aan dat die 102 vertaalde berigte uit *The Guardian* en die 64 uit *The European* met 102 en 64 nievertaalde berigte uit dieselfde koerante vergelyk word, aangesien sy gebruik maak van die eentalige vergelykende korpus-metodologie. Laviosa (2002:60) gebruik drie hipoteses ten opsigte van leksikale variasie, informasielading en sinlengte om die vertaalkenmerke te ondersoek. Die hipoteses is dat

1. die omvang van woordeskat wat in die vertaalde tekste gebruik word, nouer sal wees as die omvang van die woordeskat in die nievertaalde tekste;
2. vertaalde tekste 'n laer leksikale digtheid as die nievertaalde tekste het; en
3. vertaalde tekste 'n laer gemiddelde sinlengte as die nievertaalde tekste het.

Ten opsigte van hipotese 1 bevind sy dat daar eerstens 'n relatief hoër proporsie hoëfrekwensiewoorde is teenoor woorde wat nie so gereeld voorkom nie; tweedens dat die 108 woorde wat die meeste in die vertaalde subkorpus voorkom 'n groter deel van die korpus uitmaak (wat beteken dat die woorde wat meer gebruik word, meer dikwels herhaal word) en derdens dat die 108 mees frekwente woorde van die vertaalde tekste minder lemmas (sien 3.3.3) bevat. Vir hipotese 2 bevind sy dat vertaalde tekste 'n relatief laer persentasie inhoudswoorde teenoor funksiewoorde het (met ander woorde dat die leksikale digtheid laer is), wat moontlik 'n aanduiding van 'n laer informasielading is. Vir hipotese 3 bevind sy dat vertaalde nuusberigte 'n laer gemiddelde sinlengte het as nievertaalde berigte. (Laviosa, 2002:62). Sy herhaal die studie met narratiewe prosa (Laviosa, 1998b) en bevestig dieselfde patrone vir hipotese 1 en 2, maar nie vir hipotese 3 nie. Sy noem die bevindinge van hipotese 1 en 2 die vier 'kernpatrone van leksikale gebruik'.

Daarby bevind sy ook dat die vertaalde koerantberigte vergeleke met nievertaalde koerantberigte meer uniform is ten opsigte van tipe-tekenverhouding, leksikale digtheid en gemiddelde sinlengte. Hierdie resultate word verkry deur die berekening van variansie, die statistiese maatstaf van die variasie van 'n frekwensiedistribusie vanaf die gemiddelde. Dit dui aan hoe homogeen die data is: hoe hoër die waarde, hoe minder homogeen is die groep. (Laviosa, 2002:71).

Bam (2005) gebruik in sy studie oor Afrikaanse koerantberigte 30 vertaalde en 30 nievertaalde berigte oor die Wêreldbekerrugby uit *Die Burger*. Hy gee nie 'n aanduiding van sy totale aantal lopende woorde nie. Hy gebruik twee hipoteses, naamlik dat

1. daar waarneembare verskille tussen die leksikale strukture van vertaalde en nievertaalde tekste is; en
2. die verskille tussen die leksikale strukture van vertaalde en nievertaalde tekste nie beduidend is nie en nie die klaarprodukt sterk beïnvloed nie.

Ten opsigte van hipotese 1 bevind hy dat daar wel waarneembare verskille tussen vertaalde en nievertaalde tekste is na aanleiding van die statistiese bewerkings. Hy bevind dat die gemiddelde woordlengte, gemiddelde sinlengte en leksikale digtheid hoër is vir die vertaalde subkorpus as vir die nievertaalde subkorpus. Hoewel die tipe-teken-verhouding vir die vertaalde subkorpus hoër is in vergelyking met dié van die nievertaalde subkorpus, neig albei na die hoë kant. Hy beskou die ooreenkoms tussen die tipe-teken-verhoudings as 'n aanduiding van normalisering. Bam (2005) gebruik die statistiek van elke artikel se tipe-teken-verhouding en bepaal dat daar 'n groot mate van konvergensie tussen die vertaalde subkorpus se artikels is teenoor 'n gebrek aan soortgelykheid in die nievertaalde subkorpus. Verder kyk hy na metafore en bevind dat die vertaalde tekste gebruik maak van woorde met algemene betekenis en uitdrukkings wat gereeld in Afrikaans voorkom, wat hy as normalisering beskou.

Ten opsigte van hipotese 2 bevind Bam (2005:80) dat die statistiese ontleding (ANOVA) die hipotese bevestig dat die verskille tussen die vertaalde en nievertaalde subkorpora nie beduidend is nie. Hy skryf dit toe aan die groot mate van gedeelde inligting oor die onderwerp in die artikels.

**Tabel 2.2: Vergelyking van bevindinge van korpusondersoek van koerantberigte ten opsigte van vertaalde subkorpus (VS) en nievertaalde subkorpus (NS)**

	<b>Engels</b>	<b>Afrikaans</b>
hoëfrekwensie teenoor niefrekwente wrde	VS hoër	
herhaling van 108 wrde wat meeste voorkom	VS meer	
lemmas vir 108 wrde wat die meeste voorkom	VS minder	
leksikale digtheid	VS laer	VS hoër
gemiddelde sinlengte	VS laer	VS hoër
konvergensie	Konvergensie t.o.v. TTV, leksikale digtheid en gemiddelde sinlengte	Konvergensie t.o.v. TTV
gemiddelde woordlengte		VS hoër
tipe-teken-verhouding		VS hoër; maar albei hoog

In tabel 2.2 word die bevindinge van Laviosa en Bam vir Engelse en Afrikaanse vertaalde berigte vergelyk. Daaruit kan duidelik gesien word watter bevindinge vergelykbaar is en watter bevindinge nie ooreenstem nie. Bam (2005), anders as Laviosa, toon aan dat die leksikale digtheid en gemiddelde sinlengte hoër is vir die vertaalde subkorpus as vir die nievertaalde subkorpus.

Die metodologie wat in die vorige onderafdeling weergegee is, sal ook Laviosa se bevindinge oor die kernpatrone van leksikale gebruik en die tipe-teken-verhouding dek. Bykomend daarby sal ek ook na die statistiek vir gemiddelde woordlengte en konvergensie ten opsigte van die tipe-teken-verhouding kyk.

Bevindinge van ondersoek in koerante vir Engels en Afrikaans is in hierdie onderafdeling weergegee. Vervolgens word korpusvertaalstudies krities beskou.

### 2.5.5 'n Kritiese beskouing oor korpusvertaalstudies

Malmkjær (in Olohan, 2004:23) is van mening dat

*The use in translation studies of methodologies inspired by corpus linguistics has proved to be one of the most important gate-openers to progress in the discipline since Toury's rethinking of the concept of equivalence.*

Ook Tymoczko (1998:652) noem dat korpusstudies die inhoud en metodes van vertaalstudies as dissipline kwalitatief en kwantitatief verander het om aan te pas by die inligtingsera. Die groot voordeel van korpusnavorsing is die vermoë om groot hoeveelhede data te ontleed (Paloposki, 2001:283) en ook uit verskillende hoeke (Kenny, 2001:69). Omdat die data elektronies is, kan dit op verskillende maniere geberg, versprei en gemanipuleer word om die nut daarvan te verbeter (Kenny, 2001:69). Aangesien data maklik onttrek kan word, is dit moontlik om korpusondersoek maklik aan te vul, te herhaal en

selfs te vervang met meer geskikte ondersoekmetodes (Kenny, 2001:69). Tymoczko (1998:653) beskou die onderliggende buigzaamheid en aanpasbaarheid van korpora sowel as die *openendedness of the construction of corpora* as 'n sterkpunt.

Die groot trekkrag van korpusvertaalstudies is dat sowel ooreenkomste as verskille uitgelig kan word om taalspesifieke verskynsels van verskillende tale en kulture te bestudeer (Tymoczko, 1998:657). Tymoczko (1998:657) glo dat korpusvertaalstudies die potensiaal het om 'n middel te wees waardeur hegemoniese, kultuurgebonde sieninge van tekste en vertalings uitgedaag kan word. Hierdie gedagte sluit aan by Toury se idee van die veronderstelde vertaling en die bestudering van werklike voorbeelde van vertalings wat deur korpusvertaalstudies aanvaar word.

Hoewel die empiriese benadering en die soeke na wette, wat voortvloei uit die polisisteamteorie en Beskrywende Vertaalstudies, onderliggend is aan korpusvertaalstudies, beskou Tymoczko (1998:654) dit nie as die belangrikste eienskap van korpusvertaalstudies nie. Sy wys daarop dat objektiwiteit in die moderne eeu weêrlê is omdat die perspektief van die waarnemer in elke ondersoek ingeskryf word en dus inbreuk maak op die voorwerp en die resultate van die studie (Tymoczko, 1998:653).

Korpusvertaalstudies word inderdaad aan subjektiwiteit onderwerp. Van die samestelling van die korpus tot die interpretasie van die bevindinge kan die rol van intuïsie en menslike oordeel nie uitgesluit word nie (Tymoczko, 1998:654). Verskillende navorsers sal byvoorbeeld nie dieselfde nievertaalde berigte noodwendig as vergelykbaar vir die spesifieke vertaalde berigte beskou nie. Net so kan getalle ook verskillend deur verskillende navorsers geïnterpreteer word: wat vir die een 'n besliste "hoër" getal is, is vir 'n ander navorser dalk net 'n "effens hoër" getal. Selfs die woordsoortklassifikasie, wat mens sal dink voor die hand liggend sal wees, is ook subjektief (sien 2.4.3).

Paloposki (2001) weêrlê ook die sogenaamde objektiwiteit van korpusvertaalstudies. Sy noem dat hoewel die korpusmetodologie groot



hoeveelhede linguistiese data doeltreffend kan ontleed, daar sekere aannames oor die aard van die data is wat nie voorsiening maak vir alle tipes vertaalsituasies nie en sê dat daar 'n geneigdheid is om eksperimentele materiaal uit te laat in 'n poging om die veranderlikes te beheer (Paloposki, 2001:267). Sy wys ook daarop dat die begrip vertaling (en ook korpusseleksie) tydgebonde en lokaal is, wat daartoe kan bydra dat een korpus ander funksies van vertaling kan weergee as 'n ander korpus (Paloposki, 2001:268). Dit kan ook gebeur dat daar niks vergelykbaars in die doelkultuur is wanneer 'n nuwe genre deur vertaling ingevoer word nie (Kenny, 1998a:53). Volgens Baker (2004b:185) sal vergelykbaarheid altyd 'n relatiewe kwessie wees en mens moet die verskille in die samestelling van subkorpora in gedagte hou. Sy wys egter daarop dat sulke ongelykhede onvermydelik is en dat dit nie net waar is vir korpusgebaseerde studies nie: *It is in the nature of any type of comparison, any attempt to look for similarities and differences, that what is being compared can never be totally balanced in every respect* (Baker, 2004b:170).

Snell-Hornby (2006:150-159) beskou die gebruik van rekenaarkorpora as 'n terugbeweeg na linguistiek. Sy betreur die blote vergelyking van linguistiese items sonder insig in tekste as geheel. Die invloed van korpuslinguistiek is hierbo bespreek (sien 2.4), maar ek beskou dit nie as regressie nie. Ek glo nie dat voorstanders van korpusvertaalstudie nie bewus is van enige ontwikkeling ná die linguistiese benaderings en Beskrywende Vertaalstudies nie. Baker (2004b:183) sê byvoorbeeld dat navorsers noukeurig na die data moet kyk en 'n gevoel vir die tekste kry, om so verby laevlakbeskrywing te beweeg. Hoewel vertaling duidelik nie beperk is tot linguistiek nie, is woorde *the only means that the translator has to identify the message he has to transfer into another language* (Roberts, 1996:1).

Kenny (2001:70) waarsku dat hoewel korpusvertaalstudies dit leen tot die bestudering van globale patrone, navorsers versigtig moet wees om nie veralgemenings tot die nadeel van individuele tekste te maak nie. Reëlmatighede moet geïnterpreteer word en om dit as bewys vir die werking van norme te interpreteer, is nie so maklik nie (Kenny, 2001:71). Opvolgende



studies kan 'n aanduiding gee of die bevindinge van 'n studie bloot geldig is vir daardie spesifieke korpus en of dit wyer toegepas kan word, veral waar navorsers hulle prosedures en aannames noukeurig aanteken (Kenny, 2001:70). Indien korpusgebaseerde ondersoeke verantwoordelik gedoen word, beweer Baker (2004b:184), is die voordeel van die metodologie dat dit deursigtig is en dat ander navorsers nie net die geldigheid van die basiese bewerings wat gemaak word, kan toets nie, maar ook ander interpretasies van die data kan gee (dis nou wanneer die data vir ander beskikbaar is of gestel word).

Dit verg tyd en inspanning om 'n korpus saam te stel (Zanettin, 2000:116). Boonop is dit 'n uitdaging om vergelykbare nievertaalde tekste vir vertaalde tekste te identifiseer, soos hierbo aangedui. *While the methods of analysis are empowering to the point of pure euphoria, the collection of data can be challenging to the point of deep despair* (Laviosa, 2002:119).

Kenny (2000:100) wys daarop dat die outomatiese verwerking noodwendig 'n groot hoeveelheid geraas produseer wat tydens die ondersoek skoongemaak moet word. In die hieropvolgende bladsye word 'n aanduiding gegee van hoeveel inligting deur byvoorbeeld die konkordansiefunksie (sien 2.4.3) gegenereer word vir een soekterm, naamlik *span* in die vertaalde subkorpus.

Voorbeeld 2.1 wys die sleutelwoord *span* in die middel met die kontekste waarin dit voorkom; en mens het ook toegang tot die volledige teks waaruit die sleutelwoord-in-konteks getrek is. Voorbeeld 2.2 wys die woorde wat dikwels saam met die woord *span* voorkom. Voorbeeld 2.3 wys die patrone waarin die woord *span* voorkom. Mens het ook in die konkordansiefunksie toegang tot die hele brondokument waarin die sleutelwoord *span* vir 'n spesifieke konteks aangedui word.

Pectora robotant cultus recti

## Voorbeeld 2.1: Sleutelwoord-in-konteks

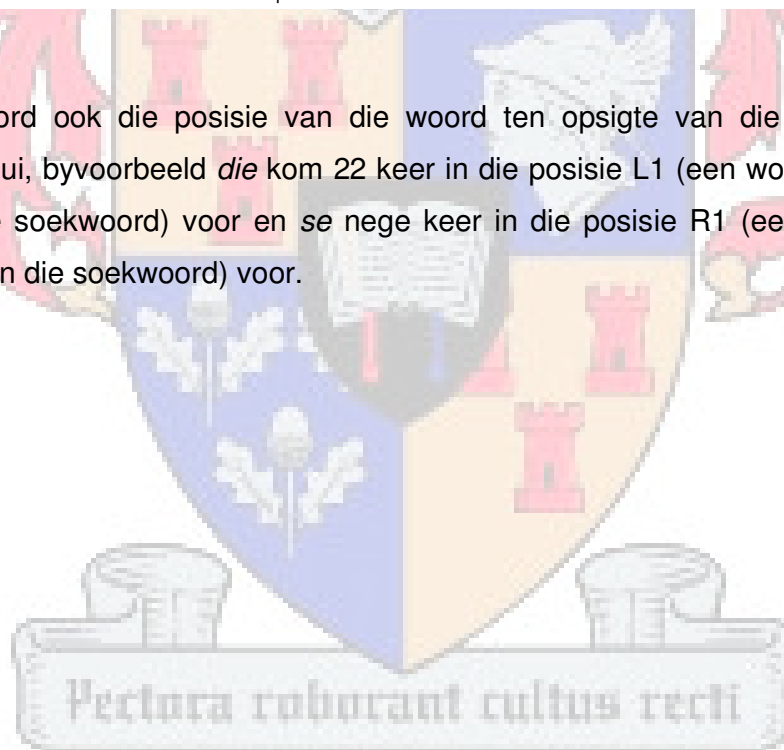
N-Concordance	Set Tag	Word #	t. #	os. a. #	os. d. #	os. t. #	os.	File
1 en dit gee my groot bevrediging. Herde <b>span</b> go in homself," meen hy. Die		446	27	8%	0	4%	0	0 4% ort-vert-rugby.txt
2 ons hom ook wil neutraliseer." Hy het sy <b>span</b> se vertoning in die tweede helfte, toe		404	25	6%	0	5%	0	0 5% ort-vert-rugby.txt
3 ondergaan voordat hy saam met die <b>span</b> na Suid-Afrika kan reis. Dit het ook		112	6	8%	0	3%	0	0 3% ort-vert-rugby.txt
4 Arrows 'n doel behaal wat sy <b>span</b> in staat gestel het om die klipharde		380	17	8%	0	6%	0	0 6% rt-vert-sokker.txt
5 van Benoni Premier United. Die Benoni- <b>span</b> het Santos met die vernederende		346	16	5%	0	8%	0	0 8% rt-vert-sokker.txt
6 Mushin Ertugral, se opsomming van sy <b>span</b> se Premierliga-sokkerwedstryd teen		31	2	4%	0	8%	0	0 8% rt-vert-sokker.txt
7 'n skoot van Ian Symons gekeer om sy <b>span</b> se neus voor te hou. Later in die		199	11	9%	0	6%	0	0 6% d-manshokkie.txt
8 gehad het, het Taake Taakema die <b>span</b> se enigste doel in die eerste helfte		156	9	2%	0	7%	0	0 7% d-manshokkie.txt
9 nie bang nie. Neu-Seeland is 'n puik <b>span</b> , maar as ons goed speel, kan ons		232	16	3%	0	6%	0	0 6% ort-vert-rugby.txt
10 nie, maar Connolly is vol vertroue dat sy <b>span</b> in die Dierasies-wedloop kan bly.		209	10	4%	0	0%	0	0 0% ort-vert-rugby.txt
11 nie, wen. Marco van Basten (afrigter) se <b>span</b> het in sy Groep G-wedstryd in die		150	7	7%	0	7%	0	0 7% rt-vert-sokker.txt
12 gevaar nie en kon slegs met 1-0 teen die <b>span</b> van Luxemburg, wat nie onder die		133	6	9%	0	1%	0	0 1% rt-vert-sokker.txt
13 wedstryd moet stopgesit word deur 'n <b>span</b> wat voel hy is ingedoen nie. Die		236	11	0%	0	2%	0	0 2% rt-vert-krieket.txt
14 se mag weg te neem om te beslis 'n <b>span</b> het 'n wedstryd prysgegee. Dit is in		30	1	6%	0	8%	0	0 8% rt-vert-krieket.txt
15 nie. "Chelsea is 'n baie goeie <b>span</b> , maar kan geklop word.		44	2	4%	0	7%	0	0 7% rt-vert-sokker.txt
16 en ek is daarvan cortuig dat die hele <b>span</b> sy bes sal doen om weer aan die		218	10	1%	0	0%	0	0 0% rt-vert-krieket.txt
17 David Graveney, gesê toe hy die <b>span</b> bekend gemaak het. Die voormalige		120	4	3%	0	4%	0	0 4% rt-vert-krieket.txt
18 het, het sy plek in die groep behou. "Die <b>span</b> het werklik goed gespeel in die twee		83	4	7%	0	0%	0	0 0% rt-vert-krieket.txt
19 Die keurders het vertroue getoon in die <b>span</b> wat die derde toets op Headingley		38	2	7%	0	4%	0	0 4% rt-vert-krieket.txt
20 toer in Engeland was 'n keerpunt vir die <b>span</b> ," het Draid gesê. "Sri Lanka het 'n		384	19	4%	0	0%	0	0 0% rt-vert-krieket.txt
21 in eendaagse krieket nie. "Enige <b>span</b> wat meer as 400 lopies suksesvol		215	12	7%	0	0%	0	0 0% rt-vert-krieket.txt
22 dowe neutre die wêreld se nommertwee <b>span</b> in eendaagse krieket nie. "Enige		209	11	3%	0	9%	0	0 9% rt-vert-krieket.txt
23 nie dat McLaren hom uit sy eerste <b>span</b> as die nuwe Engelse breier		97	4	5%	0	4%	0	0 4% rt-vert-krieket.txt
24 afrigter, het intussen gesê hy hoop sy <b>span</b> vind gou sy voete in die nuwe		353	15	9%	0	4%	0	0 4% rt-vert-sokker.txt
25 van Chelsea se sokkerspan, het gesê sy <b>span</b> sal vanaand alles in sy vermoë doen		20	1	3%	0	5%	0	0 5% rt-vert-sokker.txt
26 het. "Dit is in die beste belang van die <b>span</b> dat Inzamam uittree en (die		346	14	9%	0	9%	0	0 9% rt-vert-krieket.txt
27 in 24 jaar oor die Suid-Asiese <b>span</b> behaal het. "Dit is in die beste		335	13	4%	0	7%	0	0 7% rt-vert-krieket.txt
28 Inzamam-ul-Haq moet uittree as die <b>span</b> se kaptein ná Engeland sy eerste		321	13	2%	0	3%	0	0 3% rt-vert-krieket.txt
29 (knie), het gesê dit is belangrik vir die <b>span</b> om sy wemomentum te behou vóór		204	8	2%	0	7%	0	0 7% rt-vert-krieket.txt
30 se kaptein, Andrew Strauss, wat die <b>span</b> aanvoer in die afwesigheid van		185	8	0%	0	2%	0	0 2% rt-vert-krieket.txt
31 Ricky Ponting en sy meklers, sal die <b>span</b> teen Australië aanvoer. "As hulle		128	5	9%	0	9%	0	0 9% rt-vert-krieket.txt
32 tweede keer kan wen ná die Engelse sy <b>span</b> met 167 lopies geklop het in die		36	1	2%	0	8%	0	0 8% rt-vert-krieket.txt
33 Die All Blacks is die wêreld se voorste <b>span</b> , het 14 agtereenvolgende toetse		311	15	7%	0	0%	0	0 0% ort-vert-rugby.txt
34 ondersoek om Matt Giteau in te <b>span</b> . Hy is 'n uitstekende skrumskakel."		270	12	0%	0	1%	0	0 1% ort-vert-rugby.txt
35 dit wil voorkom of die ouer spelers in die <b>span</b> sukkel om by die nuwe		396	21	0%	0	7%	0	0 7% ort-vert-rugby.txt
36 lê om dramatiese veranderings aan die <b>span</b> aan te bring. "As Sam Cordingley		311	14	8%	0	6%	0	0 6% ort-vert-rugby.txt
37 ware omkeer te bewerkstellig nie. Die <b>span</b> is nog op dieselfde pad as die een		242	12	3%	0	9%	0	0 9% ort-vert-rugby.txt
38 diabolies en laat 'n mens wonder of die <b>span</b> sedert verlede jaar enigsins verbeter		214	10	8%	0	2%	0	0 2% ort-vert-rugby.txt
39 nie. "Die Wallaby-coonwining kon nie die <b>span</b> se gebreke verbloem nie. Dit was		199	9	4%	0	9%	0	0 9% ort-vert-rugby.txt
40 Jones oorgeneem het, het erken dat sy <b>span</b> nie die Springbokke genoeg		107	5	9%	0	6%	0	0 6% ort-vert-rugby.txt
41 (lies) se beserings het die <b>span</b> ook in die eerste toets op Lord's en		471	26	2%	0	1%	0	0 1% rt-vert-krieket.txt
42 'n mens moet baie versigtig wees as 'n <b>span</b> onder druk is, want hy kan baie		420	23	5%	0	3%	0	0 3% rt-vert-krieket.txt
43 klinies en meedobneloos wees. "Die <b>span</b> het hom nou al twee keer baie goed		391	22	4%	0	8%	0	0 8% rt-vert-krieket.txt
44 egter nog altyd sukses vir die nasionale <b>span</b> . Hy het onder meer twee		89	5	0%	0	5%	0	0 5% rt-vert-krieket.txt
45 beslis kans om die eerste Australiese <b>span</b> sedert 1986 te word wat die All		248	13	2%	0	4%	0	0 4% ort-vert-rugby.txt

Pectora roborant cultus recti

## Voorbeeld 2.2: Kollokasies - span

N	Word	With elation	Total otal	Left	al Right	L2	L1	Centre	R1	R2
1	SPAN	span 0.000	45	0	0	0	0	45	0	0
2	DIE	span 0.000	22	18	4	3	15	0	0	4
3	SY	span 0.000	12	10	2	1	9	0	1	1
4	HET	span 0.000	10	2	8	2	0	0	7	1
5	IN	span 0.000	10	3	7	3	0	0	3	4
6	SE	span 0.000	9	3	6	2	1	0	6	0
7	WAT	span 0.000	5	2	3	2	0	0	3	0
8	AS	span 0.000	4	2	2	2	0	0	1	1
9	DIÉ	span 0.000	4	4	0	0	4	0	0	0
10	OM	span 0.000	3	1	2	1	0	0	1	1
11	VAN	span 0.000	3	2	1	2	0	0	1	0
12	DAT	span 0.000	3	2	1	2	0	0	1	0
13	TE	span 0.000	2	1	1	0	1	0	0	1
14	EERSTE	span 0.000	2	2	0	1	1	0	0	0
15	TEEN	span 0.000	2	1	1	1	0	0	1	0
16	VIR	span 0.000	2	2	0	2	0	0	0	0
17	AAN	span 0.000	2	1	1	1	0	0	1	0
18	MET	span 0.000	2	1	1	1	0	0	1	0
19	SAL	span 0.000	2	1	1	1	0	0	1	0
20	NIE	span 0.000	2	1	1	1	0	0	1	0
21	SUID	span 0.000	2	1	1	1	0	0	0	1
22	MAAR	span 0.000	2	0	2	0	0	0	2	0
23	SEDERT	span 0.000	2	0	2	0	0	0	2	0

Hier word ook die posisie van die woord ten opsigte van die soekwoord aangedui, byvoorbeeld *die* kom 22 keer in die posisie L1 (een woord na links van die soekwoord) voor en se nege keer in die posisie R1 (een woord na regs van die soekwoord) voor.



### Voorbeeld 2.3: Patrone - span

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HET	DIE	WÊRELD	IN	DIE	SPAN	HET	IN	SY	DIE	IN
2	N	OM	HY	NIE	SY		SE	DIE	DIE	NIE	OM
3	OM	SE	DIE	DIE	DIÉ		IN	HET		TE	SY
4	IS	IS		VAN	N		WAT			WORD	TWEE
5	DIE	GESÊ		VIR			SEDERT			AS	DIT
6	IN	HET		WAT			HY			IN	HY
7				SE			MAAR			IS	DIE
8				HET							
9				DAT							
10				AS							

Die eerste ry in Voorbeeld 2.3 dui op tien patrone wat in die vertaalde subkorpus voorkom ten opsigte van die soekwoord *span*:

het \_\_\_ span

die \_\_\_ span

wêreld \_\_\_ span

in \_\_\_ span

die span

span het

span \_\_\_ in

span \_\_\_ sy

span \_\_\_ die

span \_\_\_ in

In hierdie afdeling oor korpusvertaalstudies is daar gekyk na vertaalkenmerke, ondersoek wat na vertaalkenmerke gedoen is asook bevindinge van korpusondersoek na koerantberigte vir Engels en Afrikaans. Laviosa-Braithwaite se metodologie van hipoteses vir die ondersoek na vertaalkenmerke in 'n eentalige verklarende korpus is ook uiteengesit. Die afdeling sluit af met 'n kritiese beskouing van korpusvertaalstudies.

Pectora robotant cultus recti

## 2.6 Samevatting

In hierdie literatuurstudie is daar gekyk na die invloed van die polisisteepteorie op Beskrywende Vertaalstudies en laasgenoemde se onderliggende beginsels in korpusvertaalstudies. Daar is ook gekyk na die invloed van korpuslinguistiek op korpusvertaalstudies en die elektroniese teksverwerkingsmiddele wat daaraan ontleen is. 'n Kort bespreking oor vertaalkenmerke, hoe dit in tekste geïdentifiseer kan word en watter bevindinge reeds vir koerantberigte verkry is, volg. Die hoofstuk sluit af met 'n kritiese beskouing oor korpusvertaalstudies.

In die volgende hoofstuk word daar gekyk na die navorsingsresultate van hierdie ondersoek en spesifiek na die vertaalkenmerke in 'n eentalige vergelykende korpus koerantberigte.





