Die ortografiese realisering van komposita met en afleidings van multiwoordeiename

Gerhard B. van Huyssteen

Gerhard B. van Huyssteen,
Sentrum vir Tekstegnologie (CTexT), Noordwes-Universiteit (Potchefstroomkampus)

Opsomming

’n Multiwoordeienaam is ’n frasekonstruksie wat na ’n benoemde entiteit verwys en wat uit twee of meer ortografiese woorde bestaan; dié ortografiese woorde word deur spasies geskei en gewoonlik met hoofletters geskryf (uitgesonderd funksiewoorde soos lidwoorde, voorsetsels en partikels), byvoorbeeld *Nelson Mandela* en *Atlantiese Oseaan*. Daar is geen spesifieke normerende reëls in Afrikaans wat betref die skryfwyse van samestellings met en afleidings van multiwoordeiename nie. Skryf ’n mens byvoorbeeld *Atlantiese Oseaan+verskynsel* as *Atlantiese Oseaanverskynsel*, *Atlantiese-Oseaanverskynsel* of *Atlantieseoseaanverskynsel*; en *Atlantiese Oseaan·ADJZ* as *Atlantiese Oseanies* of *Atlantiese-Oseanies* of *Atlantieseoseanies*?[1]

Die oogmerk van hierdie studie is om ’n beskrywing van die ortografiese realisering van hierdie twee konstruksies in Afrikaans te gee, met die oog daarop om te bepaal wat die keusemoontlikhede (Wallis, Bowie en Aarts 2012) is wat gebruikers vir die skryfwyses het, en watter keuses inderdaad uitgeoefen word. Die basiese navorsingsvraag is: Is daar ’n positiewe korrelasie tussen die ortografiese vorm van ’n multiwoordeienaam as sodanig en die ortografiese vorm daarvan in komposita daarmee of afleidings daarvan? Daar word van drie metodes gebruik gemaak in die ondersoek, naamlik ’n literatuurstudie, korpusondersoek en ’n meningspeiling onder taalpraktisyns.

Uit die kleinskaalse korpusondersoek en meningspeiling wat geloods is, blyk dit dat die ortografiese integriteit van multiwoordeiename oorwegend behoue bly in komposita daarmee en gesuffigeerde afleidings daarvan. Vier veranderlikes waaraan in toekomstige navorsing praktiese uitvoering gegee moet word, is geïdentifiseer, te wete die:

- woordsoort van die resulterende woordvorm
- voorkoms van die resulterende woordvorm in normerende bronne en woordeboeke
- mate van vestiging van multiwoordeiename
- verskil tussen geredigeerde tekste en ongeredigeerde tekste.

Trefwoorde: Afrikaans; *Afrikaanse woordelys en spelreëls*; frasekonstruksie; korpus; morfologie; multiwoordeienaam; ortografie

Abstract

The orthographic realisation of compounds with and derivations of multiword proper nouns

A multiword proper noun is a phrasal construction that refers to a named entity, and that consists of two or more orthographic words. These orthographic words are separated by spaces, and are usually written with capital letters (with the exception of function words like articles, prepositions and particles). Examples include *Nelson Mandela* (personal name consisting of first name and surname), *J. Chris Coetzee* (personal name consisting of initial, first name and surname), *J.M. Coetzee* (personal name consisting of two initials and surname), *Van Heerden* (surname); *Sunday Times* (publication name); *Pacific Ocean* (geographical name); *Organisation of African Unity* (organisation name); *Concordia Pharmacy* (business name); *Old Norse* (language name), etc. Such multiword proper nouns contrast, morphographemically speaking, with compound proper nouns such as *Johannesburg* (*Johannes + burg*), or *Hewlett-Packard* (*Hewlett + Packard*).

In Afrikaans, there are no specific normative rules for the orthographic realisation of derivations of and compounds with such multiword proper nouns. For example, should we write *Atlantiese Oseaan + verskynsel* ("Atlantic Ocean" + "phenomenon") as *Atlantiese Oseaanverskynsel*, *Atlantiese-Oseaanverskynsel* or *Atlantieseoseaanverskynsel*; and *Atlantiese Oseaan·ADJZ* (where *ADJZ* represents an adjective-forming suffix) as *Atlantiese Oseanies* of *Atlantiese-Oseanies* of *Atlantieseoseanies*?

The primary aim of this article is to present a description of the orthographic realisation of these two constructions in Afrikaans, with a view to investigating the orthographic choices that users have when writing these constructions, as well as the actual choices that they make. The basic research question is: Is there a positive correlation between the orthographic form of a bare proper noun, and the orthographic form of such a proper noun in derivations and compounds? The secondary aim is to identify variables that have an influence on the way these constructions are written in Afrikaans, so that these variables can be operationalised in future corpus studies.

Three different methods are used in the investigation: (1) a comprehensive literature study, with a specific focus on normative rules in the *Afrikaanse Woordelys en Spelreëls* (AWS) ("Afrikaans Word List and Spelling Rules" – the standard orthographic guide for Afrikaans); (2) a corpus study using three assumedly different corpus sources, viz. edited texts, semi-edited texts and unedited texts; and (3) a questionnaire sent to language practitioners to gauge their preferred way of writing these constructions.

From the literature review it transpires that there is no unambiguous, uniform set of writing rules in the AWS for the orthographic realisation of derivations of, or compounds with, multiword proper nouns. In fact, some of the rules or remarks in the AWS contradict one another regarding the possible ways of writing these constructions. A summary is provided of

all the possibilities that the AWS affords theoretically, and illustrated with derivations of and compounds with *Middellandse See* ("Mediterranean Sea").

For the purposes of the corpus study, the focus is specifically on the use of capital letters (i.e. are the capital letters of multiword proper nouns preserved in these constructions?), and the conjunctive vs disjunctive orthographical realisation of these constructions (i.e. are the white spaces in multiword proper nouns preserved in such constructions?). Within the framework of Wallis, Bowie and Aarts's (2012) so-called "model of choice", we can say that the potential choice space for derivations of and compounds with a two-part multiword proper noun consists of 72 options, i.e. a matrix of eight (2x2x2; two choices – upper vs lower case – in three positions), by nine (3x3; three choices – disjunctive, conjunctive or conjunctive with a hyphen – in two positions). From the usage-based data it emerges that the choice space actually consists of only 23 options; theoretically possible cases like **AtlantieseOseaanVerskynsel* and **atlantieseoseaan-Verskynsel* do not occur in the data at all.