## HOOFSTUK 3 NAVORSINGSRESULTATE

### 3.1 Inleiding

In die vorige hoofstuk is daar gekyk na die invloed van die polisisteamteorie, Beskrywende Vertaalstudies en korpuslinguistiek wat op korpusvertaalstudie ingewerk het. Daar is gekyk na korpusondersoeke van koerantberigte vir Engels en Afrikaans. *WordSmith Tools* is wel vir albei ondersoeke gebruik, maar slegs leksikale digtheid, gemiddelde sinlengte en die aspek konvergensie word in albei gerapporteer (sien tabel 2.2) en slegs ten opsigte van laasgenoemde is dieselfde resultate vir Engels en Afrikaans gevind.

'n Eentalige vergelykende korpus is saamgestel uit koerantberigte uit *Die Burger* uit 'n verskeidenheid domeine (sien 1.4). Die korpus bestaan uit 'n totaal van 36 733 lopende woorde (17 450 lopende woorde vir die vertaalde subkorpus en 19 283 lopende woorde vir die nievertaalde subkorpus).

### 3.2 Modus operandi

*WordSmith Tools* is gebruik om die twee subkorpora met mekaar te vergelyk volgens Laviosa-Braithwaite se metodologie (sien 2.5.3). Ek kyk ook na die statistiek vir gemiddelde woordlengte en tipe-teken-verhoudings (TTV) van die individuele tekste sodat ek my bevindinge met dié van Bam kan vergelyk (sien 2.5.4).

In die onderstaande tabelle word die praktiese bewerkings weergegee waardeur inligting deur middel van *WordSmith Tools* uit die korpus onttrek is ten opsigte van Laviosa-Braithwaite (1995) se hipotese-metodologie (tabel 3.1) en die aspekte van Laviosa en Bam se bevindinge wat nie in die hipotese-metodologie gedek word nie (tabel 3.2).

**Tabel 3.1: Praktiese bewerking van Laviosa-Braithwaite (1995) se hipotese-metodologie**

Hipotese	Praktiese bewerking
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n hoër frekwensie van die opsionele <i>that</i> in indirekte rede as nievertaalde tekste.</li> </ul>	Konkordansielys; soekwoord <i>dat</i> ; identifiseer opsionele <i>dat</i> in sleutelwoord-in-konteks
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n hoër frekwensie van herhaalde, dus oortollige, grammatikale items as nievertaalde tekste, bv. toutologiese uitdrukkings soos <i>ronde sirkel</i>; pleonasme soos <i>herhaal dit weer</i></li> </ul>	Konkordansielys; identifiseer moontlike soekwoorde; bestudeer sleutelwoord-in-konteks; moontlik ook patrone, woordgroepe en kollokasies
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n laer frekwensie voornaamwoorde as nievertaalde tekste.</li> </ul>	Gebruik geëtiketteerde data om frekwensie van voornaamwoorde te bepaal
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n laer tipe-teken-verhouding as nievertaalde tekste.</li> </ul>	Woordelys; data gegenereer deur <i>WordSmith Tools</i>
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n hoër frekwensie superordinate as nievertaalde tekste.</li> </ul>	Gebruik geëtiketteerde data om naamwoorde te identifiseer; vergelyk naamwoordlyste van die twee subkorpora
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n laer leksikale digtheid as nievertaalde tekste.</li> </ul>	Gebruik geëtiketteerde data om inhoudswoorde te identifiseer: (Leksikale digtheid = Totale aantal woorde - inhoudswoorde x 100)
<ul style="list-style-type: none"> <li>Vertaalde tekste het 'n laer gemiddelde sinlengte as nievertaalde tekste.</li> </ul>	Woordelys; data gegenereer deur <i>WordSmith Tools</i>
<ul style="list-style-type: none"> <li>Die omvang van woorde in vertaalde tekste is nouer as die omvang van woordeskat in nievertaalde tekste.</li> </ul>	Woordelys volgens frekwensie. <ol style="list-style-type: none"> <li>Wat is die verhouding tussen hoëfrekwensiewoorde teenoor niefrekwente woorde? Gebruik <i>WordSmith Tools</i>-data om in <i>MsExcel</i> te bereken:</li> </ol>

	b) In hoeveel tekste kom frekwentste woorde voor? Gebruik <i>WordSmith Tools</i> -data om in <i>MsExcel</i> te bereken: c) Hoeveel lemmas bevat frekwentste woorde? Lemmatiseer frekwentste woorde
<ul style="list-style-type: none"> <li>• Vertaalde tekste het 'n laer frekwensie kollokasiebotsings (<i>collocational clashes</i>) wat aanleiding gee tot doelbewuste ironie of ongewone metafore as nievertaalde tekste.</li> </ul>	Konkordansie; identifiseer soekterme; ondersoek kollokasies, patrone en woordgroepe

**Tabel 3.2: Praktiese bewerking van aspekte wat nie in Laviosa-Braithwaite (1995) se metodologie gedek word nie**

Aspek van bevinding	Praktiese bewerking
<ul style="list-style-type: none"> <li>• gemiddelde woordlengte</li> </ul>	Woordelys; data gegeneer deur <i>WordSmith Tools</i>
<ul style="list-style-type: none"> <li>• konvergensie</li> </ul>	Woordelys; data gegeneer deur <i>WordSmith Tools</i> vir elke artikel

Praktiese bewerkings wat deur dieselfde hulpmiddel onttrek word, groepeer ek saam (tabel 3.3) en ek rapporteer ook die bevindinge volgens hierdie stappe.



**Tabel 3.3: Praktiese bewerking stapsgewys**

Stap 1: Woordelys - Statistiek gegenereer deur <i>WordSmith Tools</i>	<ul style="list-style-type: none"><li>• tipe-teken-verhouding (TTV)</li><li>• gemiddelde sinlengte</li><li>• gemiddelde woordlengte</li><li>• standaardafwyking-data vir TTV en gemiddelde sinlengte</li></ul>
Stap 2: Woordelys - volgens frekwensie	<ul style="list-style-type: none"><li>• frekwentse teenoor niefrekwente woorde</li><li>• frekwensie van frekwentste woorde voorkoms in tekste</li></ul>
Stap 3: Woordelys - lemmatisering	<ul style="list-style-type: none"><li>• aantal lemmas in frekwentste woorde</li></ul>
Stap 4: Geëtiketteerde data	<ul style="list-style-type: none"><li>• voornaamwoorde</li><li>• leksikale digtheid</li><li>• superordinate</li></ul>
Stap 5: Konkordansie	<ul style="list-style-type: none"><li>• opsionele <i>dat</i></li><li>• kollokasiebotsings</li><li>• oortollige grammatikale items</li></ul>

Vervolgens word my bevindinge staggewys weergegee, omdat ek sodoende in my beskrywing wys hoe ek te werk gegaan het om die resultate te kry. Hoewel *WordSmith Tools* gebruik word om die data en statistiek te genereer, vind ek dit nie 'n nuttige hulpmiddel wanneer ek die resultate vir die twee subkorpora op die skerm wil vergelyk nie. In baie gevalle dra ek my data oor na *MSExcel* waar ek die data vir die twee subkorpora langs mekaar kan vergelyk. Nie alle data wat hieronder weergegee word, is outomaties deur *WordSmith Tools* gegenereer nie en persentasies is soms met die hand bereken.



### 3.3 Stap 1: *WordSmith Tools*-gegenereerde statistiek

#### 3.3.1 *Tipe-teken-verhouding, gemiddelde sinlengte en gemiddelde woordlengte*

Die woordelysfunksie is gebruik om woordelyste vir die vertaalde en die nievertaalde komponent te skep wat dan alfabeties en ook volgens frekwensie gesorteer word. Verstellings is op *WordSmith Tools* gemaak sodat daar voorsiening gemaak is dat woorde met koppeltekens nie as aparte woorde getel word nie en dat syfers nie apart gelys word nie. Omdat die vertaalde artikels korter is as die nievertaalde artikels, is 179 as die standaardiseringsbasis (sien 2.4.3) gebruik sodat die data vergelyk kon word. *WordSmith Tools* gee statistiek vir elke subkorpus as geheel (sien tabel 3.4) en vir elke artikel in die subkorpus. Ek het die data vir statistiek en die woordelyste (volgens frekwensie en alfabeties) vir elke subkorpus na *MSEXcel* oorgedra sodat ek dit makliker kon vergelyk.

**Tabel 3.4: Statistiek gegenereer deur *WordSmith Tools***

	<b>VERTAAL</b>	<b>NIEVERTAAL</b>
tokens (running words) in text	17450	19283
tokens used for word list	17038	18749
types (distinct words)	3711	4004
type/token ratio (TTR)	21.78072548	21.35580635
standardised TTR	60.76723099	60.60123062
standardised TTR std.dev.	37.37368393	38.40524292
standardised TTR basis	179	179
mean word length (in characters)	4.883847713	4.82980442
word length std.dev.	3.208684921	3.227355957
sentences	941	1007
mean (in words)	18.10627556	18.61867905
std.dev.	10.17565346	10.09426689
paragraphs	45	45
mean (in words)	378.6222229	416.6444397
std.dev.	116.924469	102.0062637



In tabel 3.4 kan gesien word dat die vertaalde tekste se gestandaardiseerde TTV (60,8 vertaal; 60,6 nievertaal) en gemiddelde woordlengte (4,9 vertaal; 4,8 nievertaal) effens hoër is as die nievertaalde tekste. Die gemiddelde sinlengte van die vertaalde subkorpus is korter as in die nievertaalde subkorpus (18,1 teenoor 18,6).

My bevindinge stem dus ooreen met Bam se bevindinge ten opsigte van die TTV en gemiddelde woordlengte. Ten opsigte van gemiddelde sinlengte stem my bevinding, anders as Bam se bevinding dat die vertaalde tekste se gemiddelde sinlengte hoër is, ooreen met Laviosa se bevinding vir Engelse koerantberigte.

### 3.3.2 Standaardafwyking-data vir TTV en gemiddelde sinlengte

Laviosa (2002:71) neem tydens die berekening van variansie van die artikels in haar korpus se TTV, leksikale digtheid en gemiddelde sinlengte waar dat die vertaalde koerante meer uniform is. Sy noem die verskynsel konvergensie. Bam (2005) neem konvergensie ten opsigte van die verspreiding van die tipe-teken-verhoudings van die artikels in sy korpus waar vir die vertaalde subkorpus. Hy maak gebruik van ANOVA as statistiese hulpmiddel om die variansie te bereken.

Volgens McEnery *et al.* (2006:54) is daar drie maniere om die verspreiding van 'n datastel te meet, te wete omvang, variansie en standaardafwyking. Omvang is die verskil tussen die hoogste en die laagste frekwensie. Dit is egter nie 'n goeie maatstaf van die verspreiding nie, omdat 'n ongewone hoë of lae telling in 'n datastel die omvang onredelik hoog kan maak en dus 'n verdraaide voorstelling van die datastel gee. Variansie meet hoe ver elke telling in die datastel van die gemiddelde telling verwyder is. Vir 1, 2, 3 en 4 is die gemiddeld 2,5. Die variansie van 1 word bereken deur 1 van die gemiddeld af te trek, naamlik  $2,5-1$  wat 1,5 is. Vir 3 word die variansie bereken deur die gemiddeld van 3 af te trek, naamlik  $3-2,5$  wat 0,5 is. Vir 'n hele datastel balanseer die verskille mekaar uit omdat party tellings bo die gemiddeld is en party onder (McEnery *et al.*, 2006:54). Dit word dus nie

gebruik om die verspreiding van 'n hele datastel te bereken nie, maar die verspreiding van die tekste in die datastel. Standaardafwyking, die vierkantswortel van variansie, is nuttig wanneer mens wil bepaal of die meeste items in die middel eerder as die laer of hoër end van die skaal gegroepeer is. 'n Lae standaardafwyking dui op 'n stel waardes wat naby die gemiddeld gegroepeer is en 'n stel waardes wat ver uitmekaar is, sal 'n hoër standaardafwyking hê. (McEnery *et al.*, 2006:54).

Tabel 3.5 gee die standaardafwykingswaardes wat deur *WordSmith Tools* gegenereer is vir die gestandaardiseerde TTV, woordlengte en sinlengte.

**Tabel 3.5: Standaardafwyking vir TTV, gemiddelde woord- en sinlengte**

	VERTAAL	NIEVERTAAL
standardised TTR	60.76723099	60.60123062
<b>standardised TTR std.dev.</b>	<b>37.37368393</b>	<b>38.40524292</b>
mean word length (in characters)	4.883847713	4.82980442
<b>word length std.dev.</b>	<b>3.208684921</b>	<b>3.227355957</b>
sentences	941	1007
mean (in words)	18.10627556	18.61867905
<b>std.dev.</b>	<b>10.17565346</b>	<b>10.09426689</b>

Hier kan gesien word dat die standaardafwykingwaardes vir die vertaalde sowel as nievertaalde subkorpora hoog is vir al drie veranderlikes. Grafieke wat die standaardafwykingdata visueel voorstel, kan in bylaag A gesien word. Hoewel dit kan voorkom asof die gemiddelde waardes naby die gemiddeld gekonsentreer is, moet 'n mens in ag neem dat die waardes vir TTV en gemiddelde woord- en sinlengte vir elke artikel reeds gemiddeldes is (Scott, 2008b). Die hoë standaardafwykingwaardes dui daarop dat die waardes vir hierdie drie veranderlikes vir die artikels in die subkorpora ver uitmekaar is en dus nie uniform is nie.

Om die standaardafwyking tussen die twee subkorpora te vergemaklik, gee tabel 3.6 die waardes persentasiegewys:

**Tabel 3.6: Standaardafwyking in verhouding met TTV, gemiddelde woord- en sinlengte**

	<b>VERTAAL</b>	<b>NIEVERTAAL</b>
standaardafwyking-gestandaardiseerde TTV-verhouding	61,5%	63,4%
standaardafwyking- gem. woordlengte-verhouding	65,79%	66,8%
standaardafwyking-gem. sinlengte-verhouding	56,2%	54,2%

In tabel 3.6 kan dus gesien word dat die standaardafwyking in verhouding met die aangeduide veranderlikes vir die vertaalde en nievertaalde subkorpora ooreenstemmend hoog is.

Ek vind dus, anders as Laviosa en Bam, dat die vertaalde subkorpus se waardes spesifiek vir TTV, gemiddelde woordlengte en gemiddelde sinlengte nie uniform is nie maar dat die nievertaalde subkorpus se waardes vir daardie drie veranderlikes ook nie uniform is nie.

### **3.4 Stap 2: Frekwensielys**

Om die omvang van woorde in die subkorpora te bepaal, laat ek my lei deur Laviosa (2002:60-62) se bevindinge vir die hipotese (sien 2.5.4) dat die omvang van woorde in vertaalde tekste nouer is as die omvang van die woorde in nievertaalde tekste. In die bespreking van haar bevindinge vir hierdie hipotese, maak sy aansprake oor die verhouding hoëfrekwensie-teenoor niefrekwente woorde, hoeveel keer die 108 frekwentste woorde in die tekste voorkom en uit hoeveel lemmas die 108 frekwentste woorde bestaan.

Laviosa verwys hier na die verhouding hoëfrekwensie- teenoor niefrekwente woorde sonder om die frekwensies duidelik te onderskei. Dit is dus onduidelik of sy verwys na woorde wat meer as 50 keer voorkom teenoor woorde wat net een keer voorkom, of na al die woorde wat meer as een keer voorkom teenoor dié wat net een keer voorkom. Tabel 3.7 gee 'n aanduiding van woorde wat meer as 30 keer in die korpus voorkom in watter persentasie tekste asook vir hapax legomena.

**Tabel 3.7: Die voorkoms van woorde in tekste**

	Vertaal 17 450 woorde			Nievertaal 19 283 woorde		
	worde	% van subkorpus	tekste	worde	% van subkorpus	tekste
woorde >20x	88	0,5%	55,4%	99	0,5%	55,9%
1x	2 225	12,8%	2,2%	2 414	12,5%	2,2%

Uit bostaande tabel kan daar dus bereken word dat die verhouding frekwentste woorde tot hapax legomena vir die vertaalde subkorpus 3,96% (88 gedeel deur 2 225 maal met 100) is en vir die nievertaalde subkorpus 4,1% (99 gedeel deur 2 414 maal met 100). My bevinding is dus - anders as Laviosa se bevinding dat daar 'n relatiewe hoër verhouding hoëfrekwensie-teenoor niefrekwente woorde in die vertaalde subkorpus is - dat die verhouding hoëfrekwensie-teenoor niefrekwente woorde vir die vertaalde subkorpus laer is as vir die nievertaalde subkorpus.

In tabel 3.7 kan mens ook sien watter persentasie die hoëfrekwensie- en niefrekwente woorde van elke subkorpus uitmaak. Die mees frekwente woorde van albei subkorpora maak 0,5% van die onderskeie subkorpora uit, maar daardie 0,5% kom in minder tekste voor in die vertaalde subkorpus (55,4%) as in die nievertaalde subkorpus (55,9%). Woorde wat net een keer voorkom, kom persentasiegewys meer in die vertaalde subkorpus voor (12,8%) wat die indruk wek dat daar 'n groter verskeidenheid woorde in die vertaalde subkorpus gebruik word as in die nievertaalde subkorpus. My bevinding is dus dat die hoëfrekwensiewoorde in die vertaalde subkorpus in minder tekste voorkom as wat die geval is vir die nievertaalde subkorpus.

Alhoewel ek vermoed dat Laviosa se 108 mees frekwente woorde waarskynlik ook gebaseer is op 'n natuurlike skeiding in die data van haar vertaalde subkorpus (soos wat daar vir hierdie vertaalde subkorpus 88 woorde is wat meer as 20 keer voorkom), gee tabel 3.8 die inligting weer ten opsigte van die 108 mees frekwente woorde vir hierdie ondersoek:

**Tabel 3.8: Voorkoms van 108 frekwentste woorde teenoor hapax legomena**

	Vertaal 17 450 woorde		Nievertaal 19 283 woorde	
	% van subkorpus	tekste	% van subkorpus	tekste
108 frekwentste wrde	0,6%	49,5%	0,6%	53%
1x	12,8%	2,2%	12,5%	2,2%

Uit bostaande tabel kan gesien word dat die 108 frekwentste woorde in beide subkorpora 0,6% van die onderskeie subkorpora uitmaak, maar dat die vertaalde subkorpus se frekwentste woorde in minder tekste (49,5%) voorkom as wat die geval is vir die nievertaalde subkorpus (53%). Die bevinding ten opsigte van die 108 frekwentste woorde in tabel 3.8 stem dus ooreen met die bevinding gegrond op die frekwentste woorde in hierdie spesifieke subkorpora. My bevinding is dus - anders as Laviosa se bevinding dat die 108 woorde wat die meeste voorkom in die vertaalde subkorpus 'n groter deel van die korpus uitmaak - dat die hoëfrekwensiewoorde in die vertaalde subkorpus 'n kleiner deel van die korpus uitmaak en dat daar meer hapax legomena in die vertaalde subkorpus voorkom.

Die derde maatstaf wat Laviosa gebruik het vir haar uitspraak oor die nouer omvang van woorde vir vertaalde berigte teenoor nievertaalde berigte is die aantal lemmas in die 108 frekwentste woorde. Lemmatisering word in die volgende onderafdeling bespreek.

### **3.5 Stap 3: Lemmatisering**

Lemmatisering is die proses waardeur die verskillende vorme van 'n woord saam gegroepeer word (Butler, 1985:14). In hierdie stap word woorde wat afsonderlik gelys word, maar dieselfde lemma is, bymekaar gevoeg. 'n Voorbeeld hiervan is *hulle* en *hul* en *Suid-Afrika* en *SA*. Sien bylaag B vir woorde wat gelemmatiseer is vir die eerste 108 woorde van die twee subkorpora.



Na lemmatisering het die vertaalde subkorpus 95 lemmas en die nievertaalde subkorpus 96 lemmas. My resultaat stem dus ooreen met Laviosa se bevinding dat die 108 mees frekwente woorde van die vertaalde tekste minder lemmas bevat as die nievertaalde subkorpus.

### 3.6 Stap 4: Etiketring van data

Ek het die korpus leksikaal laat annoteer deur die outomatiese woordsoortetiketteerder van CTeXT by die Noordwes-Universiteit (sien 2.4.3). Die woordsoortetiketteerder se verkorte etiketstel is gebruik om woordsoorte in 13 klasse te etiketteer, naamlik naamwoord, werkwoord, adjektief, voornaamwoord, bywoord, telwoord, lidwoord, setsel, tussenwerpsel, voegwoord, residu, uniek en punktuasie (sien bylaag C).

Die woordklasse is redelik voor die hand liggend behalwe "residu" en "uniek". Tabel 3.9 gee voorbeelde van wat as "residu" geklassifiseer word en tabel 3.10 gee voorbeelde van wat as "uniek" geklassifiseer word.

**Tabel 3.9: Voorbeelde van die woordklas residu**

Residu	Voorbeeld
vreemdetaalwoorde	<i>de en facto</i>
formules	a+b
simbole	%
akronieme	ABSA
afkortings	mv.
lettername	AA

(Pilon, 2005:60)

Onder residu val ook gevalle met die afstandskoppeltteken soos *media-* (in media- en uitsaaisentrum).

*Pectora robotant cultus recti*

**Tabel 3.10: Voorbeelde van die woordklas uniek**

Uniek	Voorbeeld
samesmeltings	<i>dis</i>
enklitiese vorme	<i>jy't</i>
eksistensiële <i>daar</i>	<i>daar</i>
skakelpartikel	<i>Johan sit <b>en</b> slaap</i>
genitiefpartikel	<i>Madeleen <b>se</b> boek</i>
vergelykingspartikel	<i>jy is <b>so</b> dom <b>soos</b> grond</i>
ontkenningspartikel	<i>jy is nie vet <b>nie</b></i>
deelpartikel	<i>een <b>van</b> julle is skuldig</i>
betrekklike partikel	<i><b>so/wat</b> 'n mooi dag</i>
graadpartikel	<i>dit is <b>te</b> vroeg om op te staan</i>
infinitiespartikel	<i>ek is te moeg om op <b>te</b> staan</i>
werkwoordpartikel	<i>skryf dit <b>op</b> die bord (sic)<sup>2</sup></i>

(Pilon, 2005:55-58)

Die etiketteerder is morfologies nagevors (sien Pilon, 2005) en het 'n akkuraatheid van tussen 85,87% en 93,69% volgens die skrywer van die program (Pilon, 2005:6; 118). Die data is wel gekontroleer vir ooglopende foute. Die foute hou veral verband met persone se vanne en woorde wat, afhangend van die konteks, as homonieme kan optree en wat meerkonseptueel benoem kan word, soos byvoorbeeld *geld in my beursie*, as naamwoord en *wat in dié geval geld*, as werkwoord.

### 3.6.1 Voornaamwoorde

Met behulp van die geëtiketteerde data vir albei subkorpora, bereken ek die aantal voornaamwoorde vir elke subkorpus. Tabel 3.11 gee die resultate. Daar is meer voornaamwoorde in die vertaalde subkorpus (6,7%) as in die nievertaalde subkorpus (6,8%). Die resultaat vir die vertaalde en nievertaalde subkorpus is soortgelyk. Ek skryf dit toe aan goeie subredigering waardeur

<sup>2</sup> **op** in dié voorbeeld is 'n voorsetsel; in *skryf dit **neer** op die bord* is *neer* 'n werkwoordpartikel.

die kans op dubbelsinnigheid met die gebruik van voornaamwoorde verwyder word deur eksplisiet naamwoorde te gebruik.

**Tabel 3.11: Frekwensie van voornaamwoorde**

	Aantal voornaamwoorde	Persentasie
Nievertaal	1 289	6,7%
Vertaal	1 179	6,8%

My bevinding stem dus nie ooreen met Laviosa se hipotese dat vertaalde tekste 'n laer frekwensie voornaamwoorde as nievertaalde tekste het nie.

### 3.6.2 Leksikale digtheid

Leksikale digtheid word bereken deur die aantal inhoudswoorde (naamwoorde, byvoeglike naamwoorde en werkwoorde) deur die totale aantal woorde van die korpus te deel en te vermenigvuldig met 100. Tabel 3.12 wys die aantal inhoudswoorde teenoor die totale aantal woorde en die berekening vir die leksikale digtheid.

**Tabel 3.12: Leksikale digtheid**

	Aantal inhoudswoorde	Totale aantal woorde	Leksikale digtheid
Nievertaal	9 299	19 283	48,2%
Vertaal	8 617	17 450	49,4%

My bevinding is dat die leksikale digtheid van die vertaalde subkorpus hoër is as dié van die nievertaalde korpus. My bevinding stem dus nie ooreen met Laviosa se bevinding dat vertaalde tekste se leksikale digtheid laer is as vir nievertaalde tekste nie maar stem ooreen met Bam wat bevind het dat die leksikale digtheid vir die vertaalde tekste hoër is as vir die nievertaalde tekste.

*Pectora robotant cultus recti*

### 3.6.3 Superordinate

Carstens (2003:115) verklaar superordinate as 'n algemener woord wat die hoofklas is waarin verskeie subordinate (die meer spesifieke woorde) se betekenis ingesluit kan wees. As voorbeeld gee hy *dier* (die superordinaat) waarvan die subordinate of hiponieme *hond*, *kat*, *leeu* en *perd* is. Om te kyk of daar meer superordinate in die vertaalde subkorpus as in die nievertaalde subkorpus voorkom, het ek die geëtiketteerde data in *MSEExcel* eers volgens woordsoort gesorteer en toe alfabeties. Daarna het ek die naamwoorde van die twee subkorpora met mekaar vergelyk. Dit is moeilik om superordinate te onderskei deur bloot na die woordelyste te kyk. Ek kan daaruit geen gevolgtrekking maak oor die gebruik van superordinate in die subkorpora nie.

Met die vermoede dat dit makliker sal wees om die gebruik van superordinate te identifiseer in een vergelykende artikelpaar, selekteer ek artikel 8 (vertaal en nievertaal) (bylaag D). Ek skep woordelyste vir elkeen en vergelyk veral die naamwoorde om te sien of die vertaalde artikel meer superordinate gebruik as die nievertaalde artikel.

**Tabel 3.13 Superordinate in artikel 8**

Vertaling		Nievertaling	
Superordinaat	Subordinaat	Superordinaat	Subordinaat
maatskappy	motormaatskappy	beheermaatskappy	maatskappy innoveringsmaat- skappy
personeel personeel	bestuurspersoneel werknemers hoof werknemers bestuur	bestuurder arbeid mense	hoofbestuurder werknemers
sportnutsvoertuie		kontrak	McLaren-kontrak
uitgawes	loonkoste lone	fabriek	ACG-fabriek
fabrieke aanlegte		onderdele	sitplekke kattebakke

	advertensie-afdeling bemarkingsafdeling produksiebeplannings afdeling	koolstofvesel	koolstofvesel- onderdele koolstofvesel- produkte koolstofvesel-wiele
			super-sportmotor Mercedes-Benz SLR McLaren
		resiesfietse	Shimano-fietse
			militêre vegvliegtuie helikopters

Sommige subordinate kan self ook as superordinate optree, mens sou byvoorbeeld kon sê dat *koolstofvesel-onderdele* onder *koolstofvesel* kan val, maar dan sou mens ook subordinate vir *koolstofvesel-onderdele* kon kry. Dus kan ek ook hieruit geen vaste gevolgtrekking maak oor die gebruik van superordinate in die onderskeie subkorpora nie.

### 3.7 Stap 5: Konkordansie

#### 3.7.1 Opsionele dat

In Afrikaans is *dat* 'n neutrale voegwoord wat maklik uitgelaat kan word sonder dat die semantiese waarde van 'n sin verander word (Feinauer, 1990:116). In Afrikaans word *dat* weggelaat na 'n semanties minder dominante hoofsin (Feinauer, 1990:117). Vir die doeleindes van hierdie studie, bestudeer ek al die gevalle waar *dat* voorkom, herskryf kan word sonder *dat*. Byvoorbeeld:

*Pell, die aartsbiskop van Sydney, het gesê hy is dankbaar dat daar geen geweld in Australië was in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*

kan herskryf word sonder *dat* as:



*Pell, die aartsbiskop van Sydney, het gesê hy is dankbaar daar was geen geweld in Australië in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*

In hierdie selfde sin is ook 'n voorbeeld waar 'n opsionele *dat* nie gebruik is nie:

*Pell, die aartsbiskop van Sydney, het gesê Ø hy is dankbaar dat daar geen geweld in Australië was in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.*

Om die opsionele *dat* in die vertaalde subkorpora te identifiseer, is die konkordansie met behulp van die soekwoord *dat* geskep. Daarna is die kollokasiefunksie gebruik waar *verwag*, *gesê* en *verseker* uitgewys is as werkwoorde wat saam met *dat* voorkom (bylaag E). Die sleutelwoord-in-konteks saam met *verwag*, *gesê* en *verseker* is daarna deurgewerk. Daarna is die konkordansiefunksie gebruik om voorbeelde van die werkwoorde *verwag*, *gesê* en *verseker* waar die opsionele *dat* nie gebruik is nie, op te spoor. Voorbeeldsinne hiervan word hieronder aangedui.

**Vertaalde subkorpus: *verwag* en *dat***

1. <i>Iran het ook bevestig dat hy verwag <u>dat</u> die produksielimiet dieselfde sal bly.</i>	opsioneel
2. <i>Die Asiërs moes met 'n gelykop-uitslag van 2-2 tevrede wees nadat allerweë verwag is <u>dat</u> hulle maklik sou wen.</i>	opsioneel
3. <i>Volgens Charlton het hy nooit verwag <u>dat</u> Ronaldo die klub sou verlaat nie.</i>	opsioneel
4. <i>“Ons verwag Ø Hans gaan vir tussen vier en ses weke buite aksie wees,” het Ertugral gesê.</i>	Ø

Van hierdie vier gevalle, gebruik drie die opsionele *dat* en die opsionele *dat* word in een geval weggelaat.

**Vertaalde subkorpus: gesê en dat**

1. Mnr. Shukri Ghanem, hoof van Libië se nasionale oliemaatskappy, het gesê <u>dat</u> dit geopolitieke probleme is wat die grootste invloed op die olieprys uitoefen.	opsioneel
2. Ter verdediging moet darem gesê word <u>dat</u> die Italianers, wat in Napels, gespeel het, nog nie met hul plaaslike ligasokker begin het nie en duidelik verroes was.	opsioneel
3. Pecoriello het gesê <u>dat</u> hy die hoogste vertroue het in Nooyi se bekwaamheid om die maatskappy vorentoe te lei.	opsioneel
4. Net 11% van die 1 300 werkers wat deur die indiensnemingsmaatskappy Peninsula ondervra is, het gesê <u>dat</u> hulle 'n geval van afknouery by 'n bestuurder sal aanmeld.	opsioneel
5. Kim het aan 'n internasionale forum gesê Ø Noord-Korea kan dalk sanksies omseil deur kerntegnologie aan lande soos Iran te verkoop.	Ø
6. Henry het gesê Ø dit was 'n taktiek om McCaw uit die wedstryd te kry.	Ø
7. As deel van die herstrukturierungsplan, het Ford in Januarie vanjaar gesê Ø hy sou 14 fabrieke toemaak en sowat 30 000 Amerikaanse poste teen 2012 besnoei.	Ø
8. "Ek het gesê Ø ons is beroepslui wat besig is om te doen wat ons die beste doen en vir die beampte gesê om eerder te gaan sit en met sy kinders te speel."	Ø
9. Pres. Mahmoed Ahmadinedjad van Iran het Maandag gesê Ø hy respekteer die pous.	Ø
10. Imam Sami Salem van die Groot Moskee in Rome het gesê Ø die pous het Sondag met sy verskoning "weggestap" van sy aanmerking oor die Islam en geweld.	Ø
11. Intussen het Rome se mees senior Moslem-amptenaar gister gesê Ø die pous se verskoning baan die weg vir nuwe samesprekings tussen die gelowe.	Ø
12. Die pous het vroeër die steun gekry van die aartsbiskop van Kantelberg, dr. Rowan Williams, en pres. George W. Bush van Amerika het gesê Ø die verskoning was "opreg".	Ø
13. Pell, die aartsbiskop van Sydney, het gesê Ø hy is dankbaar dat	Ø

<p>daar geen geweld in Australië was in reaksie op die pous se aanhalings uit 'n Middeleeuse teks nie, maar het die gewelddadige protes in ander lande veroordeel.</p>	
<p>14. Die hoof van Australië se Katolieke Kerk het die gedempte reaksie in dié land op pous Benedictus XVI se uitsprake oor die Islam verwelkom, maar gesê Ø die geweld elders ter wêreld “bewys dat die pous se vrese gegrond is”.</p>	Ø
<p>15. Henry het gesê Ø sy spanne gaan ook in aanstaande jaar se Drienasies-toetse én in die tuisreeks teen Frankryk, op volle sterkte wees.</p>	Ø
<p>16. In Wellington het die All Black-afrigter, Graham Henry, gesê Ø sy beleid om spanne te roteer, sal nie meer aan die einde van die jaar in die toetse teen Engeland, Wallis en Frankryk geld nie.</p>	Ø
<p>17. Die Wallaby-afrigter, John Connolly, het gesê Ø 'n sterker geestelike aanslag kan Australië help om sy swak vertoningslys in wegtoetse te oorkom.</p>	Ø
<p>18. Iran se departement van buitelandse sake het ure vroeër gesê Ø Iran beplan 'n konferensie waarin die omvang van die Holocaust bevraagteken gaan word.</p>	Ø
<p>19. Hy het gesê Ø Hair geniet sy volle steun.</p>	Ø
<p>20. Percy Sonn, president van die Internasionale Krieketraad, het gesê Ø hy het die uitvoerende raad in die lig van die gebeure byeengeroep.</p>	Ø
<p>21. Hy het gesê Ø die besonderhede van die betaling wat hy verlang, is vertroulik en dat dit op dié datum in sy rekening inbetaal moet word.</p>	Ø
<p>22. Hair het in sy brief, wat aan die media beskikbaar gestel is, gesê Ø hy is bereid om te bedank uit sy posisie op die skeidsregter se elite-paneel.</p>	Ø
<p>23. Malcolm Speed, uitvoerende hoof van die IKR, het gister hier op 'n nuuskonfersie gesê Ø die skeidsregter het onder groot spanning verkeer toe hy per brief dié aanbod aan die krieketraad gemaak het</p>	Ø
<p>24. Hisbolla het gesê Ø hy het minstens 10 sterftes veroorsaak en ses tens vernietig by Kantara en in ander gebiede, maar het,</p>	Ø

soos die Israelse weermag, nie gesê of van sy eie vegters dood is nie.	
25. Intussen het Libanese veiligheidsmagte gesê Ø die Israelse magte het wes van Taibé, sowat 5 km van die grens, na die dorpie Kantara en noord na Boerdj al-Moloek en Klaia opbeweeg.	Ø
26. Arsene Wenger, Arsenal se afrigter, het intussen gesê Ø hy hoop sy span vind gou sy voete in die nuwe Emirates-stadion, waar hy vanaand teen die Kroasië se Dinamo Zagreb te staan kom.	Ø
27. Dié verdediger het gesê Ø Chelsea benodig 'n opgeruimde spangees en verskerpde (sic) konsentrasie as hy sy droom van 'n derde Premierliga-titel agtereenvolgens wil verwesenlik.	Ø
28. John Terry, kaptein van Chelsea se sokkerspan, het gesê Ø sy span sal vanaand alles in sy vermoë doen om 'n vroeë voorsprong oor sy Engelse Premierliga-teenstanders te bewerkstellig.	Ø
29. IAAF het gesê Ø die ondersoek oor die werksaamhede van Trevor Graham sal uitgevoer word in samewerking met die Amerikaanse agentskap teen opkikkers (Usada).	Ø
30. Sarfraz het gesê Ø swak spankeuses en 'n gebrekkige spangees was die redes vir Pakistan se sukkelvertonings.	Ø
31. Die voormalige Pakistanse snelbouler Sarfraz Mawaz het intussen gesê Ø Inzamam-ul-Haq moet uittree as dié span se kaptein ná Engeland sy eerste tuisreekssege in 24 jaar oor dié Suid-Asiese span behaal het.	Ø
32. Engeland se kaptein, Andrew Strauss, wat die span aanvoer in die afwesigheid van Flintoff (enkelbesering) en Michael Vaughn (knie), het gesê Ø dit is belangrik vir die span om sy wenmomentum te behou vóór die eerste toets teen Australië op 23 November in Brisbane begin.	Ø
33. Sinyangana het gesê Ø 'n lid van die publiek het die gesteelde motor geparkeer gesien nie ver van die kapingstoneel nie en die polisie daarvan verwittig.	Ø
34. Ofskoon Brooke die All Blacks net vandeeweek bystaan, het Henry gesê Ø sy dienste kan weer gebruik word.	Ø

35. Die afrigter, Graham Henry, het gesê Ø dié probleem is groot en kan dalk nie betyds opgelos word voor die WB-toernooi nie.	Ø
36. Jones het gesê Ø die dinamiese Giteau sal 'n uitstekende plaasvervanger vir Gregan wees.	Ø
37. Jones het aan die Sydney Morning Herald gesê Ø die 32-jarige Larkham moet steeds op losskakel behou word, maar Matt Giteau moet in die plek van die 33-jarige Gregan op skrumkakel oorweeg word.	Ø
38. Jones, wat verlede jaar in Desember afgedank is ná Australië sy swakste reeks vertonings in 36 jaar gelewer het, het gesê Ø die Wallabies se veteraan-skakelpaar, Gregan en die losskakel Stephen Larkham, moet oor 'n fyn kam geskeer word.	Ø
39. Bronne na aan die Wallaby-kamp het aan die Telegraph gesê Ø dit wil voorkom of die ouer spelers in die span sukkel om by die nuwe aanvalsbenadering aan te pas.	Ø
40. Connolly het aan die Sydney Morning Herald gesê Ø sy spelers se sielkundige ingesteldheid was nie wat dit moes wees in die aanloop tot die toets nie.	Ø
41. Ofskoon Landis nie prontuit gesê het Ø iemand het met sy toetse gepeuter nie, was die implikasie duidelik.	Ø
42. Deutsche Bank het die afknouery ontken, maar gesê Ø hy respekteer die hof se besluit terwyl hy 'n appèl oorweeg.	Ø
43. Ná die uitspraak het sy gesê Ø die enigste rede waarom sy die afknouery so lank verduur het, was omdat sy haar loopbaan op prys gestel het.	Ø
44. Hy het gesê Ø hy het kans gehad om tot op sy dak te klim en het van daar af gesien hoe talle mense hulpeloos deur die stroom weggevoer word.	Ø
45. Die afrigter van Pakistan, Bob Woolmer, het gesê Ø Akhtar sal die naweek 'n klubwedstryd vir Berkswell speel.	Ø
46. Trescothick het gesê Ø hulle het altyd geweet dit sou nie veel verg om weer op die wenpad te kom nie.	Ø
47. Hy het nie juis nuwe United-ondersteuners bygekry nie, nadat hy gesê het Ø hy wil eerder vir Real Madrid gaan speel.	Ø
48. Die bank het gesê Ø die “drastiese” aksie was nodig om Swissfirst se sukses te verseker en om die maksimum	Ø



<i>beskerming aan sy kliënte, werknemers en aandeelhouers te bied.</i>	
<i>49. Die klein bank, wat in batebestuur, private- en beleggingsbanksake spesialiseer, het gesê Ø hy wil verdere skade wat deur die media veroorsaak is, beperk deur sy strategiese opsies, soos die verkoop van die groep, te bekyk.</i>	Ø
<i>50. Graham Henry, die All Blacks se afrigter, het gesê Ø Toeava is 'n talentvolle speler wat die kans verdien om hom ook op toetsvlak te bewys.</i>	Ø
<i>51. Connolly het gesê Ø die Wallabies sien beslis kans om die eerste Australiese span sedert 1986 te word wat die All Blacks op Eden Park kan klop.</i>	Ø
<i>52. Me. Lynn Franco, direkteur van die Konferensieraad se Verbruikersnavorsingsentrum (sic) het gesê Ø minder gunstige saketoestande wat na 'n minder gunstige werkscenario gelei het, het die grootste een-maandvertrouefname sedert die orkaan Katrina veroorsaak.</i>	Ø
<i>53. Hy het gesê Ø 'n verdere versekering van leenkoste is nodig om inflasie vinnig af te bring.</i>	Ø

Die opsionele *dat* word in vier van hierdie 53 gevalle gebruik en in 49 gevalle weggelaat.

**Vertaalde subkorpus: verseker en dat**

<i>1. "Ons wil seker maak dat ons dinge reg doen vóór die Wêreldbeker-toernooi in 'n poging om te verseker <u>dat</u> van ons uitblinkers drie wedstryde agtereenvolgens kan speel."</i>	opsioneel
<i>2. Stevie is steeds ons beste nommer 10, maar hy moet goed bestuur word om te verseker <u>dat</u> hy sy beste peil handhaaf.</i>	opsioneel
<i>3. Pakistan is die taak gelaat om 323 lopies vir 'n sege aan te teken, maar uitstekende boulwerk het verseker <u>dat</u> die besoekers vir net 155 lopies uitgehaal is.</i>	opsioneel

Al drie voorkomste van *verseker* gevolg deur *dat* is opsioneel. Daar is geen voorbeelde van waar die opsionele *dat* nie voorkom nie.



Vir die nievertaalde subkorpus gebruik ek die kollokasiefunksie (bylaag F) om te kyk na gevalle waar *verwag*, *gesê* en *verseker* voorkom. Daarna word die konkordansiefunksie gebruik om gevalle op te spoor waar die opsionele *dat* nie saam met die werkwoorde *verwag*, *gesê* en *verseker* gebruik word nie. Voorbeeldsinne word hieronder aangedui.

**Nievertaalde subkorpus: *verwag dat***

1. "Intussen het die winkelsentribestuurders ook goed op ons konsep gereageer en ons verwag <u>dat</u> ons internasionale reeks kossoorte binnekort wyd beskikbaar sal wees," het Draper gesê.	opsioneel
2. Sonn het by navraag gesê hy het verwag <u>dat</u> die GPI-direksie hulle sou uitskop, maar hy meen die dispuut tussen Sancino en GPI sou baie beter oplosbaar gewees het as hulle binne die GPI-direksie daarvoor kon praat.	opsioneel
3. Osrin verwag <u>dat</u> verbruikersbesteding, hoewel steeds lewendig, nie so sterk gaan wees in die tweede helfte van die boekjaar nie.	opsioneel
4. Ekonomie verwag <u>dat</u> VPIX vir die res van die jaar verder sal styg en in die eerste helfte van 2007 6% oorsteek.	opsioneel

Al vier die voorbeelde gebruik die opsionele *dat*.

**Nievertaalde subkorpus: *gesê dat***

1. Die Foschini Groep het op sy jaarvergadering gesê $\emptyset$ $\emptyset$ voortgesette sterk verbruikersbesteding is in die vyf maande tot einde Augustus ervaar.	$\emptyset$
2. Viljoen het die Statebondtitel vroeër vanjaar met 'n gooi van 60.72 m verower en het gesê $\emptyset$ sy wil ook die Afrika-titel by haar versameling voeg.	$\emptyset$
3. Pahad het gesê $\emptyset$ sy ma, Amina, het aan die optog deelgeneem.	$\emptyset$
4. Buckley glo sy weglating uit die Bafana-span het in Duitsland ontstaan, waar hy in 'n koerant aangehaal is wat geskryf het dié speler het gesê $\emptyset$ Suid-Afrika sal nie 'n suksesvolle Wêreldbeker-toernooi in 2010 kan aanbied nie.	$\emptyset$
5. Mnr. John Bolton, Amerika se ambassadeur by die VN, het gesê $\emptyset$	$\emptyset$

<i>Ø sanksies sal onmiddellik ingestel word wanneer die sperdatum verstryk.</i>	
<i>6. Hack het gister gesê Ø Safa se finansiële state is ook gister aan die uitvoerende bestuur voorgelê.</i>	Ø
<i>7. Muir het gesê Ø Ruan Pienaar en BJ Botha se insluiting in die Bok-groep laat 'n gaping, maar hy het die volste vertrouwe in die ander spelers in sy groep.</i>	Ø
<i>8. Oliphant het ook gesê Ø Parreira sal sy eie hulpafrigter en 'n fiksheidsafrigter saambring.</i>	Ø
<i>9. Mickey Arthur, afrigter van die Suid-Afrikaanse span het gister gesê Ø ofskoon die wedstryd nog nie amptelik afgestel is nie, blyk dit in ernstige gedrang te wees.</i>	Ø
<i>10. Arthur het gister gesê Ø ofskoon sake nog nie bevestig is nie, gaan hy probeer om die span vroeër na Indië te neem en daar wedstryde te speel ter voorbereiding vir die toernooi om die Kampioene-trofee.</i>	Ø
<i>11. Hy het gesê Ø net sowat R250 miljoen se voorkeuraandele sal vir eers uitgereik word, en die geld sal gebêre word vir enige verkrygingsgeleenthede wat mag opduik.</i>	Ø
<i>12. Ander kleinhandelgroepe het verlede week ook gesê Ø 'n verlansaming in verkopegroei word voorsien, maar dat goeie besteding steeds plaasvind.</i>	Ø
<i>13. Polak het egter gesê Ø 'n ineenstorting in besteding word nie voorsien nie.</i>	Ø
<i>14. Mnr. Eliot Osrin, voorsitter van die Foschini Groep, het gesê Ø in dié tydperk het verkope met 17,5% gestyg vergeleke met dieselfde tyd die vorige jaar.</i>	Ø
<i>15. De Sa, wie se span twee weke gelede erg slae gekry het toe hy met 1-4 teen Sundowns vasgeval het in 'n Supa8-wedstryd, het gesê Ø hy is opgewonde oor die nuwe seisoen.</i>	Ø
<i>16. Dié Bolander het gesê Ø hy glo sy hand is weer reg terwyl die litteken van die operasie duidelik gesien kon word.</i>	Ø
<i>17. Buckley het gesê Ø hy is gelukkig by Basel en dit wil lyk of sy klubloopbaan weer lewe gekry het nadat hy drie doele in sewe wedstryde aangeteken het.</i>	Ø
<i>18. Hy het gesê Ø hy het amper 'n kontrak met die Engelse</i>	Ø

<i>Premierliga-klub Everton gesluit.</i>	
<i>19. Villiers het gesê Ø hy het té vinnig begin.</i>	Ø
<i>20. Boucher het ook gesê Ø hy glo hulle kan enige span op hul dag klop.</i>	Ø
<i>21. Sonn het by navraag gesê Ø hy het verwag dat die GPI-direksie hulle sou uitskop, maar hy meen die dispuut tussen Sancino en GPI sou baie beter oplosbaar gewees het as hulle binne die GPI-direksie daaroor kon praat.</i>	Ø
<i>22. Jooste het gister gesê Ø hy het met die Leeus gepraat en weet hulle dink Pretorius het nog een wedstryd ná Saterdag s'n nodig.</i>	Ø
<i>23. Ithier het aan Die Burger gesê Ø hy sien dit as 'n eer om hierdie jong mans af te rig.</i>	Ø
<i>24. Die Turk Muhsin Ertugal het by Gordon Igesund as afrigter oorgeneem en gesê Ø hy sal nie dié seisoen skroom om jong spelers die kans te gee om hul staal op nasionale vlak te wys nie.</i>	Ø
<i>25. Pres. George W. Bush het gister gesê Ø hy neem kennis daarvan dat Israel, Libanon en Hisbolla teen sekere dele van die resoluksie gekant is, maar “almal wil hê die geweld moet stop”.</i>	Ø
<i>26. Die Proteas se afrigter, Mickey Arthur, het vroeër gesê Ø hy sou só 'n geleentheid met albei hande aangryp.</i>	Ø
<i>27. Buckley het gesê Ø hy is beïndruk met die “nuwe” Bafana.</i>	Ø
<i>28. Ithier het gesê Ø hy sal sy span vanaf 14 Augustus na 'n oefenkamp by 'n binnenshuise sokkerveld neem om hulle aan die speeloppervlak bloot te stel.</i>	Ø
<i>29. Igesund het laas seisoen keer op keer gesê Ø hy is besig om 'n span te bou, maar dié verskoning het afgewater geraak namate Ajax CT al verder afgeskuif het op die Premierliga-punteleer.</i>	Ø
<i>30. Hy het gister gesê Ø hy sien dit as 'n geleentheid.</i>	Ø
<i>31. Boucher het egter gesê Ø hulle gaan uit om die toernooi te wen.</i>	Ø
<i>32. Hy het gesê Ø hulle het op Maandag 31 Julie – die dag ná die spesiale aandeelhouersvergadering op Sondag 30 Julie – briewe gekry om te sê hul dienste word nie meer benodig nie.</i>	Ø
<i>33. Clark het gesê Ø hulle hoop om die bank se kapitaal van die huidige R125 miljoen tot R250 miljoen te vergroot, en om mettertyd bemagtigingsvennote te betrek.</i>	Ø
<i>34. Clark het gesê Ø hulle sal nie al die voorkeuraandele waarvoor</i>	Ø

<i>goedkeuring gekry is, uitreik nie.</i>	
<i>35. Ontleders het gister gesê Ø hoewel die prys 'n ernstige terugsakking beleef, kan dit weer terugspring as die verwagte seisoenale tekort aanstaande kwartaal ontstaan.</i>	Ø
<i>36. Mnr. Dennis Polak, besturende direkteur, het aan Sake gesê Ø hoewel 'n verlansaming in verkopegroei verwag word, is dit nie 'n groot verlansaming nie.</i>	Ø
<i>37. "Die dokters het aan my gesê Ø ek moes vanjaar heeltemal rus," het sy gister erken.</i>	Ø
<i>38. Die premier het later 'n vroeëre stelling gewysig en gesê Ø dit het geblyk dat slegs een persoon in dié spesifieke aanval omgekom het.</i>	Ø
<i>39. Oor die vooruitsig om die eerste keer saam met Butch James te speel, het De Villiers gesê Ø dit is iets waarna hy uitsien.</i>	Ø
<i>40. Op 'n vraag oor die voorkoms van anti-Semitisme het Baruch gesê Ø dit is net 'n groeiende verskynsel op die vasteland van Europa en dat die Israeliese regering dit monitor en deur diplomatieke oorlegpleging met regerings hanteer.</i>	Ø
<i>41. Die kaptein het ook gesê Ø dit is nou hoog tyd dat die span se laaste eendaagse reeksnederlaag (met 0-5) hier in Sri Lanka iets van die verlede is en dat mense nou op die toekoms moet konsentreer.</i>	Ø
<i>42. De Villiers het gesê Ø dit is wonderlik om terug te wees by die Bokke ná 'n ribbesering 'n stokkie voor sy toestrugby-deelname (sic) gesteek het.</i>	Ø
<i>43. Clark het gister in 'n verklaring gesê Ø dit kos tyd om die nie-skeepsbedrywighede volledig te ontwikkel, maar dit het groot potensiaal.</i>	Ø
<i>44. Mnr. Alan Olivier, aangewese uitvoerende hoof wat teen einde vanjaar by Clark sal oorneem, het in 'n verklaring gesê Ø die skeepsmarkte het vroeër as verwag ná die noordelike somer herstel.</i>	Ø
<i>45. Mnr. Trevor Manuel, minister van finansies, het in reaksie daarop gesê Ø die IMF sal altyd standpunt teen arbeidswetwetgewing (sic) inneem, anders sal hy nie die IMF wees nie.</i>	Ø
<i>46. Williams het gister gesê Ø die pous moet geoordeel word na</i>	Ø

aanleiding van sy hele geskiedenis, waarin hy “baie positief gepraat het oor dialoog”.	
47. Shaun Pollock, Suid-Afrika se deurwinterde bouler, het egter gesê Ø die reeks gee die Suid-Afrikaanse spanbestuur 'n uitstekende geleentheid om te sien hoe die minder gevestigde spelers in die Protea-groep gaan vaar.	Ø
48. Bush het Iran 'n ondersteuner van terroriste genoem en gesê Ø die wêreld word bedreig deur dié radikale regime.	Ø
49. Siniora het reeds gesê Ø die resoluëie is “nie voldoende nie”, maar hy het dit nog nie verwerp nie.	Ø
50. Raymond Hack, Sufa se uitvoerende hoof, het gesê Ø die vereniging gaan alles in sy vermoë doen om die afrigter so spoedig moontlik in Suid-Afrika te kry.	Ø
51. GPI het gesê Ø die uitreiking dien juis bemagtiging omdat nuwe aandele aan bestaande bemagtigingsvennote en voorheen benadeelde individuele GPI-aandeehouers aangebied sal word.	Ø
52. Clark het met die bekendmaking van die 2005-jaarresultate vroeër vanjaar gesê Ø die bydrae kan in die lopende boekjaar tot R240 miljoen groei.	Ø
53. GPI het Donderdag in 'n persverklaring gesê Ø die aandeleuitreiking sal GPI help om sy belang in SunWest van 19% tot 30% uit te brei.	Ø
54. Swarup, wat pas begin werk het as Indië se onderhoëkommissaris in Pretoria, het gesê Ø die bekroning is naas die publikasie van sy roman die gelukkigste oomblik van sy lewe.	Ø
55. Baruch het gesê Ø die Midde-Ooste het nou 'n “positiewe agenda” nodig.	Ø
56. Mnr. Jan van Niekerk, beleggingshoof van Citadel, het gesê Ø die JSE is tans op 'n sikliese hoogtepunt.	Ø
57. Mnr. Leon van Wyk, finansiële direkteur van Sizafika, het gesê Ø die verkryging behels tussen 20 en 25 eiendomme, met nog 30 in die vooruitsig.	Ø
58. Wat die weglating van die ervare Breyton Paulse uit die groep betref, het Jooste gesê Ø die keurders voel dié kwiksilwer-vleuel het 'n ruskans nodig nadat hy vroeër vanjaar klubrugby in	Ø



<i>Frankryk gespeel het.</i>	
<i>59. Paarwater het gesê Ø die rol wat Melck in die WP-span speel, durf nie onderskat word nie.</i>	Ø
<i>60. Die speler, wat vir Basel in Switserland speel, het gesê Ø die wedstryd teen Zambië gaan moeilik wees.</i>	Ø
<i>61. Annan het gesê Ø die sanksies is in die hande van die Veiligheidsraad.</i>	Ø
<i>62. Mnr. Abdullah Gul, Turkse minister van buitelandse sake, het aan die BBC gesê Ø die beplande besoek sal voortgaan soos beplan.</i>	Ø
<i>63. Die aartsbiskop van Kantelberg, dr. Rowan Williams, het aan die BBC gesê Ø die pous se uitspraak word buite konteks gesien en hy verdien nie verdere kritiek nie.</i>	Ø
<i>64. Gerald Majola, uitvoerende hoof van Krieket Suid-Afrika (KSA), het gister gesê Ø die beleidskomitee sal die eerste week in Oktober byeenkom om aspekte soos die presiese datums, velde en die formaat van die toernooi te bespreek.</i>	Ø
<i>65. Mnr. Richard Brown, finansiële bestuurder en waarnemende hoofbestuurder van ACG SA, het gesê Ø die Britse beheermaatskappy was al lank op soek na 'n manier om toegang tot die Suid-Afrikaanse mark te verkry.</i>	Ø
<i>66. Hy het gesê Ø die eiendomme wat verkry word, asook dié in die onderhandelingsfase, voldoen aan die maatskappy se beleggingskriteria van B- en C-graad-kantoorgeboue, kleinhandel- en nywerheidseiendomme met herontwikkelings en opgraderingspotensiaal.</i>	Ø
<i>67. Met verwysing na pres. Mahmoed Ahmadinedjad van Iran se wilde uitsprake, soos dat die Joodse volksmoord nooit plaasgevind het nie en dat die staat Israel geen bestaansreg het nie, het Baruch gesê Ø die ondervinding van die verlede leer dat stellings van kolossale omvang nie ligtelik opgeneem moet word nie.</i>	Ø
<i>68. Oor sy eie wedstrydfiksheid het De Villiers gesê Ø daar is ruimte vir 'n verbetering van tussen 10 en 20%.</i>	Ø
<i>69. Olifant het gister gesê Ø daar is 'n proses waardeur die vereniging gaan voordat die afrigter met sy werk kan begin.</i>	Ø
<i>70. Hy het by navraag gesê Ø daar is nie tans sprake van 'n</i>	Ø



<i>oorname in die vrugtehanteringsbedryf nie.</i>	
71. <i>Mnr. Kofi Annan, die VN se sekretaris-generaal, het gister gesê Ø daar is “internasionale pogings” om ’n ontmoeting tussen Iran en die ses nasies wat die aansporingspakket saamgestel het, te bemiddel.</i>	Ø
72. <i>Die Proteas het wel die toetsreeks teen Sri Lanka met 0-2 verloor, maar die span se kriketafrigter, Mickey Arthur, het gister gesê Ø daar was baie positiewe aspekte vir die toekoms.</i>	Ø
73. <i>Mohr het gesê Ø Citadel se grondliggende ondersoek dui daarop dat die mark en die rand tans nie die risiko’s in die Suid-Afrikaanse ekonomie inprys nie.</i>	Ø
74. <i>Mohr het gesê Ø Citadel se siening, wat gewoonlik meer pessimisties as die heersende konsensus is, behels dat die reuse-tekort op die lopende rekening van 6,4% van die bruto binnelandse produk (BBP) een of ander tyd ’n knelpunt in die ekonomie sal word.</i>	Ø
75. <i>Hy het gesê Ø as die verkryging van R225 miljoen ingesluit word, is die blootstelling aan nywerheidseiendomme 45%, terwyl die balans tussen kleinhandel- en kantooreiendom gedeel word.</i>	Ø

Van die 75 voorbeelde, is almal sonder die opsionele *dat*.