From the corpus study it appears as if the orthographic integrity of multiword proper nouns is mostly preserved in these kinds of morphological constructions. This finding is confirmed by the results of the questionnaire that was sent to language practitioners. 76% of respondents indicated a clear choice for rules that would preserve the orthographic form of such multiword proper nouns in derivations and compounds, e.g. *Middellandse Seegebied* ("Mediterranean Sea area"), *Konstitusionele Hofregter* ("Constitutional Court judge") and *Derde Wêreldland* ("Third World country"); and *Nabye Oosterse* ("Near Eastern"), *Verre Oosterse* ("Far Eastern"), *Ou Testamenties* ("Old Testament·ADJZ") and *Derde Wêrelds* ("Third World·ADJZ").

Four variables that could be operationalised in future research are identified, viz.:

- part of speech of the resulting word form
- institutionalisation of the resulting word form in normative literature and dictionaries
- level of entrenchment of the multiword proper noun
- the difference between edited and unedited texts.

The article concludes with a summary of all findings, and some recommendations for normative rules in an orthographic guide such as the AWS.

Keywords: Afrikaans; *Afrikaanse Woordelys en Spelreëls*; corpus; morphology; multiword orthography; phrasal construction; proper noun

1. Inleiding

Geen spesifieke normerende reëls wat betref die skryfwyse van komposita met en afleidings van multiwoordeiename bestaan in Afrikaans nie. Skryf 'n mens byvoorbeeld die kompositum *Atlantiese Oseaan+verskynsel* as *Atlantiese Oseaan-verskynsel*, *Atlantiese Oseaanverskynsel* of *Atlantiese-Oseaanverskynsel*? Hoe gemaak met die afleiding *Atlantiese Oseaan ·ADJZ*?[1] Word dit *Atlantiese Oseanies* of *Atlantiese-Oseanies*? Die *Afrikaanse*

Woordelys en Spelreëls (Taalkommissie (samest.) 2009) (voorts AWS) verskaf in dié verband weinig gebruiksleiding aan eindgebruikers.

Die oogmerk van hierdie studie is om ’n beskrywing van die ortografiese realisering van hierdie twee konstruksies in Afrikaans te gee, met die oog daarop om te bepaal wat die keusemoontlikhede (Wallis, Bowie en Aarts 2012) is wat gebruikers vir die skryfwyses het, en watter keuses inderdaad uitgeoefen word.

Drie metodes word in hierdie navorsing gebruik:

1. Literatuurondersoek. Taalkundige literatuur oor die ortografie en morfologie van multiwoordeiename, komposita, afleidings, samestellende komposita en samestellende afleidings sou lig kon werp op die verskillende skryfwysemoontlikhede en hoe dit sou inpas in die huidige ortografie van Afrikaans. Die doel van die literatuurondersoek is in wese om die aard en omvang van die keuseruimte te omskryf.
2. Korpusondersoek. ’n Kleinskaalse korpusondersoek is gedoen, met die doel om te bepaal watter van die keuses in werklike taalgebruik deur taalgebruikers uitgeoefen word. Vir dié doeleindes is van veronderstelde geredigeerde, semigeredigeerde en ongeredigeerde korpusmateriaal gebruik gemaak.
3. Meningspeiling. Die opinie van taalpraktisyns is in 2012 deur die Afrikaanse Taalkommissie (TK) van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns ingewin om te bepaal hoe diesulkes intuïtief oor die saak dink. Die resultate van dié peiling is vir hierdie artikel beskikbaar gestel en word betrek om ’n gebruikersgebaseerde perspektief te bied.

Gegewe bostaande metodes is hierdie navorsing gebruiks- en gebruikersgebaseerd en dusdanig toepaslik deskriptief. Aangesien die korpusondersoek van beperkte omvang is, kan dit nie aanspraak op volledigheid maak nie. Dit bied egter wel die geleentheid om veranderlikes te identifiseer waaraan in verdere navorsing praktiese uitvoering gegee kan word.

In afdeling 2 word die literatuurondersoek bespreek, waarna die korpusondersoek (afdeling 3) en meningspeiling (afdeling 4) bespreek word. Die artikel sluit af met ’n uitvoerige bespreking oor die voorgestelde reëls en die implikasies wat dit op ander, bestaande reëls in die AWS sou hê.

2. Literatuurondersoek

Die eerste oogmerk van die literatuurondersoek is om die konsep *multiwoordeiename* beter te verstaan. Om hierdie doel te bereik, word daar in afdeling 2.1 gekyk na die aard van eiename en multiwoordeenhede, en in afdeling 2.2 word aandag geskenk aan komposita, afleidings, samestellende komposita en samestellende afleidings en spesifiek hoe die algemene skryfwyse hiervan die skryfwyse van komposita met en afleidings van multiwoordeiename sou raak.

2.1 Multiwoordeiename

’n Eienaam is ’n leksikale eenheid (d.i. woord of woorde) wat na ’n benoemde entiteit verwys en as sodanig as ’n onomastiese eenheid/woord (Kirk 2006; McArthur 1992) optree. Eiename sluit in “vanne, voorname (dit sluit vanne in wat as voorname optree), byname, toename (dit wil sê name wat ’n beskrywende element bygevoeg kry), name van bevolkingsgroepe, taalbenamings, geografiese name, straatname, geologiese tydperke, maand- en dagname, name vir geskiedkundige gebeurtenisse, ensovoorts” (AWS 2009:46; sien ook figuur 2).