**Nievertaalde subkorpus: verseker dat**

1. <i>Die Suid-Afrikaanse Sokkervereniging (Safa) gaan alles moontlik doen om te verseker <u>dat</u> die nuwe nasionale breier, Carlos Alberto Parreira, op sy pos is vir Bafana Bafana se eerste kwalifiserende wedstryd om die Afrikabeker-toernooi in September.</i>	opsioneel
2. <i>Hulle het vir ongeveer 80 km saamgewerk en so verseker Ø die gaping is só groot dat die bondel hulle nie kan inhaal nie.</i>	Ø
3. <i>Dit het verseker Ø hulle kom reeds vandag in Suid-Afrika aan en nie Maandag (vandag) nie.</i>	Ø
4. <i>Amerika het Japan verseker Ø hy sal die volle reeks van sy afskrik- en veiligheidsverpligtinge teenoor die land nakom sou Noord-Korea hom aanval.</i>	Ø

Van die vier voorbeelde is drie sonder die opsionele *dat* en een met die opsionele *dat*.

Tabel 3.14 gee 'n opsomming van die voorkoms van die opsionele *dat*.

**Tabel 3.14: Voorkoms van opsionele dat**

Werkwoord	Vertaal	%	Nievertaal	%
<i>verwag + dat</i>	3 van 4	75%	4 van 4	100%
∅	1 van 4	25%	geen	0
<i>gesê + dat</i>	4 van 53	7,5%	geen	0
∅	49 van 53	92,5%	75 van 75	100%
<i>verseker dat</i>	3 van 3	100%	1 van 4	25%
∅	geen	0	3 van 4	75%

In tabel 3.14 kan ons sien dat die opsionele *dat* soms saam met *gesê* gebruik word in die vertaalde subkorpus (7,5% wel gebruik; 92,5% weggelaat), maar dat die opsionele *dat* weggelaat word in die nievertaalde subkorpus (100%). Saam met *verwag* word *dat* gebruik in die nievertaalde subkorpus (100%), maar nie altyd in die vertaalde subkorpus nie (75% wel gebruik; 25% weggelaat). Daarteenoor word *dat* in die vertaalde subkorpus nooit saam met *verseker* weggelaat nie (100%), maar wel soms in die nievertaalde subkorpus (25% gebruik; 75% weggelaat).

In hierdie ondersoek kan daar dus gesê word dat die opsionele *dat* saam met die werkwoorde *gesê* en *verseker* meer in die vertaalde subkorpus voorkom as in die nievertaalde subkorpus. Vir hierdie twee werkwoorde stem my bevinding ooreen met Laviosa se hipotese dat daar 'n hoër frekwensie van die opsionele *dat* in vertaalde tekste voorkom.

### 3.7.2 Kollokasiebotsings

*WordSmithTools* beskou kollokasies as woorde wat in die omgewing van soekwoorde voorkom en onderskei naasliggende kollokasies (woorde wat langs mekaar voorkom - *coherence collocates*) en omgewingskollokasies

(woorde wat nie noodwendig langs mekaar voorkom nie, maar in die omgewing voorkom - *neighbourhood/horizon collocates*). Om die voorkoms van kollokasiebotsings te ondersoek, gebruik ek die woordelyste wat vir elke subkorpus gegenereer is volgens frekwensie. Die frekwentste woorde vir die vertaalde subkorpus is *die, het, in, van, en, 'n, is, nie, te* en *sy*. Die frekwentste woorde is dus funksiewoorde wat nie in hierdie ondersoek na kollokasiebotsings van nut sal wees nie.

Daarna kyk ek na ander frekwente woorde wat in albei subkorpora voorkom sodat ek dit as soekterm vir my konkordansies kan gebruik en die gebruik daarvan in vertalings en nievertalings kon vergelyk. Frekwente woorde vir albei subkorpora is *span, mense, ander* en *toets*. Konkordansies is vir hierdie woorde geskep en daarna is kollokasies wat twee of meer keer in die subkorpora voorkom, geïdentifiseer (sien bylaag G).

Tabel 3.15 wys hoeveel keer die soekwoorde persentasiegewys in die onderskeie subkorpora voorkom.

**Tabel 3.15: Soekwoorde se frekwensie in onderskeie subkorpora**

<b>Vertaal 17 450 woorde</b>	<b>Nievertaal 19 283 woorde</b>
span (45 - 0,26%)	span (53 - 0,27%)
mense (27 - 0,15%)	mense (12 - 0,06%)
ander (29 - 0,17%)	ander (22 - 0,11%)
toets (36 - 0,21%)	toets (18 - 0,09%)

In tabel 3.15 kan gesien word dat die soekwoord *span* ongeveer dieselfde persentasie uitmaak in albei subkorpora en word nou in meer besonderhede bespreek. Vir die soekwoord *span* is daar meer verbindings met ander woorde in die nievertaalde subkorpus vóór *span* (in die L1-posisie) en meer in die vertaalde subkorpus ná *span* (in die R1-posisie), maar geen kollokasiebotsings kan opgemerk word nie (bylaag G).

**Tabel 3.16: Kollokasiefrekwensietabel vir span in vertaalde subkorpus**

N	Word	With elation]	Total otal Left]al Right]	L2	L1	Centre	R1	R2
1	SPAN	span 0.000	45 0 0	0	0	45	0	0
2	DIE	span 0.000	22 18 4	3	15	0	0	4
3	SY	span 0.000	12 10 2	1	9	0	1	1
4	HET	span 0.000	10 2 8	2	0	0	7	1
5	IN	span 0.000	10 3 7	3	0	0	3	4
6	SE	span 0.000	9 3 6	2	1	0	6	0
7	WAT	span 0.000	5 2 3	2	0	0	3	0
8	AS	span 0.000	4 2 2	2	0	0	1	1
9	DIÉ	span 0.000	4 4 0	0	4	0	0	0
10	OM	span 0.000	3 1 2	1	0	0	1	1
11	VAN	span 0.000	3 2 1	2	0	0	1	0
12	DAT	span 0.000	3 2 1	2	0	0	1	0
13	TE	span 0.000	2 1 1	0	1	0	0	1
14	EERSTE	span 0.000	2 2 0	1	1	0	0	0
15	TEEN	span 0.000	2 1 1	1	0	0	1	0
16	VIR	span 0.000	2 2 0	2	0	0	0	0
17	AAN	span 0.000	2 1 1	1	0	0	1	0
18	MET	span 0.000	2 1 1	1	0	0	1	0
19	SAL	span 0.000	2 1 1	1	0	0	1	0
20	NIE	span 0.000	2 1 1	1	0	0	1	0
21	SUID	span 0.000	2 1 1	1	0	0	0	1
22	MAAR	span 0.000	2 0 2	0	0	0	2	0
23	SEDERT	span 0.000	2 0 2	0	0	0	2	0

Volgens *WordSmith Tools* se kollokasiefunksie kom *span* in die vertaalde subkorpus saam met *die, sy, het, in, se, wat, as en dié*, vier keer of meer voor in die twee posisies voor (L1 en L2) en ná (R1 en R2) *span* (tabel 3.16).

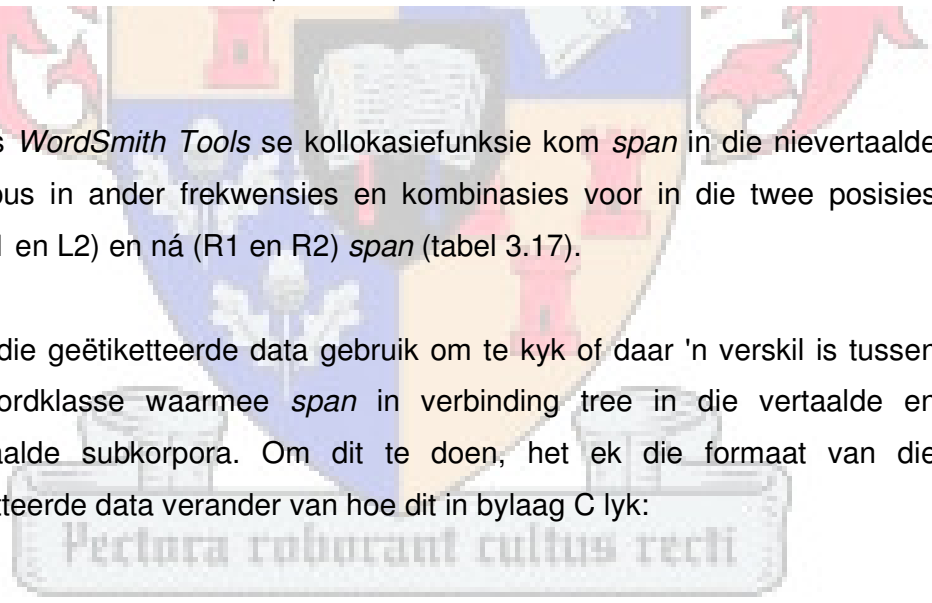


**Tabel 3.17: Kollokasiefrekwensietabel vir *span* in vertaalde subkorpus**

N	Word	With elation]	Total otal Left]al Right]	L2	L1	Centre	R1	R2
1	SPAN	span 0.000	53 0 0	0	0	53	0	0
2	DIE	span 0.000	32 28 4	7	21	0	2	2
3	SE	span 0.000	10 4 6	1	3	0	6	0
4	IN	span 0.000	8 3 5	3	0	0	4	1
5	HET	span 0.000	8 3 5	3	0	0	3	2
6	SY	span 0.000	6 6 0	1	5	0	0	0
7	MET	span 0.000	5 3 2	3	0	0	1	1
8	DAT	span 0.000	3 3 0	3	0	0	0	0
9	BESTE	span 0.000	3 3 0	0	3	0	0	0
10	OP	span 0.000	3 0 3	0	0	0	2	1
11	IS	span 0.000	3 1 2	1	0	0	2	0
12	NA	span 0.000	3 0 3	0	0	0	2	1
13	OM	span 0.000	3 3 0	3	0	0	0	0
14	SAL	span 0.000	2 1 1	1	0	0	1	0
15	VAN	span 0.000	2 0 2	0	0	0	2	0
16	VIR	span 0.000	2 2 0	2	0	0	0	0
17	WAT	span 0.000	2 0 2	0	0	0	2	0
18	TWEE	span 0.000	2 0 2	0	0	0	1	1
19	SPEEL	span 0.000	2 0 2	0	0	0	1	1
20	TANS	span 0.000	2 1 1	1	0	0	0	1
21	TE	span 0.000	2 0 2	0	0	0	2	0
22	EN	span 0.000	2 1 1	1	0	0	1	0
23	ENIGE	span 0.000	2 2 0	1	1	0	0	0
24	EENDAAGSE	span 0.000	2 2 0	0	2	0	0	0
25	BAIE	span 0.000	2 1 1	1	0	0	1	0
26	DIÉ	span 0.000	2 2 0	0	2	0	0	0
27	GAAN	span 0.000	2 0 2	0	0	0	2	0
28	KAN	span 0.000	2 1 1	1	0	0	1	0
29	KAPTEIN	span 0.000	2 1 1	1	0	0	0	1
30	HUL	span 0.000	2 1 1	0	1	0	0	1
31	GOEIE	span 0.000	2 1 1	0	1	0	0	1
32	GROEP	span 0.000	2 0 2	0	0	0	0	2

Volgens *WordSmith Tools* se kollokasiefunksie kom *span* in die nievertaalde subkorpus in ander frekwensies en kombinasies voor in die twee posisies voor (L1 en L2) en ná (R1 en R2) *span* (tabel 3.17).

Ek het die geëtiketteerde data gebruik om te kyk of daar 'n verskil is tussen die woordklasse waarmee *span* in verbinding tree in die vertaalde en nievertaalde subkorpora. Om dit te doen, het ek die formaat van die geëtiketteerde data verander van hoe dit in bylaag C lyk:



Straf	N
'	Z
sal	V
N-Korea	N
in	S
Iran	N
se	U
arms	N
dryf	V
'	Z

na

Straf <N> ' <Z> sal <V> N-Korea <N> in <S> Iran <N> se <U> arms <N> dryf <V>'Z.

Sodoende kon ek die woordklasetiket saam met *span* inspan om woordklasse voor en ná *span* te identifiseer, deur byvoorbeeld <A> *span* as soekwoord in te tik vir byvoeglike naamwoorde vóór *gebruik* en die soekwoord *span* <N> \* <V> vir werkwoorde ná *span*.

In tabel 3.18 word die resultate van die soektog weergegee met voorbeelde uit die onderskeie subkorpora.

**Tabel 3.18: Woordsoorte se kombinasie met *span***

	<b>Vertaal</b>		<b>Nievertaal</b>
	45 - 0,26%		53 - 0,27%
<b>byvoeglike naamwoord + <i>span</i></b>			
7/45 15,5%	<i>goeie span</i> <i>hele span</i> <i>Suid-Asiese span</i> <i>vorste span</i> <i>nasionale span</i> <i>Australiese span</i> <i>puik span</i>	11/53 20,8%	<i>ander span</i> <i>beste span</i> <i>Suid-Afrikaanse span</i> <i>jong span</i> <i>Indiese span</i> <i>eendaagse span</i> <i>goeie span</i>



<b>span + byvoeglike naamwoord</b>			
1/45 2,2%	<i>span bekend gemaak</i>	geen	
<b>span + werkwoord</b>			
10/45 22,2%	<i>span het</i> <i>span vind</i> <i>span sal</i> <i>span behaal</i> <i>span aanvoel</i> <i>span is</i> <i>span het</i>	17/53 32,1%	<i>span word</i> <i>span wil</i> <i>span voeg</i> <i>span takel</i> <i>span speel</i> <i>span sal</i> <i>span oefen</i> <i>span kan</i> <i>span is</i> <i>span het</i> <i>span gekies</i> <i>span gedreig</i> <i>span gaan</i>
<b>voornaamwoord + span</b>			
14/45 31,1%	<i>sy span</i> <i>dié span</i> <i>hierdie span</i>	9/53 17%	<i>sy span</i> <i>my span</i> <i>dié span</i> <i>hul span</i>
<b>span + voornaamwoord</b>			
4/45 8,9%	<i>span wat</i> <i>span sy</i>	2/53 3,8%	<i>span wat</i>
<b>lidwoord + span</b>			
18/45 40%	<i>die span</i> <i>'n span</i>	22/53 41,5%	<i>die span</i> <i>'n span</i>
<b>uniek + span</b>			
2/45 4,4%	<i>se span</i> <i>te span</i>	3/53 5,7%	<i>se span</i>
<b>span + uniek</b>			
4/45 8,9%	<i>span se</i>	8/53 15,1%	<i>span se</i> <i>span te</i>
<b>span + voorsetsel</b>			
10/45	<i>span van</i>	9/53	<i>span na</i>

22,2%	<i>span in</i> <i>span om</i> <i>span teen</i> <i>span aan</i> <i>span sedert</i> <i>span onder</i> <i>span na</i>	17%	<i>span in</i> <i>span ter</i> <i>span van</i> <i>span op</i> <i>span vanaf</i>
<b>span + voegwoord</b>			
2/45 4,4%	<i>span as</i> <i>span dat</i>	geen	

Hoewel ek geen uitsprake kan lewer oor die voorkoms van kollokasiebotsings nie, kan mens tog in tabel 3.18 sien dat *span* anders met woordklasse in verbinding tree as in die nievertaalde subkorpora, al kom *span* persentasiegewys amper ewe veel in die onderskeie subkorpora voor. So kom *span* byvoorbeeld in die vertaalde subkorpus voor byvoeglike naamwoorde (*span bekend maak*) en voegwoorde (*span as, span dat*) voor maar nie in die nievertaalde subkorpus nie. *Span* kom in die vertaalde subkorpus meer voor en na voornaamwoorde voor (*sy span, hierdie span; span wat, span sy*) as in die nievertaalde subkorpus. So ook kom *span* meer voor voorsetsels (*span van, span in*) in die vertaalde subkorpus voor as in die nievertaalde subkorpus.

### 3.7.3 Oortollige grammatikale items

Om oortollige grammatikale items te identifiseer, kyk ek na die gebruik van toutologie en pleonasme in die onderskeie subkorpora. Carstens (2003:125) verklaar toutologie as *die opeenstapeling van woorde waarvan die betekenis dieselfde is*. Hy gee *kabeltou* (kabel = tou) as een van sy voorbeelde. Pleonasme verduidelik hy as 'n woord of gedagte wat in 'n sekere konstruksie onnodig herhaal word (Carstens, 2003:126). As voorbeeld noem hy onder andere *die skoolhoof van 'n skool*.

Ek kon nie 'n korpusvertaalstudie-metode suksesvol toepas in die bestudering van oortollige grammatikale items in hierdie eentalige vergelykende korpus nie. En hoewel ek elke artikel in die vertaalde en nievertaalde subkorpora deurgelees het, kon ek nie oortollige grammatikale items identifiseer nie. Hierdie verskynsel is waarskynlik toe te skryf aan goeie subredigering om die berigte so ekonomies moontlik te hou.

### 3.8 Samevatting

In hierdie hoofstuk is *WordSmith Tools* gebruik om 'n eentalige vergelykende korpus van koerantberigte te ondersoek vir vertaalkenmerke. Die vertaalkenmerke is ondersoek na aanleiding van Laviosa-Braithwaite (1995) se hipotese-metodologie saam met kenmerke wat deur Bam (2005) vir Afrikaans gevind is.

Die ondersoek is stapsgewys gedoen sodat dit duidelik kon blyk watter funksie van *WordSmith Tools* vir watter kenmerk gebruik is. Die tipe-tekenverhouding, gemiddelde sinlengte en gemiddelde woordlengte asook standaardafwykingdata is deur die program gegenereer as deel van die woordelysfunksie. Die woordelysfunksie is ook gebruik om die lysse volgens frekwensie te sorteer en dan berekeninge volgens die frekwensies van woorde te doen. Die frekwentste woorde is ook volgens lemma saamgegroepeer.

Die inhoud van die eentalige vergelykende korpus is deur middel van 'n outomatiese woordsoortetiketteerder leksikaal geannoteer. Dit was nuttig vir die bestudering van voornaamwoorde, leksikale digtheid en ten opsigte van kollokasieverbindings. Die konkordansiefunksie se identifikasie van kollokasies en patrone is gebruik om die opsionele *dat* en moontlike kollokasiebotsings te bestudeer.

In die volgende hoofstuk word gevolgtrekkings oor die navorsingsdata gemaak en aanbevelings vir verdere navorsing gedoen.

## HOOFSTUK 4 GEVOLGTREKKINGS EN AANBEVELINGS

### 4.1 Inleiding

Hierdie studie het ontstaan weens die teenstrydige bevindinge wat Bam (2005) in sy korpusstudie oor Afrikaanse koerantberigte gekry het teenoor Laviosa (1998) se bevindinge en hipoteses (Laviosa-Braithwaite, 1995) vir Engels (sien 2.5.4).

In hoofstuk 3 is beskryf hoe die hipotese-metodologie en bykomende aspekte wat nie in die metodologie gedek is nie, gebruik is om die data in die vertaalde en nievertaalde subkorpus stapsgewys te vergelyk.

Tabel 4.1 dui die bevindinge van hierdie ondersoek aan.

**Tabel 4.1: Bevindinge van hierdie ondersoek ten opsigte van vertaalde subkorpus (VS) teenoor nievertaalde subkorpus (NS)**

Aspek	Hierdie studie	Laviosa	Bam
<b>tipe-teken-verhouding</b>	VS effens hoër (60,7 teenoor NS 60,6); albei hoog	VS laer	VS hoër (50,5 teenoor NS 49,4); albei hoog
<b>gemiddelde sinlengte</b>	VS korter (18,1) teenoor NS (18,6)	VS korter	VS langer (24,9) teenoor NS 20,9)
<b>gemiddelde woordlengte</b>	VS effens hoër (4,9) teenoor NS 4,8)		VS effens hoër (4,6) teenoor NS 4,5)
<b>konvergensie</b>	Nee vir VS t.o.v. TTV en gemiddelde woord- en sinlengte; maar ook nee vir NS	Ja vir VS t.o.v. TTV, leksikale digtheid & gemiddelde sinlengte)	Ja vir VS t.o.v. TTV
<b>frekwentste teenoor niefrekwente woorde</b>	VS laer	VS hoër	
<b>frekwensie van</b>	VS minder	VS meer	

<b>frekwentste woorde voorkoms in tekste</b>			
<b>aantal lemmas in frekwentste woorde</b>	VS minder	VS minder	
<b>voornaamwoorde</b>	VS effens meer (6,8% teenoor NS 6,7%)	VS minder	
<b>leksikale digtheid</b>	VS hoër (49,4% teenoor NS 48,2%)	VS laer	VS hoër (12,2% teenoor NS 8,6%)
<b>superordinate</b>	Onbeslis	VS meer	
<b>opsionele dat</b>	VS hoër saam met werkwoord <i>gesê</i> en <i>verseker</i> , maar nie saam met <i>verwag</i> nie	VS hoër	
<b>kollokasiebotsings</b>	geen kollokasiebotsings, wel ander kombinasies met <i>span</i> in vertaalde teenoor nievertaalde subkorpus	VS minder	
<b>oortollige grammatikale items</b>	Onbeslis	VS hoër	

## 4.2 Bespreking van bevindinge

Ten opsigte van Bam (2005) se bevindinge vir Afrikaans stem die bevindinge van hierdie ondersoek dus ooreen ten opsigte van tipe-teken-verhouding, gemiddelde woordlengte en leksikale digtheid, maar nie ten opsigte van gemiddelde sinlengte en konvergensie nie.

Die tipe-teken-verhouding is 'n aanduiding van die leksikale kompleksiteit van 'n teks en van die verskeidenheid van woorde wat gebruik word. Hoe hoër die persentasie, hoe meer variasie is daar in die woordeskat. Hoe laer die

persentasie, hoe meer word die woorde in die subkorpus herhaal. In my ondersoek is die tipe-teken-verhouding vir die onderskeie subkorpora feitlik dieselfde, maar hoog. Die vertaalde en nievertaalde berigte gebruik dus dieselfde mate van woordverskeidenheid. Bam bevind dat die tipe-teken-verhouding van die vertaalde berigte effens hoër is as die nievertaalde berigte (50,5 vertaalde tekste en 49,4 nievertaalde tekste). Ek skryf my hoër persentasie (60,7 en 60,6) daaraan toe dat my standaardiseringsbasis (sien 2.4.3) die lengte van my kortste artikel (179) was en Bam se standaardiseringsbasis 500 (Bam, 2005:57). Lengte is nou verwant aan die berekening van die tipe-teken-verhouding. Scott (2008a) beskou TTV egter as slegs 'n rowwe aanduiding vir variasie. Zanettin (2000:113) is van mening dat leksikale digtheid en die hapax legomena-tot-totale aantal woorde saam met die TTV 'n beter aanduiding van variasie is.

Leksikale digtheid is die aanduiding van hoeveel inhoudswoorde (naamwoorde, byvoeglike naamwoorde) in 'n korpus voorkom en gee 'n aanduiding van die inhoudslading van die tekste. Moeiliker tekste se leksikale digtheid is hoog en makliker tekste se leksikale digtheid is laag. Bam (2005:71) vind die leksikale digtheid van die vertaalde tekste hoër (12,2%) as die nievertaalde tekste (8,6%). Hy dui aan dat die vertaalde tekste se inligtingslading te groot is en bevestig dit met 'n leesbaarheidstoets.

In my ondersoek vind ek dat die leksikale digtheid ook hoër is vir die vertaalde subkorpus (49,4%) as vir die nievertaalde subkorpus (48,2%), maar my resultate is aansienlik hoër as Bam se resultate (12,2% teenoor 8,6%), wat dus daarop dui dat die inhoud van die artikels moeilik is. Aangesien Bam (2005) nie 'n uiteensetting gee van watter woordklasse hy as funksie- teenoor inhoudswoorde beskou nie, is dit nie moontlik om 'n redelike verduideliking te gee oor hierdie verskil in bevindinge nie. Ek reken egter dat die gebruikmaking van die outomatiese etiketteerder daartoe sal bydra om foute in die berekening van leksikale digtheid te verminder.

Die verhouding tussen hapax legomena en totale aantal woorde is 12,8% vir die vertaalde tekste en 12,5% vir die nievertaalde tekste. Dus is die tipe-



teken-verhouding, die leksikale digtheid en die hapax legomena hoër vir die vertaalde tekste as vir die nievertaalde tekste. Indien mens Zanettin (2000) se siening in ag neem dat hapax legomena, leksikale digtheid en tipe-teken-verhouding saam as aanduiding van variasie moet beskou, sou mens dus kon sê dat die vertaalde tekste soortgelyk is aan die nievertaalde tekste.

Bam vind die gemiddelde woordlengte hoër in die vertaalde tekste (4,6) as in die nievertaalde tekste (4,5) en skryf dit daaraan toe dat die rykheid van die woordeskat van die vertaalde tekste gelykstaande is met dié van die nievertaalde tekste. Ek sou langer woordlengtes in Afrikaans eerder toeskryf aan die feit dat samestellings vas aan mekaar geskryf word. In Engels is *internet sport betting activities* vier kort woorde, maar in Afrikaans een lang woord, *internetsportweddenskapbedrywighede*. Ek vind ook dat die gemiddelde woordlengte vir die vertaalde tekste effens hoër is as vir die nievertaalde tekste (4,9 teenoor 4,8).

In hierdie ondersoek na Afrikaanse koerantberigte, het die insluiting van ander domeine uit *Die Burger* gelei tot ooreenstemmende bevindinge ten opsigte van tipe-teken-verhouding, gemiddelde woordlengte en leksikale digtheid, maar teenoorgestelde bevindinge ten opsigte van gemiddelde sinlengte en konvergensie as wat Bam (2005) bevind het.

My bevindinge verskil van Laviosa-Braithwaite (1995) se hipotese-metodologie ten opsigte van die tipe-teken-verhouding en frekwensie van frekwentste woorde. Hierdie aspekte lys sy onder vereenvoudiging - die veronderstelde kenmerk dat vertalers tekste eenvoudiger stel as die oorspronklike. Sy veronderstel dat die tipe-teken-verhouding laer sal wees vir vertaalde tekste as vir nievertaalde tekste. In hierdie ondersoek is die tipe-teken-verhouding van die twee subkorpora feitlik dieselfde. Die feitlik dieselfde persentasie tipe-teken-verhouding hang daarmee saam dat vertaalde berigte nie as vertalings uitgeken kan word behalwe deur ekstralinguistiese aanduidings nie.

Ten opsigte van die hipotese oor die omvang van die vertaalde tekste se woordeskat wat nouer is, word die betekenis van 'omvang' afgelei uit haar beskrywing van resultate vir die hipotese, naamlik dat sy gevind het dat daar 'n hoër proporsie hoëfrekwensiewoorde teenoor niefrekwente woorde in die vertaalde tekste voorkom, dat die 108 hoëfrekwensiewoorde 'n groter deel van die korpus uitmaak en dat daardie 108 woorde uit minder lemmas bestaan as in die nievertaalde tekste. Ek bevestig net die laaste bevinding, dat die verhouding hoëfrekwensiewoorde uit minder lemmas bestaan vir die vertaalde tekste, in hierdie ondersoek. Die eerste twee bevindinge is nie waar vir my ondersoek nie. Om daardie rede beweer ek dat die omvang van die woordeskat van my vertaalde subkorpus nie nouer is as dié van die nievertaalde subkorpus nie.

Hipoteses wat verder onder vereenvoudiging gelys word, is die veronderstelde laer gemiddelde sinlengte, laer leksikale digtheid en hoër frekwensie van superordinate. Ek vind wel die gemiddelde sinlengte effens korter, maar die leksikale digtheid is hoër vir die vertaalde tekste as vir die nievertaalde tekste. Die voorkoms van superordinate is egter moeilik om te bepaal met korpusmetodes en selfs met die bestudering van die volledige tekste is die aanduiding van superordinate 'n subjektiewe proses wat moeilik vasgevat word.

Vir die aspek vereenvoudiging soos deur die hipotese-metodologie veronderstel, kry ek dus teenstellende bevindinge. As laer gemiddelde sinlengte 'n aspek is wat op vereenvoudiging dui, dan vind daar vereenvoudiging in my vertaalde subkorpus plaas. As 'n laer tipe-tekenverhouding, 'n laer leksikale digtheid en nouer omvang van woorde aspekte is wat dui op vereenvoudiging, vind daar nie vereenvoudiging in my vertaalde subkorpus plaas nie. Die tekskategorie koerantberigte en die leserskap van die koerant speel duidelik ook 'n rol en om na net die genoemde aspekte in isolasie te kyk, is nie werklik 'n aanduiding van vereenvoudiging nie.

Vir die vertaalkenmerk eksplisitering veronderstel Laviosa-Braithwaite (1995) 'n hoër frekwensie opsionele *dat*, meer oortollige grammatikale items en

minder voornaamwoorde. As hierdie eienskappe wel op eksplisitering - die veronderstelde kenmerk dat vertalers dinge in die bronteks duideliker stel in die vertaalde teks - dui, kan ek dit slegs beaam met die bevinding van die opsionele *dat* vir die werkwoorde *gesê* en *verseker*, maar nie vir *verwag* nie. Die opsionele *dat* word wel meer saam met *verwag* en *verseker* gebruik omdat hierdie werkwoorde semanties ryker is as *gesê* en dus bydra tot die semantiese dominansie van die hoofsin. *Dat* word in Afrikaans weggelaat na 'n semanties minder dominante hoofsin. (Feinauer, 1990:117-119).

Ook ten opsigte van die oortollige grammatikale items (in die vorm van tautologie en pleonasmes) is my bevinding onbeslis. Eksplisitering kom dan bloot in my vertaalde subkorpus voor op grond van die voorkoms van die opsionele *dat* by genoemde werkwoorde in die vertaalde subkorpus.

Laviosa-Braithwaite (1995) gee slegs een hipotese onder normalisering, naamlik dat vertaalde tekste minder kollokasiebotsings sal vertoon. Normalisering is die veronderstelde kenmerk dat vertaalde tekste so 'normaal' moontlik vertaal word. As min of geen kollokasiebotsings wel 'n aanduiding van normalisering is, kom daar normalisering in my vertaalde en ook nievertaalde subkorpus voor, ten opsigte van die naamwoord *span*.

Dat vertaalde koerantberigte nie maklik van nievertaalde koerantberigte onderskei kan word nie, kan gesien word in die bevindinge van hierdie ondersoek ten opsigte van die tipe-teken-verhouding, gemiddelde woordlengte en kollokasiebotsings. In die koerant is dit slegs die ekstralinguistiese merker (die bron onderaan die artikels) wat 'n aanduiding gee van die berig se status as vertaalde teks. Tog is dit so dat daar deur middel van *WordSmith Tools* wel verskeie opgemerk kon word ten opsigte van korter sinne en kombinasiepatrone. Verder is die verhouding van woorde wat baie gebruik word teenoor dié wat nie baie gebruik word nie, laer by vertaalde tekste; kom daardie woorde in vergelyking met die nievertaalde subkorpus in minder tekste voor en het daardie woorde minder lemmas as dié van die nievertaalde tekste.

### 4.3 Bevindinge ten opsigte van hierdie studie se hipoteses

Twee hipoteses is aan die begin van die studie vasgestel, naamlik dat 1) vertaalde taal van nievertaalde taal verskil ten opsigte van leksikale digtheid en die omvang van woordeskat en 2) dat die hipotese-metodologie vir vertaalkenmerke op Afrikaans van toepassing is ten opsigte van die opsionele *dat*, oortollige grammatikale strukture, voornaamwoorde, die tipe-teken-verhouding, gemiddelde sinlengte en kollokasiebotsings.

Vir hipotese een vind ek wel dat vertaalde taal van nievertaalde taal verskil ten opsigte van die hoeveelheid funksie- teenoor inhoudswoorde, die verhouding hoëfrekwensiewoorde teenoor niefrekwente woorde, die hoeveelheid kere wat hoëfrekwensiewoorde in die tekste voorkom en die hoeveelheid lemmas waaruit die hoëfrekwensiewoorde bestaan.

Vir hipotese twee vind ek dat die hipotese-metodologie vir vertaalkenmerke wel van toepassing is op Afrikaans vir die bestudering van die opsionele *dat* in die indirekte rede, die hoeveelheid voornaamwoorde, die tipe-teken-verhouding, die gemiddelde sinlengte en die voorkoms van kollokasiebotsings, hoewel die hipoteses nie noodwendig vir Afrikaans bevestig word nie. Die bestudering van oortollige grammatikale strukture en superordinate deur middel van korpusmetodes vind ek moeilik. Daarom vind ek ten opsigte van laasgenoemde twee aspekte dat die hipotese-metodologie nie sonder aanpassing op Afrikaans van toepassing is nie.

### 4.4 Verdere navorsing

Hoewel die eentalige vergelykende korpus in teorie sin maak, kan die praktiese uitvoerbaarheid daarvan ingevolge vergelykbaarheid problematies wees omdat data wat volkome vergelykbaar is nie altyd beskikbaar is nie.

Tydens en ná afloop van hierdie ondersoek was ek geweldig onder die indruk van hoevéél data *WordSmith Tools* kan onttrek; soveel data, dat mens met tye verlei word deur interessantheide wat ure van jou tyd in beslag neem en

nie van toepassing is op jou ondersoek nie en dat mens deur "bladsye en bladsye" (by gebrek aan 'n visuele elektroniese metafoor) data moet werk vir min toepaslike resultate.

'n Moontlike rede vir bogenoemde negatiewe gesindheid is waarskynlik omdat hierdie studie datagebaseerd benader is met die doel om bestaande bevindinge en teorieë vir Afrikaans te toets. Vir 'n datagedrewe benadering sou die interessantheid in die data dus nie 'n verleiding wees nie, maar die doel van die ondersoek. Die opsionele *dat* by verskillende werkwoorde en woordsoortkombinasies kan deeglik ondersoek word met behulp van geëtiketteerde data.

Die invloed van norme in vertaalwerk sou ondersoek kon word met die eentalige vergelykende korpusbenadering deur byvoorbeeld vertalings van kinderliteratuur in die vorige dekade te vergelyk met meer onlangse vertalings van kinderliteratuur.

Die nut van korpusstudies vir die opleiding van vertaalstudente sou getoets kon word deur 'n groep studente aan die konkordansiefunksie van *WordSmith Tools* bloot te stel en 'n kontrolegroep wat nie blootgestel word daaraan nie (of byvoorbeeld die hele groep bloot te stel aan die konkordansiefunksie vir mediese vertaalwerk, maar nie vir tegniese vertaalwerk nie). Hierdie studente sal 'n soektog kan doen op 'n spesialiskorpus om te kyk met watter woorde 'n spesifieke soekwoord in kollokasies voorkom. Die doel van die ondersoek sal wees om te bepaal of die gehalte van vertaalwerk verbeter vir dié wat blootgestel is aan die konkordansiefunksie teenoor die groep wat nie daaraan blootgestel is nie.

*WordSmith Tools* is 'n korpuslinguistiek-hulpmiddel wat 'n verskeidenheid inligting genereer waarvan die korpusvertaalstudie-navorsers dan moet kies wat hy of sy gebruik. Soos korpusvertaalstudies groei en die fokus van sodanige studies verfyn, sou die ontwikkeling van 'n hulpmiddel spesifiek vir korpusvertaalstudies van nut kon wees. Die fokus sou in die toekoms dalk glad nie meer op die statistiek en woordelyste val vir 'n vergelykende



eentalige studie nie, maar eerder op die omgewings waarin spesifieke soekwoorde voorkom. In *WordSmith Tools* is die eerste resultaat van 'n soektog in die konkordansiefunksie die sleutelwoord-in-konteks, met dan ook 'n opsie om na patrone, kollokasies en ook die volledige teks te kyk. Dalk word die fokus van korpusvertaalstudies so verfyn dat die sleutelwoord-in-konteks nie eens nodig is nie maar net die kollokasie- of patroonontleding en toegang tot die volledige teks. Die ontwikkeling van sodanige hulpmiddel sal afhang van verdere verfyning van die behoeftes van korpusvertaalstudies.

#### **4.5 Samevatting**

Hierdie studie het voortgebou op die werk van Bam (2005) wat as 'n loodsstudie van Afrikaanse koerantberigte deur middel van 'n eentalige vergelykende korpus beskou kan word. Met behulp van 'n uitbreiding van domeine is van Bam se bevindinge bevestig en ander weerlê. Net so met Laviosa-Braithwaite (1995) se hipotese-metodologie: sommige hipoteses is bevestig, ander is weerlê. Hierdie resultaat sluit na my mening aan by Tymozcko (1998:653) se kritiek teen korpusvertaalstudies se soeke na universele kenmerke (sien 2.5.5) en die verwysing na die subjektiwiteit waaraan enige korpusondersoek onderworpe is.

In 2.5.5 het ek ook verwys na Snell-Hornby (2006:150-159) wat die gebruik van rekenaarkorpora as 'n terugbeweeg na linguistiek beskou en wat die blote vergelyking van linguistiese items sonder insig in tekste as geheel betreur. Maar miskien is dit nie soseer regressie, terug na 'n oorspronklike punt soos in 'n sirkel nie, maar eerder 'n aanbeweeg soos in 'n spiraal: terug na linguistiek, ja, maar nou met 'n ander perspektief en ander hulpmiddels. Maar dit is nietemin baie maklik om vas te val in die interessantheid van die leksikale patrone selfs al het mens toegang tot die volledige tekste in die konkordansiefunksie van *WordSmith Tools*. Die voordeel van korpusvertaalstudies is ook die nadeel: 'n mens het toegang tot 'n oormag inligting waarin mens maklik kan verdrink.



## BRONNELYS

### BRONNE

- Agence France-Presse (AFP) [Aanlyn] Beskikbaar: [www.afp.com](http://www.afp.com) [2008, 3 Maart].
- Associated Press (AP). [Aanlyn] Beskikbaar: [www.ap.org](http://www.ap.org) [2008, 3 Maart].
- Baker, M. 1993. Corpus Linguistics and Translation Studies. Implications and Applications. In: Baker, M, Francis, G & Tognini-Bonelli, E (reds.) *Text and Technology In honour of John Sinclair*. Philadelphia: John Benjamins.
- Baker, M. 1995. Corpora in Translation Studies: An Overview and some suggestions for future research. *Target* 7(2):223-243.
- Baker, M. 2004b. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics* 9:2:167-193 [Aanlyn] Beskikbaar: <http://researchschool.org/documents/IJCL%20Paper.pdf> [2008, 25 Januarie].
- Bam, RG. 2005. *A corpus-driven quantitative analysis of translated and original texts in "Die Burger"*. Universiteit Stellenbosch: Ongepubliseerde MPhil-tesis.
- Barlow, M. 2004. Software for corpus access and analysis. In: Sinclair, J (red.) *How to use corpora in language teaching*. Amsterdam: John Benjamins.
- Baroni, M. en Bernardini, S. 2005. A new approach to the study of translationese: machine-learning the difference between original and translated text. *Literary and Linguistic Computing Advance Access* [Aanlyn] Beskikbaar: <http://llc.oxfordjournals.org/cgi/content/full/21/3/259?ijkey=x8FEqHDYhioBOyC&keytype=ref> [2007, 16 Februarie].
- Butler, C. 1985. *Computers in linguistics*. Oxford: Basil Blackwell.
- Carstens, WAM. 2003. *Norme vir Afrikaans*. Vierde uitgawe. Pretoria: Van Schaik.
- Chesterman, A. 1997. *Memes of translation. The spread of ideas in translation theory*. Amsterdam: John Benjamins.

- Chesterman, A. 2000. A Causal Model for Translation Studies. In: Olohan, M (red.) *Intercultural Faultlines. Research Models in Translation Studies I Textual and Cognitive Aspects*. Manchester: St Jerome.
- Chipere, N, Malvern, D, Richards, B en Duran, P. 2004. *Using a corpus of school children's writing to investigate the development of vocabulary diversity*. [Aanlyn] Beskikbaar: <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/chipere.pdf> [2008, 26 Junie].
- De Villiers, M. 1983. *Afrikaanse grammatika vir volwasse nes*. Goodwood: Nasou.
- Deutsche Presse-Agentur (DPA) [Aanlyn] Beskikbaar: [www.dpa.de](http://www.dpa.de) [2008, 3 Maart].
- Die Burger [Aanlyn] Beskikbaar: [www.dieburger.com](http://www.dieburger.com) [2006, 30 Oktober].
- Du Plessis, T. 2000. South Africa: from two to eleven official languages. In: Deprez, K & Du Plessis, T. (reds.) *Multilingualism and government: Studies in language policy in South Africa*. Pretoria: Van Schaik.
- Even-Zohar, I. 1973. The relations between primary and secondary systems in the literary polysystem. In: Even-Zohar, I. 1978. Papers in Historical Poetics. *Papers on Poetics and Semiotics No.8*. Tel Aviv: The Porter Institute for Poetics and Semiotics. [Aanlyn] Beskikbaar: <http://www.tau.ac.il/~itamarez/works/books/php1978.pdf> [2008, 20 Junie].
- Even-Zohar, I. 1976. The position of translated literature within the literary polysystem. In: Even-Zohar, I. 1978. Papers in Historical Poetics. *Papers on Poetics and Semiotics No.8*. Tel Aviv: The Porter Institute for Poetics and Semiotics. [Aanlyn] Beskikbaar: <http://www.tau.ac.il/~itamarez/works/books/php1978.pdf> [2008, 20 Junie].
- Even-Zohar, I. 1978. Papers in Historical Poetics. *Papers on Poetics and Semiotics No.8*. Tel Aviv: The Porter Institute for Poetics and Semiotics. [Aanlyn] Beskikbaar: <http://www.tau.ac.il/~itamarez/works/books/php1978.pdf> [2008, 20 Junie].
- Even-Zohar, I. 2005. Polysystem Theory (revised). In: Even-Zohar, I. *Papers in Culture Research*. Tel Aviv: Porter Chair of Semiotics (Temporary electronic book) [Aanlyn] Beskikbaar:

- <http://www.tau.ac.il/~itamarez/works/papers/papers/ps-revised.pdf> [2008, 29 Januarie].
- Feinauer, AE. 1990. Skoon afhanklike sinne in Afrikaanse spreektaal. *SA Tydskrif vir Taalkunde*. 8(3):116-120.
- Halverson, S. 1998. Translation Studies and representative corpora: establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta* 43 (4) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/1998/v43/n4/003000ar.html> [2008, 29 Januarie].
- Hermans, T. 1996. Norms and the determination of translation. A theoretical framework. In: *Translation, Power, Subversion. Topics in Translation*. Clevedon: Multilingual matters. [Aanlyn] Beskikbaar: [http://eprints.ucl.ac.uk/2005/1/Microsoft\\_Word\\_-\\_96\\_Norms\\_Determination.pdf](http://eprints.ucl.ac.uk/2005/1/Microsoft_Word_-_96_Norms_Determination.pdf) [2008, 25 Januarie].
- Hermans, T. 1999a. Translation and Normativity. In: *Translation and Norms. Current Issues in Language and Society*. [Aanlyn] Beskikbaar: [http://eprints.ucl.ac.uk/2006/1/Microsoft-Word\\_-\\_98\\_Aston\\_Norms.pdf](http://eprints.ucl.ac.uk/2006/1/Microsoft-Word_-_98_Aston_Norms.pdf) [2008, 25 Januarie].
- Hermans, T. 1999b. *Translation in Systems. Descriptive and System-oriented approaches explained*. Manchester: St Jerome.
- Heugh, K. 1995. Disabling and enabling: implications of language policy trends in South Africa. In: Mesthrie, R (red.) *Language and Social History: Studies in South African Sociolinguistics*. Kaapstad: David Philip.
- Hodges, P. s.d. Compare and contrast two theoretical approaches to translation. *TranslationDirectory.com* [Aanlyn] Beskikbaar: <http://www.translationdirectory.com/articles/article1542.php> [2008,9 Mei].
- Kang, J-H. 2007. Recontextualization of News Discourse. A case study of translation of news discourse on North Korea. *The Translator* 13 (2):219-242.
- Kennedy, G. 1998. *An introduction to corpus linguistics*. Londen: Longman.
- Kenny, D. 1998a. Corpora in Translation Studies. In: Baker, M (red.) *Routledge Encyclopedia of Translation Studies*. Londen: Routledge.
- Kenny, D. 1998b. Creatures of Habit? What translators usually do with words. *Meta* 43 (4) [Aanlyn] Beskikbaar:

- <http://www.erudit.org/revue/meta/1998/v43/n4/003302ar.html> [2008, 29 Januarie].
- Kenny, D. 2000. Lexical Hide-and-Seek: Looking for creativity in a parallel corpus. In: Olohan, M (red.). *Intercultural Faultlines. Research Models in Translation Studies I Textual and Cognitive Aspects*. Manchester: St Jerome.
- Kenny, D. 2001. *Lexis and Creativity in Translation. A corpus-based study*. Manchester: St Jerome.
- Kruger, A. 2002. Corpus-based Research: its development and implications for general, literary and Bible Translation. In: Naudé, JA & Van der Merwe, CHJ (reds.) *Contemporary Translation Studies and Bible Translation: A South African Perspective. Acta Theologica*. Supplement 2:70-107.
- Laviosa, S. 1997. How comparable can 'comparable corpora' be? *Target* 9(2):289-319.
- Laviosa, S. 1998a. The Corpus-based approach: A new paradigm in Translation Studies. *Meta* 43 (3) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/1998/v43/003424ar.html> [2008, 25 Januarie].
- Laviosa, S. 1998b. Core patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* 43 (4) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/1998/v43/003425ar.pdf> [2007, 16 April].
- Laviosa, S. 2002. *Corpus-based Translation Studies. Theory, findings, applications*. Amsterdam: Rodopi.
- Laviosa, S. 2004. Corpus-based translation studies: where does it come from? Where is it going? *Language matters* 35(1):6-18.
- Laviosa-Braithwaite, S. 1995. Comparable corpora: towards a corpus linguistic methodology for the empirical study of translation. In: Thelen, M & Lewandowska-Tomaszczyk, B (reds.). *Translation and Meaning Part 3*. Maastricht: UPM:153-163.
- Lind, S. 2007. Translation Universals (or laws, or tendencies, or ...?). *TIC Talk* 63 United Bible Societies [Aanlyn] Beskikbaar: [http://www.ubs-translations.org/tic\\_talk\\_bibliographies/past\\_issues/tic\\_talk\\_63\\_2007/](http://www.ubs-translations.org/tic_talk_bibliographies/past_issues/tic_talk_63_2007/) [2008, 29 Januarie].

- Malmkjær, K. 1998. Love thy neighbour: will parallel corpora endear linguists to translators? *Meta* 43 (4) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/1998/v43/n4/003545ar.html> [2008, 29 Januarie].
- Mauranen, A. 2000. Stange strings in translated language. A study on Corpora. In: Olohan, M. (red.). *Intercultural faultlines. Research models in Translation Studies I Textual and Cognitive aspects*. Manchester: St Jerome.
- McEnery, T, Xiao, R en Tono, Y. 2006. *Corpus-based Language Studies: an advanced resource book*. Londen: Routledge.
- Munday, J. 1998. A computer-assisted approach to the analysis of translation shifts. *Meta* 43 (4) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/1998/v43/n4/003680ar.html> [2008, 25 Januarie].
- Munday, J. 2001. *Introducing Translation Studies. Theories and Applications*. Londen: Routledge.
- Naudé, JA. 2003. Psalm 1-50 in Afrikaans. 'n Korpusgebaseerde vertaalkundige analise. *Acta Theologica* 1: 83-114.
- New Zealand Press Association (NZPA) [Aanlyn] Beskikbaar: [www.nzpa.co.nz](http://www.nzpa.co.nz) [2008, 3 Maart].
- Noord-Wes Universiteit. 2006. *Navorsingshoogtepunte vir CText*. [Aanlyn] Beskikbaar: [http://ctext.nwu.ac.za/Nuus\\_Mediaverklaring\\_Navorsingshoogtepunte\\_2006.html](http://ctext.nwu.ac.za/Nuus_Mediaverklaring_Navorsingshoogtepunte_2006.html) [2008. 14 Maart].
- Noord-Wes Universiteit. 2008. *A computer world opens up for Afrikaans*. [Aanlyn] Beskikbaar: [http://www.puk.ac.za/opencms/export/PUK/html/nuus/nuus254\\_e.html](http://www.puk.ac.za/opencms/export/PUK/html/nuus/nuus254_e.html) [2008, 13 Maart].
- O'Connor, M. 2008. Re: Vraag oor name bo of onder artikels, E-pos aan D. Roos [Aanlyn], 1 Oktober. Beskikbaar: [dbnred@dieburger.com](mailto:dbnred@dieburger.com).
- Olohan, M. 2004. *Introducing corpora in Translation Studies*. Routledge: Londen.
- Øverås, L. 1998. In search of the third code: an investigation of norms in literary translation. *Meta* 43 (4) [Aanlyn] Beskikbaar:



- <http://www.erudit.org/revue/meta/1998/v43/n4100377sar.html> [2008, 25 Januarie].
- Paloposki, O. 2001. Enriching translations, simplified language? An alternative viewpoint to lexical simplification. *Target* 13(2):265-288.
- Partington, A. 2001. Corpora and their uses in language research. In: Aston, G (red.). *Learning with corpora*. Bologna: Clueb.
- Pilon, S. 2005. *Outomatiese Afrikaanse Woordsoortetikettering*. Ongepubliseerde MA-verhandeling. Potchefstroom: Noordwes-Universiteit.
- Prabhakaran, V. 2001. *Monolingualism in a multilingual and multicultural context: The case of Indians in South Africa*. Summary Report. PANSALB Occasional Papers No. 7. Pretoria: PANSALB.
- Pym, A. 2007. *On Toury's laws of how translators translate*. [Aanlyn] Beskikbaar: [http://www.tinet.org/~apym/online/translation/2007\\_Toury\\_laws.pdf](http://www.tinet.org/~apym/online/translation/2007_Toury_laws.pdf) [2008, 29 Januarie].
- Reah, D. 1998. *The language of newspapers*. Londen: Routledge.
- Reppen, R; Fitzmaurice, SM & Biber, D. 2002. *Using corpora to explore linguistic variation*. Amsterdam: John Benjamins.
- Reuters [Aanlyn] Beskikbaar: [www.reuters.com](http://www.reuters.com) [2008, 3 Maart].
- Roberts, RP. 1996. *Dictionaries, Dictionary-Making and Translation*. [Aanlyn] Beskikbaar: [www.dico.uottawa.ca/articles-en.htm](http://www.dico.uottawa.ca/articles-en.htm) [2008, 5 Januarie].
- Scott, M. 2008a. *WordSmith Tools 4.0 On-line Manual*. [Aanlyn] Beskikbaar: <http://www.lexically.net/downloads/version4/html/index.html> [2008, 26 Junie].
- Scott, M. 2008b. *WordSmith Tools Discussion Group*. [Aanlyn] Beskikbaar: <http://groups.google.com/group/WordSmithTools> [onderwerp: 'Standard deviation generated by WordSmith Tools'] [2008, 15 Augustus].
- Snell-Hornby, M. 2006. *The turns of Translation Studies. New paradigms or shifting viewpoints?* Amsterdam: John Benjamins.
- South African Press Association [Aanlyn] Beskikbaar: [www.sapa.co.za](http://www.sapa.co.za) [2008, 3 Maart].
- Teubert, W en Cermáková, A. 2007. *Corpus Linguistics. A short introduction*. Londen: Continuum.



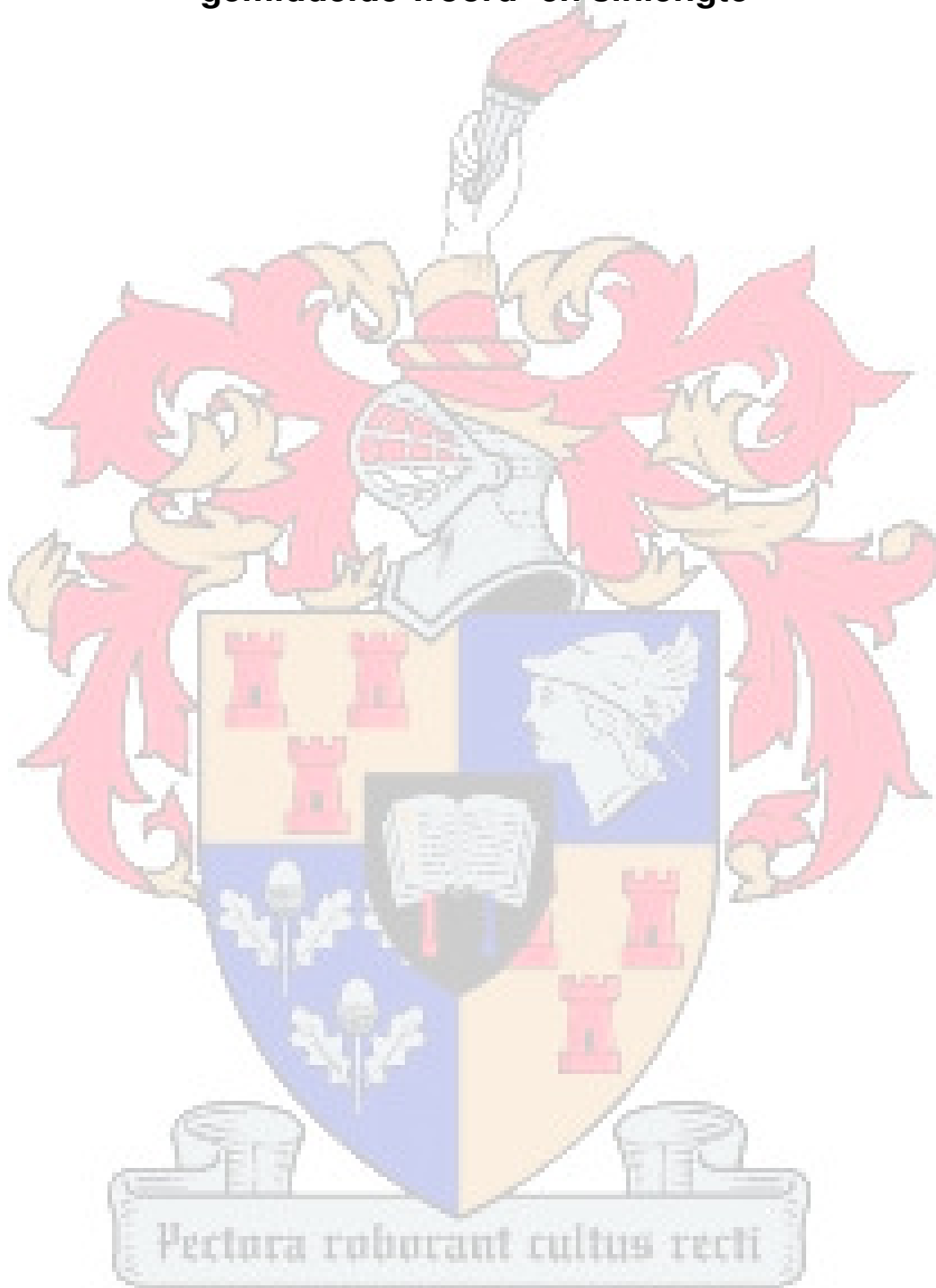
- Tirkkonen-Condit, S. 2002. Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14(2):207-220.
- Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Toury, G. 1978. The nature and role of norms in literary translation. In: Holmes, JS *et al.* (reds.). *Literature and Translation: New perspectives in Literary Studies*. Leuven: Acco. (Revised version in Toury 1995:53-69).
- Toury, G. 1980. *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.
- Toury, G. 1995. *Descriptive Translation Studies and beyond*. Amsterdam: Benjamins.
- Toury, G. 1998. A Handful of Paragraphs on 'Translation' and 'Norms'. *Current Issues in Language & Society* 5(1&2) [Aanlyn] Beskikbaar: <http://www.tau.ac.il/~toury/works/gt-tr&no.htm> [2007, 5 Oktober].
- Tymozcko, M. 1998. Computerized corpora and the future of translation studies. *Meta* 43(4):652-660.
- Ulrych, M en Bollettieri Bosinelli, RM. 1999. The State of the Art in Translation Studies: An Overview. *Textus English studies in Italy*. XII(2):219-242.
- Van Doorslaer, L. 1995. Quantitative and Qualitative Aspects of Corpus Selection in Translation Studies. *Target* 7(2):245-260.
- Van Rooyen, M. 2005. *'n Deskriptiewe analise van SAPA-nuus soos vertaal deur OFM*. Ongepubliseerde MA-skripsie. Bloemfontein: Universiteit van die Vrystaat.
- Venuti, L. 2005. Translation, History, Narrative. *Meta* 50 (3) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/2005/v50/n3/011597ar.pdf> [2008, 9 April].
- Wehrmeyer, E. 2004. CTS and Bible Translation: A study in belling the cat? *Language Matters* 35:214-312.
- Zanettin, F. 2000. Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. In: Olohan, M (red.). *Intercultural Faultlines. Research Models in Translation Studies I Textual and Cognitive Aspects*. Manchester: St Jerome: 105-118.