Puttkammer (2005:25) definieer ’n benoemde entiteit soos volg:

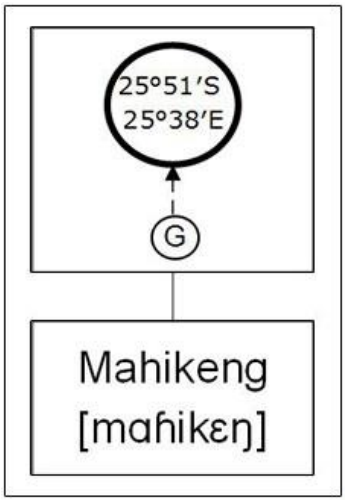
> ’n Benoemde entiteit is ’n aansyn wat binne die konseptuele ruimte aan ’n enkele instansiëring veranker word deur middel van konvensie, ’n geïnstitusionaliseerde proses of outoriteit, en waarvan die skryfwyse óf wetlik, óf deur een of ander outoriteit bepaal word.

Uit dié definisie is drie aspekte veral belangrik:

- ’n *Aansyn* (d.i. ’n “thing” in die terminologie van die kognitiewe grammatika) is “enigiets wat beskou kan word of na verwys kan word, soos byvoorbeeld konkrete dinge, verhoudings, plekke, punte op ’n skaal, sensasies, interkonneksies, waardes, ensovoorts” (Puttkammer 2005:22).
- In die geval van benoemde entiteite is die aansyn ’n enkele, *unieke instansiëring*(sien ook AWS 2009:559). Die leksikale eenheid *Verenigde Nasies* is ’n multiwoordeenheid wat na een spesifieke instansiëring verwys – die organisasie waarvan die hoofkantoor in New York is. Daarteenoor kan die woordgroep *verenigde nasies* na verskeie instansiërings verwys, soos byvoorbeeld by groot rockkonserte, of nasies wat verenig is oor ’n bepaalde politieke kwessie. Hierdie beskouing sluit aan by Lehrer (2006:140) se beskouing dat ’n eienaam aan ’n spesifieke individu of groep gegee word wat “belangrik” en identifiseerbaar” is.
- *Benoemde* verwys na die proses waardeur ’n entiteit veranker word, hetsy deur ’n naamgewingsritueel (soos in die geval van persoonsname, byvoorbeeld tydens ’n doop), ’n outoritêre instelling (soos in die geval van plekname wat in Suid-Afrika deur pleknaamkomitees bereël word), ’n registrasieproses (soos in die geval van besigheids- en produkname, asook webadresse en e-pos-adresse), of konvensie (soos by titels, waar konvensie byvoorbeeld dikteer dat ’n universiteitsdosent wat op ’n bepaalde posvlak aangestel is, normaalweg as *Professor* aangespreek word) (Puttkammer 2005:22). Hanks (2006:134) sluit hierby aan as hy daarop wys dat die gee en gebruik van eiename die manier is waarop sprekers unieke sake identifiseer en dit dan deur taal veranker aan die fisiese en konseptuele ruimte waarin hulle lewe.

So gesien kan ons dus sê dat ’n eienaam ’n konstruksie (d.i. ’n vorm-betekenis-paar) is wat na ’n konseptuele aansyn verwys. ’n Eienaam soos *Mahikeng* verwys dus konseptueel na ’n aansyn (’n plek in die fisiese ruimte by 25°51'S 25°38'E) wat uniek is (daar is nie ’n ander plek in die wêreld met dieselfde naam nie) en waarvan die naam in Februarie 2010 deur die minister van die Departement Kuns en Kultuur goedgekeur is. In figuur 1 word die vorm-betekenis-paar [MAHIKENG/Mahikeng] visueel voorgestel: die boonste blokkie beeld die eienaam op die konseptuele pool (oftewel semantiese pool) uit, terwyl die onderste blokkie die vorm op die realiseringspool (oftewel fonologiese pool) verteenwoordig. Die ortografiese

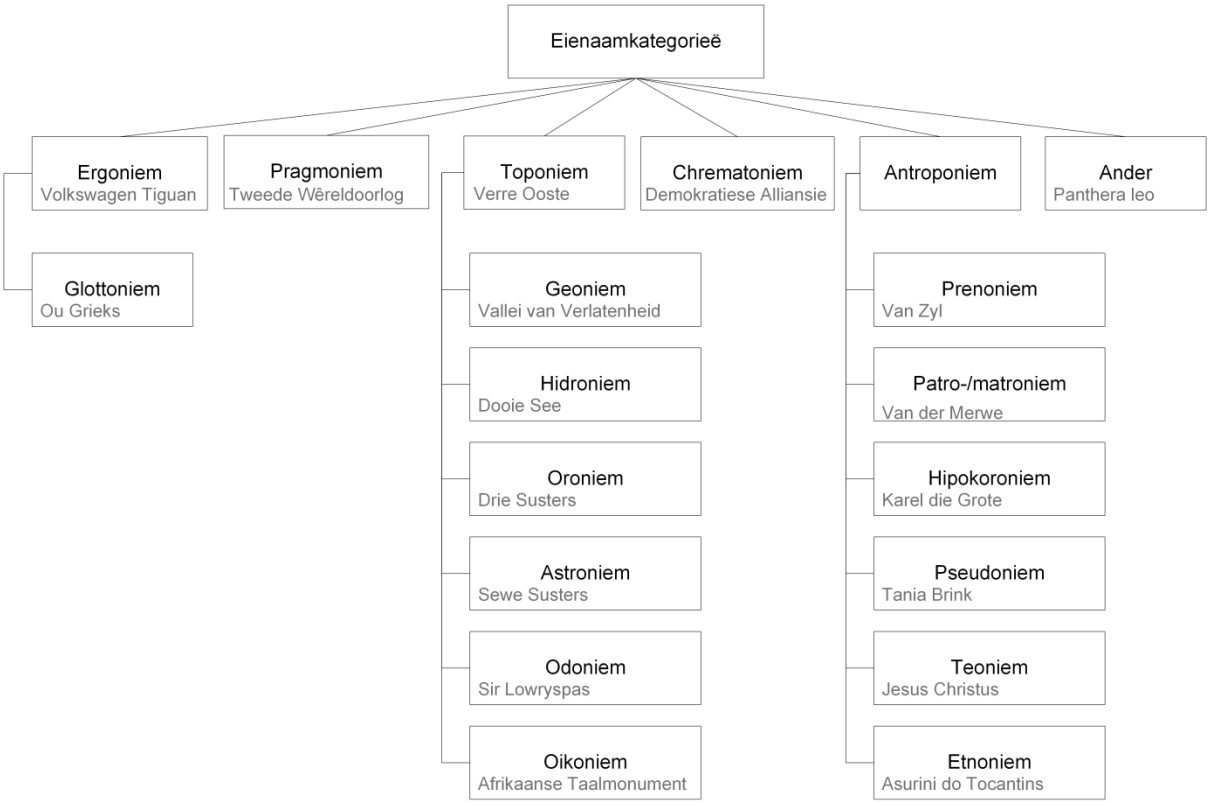
vorm *Mahikeng* of fonetiese vorm [mɑɦikɛŋ] word gekoppel aan ’n unieke aansyn (aangedui met die sirkel met die dik lyn) wat in die ruimtedomein veranker is (aangedui met die G, vir “grounding”).



Figuur 1. [MAHIKENG/Mahikeng] as simboliese eenheid

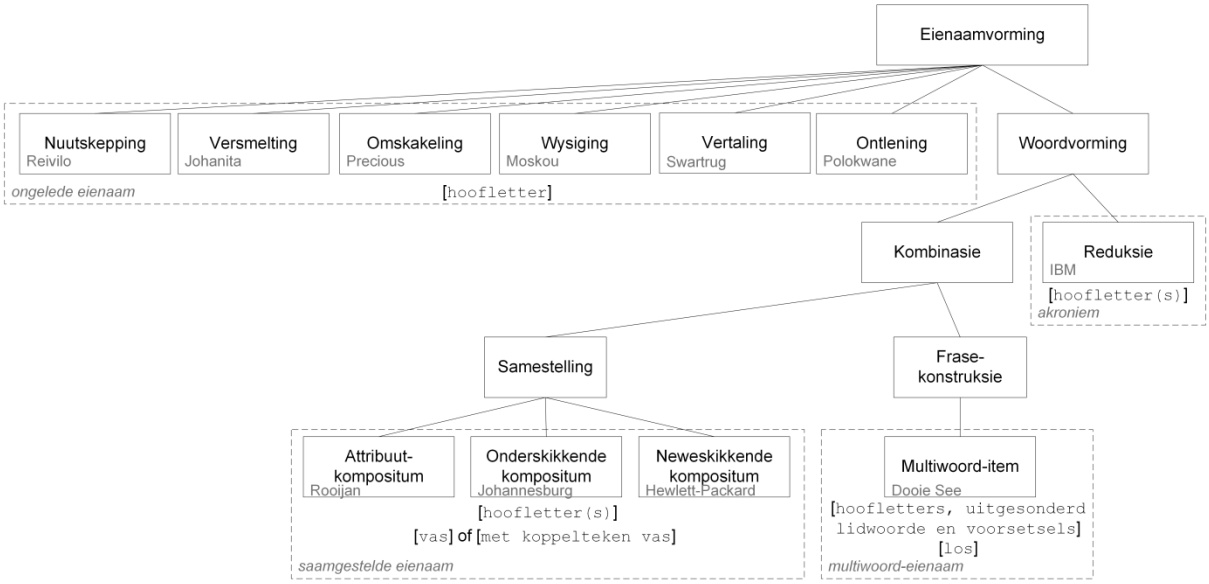
Met betrekking tot die unieke instansiërings waarna eiename verwys, moet daarop gelet word dat verskeie aansyne dieselfde naam kan hê; dit is normale homonimiese en polisemiese verhoudinge tussen leksikale items. Die dorpie Johannesburg in Kalifornië is byvoorbeeld genoem na die stad Johannesburg in Suid-Afrika, en dié twee plekname staan dus in ’n polisemiese verhouding. As twee onverwante persone Susan heet, staan dié twee voorname in ’n homonimiese verhouding.

Die semantiese kategorisering van eiename (d.i. karakterisering op die semantiese pool) is die onderwerp van omvattende navorsing in veral die rekenaarlinguistiek. Bergman (2016) toon aan dat benoemde-entiteit-herkenning eintlik na regte afgeskop het by die 1995- en 1997-*Message Understanding Conference* (MUC), met die taak om persoons-, plek- en organisasiename te herken. Oor die afgelope 20 jaar het die aantal kategorieë logaritmies meer geword: die DBpedia-ontologie (Bizer e.a. 2009) bevat byvoorbeeld meer as 700 eindnodes. Die hoëvlak semantiese kategorieë in figuur 2 is gebaseer op die Prolex-tipologie (Maurel 2008; Krstev e.a. 2005; Vitas e.a. 2007), met eie aanpassings uit die werk van onder andere Desmet en Hoste (2014), Gillick e.a. (2014), Ling en Weld (2012), Sekine e.a. (2002) en Sekine (2010).



Figuur 2. Hoëvlakkarakterisering van die konseptuele pool (met multiwoordeienaamvoorbeelde)

Op die realiseringspool kan eiename in terme van woordvormingsprosesse en die ortografiese realiserings daarvan gekarakteriseer word. In figuur 3 word die prototipiese prosesse met verteenwoordigende voorbeelde aangedui.



Figuur 3. Hoëvlakkarakterisering van die realiseringspool (met prototipiese voorbeelde)

Multiwoordeiename is eiename wat uit minstens twee dele bestaan wat deur middel van ’n spasie geskei word. Voorbeelde is:

(1) *Nelson Mandela* (persoonsnaam bestaande uit voornaam en van);
J. Chris Coetzee (persoonsnaam bestaande uit voorletter, voornaam en van);
T.T. Cloete (persoonsnaam bestaande uit voorletters en van)
Van Heerden (van)
Die Burger (publikasienaam)
Stille Oseaan (geografiese eienaam)
Organisasie vir Afrika-eenheid (organisasienaam)
Concordia Apteek (besigheidsnaam)
Ou Noors (taalnaam).

As sodanig word multiwoordeiename gesien as *frasekonstruksies*, d.i. “units that are functionally identical to complex words, but do not form one grammatical word, but two [or more]” (Booij 2002; my byvoeging). Ofskoon Booij nie aandui wat hy in dié konteks met “grammatical word” bedoel nie, is ek van mening dat die definisie eerder na “orthographic word” moet verwys, d.i. “a written sequence which has a white space at each end but no white space in the middle” (Trask 2004); dit is in ieder geval die bedoeling wat relevant is binne die konteks van hierdie artikel.