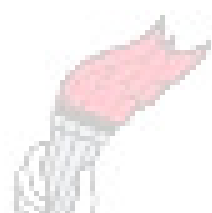
## ANDER BRONNE GERAADPLEEG

- Baker, M. 2004a. The treatment of variation in corpus-based translation studies. *Language matters* 35(1): 28-35.
- Bedeker, LM. 2004. *Interferensie in die Afrikaanse Harry Potter*. Ongepubliseerde MPhil-tesis. Stellenbosch: Universiteit Stellenbosch.
- Blum-Kulka, S en Levenston, EA. 1983. Universals of lexical simplification. In: Faerch, C en Kasper, G. *Strategies in Interlanguage Communication*. Londen: Longman.
- Carter, R, Goddard, A, Reah, D, Sanger, K en Bowring, M. 2001. *Working with Texts. A core introduction to language analysis*. Londen: Routledge.
- Chesterman, A. 2001. Empirical research methods in Translation Studies. *Erikoiskielet ja käännösteoria* (VAKKI-symposiumi XX) 27 [Aanlyn] Beskikbaar: <http://www.helsinki.fi/~chester/20001Vakki.html> [2008, 25 Januarie].
- De Mönnink, I. 1997. Combining corpus and experimental data: methodological considerations. ACHALLC'97. [Aanlyn] Beskikbaar: <http://www.cs.queensu.ca/achallc97/papers/p005.html> [2007, 5 Oktober].
- Gentzler, E. 2001. *Contemporary Translation Theories*. Clevedon: Multilingual Matters.
- Hansen, S en Teich, E. s.d. *Multi-layer analysis of translation corpora: methodological issues and practical implications*. [Aanlyn] Beskikbaar: <http://www.aramaicnt.com/files/Source-Translation%20Analysis.pdf> [2008, 29 Januarie].
- Helgegren, S. 2005. *Tracing Translation Universals and Translator Development by Word Aligning a Harry Potter Corpus*. Undergraduate thesis, Linköping University: Sweden [Aanlyn] Beskikbaar: [http://www.diva-portal.org/diva/getDocument?urn\\_nbn\\_se\\_liu\\_diva-4579-1\\_fulltext.pdf](http://www.diva-portal.org/diva/getDocument?urn_nbn_se_liu_diva-4579-1_fulltext.pdf) [2008, 29 Januarie].
- Klaudy, K. 1993. On Explication Hypothesis. *Transfere necesse est. Current issues of translation theory*. Acta Germanistica: Savariensia Szombathely.
- Laviosa, S. 2003. Corpora and the translator. In: Somers, H (red.). *Computers and translation. A translator's guide*. Amsterdam: John Benjamins.

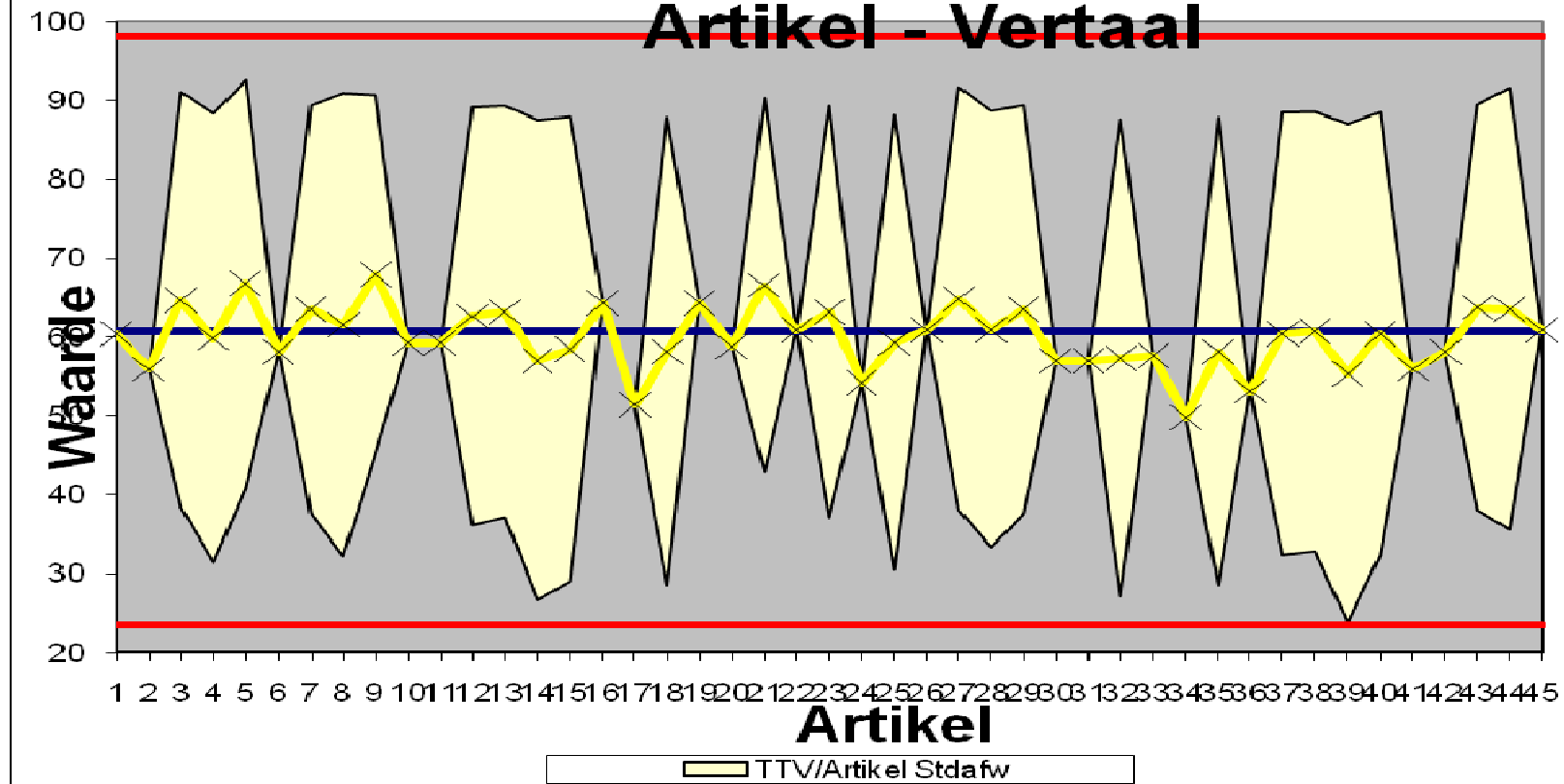
- Lewandowska-Tomaszczyk, B. 1995. Language Corpora and the translator. In: Thelen, M & Lewandowska-Tomaszczyk, B (reds.). *Translation and Meaning Part 3*. Maastricht: UPM.
- Lourens, A. 2007. *Scientific Writing Skills*. Stellenbosch: SUN PReSS.
- Machniewski, M. 2004. Contrastive functional analysis as a starting point for corpus-based translation studies (CTS): Methodological considerations for analysing small translational corpora. *Language matters* 35 (1):102-118.
- Meyer, PS, Swanepoel, WJ, Kempen, W en Pretorius, ZJ. 1984. *Ons Moedertaal*. Goodwood: Nasou.
- Olohan, M. 2002. Comparable corpora in translation research: overview of recent analyses using the Translational English Corpus. In: Yuste, E (red.). *Language Resources and the Language Professional. LREC Workshop 8 Proceedings* [Aanlyn] Beskikbaar: [www.ifi.unizh.ch/cl/yuste/postworkshop/repository/proceedings.pdf](http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/proceedings.pdf) [2008, 11 Februarie].
- Olohan, M. 2003. How frequent are the contractions? A study of contracted forms in the Translational English Corpus. *Target* 15(1):59-89.
- Pym, A. 1998. *Method in Translation History*. Manchester: St Jerome.
- Pym, A. 2005. Explaining Explicitation. Draft paper to be published in: Károly, K (red.). *New Trends in Translation Studies. In Honour of Kinga Klauy*. [Aanlyn] Beskikbaar: [http://www.tinet.org/~pym/online/translation/explicitation\\_web.pdf](http://www.tinet.org/~pym/online/translation/explicitation_web.pdf) [2008, 29 Januarie].
- Qvale, P. 1998. *From St. Jerome to Hypertexts. Translation in Theory and Practice*. Manchester: St Jerome.
- Stewart, D. 2000. Conventionality, creativity and translated text: the implications of electronic corpora in translation. In: Olohan, M (red.). *Intercultural Faultlines Research Models in Translation Studies. Textual and Cognitive Aspects*. Manchester: St Jerome.
- Tymozcko, M. 2005. Trajectories of Research in Translation Studies. *Meta* 50(4) [Aanlyn] Beskikbaar: <http://www.erudit.org/revue/meta/2005/v50/n4/012062ar.html> [2008, 9 April].

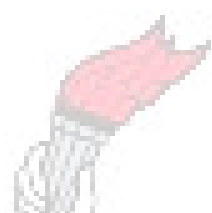
**BYLAAG A**  
**Grafieke vir standaardafwyking van TTV en**  
**gemiddelde woord- en sinlengte**



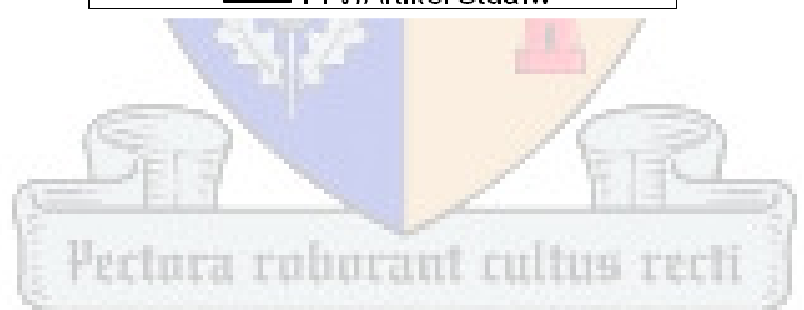
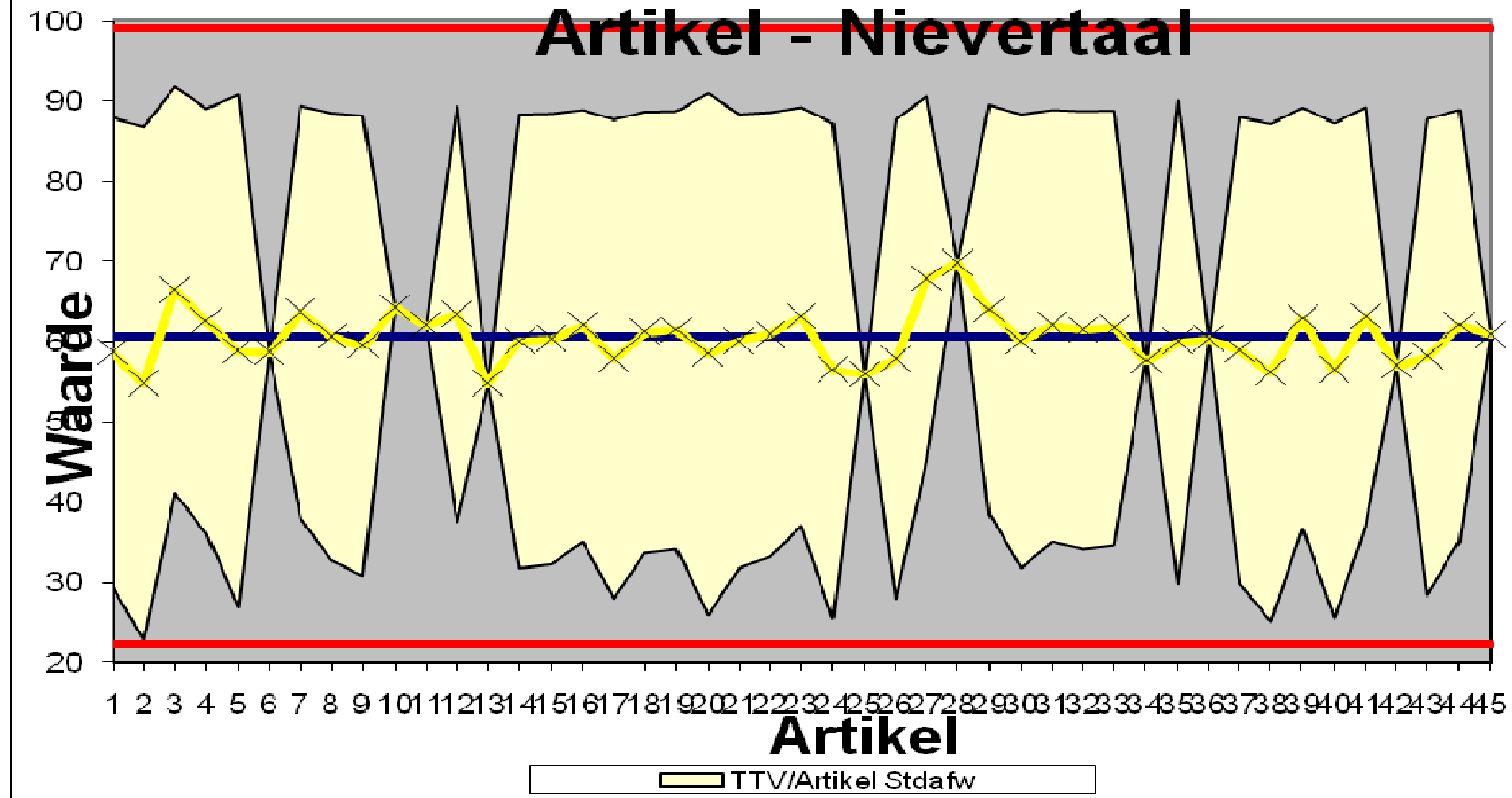


## Gestandaardiseerde TTV per Artikel - Vertaal

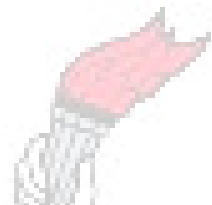




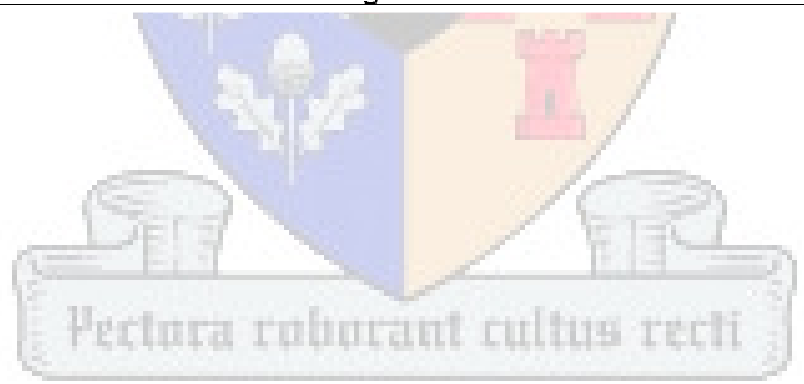
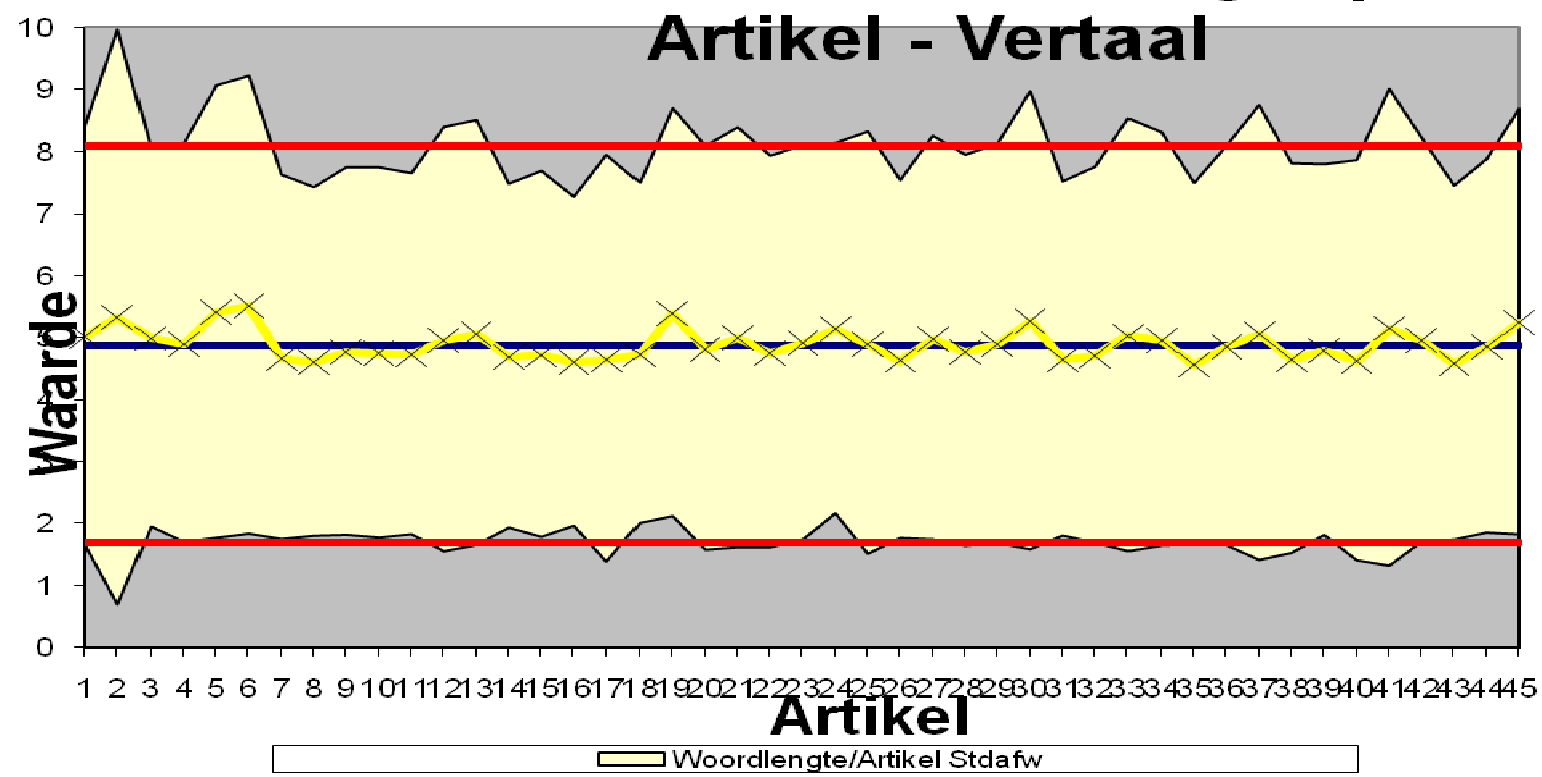
## Gestandaardiseerde TTV per Artikel - Nievertaal

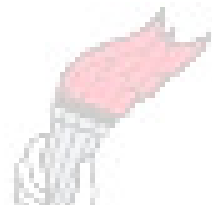




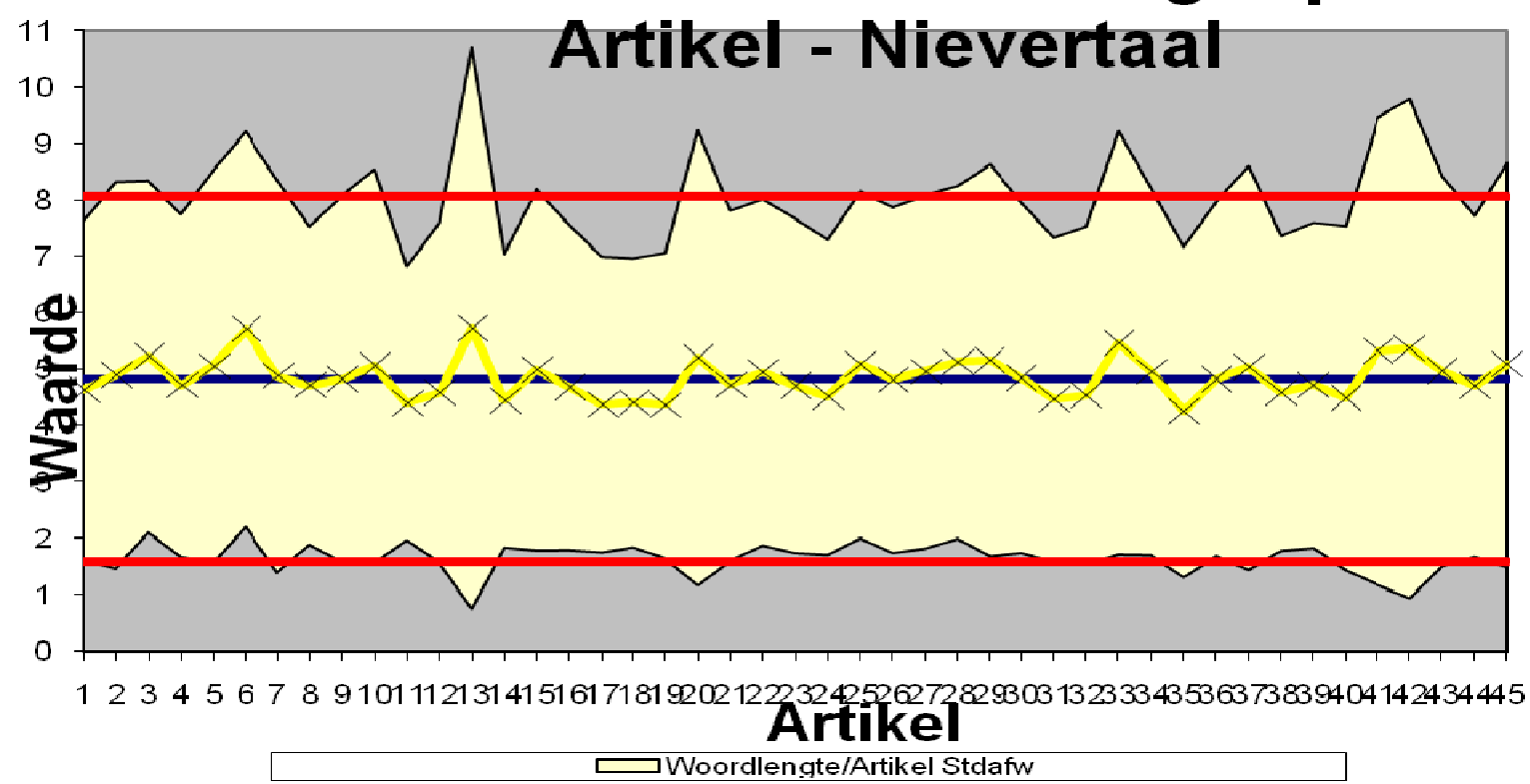


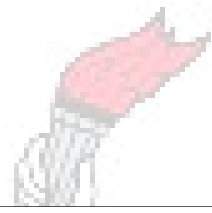
## Gemiddelde Woordlengte per Artikel - Vertaal



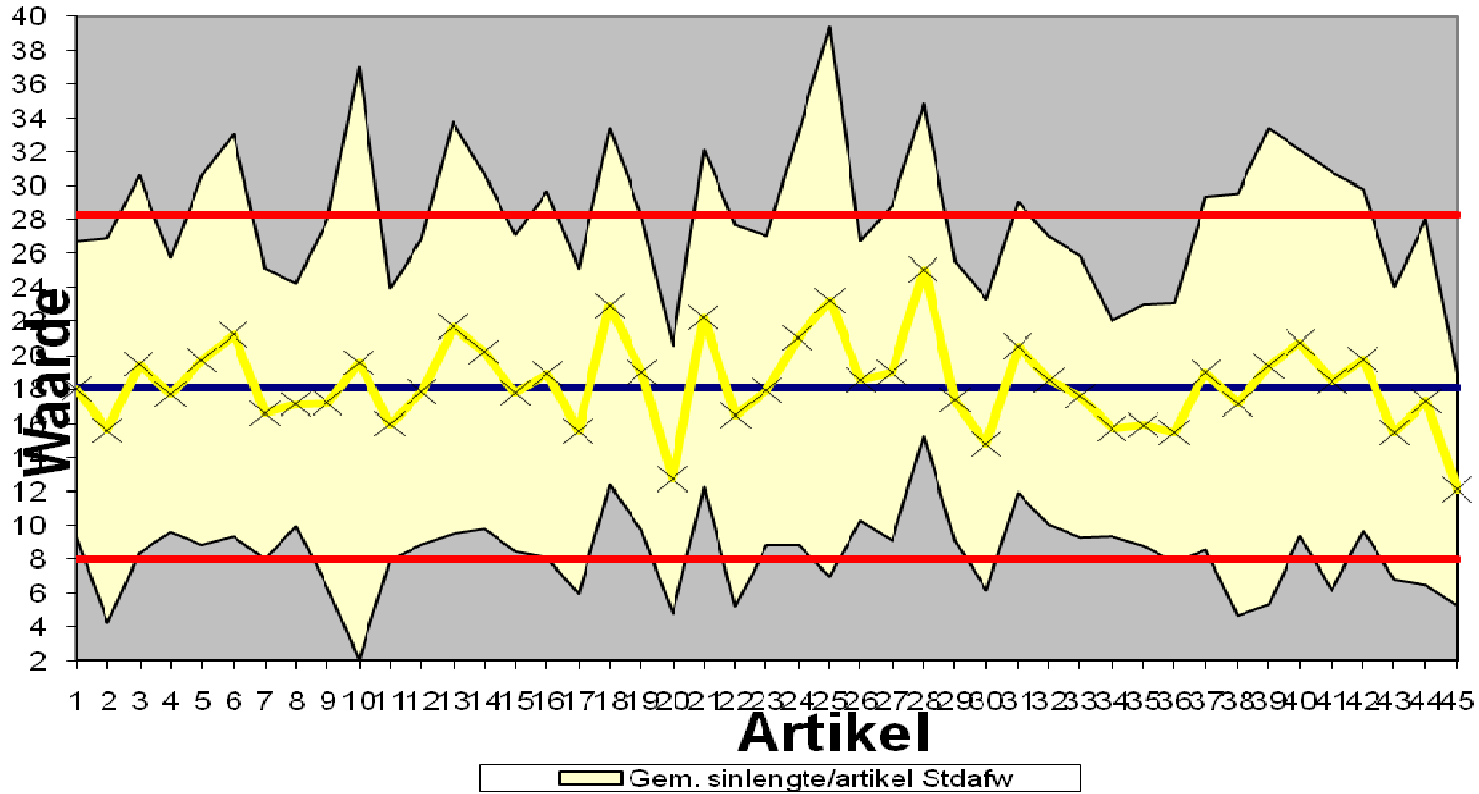


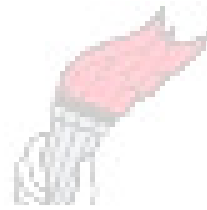
## Gemiddelde Woordlengte per Artikel - Nievertaal