Lehrer (2006:143) wys daarop dat byvoeglike naamwoorde (insluitende telwoorde wat as byvoeglike naamwoorde funksioneer) en soortname gekombineer kan word om na “belangrike” en “identifiseerbare” sake te verwys: *Dooie See* bestaan uit ’n byvoeglike naamwoord (*dooie*) en soortnaam (*see*), maar gesamentlik vorm dit ’n eienaam wat na ’n unieke saak verwys. Dieselfde geld ook gevalle soos *Nabye Ooste*, *Verre Ooste*, *Stille Oseaan*, *Persiese Golf* en *Derde Wêreld*. In navolging van Hanks (2006:134) kan dié voorbeelde dus gesien word as *multiwoordeenhede* en nie as *woordgroepe* nie.

Gralinski e.a. (2010:2; sien ook Gouws 1989:97; Louw 2006:97) karakteriseer multiwoordeenhede soos volg:

> Multi-word units (MWU) are linguistic objects placed between morphology and syntax: their general syntactic behavior makes them similar to free phrases, while some of their idiosyncratic (notably from the morphological point of view) properties call for a lexicalized approach in which they are treated as units of description. Moreover, MWUs, which encompass such classes as compounds, complex terms, multi-word named entities, etc., often have unique and constant references, thus they are seen as semantically rich objects […].

Ofskoon sommige skrywers *woordgroep* en *multiwoordeenheid* as sinonieme sien (De Haas en Trommelen 1993:3; AWS 2009:576), verduidelik Philip (2005) dat ’n multiwoordeenheid anders as normale kombinasies van woorde is, aangesien die multiwoordeenheid nie soos woordgroepe in komponente opgebreek kan word nie. Daar kan met ander woorde nie ander woorde tussen die dele van die multiwoordeenheid gevoeg word nie (vgl. **Dooie, kalm See*), en die multiwoordeenheid is ook nie omstelbaar nie (**die See is Dood*). ’n Multiwoordeenheid is dus ’n spesifieke soort woordgroep, naamlik ’n frasekonstruksie (sien figuur 3).

Multiwoordeiename word onderskei van *saamgestelde eiename* (d.i. komposita) wat ook uit minstens twee dele bestaan, maar wat vas, of met ’n koppelteken vas geskryf word. Voorbeelde sluit in:

(2) *Olifantsfontein* (pleknaam, bestaande uit *olifant+s+fontein*)
Oranjerivier (riviernaam, bestaande uit *oranje+rivier*)
Spektrumdrukkers (besigheidsnaam, bestaande uit *spektrum+drukkers*)
Standaardafrikaans (taalbenaming, bestaande uit *standaard+Afrikaans*).

(3) *Pieter-Willem* (voornaam)
Coetzee-Van Rooy (van)
Pretoria-Oos (geografiese eienaam)
Coca-Cola (merknaam)
Noordwes-Universiteit (besigheidsnaam)
Nieu-Afrikaans (taalnaam).

Oorkoepelend kan multiwoordeiename en saamgestelde eiename *meerledige eiename* genoem word, d.i. eiename wat gevorm word deur twee of meer woorde of woorddele te kombineer. *Meerledige eiename* staan teenoor *ongelede eiename* soos *Durban* (pleknaam), *Botha* (van), *Beeld* (publikasienaam) en *Tamil* (taalnaam).

Met betrekking tot die ortografiese realisering van eiename is veral twee aspekte belangrik, te wete hooflettergebruik en los-/vasskryfwyse (aangedui in figuur 3). Die belangrikste hoofstukke in die AWS in dié verband is hoofstuk 9 (“Hoofletters”), 12 (“Koppeltekens”) en 14 (“Skryfwyse – los en vas”). Met betrekking tot multiwoordeiename is die reëls in tabel 1 tersaaklik; die reëls word nie herhaal of bespreek nie, maar enkele verteenwoordigende voorbeelde uit die AWS (sonder aanduiding van moontlike wisselvorme) word verskaf.

Tabel 1. Ortografiese realisering van multiwoordeiename (AWS)

Verwysing (Reëlnommer)	**Voorbeelde**	**Opmerkings**
9.8	*Loxodonta africana*; *Brassica oleracea var. acephala*	
9.9	*Tweede Wêreldoorlog*; *Alexander die Grote*; *Kommissie vir Waarheid en Versoening*	
9.13	*Nabye Ooste; Persiese Golf*; *Stille Oseaan*	
9.15	*5de Laan*; *26e Straat*	Identies aan 14.35; sien ook 14.34
12.33	*Dooie See*; *Verenigde Arabiese Emirate*; *Verre Wes-Rand*; *Groter Kaapstad*	Word hier meerledige geografiese name genoem
Oriëntasie voor 14.6	*die President Steyn-goudmyn*; *die Victoria & Albert-waterkant*	
14.6	*Bloemfontein Museum*; *Hertzog Prys*; *Nelspruit Hospitaal; Pasella Drukkers*; *Arcadia Straat*; *Bloemfonteinse Museum*; *Kaapse Klopse*; *Randse Afrikaanse Universiteit*	
14.16.a	*Laat Latyn*; *Wit Russies*; *Ou Noors*	Sien ook 14.23.a en 14.23.b

14.23	*Kaapse Afrikaans*; *Moderne Grieks*; *Nuwe Hoogduits*	
14.23.a	*Laat Latyn*; *Wit Russies*; *Ou Noors*	Dieselfde as 14.16.a
14.24	*Derde Wêreld*; *Ou Testament*	Word hier as woordgroepe beskou
14.34	*Derde Straat*; *Vyfde Laan*; *Drie-en-twintigste Straat*	Sien ook 9.15 en 14.35
14.35	*5de Laan*; *26e Straat*	Identies aan 9.15; sien ook 14.34
14.37	*Drie Susters*; *Twee Riviere*; *Veertien Strome*	Staan teenoor 14.38 se *Driefontein* of *Vyfhoek*

Samevattend kan ons dus sê dat ’n multiwoordeienaam ’n frasekonstruksie is wat na ’n benoemde entiteit verwys en wat uit twee of meer ortografiese woorde bestaan; dié ortografiese woorde word deur spasies geskei en gewoonlik met hoofletters geskryf (uitgesonderd funksiewoorde soos lidwoorde, voorsetsels en partikels). Multiwoordeiename staan aan die een kant ortografies teenoor saamgestelde eiename (wat vas of met koppeltekens vas geskryf word), en aan die ander kant semanties en ortografies teenoor woordgroepe (wat nie na benoemde entiteite verwys nie en nie met hoofletters geskryf word nie).

2.2 Komposita, samestellende komposita, afleidings en samestellende afleidings

Vir doeleindes van hierdie artikel is dit nodig om te ondersoek hoe woordgroepe aan die een kant en eiename aan die ander kant ortografies gerealiseer word in komposita en afleidings. Die begrippe *samestellende komposita* (ook bekend as samestellende samestellings) en *samestellende afleiding* kom dus ook aan bod.

2.2.1 Met woordgroepe as komponente

Wanneer woordgroepe as linkerlid in komposita voorkom, is daar drie reëls in die AWS wat die skryfwyse (ortografiese realisering) daarvan bepaal, naamlik reëls 12.4, 12.21 en 14.26. Vergelyk die eerste drie rye van tabel 2 vir ’n opsomming.

Tabel 2. Ortografiese realisering van komposita met en afleidings van woordgroepe (AWS)

Verwysing (Reëlnommer)	**Voorbeelde**	**Opmerkings**
12.4	*ad hoc-komitee* *pro Deo-advokaat* *ex post facto-goedkeuring*	M.b.t. ’n woordgroep van vreemde herkoms wat linkerlid is
12.21	*dag vir dag > dag-vir-dagroetine* *laat maar loop > laat-maar-loophouding* *hand om die nek > hand-om-die-nekhek* *kry my baie jammer > kry-my-baie-jammergesig* [my voorbeeld]	M.b.t. ’n woordgroep van drie of meer woorde wat linkerlid is (*woordgroepkompositum*)
14.26	*breë rand > breërandhoed* *lang been > langbeenspinnekop* *twintig liter > twintigliterhouer* *vyf jaar > vyfjaarplan*	M.b.t. ’n woordgroep wat bestaan uit adjektief/telwoord + naamwoord, wat verbind met ’n naamwoord

		(*samestellende kompositum*)
12.23	*heen en weer > heen-en-weertjie* *traak my nie > traak-my-nieagtig* *oor en weer praat > oor-en-weer-pratery* *laag by die grond > laag-by-die-gronds*	"Indien 'n woordgroep met 'n agtervoegsel verbind, word koppeltekens tussen al die ander woorddele geplaas, maar nie voor die agtervoegsel nie." (AWS 2009:88)
14.25	*alleen loop > alleenloper* *binne aar > binneaars* *tien dele > tiendelig* *vyf jaar > vyfjaarliks* *in diens neem > indiensneming* *ten toon stel > tentoonstelling* *ter tafel lê > tertafellegging*	"Afleidings wat uit woordgroepe, met inbegrip van woordgroepwerkwoorde, gevorm word (samestellende afleiding), word vas geskryf." (AWS 2009:128)

Die AWS verwys na die gevalle by reël 12.21 as *woordgroepsamestellings*, terwyl die gevalle in 14.26 *samestellende samestellings* genoem word. Sonder dat dit in die AWS geëkspliseer word, lyk dit asof die verskil tussen dié twee konstruksies is dat 14.26 geld as die woordgroep uit slegs twee dele bestaan (adjektief/telwoord + naamwoord), terwyl 12.21 geld as die woordgroep uit drie of meer dele bestaan. Die belangrikste verskil tussen 12.4 en die ander twee reëls is dat die ortografiese integriteit van die woordgroep behoue bly wanneer só 'n woordgroep 'n leenwoordgroep (bv. *ad hoc*; *pro Deo*; *ex post facto*) is.

Hieruit volg dat daar drie prototipiese skryfwyses vir samestellings met woordgroepe is; as ons vir die oomblik veronderstel dat *Nuwe Testament* en *Departement Kuns en Kultuur* gewone woordgroepe is, dan sou daar hiervolgens drie moontlike skryfwyses wees:

(4) Reël 12.4: *Nuwe Testamentboek* en *Departement Kuns en Kultuurbeleid*

(5) Reël 12.21: *Departement-Kuns-en-Kultuurbeleid* (geld nie *Nuwe Testament* nie, aangesien dit uit minder as drie dele bestaan)

(6) Reël 14.26: *Nuwetestamentboek* (geld nie *Departement Kuns en Kultuur* nie, aangesien dit uit meer as twee dele bestaan).

Wanneer woordgroepe as basis dien vir afleidings, is daar in die AWS twee toepaslike reëls wat die ortografiese realisering daarvan bepaal, te wete reël 12.23 en reël 14.25. Vergelyk die onderste twee reëls van tabel 2 vir 'n opsomming.

Die AWS maak geen eksplisiete onderskeid tussen dié twee reëls nie: dit is volkome onduidelik wanneer watter een van die reëls sou geld wanneer 'n woordgroep as basis vir afleiding dien. Na analogie van reël 12.21 (sien (4) hier bo) sou 'n mens geneig wees om te dink dat reël 12.23 ook geld vir woordgroepe wat uit drie of meer woorde bestaan, maar voorbeelde soos *indiensneming* (< *in diens neem*) en *tentoonstelling* (< *ten toon stel*) repudieer so 'n aanname. Verdere oorweging van hierdie en ander diskordansies in die AWS[2] val buite die bestek van hierdie artikel.