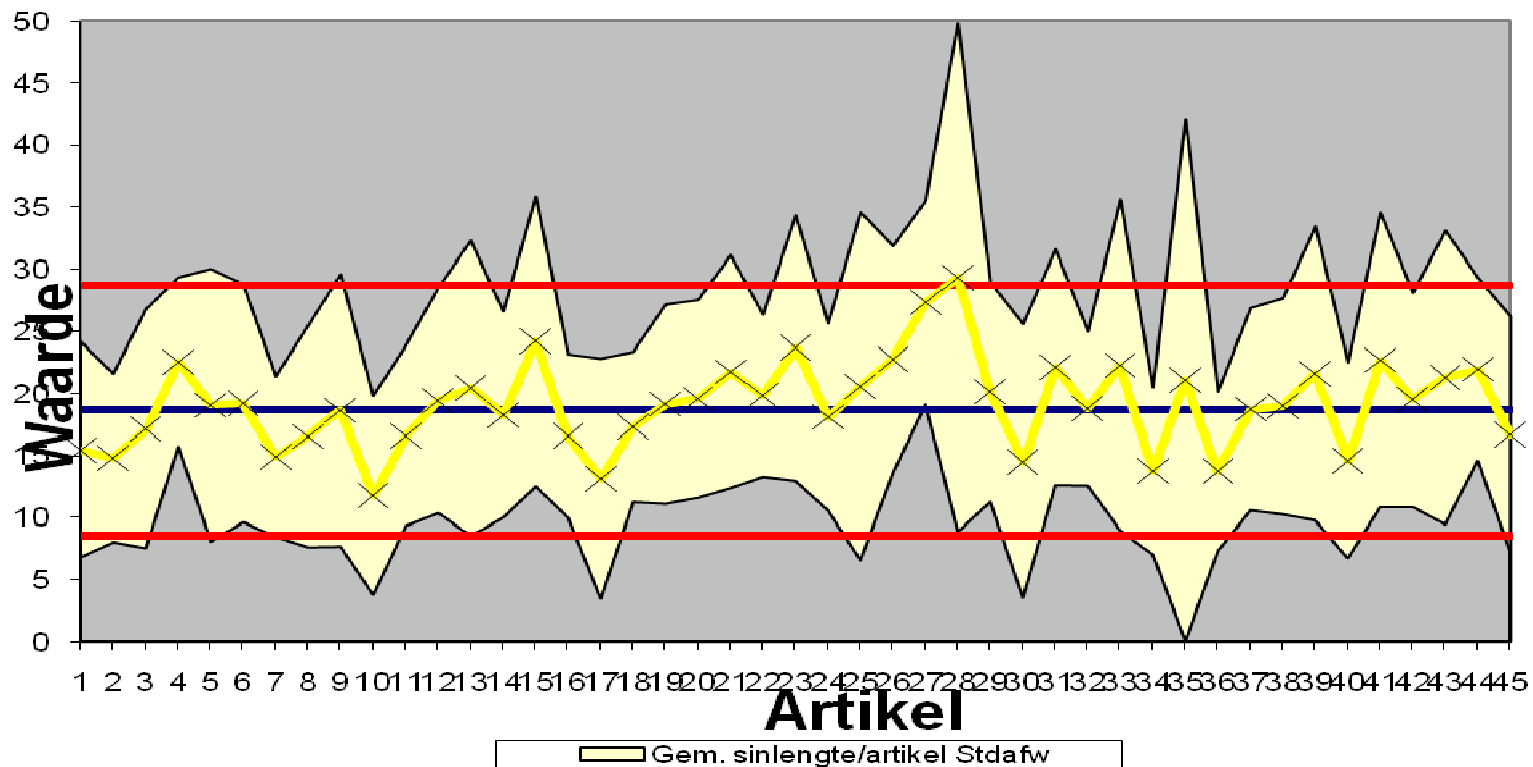


## Gemiddelde sinlengte - Vertaal





## Gemiddelde sinlengte - Nievertaal



## BYLAAG B

### Lemmatisering van subkorpora

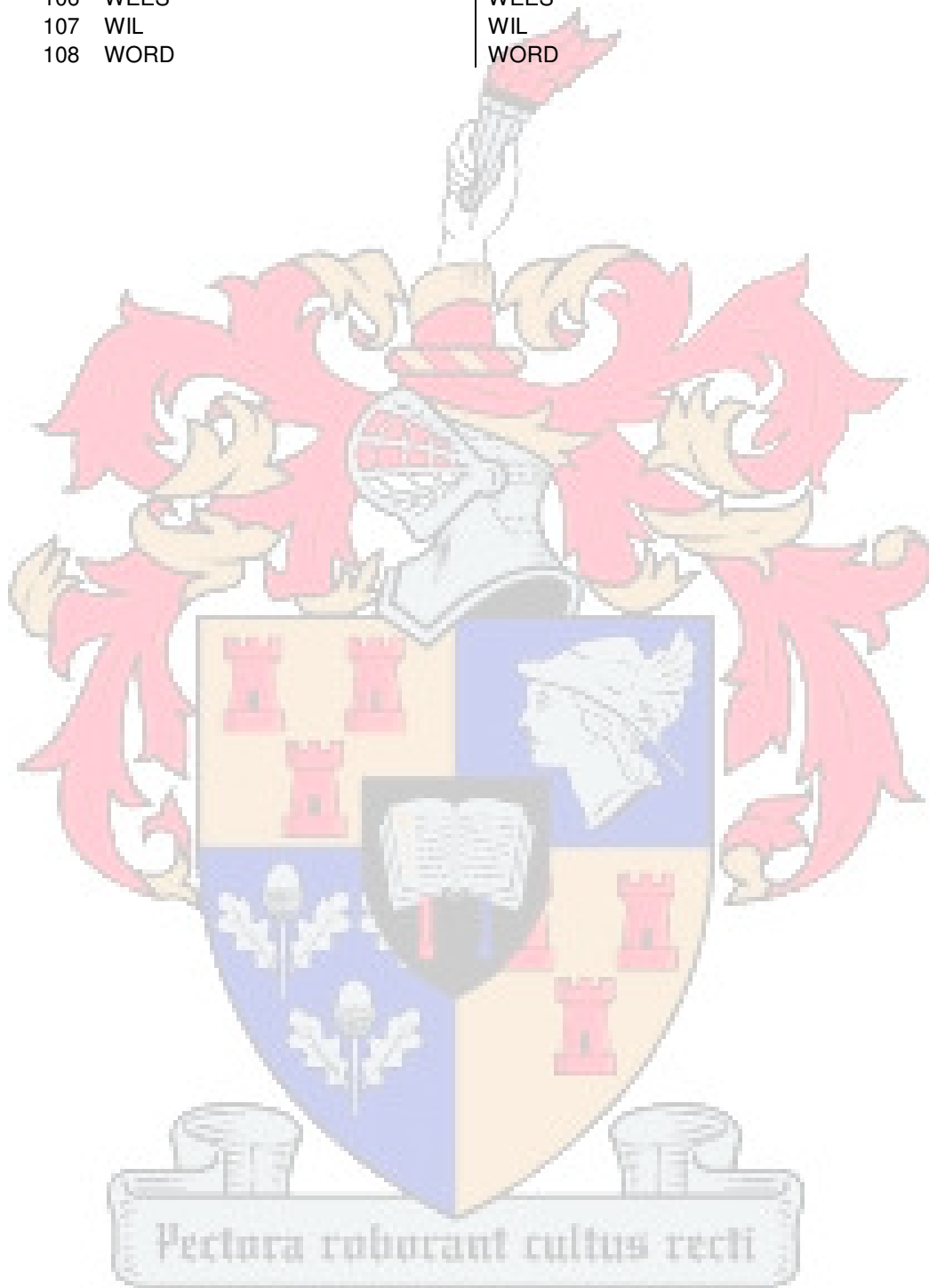
VERTAAL	NIEVERTAAL
1 #	#
2 AAN	AAN
3 AFRIGTER	AFRIGTER
4 AL	AL
5 ALL	ANDER
6 AMERIKAANSE	AS
7 ANDER	BAIE
8 AS	BEGIN
9 AUSTRALIË	BY
10 BAIE	DAAR
11 BEGIN	DAT
12 BLACKS	DE
13 BY	DEUR
14 DAAR	DIE 'n, dié
15 DAT	DIÉ
16 DEUR	DIT
17 DIE 'n, dié	DRIE
18 DIÉ	EEN eerste
19 DIT	EERSTE
20 DRIE	EGTER
21 EEN eerste	EK my
22 EERSTE	EN
23 EGTER	GAAN
24 EK my	GESÊ
25 EN	GISTER
26 ENGELAND Engelse	GOED goeie
27 ENGELSE	GOEIE
28 GAAN	GROEP
29 GEMAAK	GROOT
30 GESÊ	HET
31 GISTER	HIER daar
32 GOED	HOM
33 GROEP	HUL
34 GROOT	HULLE hul
35 HET	HÛ
36 HIER	IN
37 HOM	INDIË
38 HOOF	IS
39 HUL	ISRAEL
40 HULLE hul	JAAR
41 HÛ	KAN
42 IN	KEER
43 IRAN	LAASTE
44 IS	LAND
45 JAAR	LANKA
46 KAN kon	MAAK
47 KAPTEIN	MAAR
48 KEER	MEER

49	KON				MET
50	LAASTE				MILJOEN
51	LAAT				MNR
52	MAAR				MOET
53	MEER				MY
54	MENSE				N
55	MET				NA
56	MNR				NÁ
57	MOET				NADAT
58	MY				ná
59	N				NET
60	NA				NIE
61	NÁ				NOG
62	NADAT		ná		NOU
63	NET				NUWE
64	NIE				OF
65	NOG				OM
66	OF				ONDER
67	OM				ONS
68	OMDAT				OOK
69	ONDER				OOR
70	ONS				OP
71	OOK				PLEK
72	OOR				PROTEAS
73	OP				REEKS
74	PAKISTAN				SA
75	REEDS				SAL
76	REEKS				SANCINO
77	SAL		sou		SE
78	SAPA-AFP				SÊ
79	SE				gesê
80	SEDERT				SEISOEN
81	SOU				SOOS
82	SPAN				SOWAT
83	SPEEL				SPAN
84	SUID-AFRIKA				SPEEL
					SPELERS
					spelers
					SRI
85	SY		hy		SUID-AFRIKA
					Suid-
					Afrikaanse, SA
					SUID-
					AFRIKAANSE
					SY
					hy
86	TE				TANS
87	TEEN				TE
88	TOE				TEEN
89	TOETS				TOE
90	TOETSE				TOERNOOI
91	TOT				TOT
92	TUSSEN				TUSSEN
93	TWEE				TWEE
94	UIT				UIT
95	UNITED				VAN
96	VAN				VANJAAR
97	VIR				VERLEDE
98	VOLGENS				VIR
99	VOOR				VOLGENDE
100	WALLABIES				
101	WAS				



102 WAT  
103 WEDSTRYD      wedstryde  
104 ~~WEDSTRYDE~~  
105 WEER  
106 WEES  
107 WIL  
108 WORD

WAS  
WAT  
WEDSTRYD  
WEER  
WEES  
WIL  
WORD



# BYLAAG C

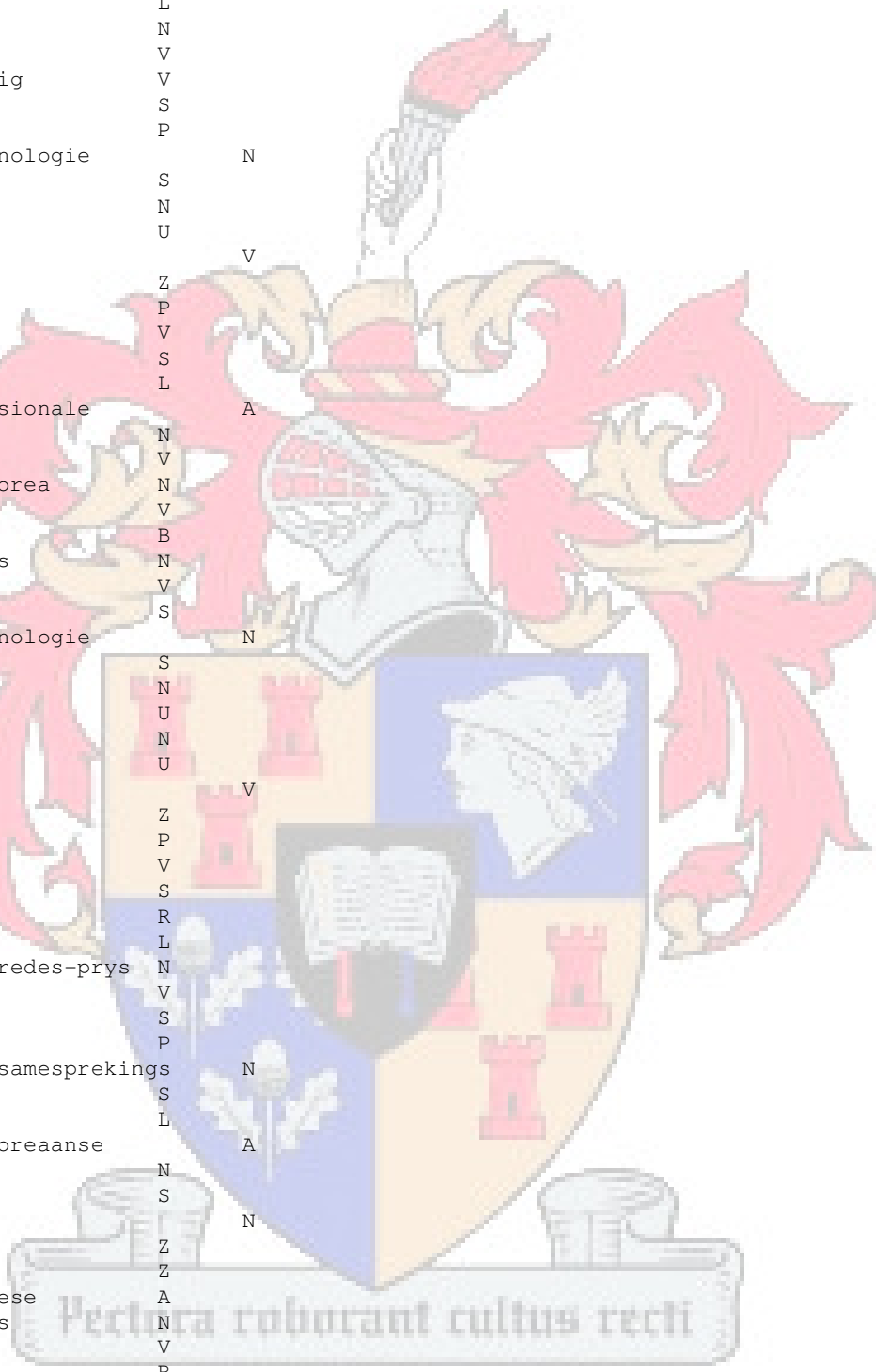
## Artikel wat deur outomatiese woordsoortetiketteerder deur CText bewerk is

### Kodes:

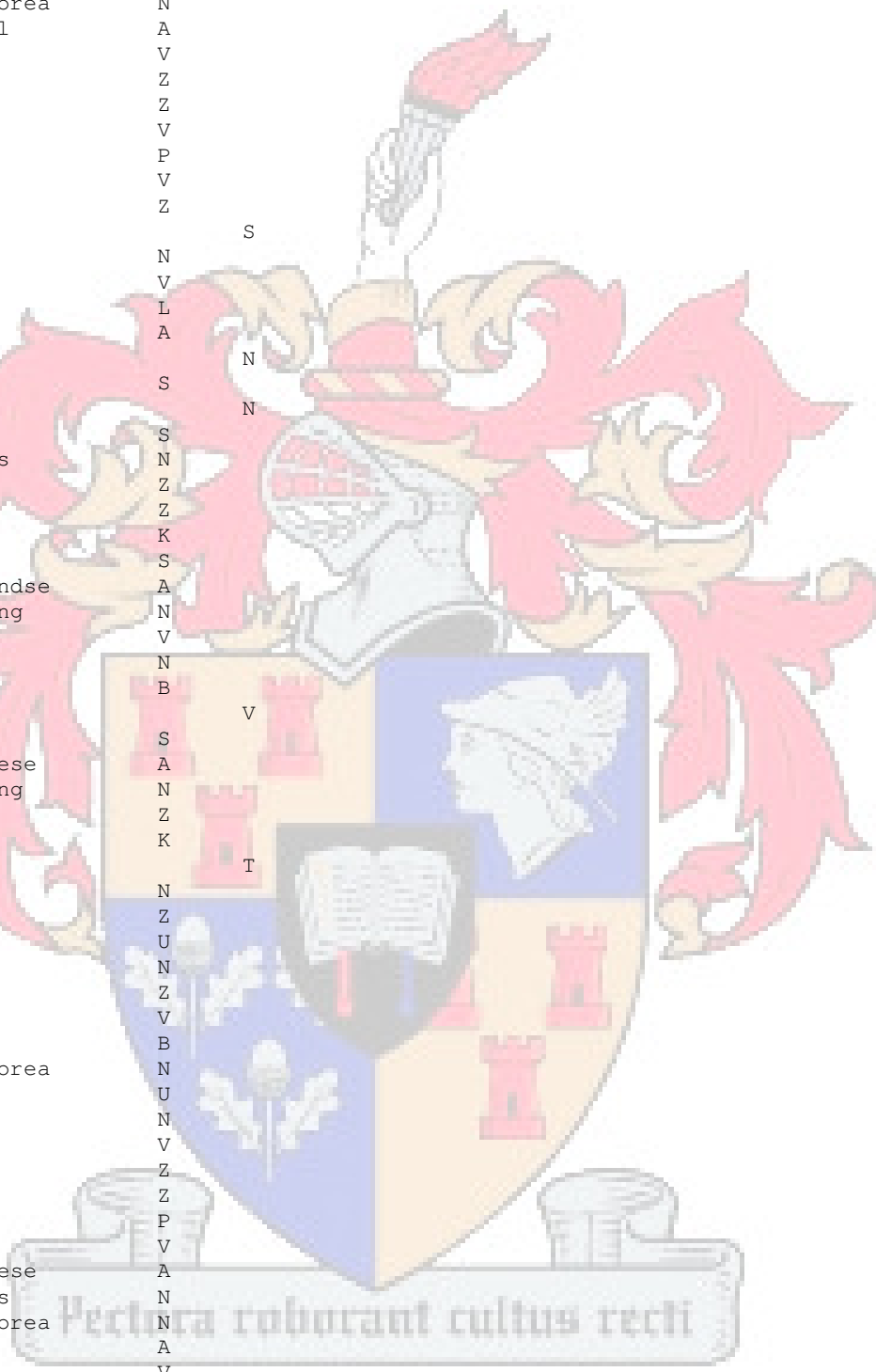
naamwoord - N  
werkwoord - V  
adjektief - A  
voornaamwoord - P  
bywoord - B  
telwoord - T  
lidwoord - L  
setstel - S  
tussenwerpsel - W  
voegwoord - K  
residu- R  
uniek - U  
puntuasie - Z

```
% InT statistically tagged file, Mon Apr 14 13:24:16 2008
% lexicon      : Afrikaans13.lex
% ngrams      : Afrikaans13.123
% corpus      : 97-2006-10-19-buiteland-vert-N-Korea.txt
% model       : trigrams
% sparse data : linear interpolation
% lambda1 = 1.714790e-01  lambda2 = 2.813135e-01  lambda3 =
5.472075e-01
% unknown mode: statistics of singletons
% case of characters is significant
% using suffix trie up to length 10
% suffix backoff with theta = 1.464597e-01
% Thorsten Brants, thorsten@brants.net
Straf      N
'          Z
sal        V
N-Korea    N
in         S
Iran       N
se         U
arms       N
dryf       V
'          Z
SEOEL      N
.          Z
-          R
Oudpres.   R
Kim        R
Dae-Djoeng N
van        S
Suid-Korea N
het        V
gister     B
gewaarsku  V
dat        K
VN-sanksies N
teen       S
Noord-Korea N
tot        S
```

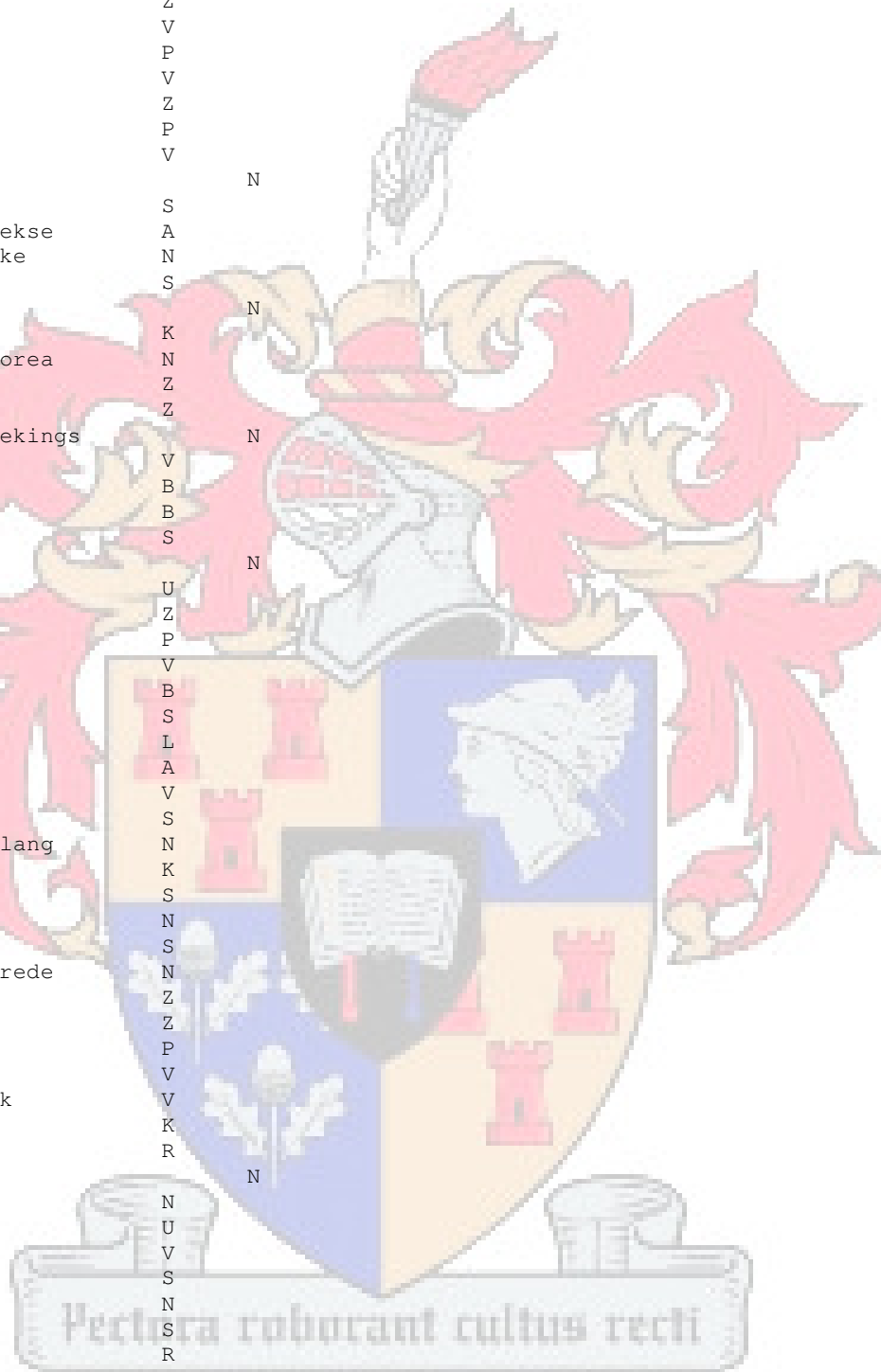
ernstige	A	
botsings	N	
kan	V	
lei	V	
en	K	
die	L	
land	N	
kan	V	
aanmoedig	V	
om	S	
sy	P	
kerntegnologie		N
aan	S	
Iran	N	
te	U	
verkoop		V
.	Z	
Kim	P	
het	V	
aan	S	
'n	L	
internasionale		A
forum	N	
gesê	V	
Noord-Korea	N	
kan	V	
dalk	B	
sanksies	N	
omseil	V	
deur	S	
kerntegnologie		N
aan	S	
lande	N	
soos	U	
Iran	N	
te	U	
verkoop		V
.	Z	
Kim	P	
het	V	
in	S	
2000	R	
die	L	
Nobel-vredes-prys	N	
gewen	V	
ná	S	
sy	P	
mylpaalsamesprekings		N
met	S	
die	L	
Noord-Koreaanse		A
leier	N	
Kim	S	
Jong-Il		N
.	Z	
"	Z	
Ekonomiese	A	
sanksies	N	
sal	V	
nie	B	
so	B	
doeltreffend	A	
wees	V	
nie	U	



omdat	K	
verskeie	A	
sanksies	N	
reeds	B	
teen	S	
Noord-Korea	N	
ingestel	A	
is	V	
,	Z	
"	Z	
het	V	
hy	P	
gesê	V	
.	Z	
Volgens		S
Kim	N	
is	V	
die	L	
Noorde	A	
gewoond		N
aan	S	
armoede		N
weens	S	
sanksies	N	
.	Z	
"	Z	
Wrewel	K	
teen	S	
buitelandse	A	
inmenging	N	
kan	V	
mense	N	
dalk	B	
verenig		V
teen	S	
ekonomiese	A	
ontbering	N	
.	Z	
En	K	
sommige		T
lande	N	
,	Z	
soos	U	
China	N	
,	Z	
sal	V	
dalk	B	
Noord-Korea	N	
te	U	
hulp	N	
snel	V	
.	Z	
"	Z	
Boonop	P	
sal	V	
ekonomiese	A	
sanksies	N	
Noord-Korea	N	
wrokke	A	
laat	V	
koester		V
wat	P	
op	S	
verskeie	A	

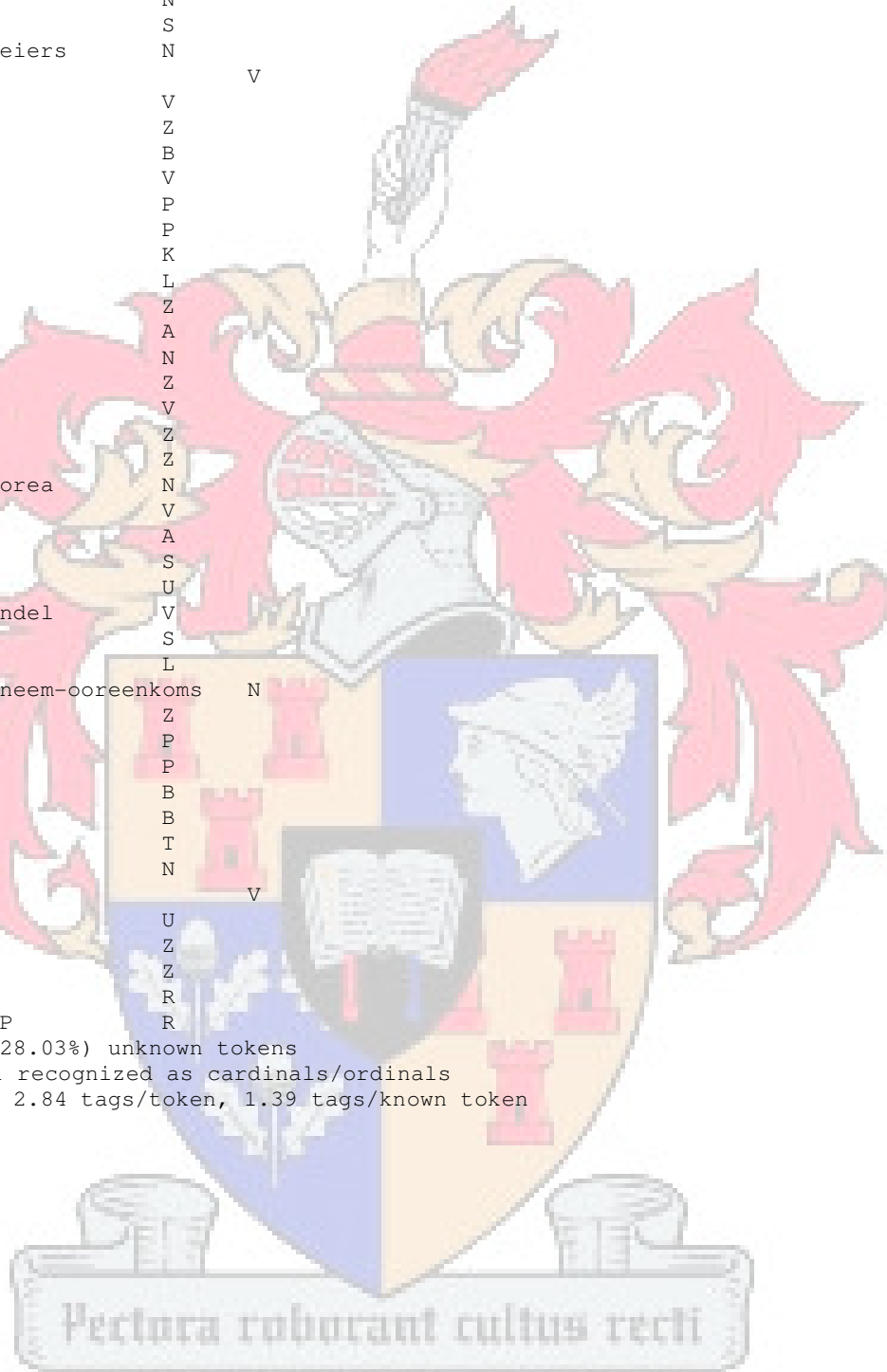


gevaarlike	A	
botsings	N	
kan	V	
afstuur		V
,	Z	
"	Z	
het	V	
Kim	P	
gesê	V	
.	Z	
Kim	P	
het	V	
gepleit		N
vir	S	
regstreekse	A	
gesprekke	N	
tussen	S	
Amerika		N
en	K	
Noord-Korea	N	
.	Z	
"	Z	
Samesprekings		N
is	V	
nie	B	
net	B	
vir	S	
vriende		N
nie	U	
.	Z	
Ons	P	
moet	V	
ook	B	
met	S	
die	L	
bose	A	
praat	V	
in	S	
landsbelang		N
en	K	
ter	S	
wille	N	
van	S	
wêreldvrede	N	
.	Z	
"	Z	
Hy	P	
het	V	
opgemerk	V	
dat	K	
pres.	R	
Richard		N
Nixon	N	
se	U	
besoek	V	
aan	S	
China	N	
in	S	
1972	R	
tot	S	
openheid	N	
en	K	
hervormings	N	
gelei	V	



het	V	
,	Z	
terwyl	K	
pres.	R	
Ronald	N	
Reagan	N	
met	S	
Sowjetleiers	N	
gepraat		V
het	V	
,	Z	
al	B	
het	V	
hy	P	
dit	P	
as	K	
'n	L	
"	Z	
bose	A	
ryk	N	
"	Z	
beskou	V	
.	Z	
"	Z	
Noord-Korea	N	
is	V	
bereid	A	
om	S	
te	U	
onderhandel	V	
oor	S	
'n	L	
gee-en-neem-ooreenkoms		N
.	Z	
Hoekom	P	
dit	P	
nie	B	
nóg	B	
een	T	
keer	N	
probeer		V
nie	U	
?	Z	
"	Z	
-	R	
Sapa-AFP	R	

%% 74 (28.03%) unknown tokens  
%% 1 recognized as cardinals/ordinals  
%% avg. 2.84 tags/token, 1.39 tags/known token





## BYLAAG D

### Artikel 8 in die vertaalde en nievertaalde subkorpus

[8-2006-08-21-sake-vert-motorfabriek.txt]

Ford wil nog Amerikaanse personeel aflê  
NEW YORK. – Die Amerikaanse motormaatskappy Ford, wat tans onder druk is om sy verliese te verminder, gaan nóg meer fabriekke toemaak. Die Wall Street Journal berig Ford in Noord-Amerika is besig om bestuurspersoneel af te dank en maatskappyvoordele te besnoei as maatreëls om sy loonkoste met tussen 10% en 30% te besnoei. Luidens die Amerikaanse nuusagentskap, Bloomberg, beplan Ford om sy uitgawes aan lone en voordele met sowat 16% te besnoei wat afliggings van sowat 6 8201;000 poste is. Die maatskappy sal eers vrywillige uitredingspakkette oorweeg voordat hulle werknemers sal afdank. Mnr. Bill Ford, hoof van Ford, is onder druk om 'n herstrukturierungsplan, Way Forward, te bespoedig ná 'n verlies van \$254 miljoen (sowat R1,7 miljard) in sy tweede kwartaal en laer as verwagte verkope in Julie vanjaar. As deel van die herstrukturierungsplan, het Ford in Januarie vanjaar gesê hy sou 14 fabriekke toemaak en sowat 30 000 Amerikaanse poste teen 2012 besnoei. Die oorspronklike plan het behels dat vier fabriekke tussen 2007 en 2008 gesluit sal word, maar die Wall Street Journal berig dié planne word nou bespoedig en daar word gekyk of nóg aanlegte gesluit sal moet word. Ford wil ook geld in sy advertensie-, bemarkings- en produkbeplanningsafdelings bespaar. Die bestuur sal op 14 September ontmoet om die planne te hersien. Ford se dalende verkope, veral sy sportnutsvoertuie, is grootliks vanweë petrolprysstygings en 'n beter vertoning van oorsese mededingers. – (Sapa)

[8-2006-08-21-sake-nievert-motorfabriek.txt]

Helderbergers help bou super-sportmotor  
CARIN SMITH  
DANKSY 'n Britse maatskappy se belegging van R20 miljoen (sowat £1,5 miljoen) is 'n 21-jaar oue maatskappy van Somerset-Wes onlangs van ondergang gered. Deur Aerodyne Technologies, wat in 'n finansiële verknorsing was, vroeër vanjaar te koop, het die Advanced Composites Group UK sy eerste tak in Suid-Afrika geopen en eersgenoemde se naam is verander na Advanced Composites Group SA (ACG SA). Mnr. Richard Brown, finansiële bestuurder en waarnemende hoofbestuurder van ACG SA, het gesê die Britse beheermaatskappy was al lank op soek na 'n manier om toegang tot die Suid-Afrikaanse mark te verkry. ACG UK is 'n groot verskaffer van koolstofvesel – onder meer aan bykans al die Formule 1-spanne. Dit het ook sterk handelsbande met McLaren Mercedes in Europa. ACG behoort op sy beurt aan UMECO, 'n innoveringsmaatskappy wat in die boekjaar tot Maart 2005 'n omset van £242,4 miljoen (sowat R3,2 miljard) gehad het en op die Londense beurs genoteer is. ACG SA gebruik koolstofvesel om 33 onderdele vir die Mercedes-Benz SLR McLaren te vervaardig – onder meer vir die sitplekke en katebakke. Dié sogenaamde “super-sportmotor” van sowat R4 miljoen waarvan 1 400 al wêreldwyd verkoop is, is nie in Suid-Afrika beskikbaar nie. Behalwe Aerodyne se bestaande kontrak met McLaren wat in die weegskaal was toe eersgenoemde voorlopig gesekwestreer is, was die bekostigbaarheid van arbeid in Suid-Afrika nog 'n belangrike faktor wat die orname gunstig laat lyk het, aangesien die vervaardiging van die koolstofvesel-onderdele baie arbeidsintensief is. Daar is tans 105 werknemers by die ACG-fabriek van 4 000 m<sup>2</sup> in die Olive Grove-nywerheidspark aan die N2. Ten einde ACG SA uit te brei om enersyds nie net van die McLaren-kontrak afhanklik te wees nie en andersyds te strewen na groter werkskepping en maatskaplike verantwoordelikheid, soek die maatskappy nog ander koolstofvesel-projekte – solank dit darem

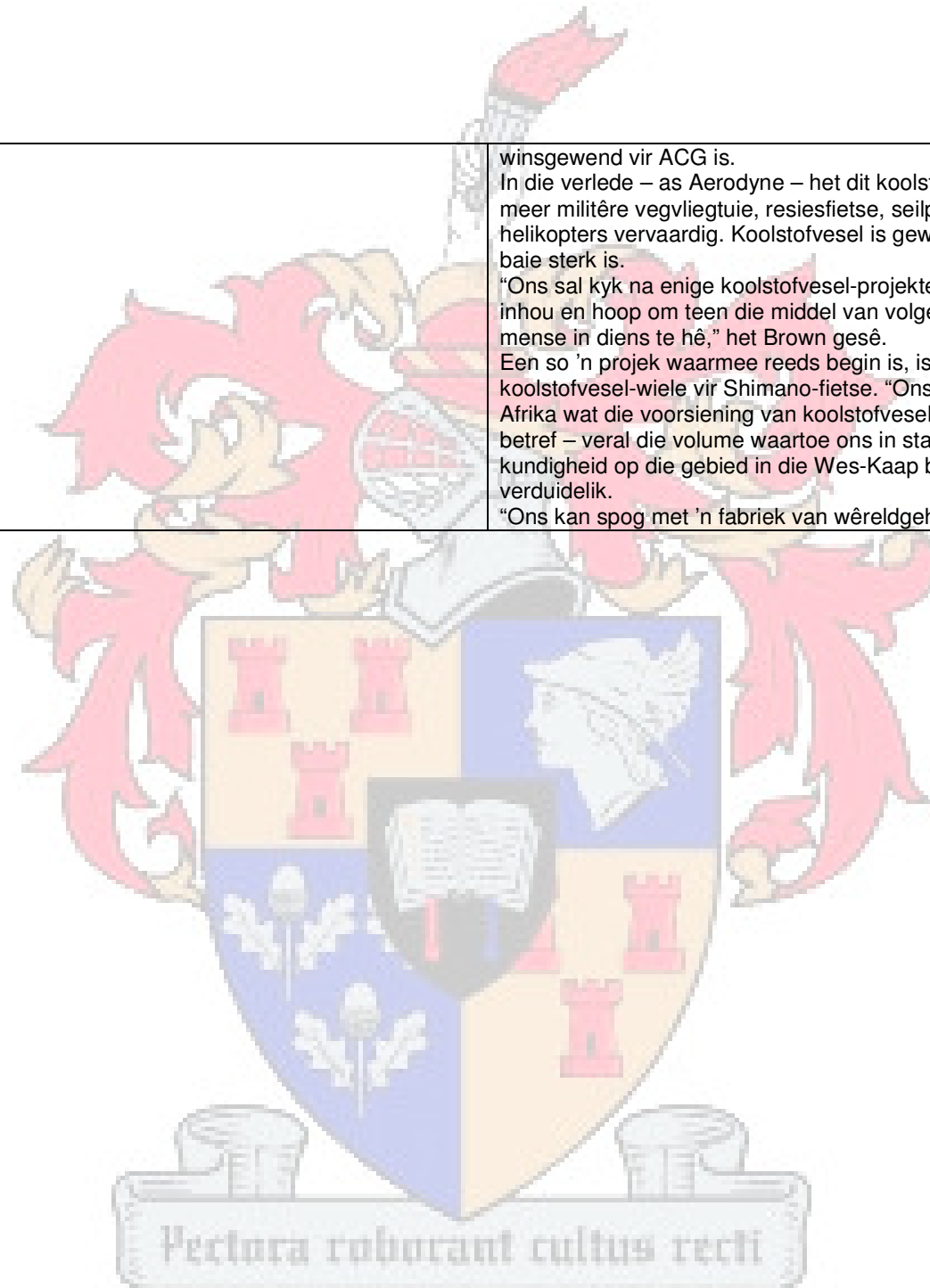
winsgewend vir ACG is.

In die verlede – as Aerodyne – het dit koolstofvesel-onderdele vir onder meer militêre vegvliegtuie, resiesfietse, seilplanke, satellietskottels en helikopters vervaardig. Koolstofvesel is gewild omdat dit so lig maar tog baie sterk is.

“Ons sal kyk na enige koolstofvesel-projekte wat ekonomiese voordeel inhou en hoop om teen die middel van volgende jaar sowat 150 tot 170 mense in diens te hê,” het Brown gesê.

Een so 'n projek waarmee reeds begin is, is die ontwikkeling van nuwe koolstofvesel-wiele vir Shimano-fietse. “Ons maatskappy is uniek in Suid-Afrika wat die voorsiening van koolstofvesel-onderdele vir die motorbedryf betref – veral die volume waartoe ons in staat is. “Boonop is die meeste kundigheid op die gebied in die Wes-Kaap bevestig,” het Brown verduidelik.

“Ons kan spog met 'n fabriek van wêreldgehalte.”



## BYLAAG E

### Kollokasies met *dat* in Vertaalde subkorpus

Kollokasiefunksie identifiseer *verwag*, *gesé* en *verseker* as werkwoorde wat saam met *dat* gebruik word

N	Word	With elation	Total otal	Left	al Right	L2	L1	Centre	R1	R2
1	DAT	dat 0.000	107	0	0	0	0	107	0	0
2	DIE	dat 0.000	53	14	39	14	0	0	32	7
3	HY	dat 0.000	18	1	17	1	0	0	17	0
4	HET	dat 0.000	16	16	0	11	5	0	0	0
5	HULLE	dat 0.000	9	0	9	0	0	0	9	0
6	FEIT	dat 0.000	8	8	0	0	8	0	0	0
7	NIE	dat 0.000	7	3	4	2	1	0	0	4
8	IS	dat 0.000	7	7	0	3	4	0	0	0
9	EN	dat 0.000	6	5	1	0	5	0	0	1
10	DIT	dat 0.000	6	2	4	2	0	0	3	1
11	ONS	dat 0.000	5	0	5	0	0	0	3	2
12	SY	dat 0.000	5	0	5	0	0	0	3	2
13	VAN	dat 0.000	4	1	3	1	0	0	1	2
14	TE	dat 0.000	4	4	0	4	0	0	0	0
15	DAARVAN	dat 0.000	4	4	0	2	2	0	0	0
16	VERWAG	dat 0.000	4	3	1	1	2	0	0	1
17	GESÉ	dat 0.000	4	4	0	1	3	0	0	0
18	WORD	dat 0.000	4	4	0	2	2	0	0	0
19	INFLASIE	dat 0.000	3	0	3	0	0	0	2	1
20	BAIE	dat 0.000	3	2	1	2	0	0	1	0
21	SPAN	dat 0.000	3	1	2	0	1	0	0	2
22	FED	dat 0.000	3	0	3	0	0	0	0	3
23	VERSEKER	dat 0.000	3	3	0	0	3	0	0	0
24	OORTUIG	dat 0.000	2	2	0	1	1	0	0	0
25	SEKER	dat 0.000	2	2	0	2	0	0	0	0
26	ONTKEN	dat 0.000	2	2	0	0	2	0	0	0
27	WYS	dat 0.000	2	2	0	0	2	0	0	0
28	SO	dat 0.000	2	1	1	1	0	0	0	1
29	VOL	dat 0.000	2	2	0	2	0	0	0	0
30	UITTREE	dat 0.000	2	0	2	0	0	0	0	2
31	VERTROUE	dat 0.000	2	2	0	0	2	0	0	0
32	VERSLAP	dat 0.000	2	1	1	1	0	0	0	1
33	HELE	dat 0.000	2	0	2	0	0	0	1	1
34	GROOT	dat 0.000	2	0	2	0	0	0	0	2
35	HUL	dat 0.000	2	0	2	0	0	0	1	1
36	HOOP	dat 0.000	2	2	0	1	1	0	0	0
37	EK	dat 0.000	2	0	2	0	0	0	2	0
38	EIS	dat 0.000	2	2	0	0	2	0	0	0
39	GEWAARSKU	dat 0.000	2	2	0	1	1	0	0	0
40	DIÉ	dat 0.000	2	2	0	2	0	0	0	0
41	AMERIKAANSE	dat 0.000	2	0	2	0	0	0	0	2
42	MET	dat 0.000	2	0	2	0	0	0	0	2
43	NOU	dat 0.000	2	2	0	1	1	0	0	0
44	NOOIT	dat 0.000	2	1	1	1	0	0	0	1
45	KEER	dat 0.000	2	2	0	1	1	0	0	0
46	BESEF	dat 0.000	2	2	0	0	2	0	0	0
47	MENSE	dat 0.000	2	0	2	0	0	0	1	1
48	MEEN	dat 0.000	2	2	0	1	1	0	0	0

# BYLAAG F

## Kollokasies met *dat* in Nievertaalde subkorpus

*verwag, gesê en verseker* as werkwoorde wat saam met *dat* gebruik word

N	Word	With elation	Total	Left	Right	L2	L1	Centre	R1	R2
1	DAT	dat 0.000	122	0	0	0	0	122	0	0
2	DIE	dat 0.000	37	7	30	7	0	0	27	3
3	IS	dat 0.000	15	15	0	11	4	0	0	0
4	NIE	dat 0.000	10	5	5	4	1	0	0	5
5	HY	dat 0.000	10	0	10	0	0	0	10	0
6	HET	dat 0.000	10	10	0	6	4	0	0	0
7	DIT	dat 0.000	9	6	3	6	0	0	3	0
8	EN	dat 0.000	9	6	3	2	4	0	0	3
9	ONS	dat 0.000	9	1	8	1	0	0	8	0
10	IN	dat 0.000	8	0	8	0	0	0	0	8
11	SY	dat 0.000	8	1	7	1	0	0	5	2
12	TE	dat 0.000	6	6	0	6	0	0	0	0
13	DAAROP	dat 0.000	5	5	0	1	4	0	0	0
14	EGTER	dat 0.000	4	4	0	2	2	0	0	0
15	VERWAG	dat 0.000	4	4	0	0	4	0	0	0
16	KEER	dat 0.000	4	4	0	1	3	0	0	0
17	GISTER	dat 0.000	4	4	0	4	0	0	0	0
18	BEVESTIG	dat 0.000	4	4	0	0	4	0	0	0
19	DAAR	dat 0.000	4	0	4	0	0	0	4	0
20	OOK	dat 0.000	3	3	0	3	0	0	0	0
21	HULLE	dat 0.000	3	0	3	0	0	0	3	0
22	BEKLEMTON	dat 0.000	3	3	0	0	3	0	0	0
23	BLYK	dat 0.000	3	3	0	1	2	0	0	0
24	BESLUIT	dat 0.000	3	2	1	1	1	0	0	1
25	BETEKEN	dat 0.000	3	3	0	0	3	0	0	0
26	SPAN	dat 0.000	3	0	3	0	0	0	0	3
27	VOORSIEN	dat 0.000	3	3	0	1	2	0	0	0
28	EEN	dat 0.000	3	1	2	1	0	0	1	1
29	EK	dat 0.000	3	0	3	0	0	0	3	0
30	FEIT	dat 0.000	3	3	0	0	3	0	0	0
31	VROEËR	dat 0.000	2	1	1	1	0	0	0	1
32	SÊ	dat 0.000	2	2	0	0	2	0	0	0
33	BESLIS	dat 0.000	2	2	0	1	1	0	0	0
34	SUID	dat 0.000	2	0	2	0	0	0	1	1
35	SOMMIGE	dat 0.000	2	0	2	0	0	0	2	0
36	SE	dat 0.000	2	0	2	0	0	0	0	2
37	SANLAM	dat 0.000	2	0	2	0	0	0	2	0
38	STAAT	dat 0.000	2	1	1	0	1	0	0	1
39	WAT	dat 0.000	2	1	1	1	0	0	0	1
40	WYS	dat 0.000	2	2	0	2	0	0	0	0
41	VAN	dat 0.000	2	0	2	0	0	0	0	2
42	AARDIGHEIDSCHEFFING	dat 0.000	2	0	2	0	0	0	0	2
43	VERDUIDELIK	dat 0.000	2	2	0	0	2	0	0	0
44	VOORGESTEL	dat 0.000	2	2	0	0	2	0	0	0
45	VERLEDE	dat 0.000	2	2	0	2	0	0	0	0
46	UITSPRAKE	dat 0.000	2	1	1	1	0	0	0	1
47	TOT	dat 0.000	2	0	2	0	0	0	1	1
48	BEKOMMERD	dat 0.000	2	2	0	0	2	0	0	0
49	TWEE	dat 0.000	2	0	2	0	0	0	1	1
50	TWEEDE	dat 0.000	2	2	0	2	0	0	0	0
51	VPIX	dat 0.000	2	0	2	0	0	0	2	0
52	REGULASIES	dat 0.000	2	0	2	0	0	0	0	2
53	GLO	dat 0.000	2	2	0	2	0	0	0	0
54	DAARVAN	dat 0.000	2	2	0	0	2	0	0	0
55	GEWAARSKU	dat 0.000	2	2	0	0	2	0	0	0
56	HOOG	dat 0.000	2	2	0	2	0	0	0	0
57	HOM	dat 0.000	2	1	1	1	0	0	0	1
58	DAARTOE	dat 0.000	2	2	0	1	1	0	0	0
59	AMERIKA	dat 0.000	2	0	2	0	0	0	2	0
60	DUIDELIK	dat 0.000	2	2	0	1	1	0	0	0
61	DUI	dat 0.000	2	2	0	2	0	0	0	0
62	GENOEG	dat 0.000	2	1	1	0	1	0	0	1
63	AS	dat 0.000	2	0	2	0	0	0	0	2
64	EKONOME	dat 0.000	2	2	0	2	0	0	0	0
65	ONTKEN	dat 0.000	2	2	0	0	2	0	0	0
66	NOU	dat 0.000	2	0	2	0	0	0	0	2
67	MY	dat 0.000	2	2	0	2	0	0	0	0
68	PITSO	dat 0.000	2	0	2	0	0	0	1	1
69	OP	dat 0.000	2	1	1	1	0	0	0	1
70	OORTUIG	dat 0.000	2	2	0	1	1	0	0	0
71	22	dat 0.000	2	0	2	0	0	0	1	1
72	INFLASIE	dat 0.000	2	0	2	0	0	0	1	1
73	HOOR	dat 0.000	2	2	0	0	2	0	0	0
74	MAAR	dat 0.000	2	2	0	1	1	0	0	0
75	KENNIS	dat 0.000	2	2	0	2	0	0	0	0
76	ISRAEL	dat 0.000	2	0	2	0	0	0	2	0

## BYLAAG G

### Soekterme gebruik vir konkordansies

Vertaal 17 450 woorde	Nievertaal 19 283 woorde
<b>1. span (45 - 0,26%)</b>	<b>span (53 - 0,27%)</b>
<p><b>(L1)</b>            voorste span            te span            sy span (9)            puik span            nommertwee-span            nasionale span            'n span            hierdie span            hele span            goeie span            enige span            eerste span (1)            dié span (4)            die span (15)            Benoni-span            Australiese span            Suid-Asiese span</p>	<p><b>(L1)</b>            WP-span            wêreldkampioen-span            voorste span            sy span (5)            se span (3)            'n span            my span            jong span            Indiese span            hul span            goeie span            enigste span            enige span            eendaagse span (2)            dié span (2)            die span (21)            CSV/CVT-span            Bok-span            beste span (3)            Bafowethu-span            Bafana-span            ander span            Suid-Afrikaanse span</p>
<p><b>span (R1)</b>            span wat (3)            span vind            span van            span teen            span sy            span sukkel            span sedert (2)            span se (6)            span sal            span ook            span onder            span om            span nie            span na            span met            span, maar            span is            span in (3)            span het (7)            span glo            span dat            span bekend</p>	<p><b>span (R1)</b>            span sal            span op (2)            span oefen            span na (2)            span 'n (3)            span met            span kan            span is (2)            span in (4)            span het (3)            span gekies            span gedreig            span gaan (2)            span en            span die (2)            span bloedneus            span baie</p>

span behaal span as span aanvoer span aan	
<b>2. mense (27 - 0,15%)</b>	<b>mense (12 - 0,06%)</b>
<b>(L1)</b> wit mense waarop mense vyf mense verskeie mense twee mense tussen mense talle mense swart mense (2) sodat mense professionele mense na mense miljoen mense kan mense het mense grawe mense fantastiese mense ek mense die mense dat mense baie mense (2) # mense ( )	<b>(L1)</b> tientalle mense miljoene mense kon mense die mense (2) dat mense # mense (6)
<b>mense (R1)</b> mense wat mense vermis mense van mense te mense steeds mense sien mense om mense kort mense is (4) mense in (4) mense hulpeloos mense hoor mense het mense gehuur mense eerder mense die mense dalk mense dakloos mense betyds mense aangestel	<b>mense (R1)</b> mense wêreldwyd mense van mense teen mense op mense oorwen mense nou mense is (3) mense in (2) mense gesterf
<b>3. ander (29 - 0,17%)</b>	<b>ander (22 - 0,11%)</b>
<b>(L1)</b> verskeie ander twyfel ander se ander (2) ook ander	<b>(L1)</b> vir ander van ander toenemend ander steeds ander



<p>of ander (2)  'n ander  in ander (4)  het ander  enige ander  en ander (3)  die ander (7)  baie ander  as ander</p>	<p>of ander  nog ander  'n ander (3)  maatskappye ander  in ander  en ander  die ander (9)</p>
<p><b>ander (R1)</b>  ander wedstryde (2)  ander verteenwoordigers  ander veranderings  ander tyd  ander twee  ander spanne (5)  ander professionele  ander planne  ander oorlewende(s) (2)  ander ontluikende  ander oefeninge  ander natuurlike  ander manier  ander lande (3)  ander keuses  ander kant  ander gebiede (2)  ander foto's  ander drie  ander buitelanders</p>	<p><b>ander (R1)</b>  ander werkers  ander vorm  ander tyd  ander spelers (2)  ander span(ne) (3)  ander SA  ander ooreenkomste  ander ontluikende  ander lede  ander lande  ander koolstofvesel-projekte  ander kolwers  ander kleinhandelgroepe  ander kerkmense  ander kant  ander dorpe  ander boeke  ander atleet</p>
<p><b>4. toets (36 - 0,21%)</b></p>	<p><b>toets (18 - 0,09%)</b></p>
<p><b>(L1)</b>  vorige toets  tweede toets  te toets  sy toets  positiewe toets (3)  'n toets (3)  laaste toets (7)  eerste toets (2)  Drienasies-toets (2)  dié toets  die toets (7)  derde toets (3)  B-toets  A-toets  50ste toets</p>	<p><b>(L1)</b>  verloor-toets  tweede toets (3)  te toets  sal toets  'n toets (2)  laaste toets  eerste toets (5)  dié toets  die toets (2)</p>
<p><b>toets (R1)</b>  toets vir  toets tussen (3)  toets teen (4)  toets te  toets op (5)  toets nie</p>	<p><b>toets (R1)</b>  toets weens  toets vanweë  toets teen (3)  toets te (2)  toets op (2)  toets oneffektief</p>

toets misgelopen toets met toets kon toets in (4) toets en (2) toets eers toets die toets begin toets aanstaande	toets met toets in toets goeie toets gekos toets en toets begin
--	--