Op sterkte van dié twee reëls sou ons dus ook kon sê dat daar twee moontlike skryfwyses vir afleidings met *Nuwe Testament* en *Departement Kuns en Kultuur* is (as ons vir doeleindes van die argument beide as gewone woordgroepe aanvaar):

(7) Reël 12.23: *Nuwe-Testamentikus* en *Departement-Kuns-en-Kultuuragtig*

(8) Reël 14.25: *Nuwetestamentikus* en *Departementkunsenkultuuragtig*

In die vyf bogenoemde reëls word daar slegs op twee plekke multiwoordeiename as voorbeelde aangebied, te wete by:

(9) Reël 14.25: *Derde Wêreld > Derdewêrelds; Ou Testament > Ou-Testamenties* (Dit is onduidelik waarom *Derdewêrelds* vas geskryf word, maar *Ou-Testamenties* met ’n koppelteken vas geskryf word.)

(10) Reël 14.26: *Derde Wêreld > Derdewêreldlande; sewende dag > Sewendedagadventis*

Hiermee impliseer die samestellers van die AWS dat multiwoordeiename ook gesien moet word as woordgroepe en dus beskikbaar is vir die prosesse van samestellende samestelling en samestellende afleiding; dit is dus potensieel moontlik dat die ander genoemde reëls ook op multiwoordeiename van toepassing sou kon wees.

2.2.2 Met multiwoordeiename as komponente

In die AWS is talle reëls wat handel oor die ortografiese realisering van komposita met en afleidings van eiename; ’n opsomming hiervan verskyn in Tabel 3. Aangesien die AWS nie formeel erkenning gee aan die kategorie “multiwoordeienaam” nie, word voorbeelde in dié tabel ingesluit waar multiwoordeiename wel as voorbeelde aangedui word. Sulke voorbeelde word in vetdruk aangedui. Die eerste vier rye van die tabel het betrekking op komposita, terwyl die res oor afleidings handel; die laaste twee rye het spesifiek betrekking op verbaliserings.

Tabel 3. Ortografiese realisering van komposita met en afleidings van (multiwoord)eiename (AWS)

Verwysing (Reëlnommer)	**Voorbeelde**	**Opmerkings**
9.7.b	*pietretiefbroodboom*; teenoor uitsonderings soos *Geoffroy se saalneusvlermuis; Moses-se-kierie; Major Mitchell-kaketoe; Père David-hert; Prins van Wallisheide*	Het betrekking op gewone name van plante en diere
14.7	*Mercedes-Benztoerbus*; *Nissanonderdele*	Geen melding word van multiwoordeiename gemaak nie; ook geen kruisverwysing na 14.26 nie
14.26	*Derde Wêreld > Derdewêreldland sewende dag > Sewendedagadventis*	Geen verwysing na 14.7 nie
14.26.d	*Demokratiese Partykandidaat*; *Historiese Monumentekommissie*; *Middellandse Seegebied*; *Nasionale Vrouemonumentkommissie*; *Tweede Konsiliebesluit*	Word as uitsonderings beskou wat a.g.v. tradisie los geskryf word
9.12	*anti-Engels*; *pro-Afrikaans*	Het betrekking op afleidings

		van taalbenamings met prefikse van Klassieke oorsprong
9.19	*Dantesk* of *dantesk*; *Hertzogiet* of *hertzogiet*	Uitgesonderd geografiese eiename en taalbenamings
9.20	*Hellenis of hellenis*; *Germanisme of germanisme*; *Neerlandistiek of neerlandistiek*; *Anglisisties of anglisisties*	Het betrekking op afleidings met *·is*, *·isme*, *·istiek* of *·isties*
9.21	*Gautenger*; *KwaZulu-Nataller*; *Anglo-Amerikaans*	Het betrekking op geografiese eiename en taalbenamings, uitgesonderd afleidings met *·is*, *·isme*, *·istiek* of *·isties*
9.22	*onafrikaans*; *onbybels*; *onskandinawies*	Het betrekking op bestaande afleidings wat deur *on·* geprefigeer word. Staan in kontras met 9.12, 12.13 en 12.28
12.13	*pan-Amerikanisme*	Hoofletterwoord behou as woorddeel hoofletter ná 'n koppelteken; staan in kontras met 9.22
12.28	*eks-Blou Bul*; *nie-Suid-Afrikaans*; *non-Calvinisme*; *oud-Tukkie*	Ná *eks·*, *nie·*, *non·* en *oud·* behou eiename hoofletters
12.33.c	*Groter-Kaapstadse*; *Nabye-Oosterse*; *Verre-Oosterse*; *Verre-Wes-Randse*	Word eksplisiet afleidings van meerledige geografiese name genoem; sien ook 14.25
12.34.b	*Jan Kempdorper*; *Noordsese*; *Piet Retiefse*	In direkte kontras met 12.33.c/14.25
14.16.b	*Ou Noors* maar *Ou-Noorse woord*; *Nuwe Nederlands* maar *Nuwe-Nederlandse vorm*	14.25 word as rede verskaf
14.23.b	*Ou Noors* maar *Ou-Noorse woord*; *Nuwe Nederlands* maar *Nuwe-Nederlandse vorm*	14.25 word as rede verskaf
14.25	*Derde Wêreld > Derdewêrelds* *Ou Testament > Ou-Testamenties*	Volgens 12.33.c geld dié reël ook *Groter-Kaapstadse; Nabye-Oosterse*
14.25.d	*New York > New Yorkse* *Sri Lanka > Sri Lankaan* *Nova Scotia > Nova Scotiaan*	"Wanneer 'n eienaam uit twee los geskrewe dele bestaan, word die skryfwyse in afleidings daarvan gehandhaaf …" (AWS 2009:129)
9.23	*amerikaniseer*; *verafrikaans*; *ontengels*	Het betrekking op verbaliserings
9.24	*geamerikaniseerde*; *verafrikaanste*; *ontengelsing* [my voorbeeld]; *afrikanisasie*	Het betrekking op verdere afleidings op basis van geverbaliseerde vorm

Uit dié opsomming kan ’n aantal algemene opmerkings met betrekking tot komposita gemaak word:

- **Behou ortografiese integritei**t: In sommige gevalle word die ortografiese integriteit van multiwoordeiename in komposita behou, byvoorbeeld die uitsonderings by reël 9.7.b (*Prins van Wallisheide*), die oriëntasie voor 14.6 (*die President Steyn-goudmyn*; sien tabel 1) en die “tradisionele gevalle” by 14.26.d (*Middellandse Seegebied*). Ofskoon geen voorbeelde met multiwoordeiename by 14.7 ingesluit is nie, is daar geen rede om te glo dat die bedoeling daar anders sou wees as om die ortografiese integriteit van die eienaam te behou nie (vergelyk byvoorbeeld *Mercedes-Benztoerbus*).
- **Behou aanvangshoofletter, maar skryf vas**: Ingevolge reël 14.26 blyk die verstekskryfwyse vir multiwoordeiename wat uit ’n adjektief/telwoord en naamwoord bestaan (*Derde Wêreld*) en wat dan met ’n naamwoord verbind, vas en met ’n hoofletter te wees (*Derdewêreldland*).
- **Met kleinletter, vas**: Ingevolge reël 9.7 word multiwoordeiename wat in algemene plant- en diername voorkom, vas en met ’n kleinletter geskryf (*pietretiefbroodboom*).

Insgelyks merk ’n mens met betrekking tot afleidings die volgende op:

- **Behou ortografiese integriteit**: Die enigste eksplisiete vermelding van die skryfwyse van afleidings met multiwoordeiename is in 14.25.d (*New Yorkse*), terwyl voorbeelde in 12.28 (*eks-Blou Bul*) en 12.34.b (*Jan Kempdorper*) dieselfde skryfwyse suggereer. Hierdie beginsel word ook deurgaans in die lys met buitelandse geografiese eiename en hulle afleidings (AWS 2009:522-545) gehandhaaf; vergelyk onder andere:

(11) Ho Tsji Minhstad – Ho Tsji Minhstadse kultuur – Ho Tsji Minhstadter

(12) Port Said – Port Saidse kultuur – Port Saider

(13) Maagde-eilande – Maagde-eilandse kultuur – Maagde-eilander

(14) Nieu-Delhi – Nieu-Delhiese kultuur – Nieu-Delhiër.

Let egter daarop dat talle multiwoord buitelandse geografiese eiename nie beskikbaar is vir afleiding nie, maar dat daar eerder van omskrywing gebruik gemaak moet word, byvoorbeeld:

(15) Las Palmas – kultuur van Las Palmas – inwoner van Las Palmas

(16) Port-au-Prince – kultuur van Port-au-Prince – inwoner van Port-au-Prince.

Insgelyks ondergaan ’n groot aantal van sulke multiwoordeiename inkorting tydens afleiding, byvoorbeeld:

(17) Verenigde State van Amerika – Amerikaanse – Amerikaner

(18) Groot-Brittanje – Britse – Brit.

- **Behou hoofletters, maar skryf dele met koppelteken vas**: In 12.33.c word voorbeelde soos *Groter Kaapstad* en *Verre Ooste* as *meerledige eiename* beskryf, en die skryfwyse van afleidings met sulke eiename word aangedui as *Groter-Kaapstadse* en *Verre-Oosterse*. Die voorbeelde by 14.16.b en 14.23.b (*Ou-Noorse woorde*) ondersteun dié skryfwyse en verwys na reël 14.25 (*Ou-Testamenties*); dáár staan *Derdewêrelds* as voorbeeld wel in teenstelling met hierdie skryfwyse, net soos wat *eks-Blou Bul* in 12.28 nie hierdie patroon ondersteun nie.
- **Behou hoofletters**: Ook in die geval van prefigering behou eiename hulle hoofletters, soos bereël word in reël 12.13 (*pan-Amerikanisme*); geen voorbeelde van multiwoordeiename word egter verskaf nie. Hierdie skryfwyse word ondersteun deur 9.12 (*pro-Afrikaans*) en 12.28 (*eks-Blou Bul*), asook 9.21 met betrekking tot geografiese eiename en taalname (*Anglo-Amerikaans*). Daar blyk egter twee uitsonderings hierop te wees, naamlik prefigering met *on·*, wat altyd met ’n kleinletter geskryf word (reël 9.22, byvoorbeeld *onafrikaans*), asook hoof-/kleinletterwisselvorme in afleidings met *·is*, *·isme*, *·istiek* en *·isties* (reël 9.20, byvoorbeeld *Anglisisties* of *anglisisties*). In geen van dié twee reëls word voorbeelde van saamgestelde of multiwoordeiename aangebied nie, maar die implikasie is dat ’n mens ?*onnew yorks/onnew-yorks/onnewyorks* moet skryf, en dat ?*New Yorkisme/New-Yorkisme/Newyorkisme* én ?*new yorkisme/new-yorkisme/newyorkisme* aanvaarbaar is.[3]
- **Wisseling tussen hoof- en kleinletters**: Ingevolge reël 9.19 word afleidings van eiename – uitgesonderd geografiese eiename en taalbenamings – met hoof- of kleinletters geskryf: *Dantesk* of *dantesk*. Geen voorbeelde van multiwoordeiename word ingesluit nie, maar dit impliseer wel dat *Tweede Wêreldoorlogse/Tweede-Wêreldoorlogse/Tweedewêreldoorlogse* én ?*tweede wêreldoorlogse/tweede-wêreldoorlogse/tweedewêreldoorlogse* korrek sou kon wees. Die skryfwyse met ’n koppelteken word in die woordelysgedeelte van die AWS bevestig met die opname van *nuwe-testamenties*, *nuwe-testamentikus*, *ou-testamenties* en *ou-testamentikus*; die reëlverwysing by hierdie inskrywings impliseer dat die hooflettervorme ook korrek is (byvoorbeeld *Nuwe-Testamenties*). Soos hier bo aangetoon, geld dieselfde wisseling enige afleidings met *·is*, *·isme*, *·istiek* en *·isties* (reël 9.20).
- **Slegs met kleinletters**: Soos hier bo aangedui, bepaal reël 9.22 dat afleiding met *on·* altyd met ’n kleinletter geskryf word (*onafrikaans*). Ook verbaliserings word ingevolge reël 9.23 slegs met kleinletters geskryf, byvoorbeeld *amerikaniseer* en *verafrikaans*. Dit geld ook verdere afleidings van sulke geverbaliseerde vorme, byvoorbeeld *geamerikaniseerde* en *verafrikaanste* (sien reël 9.24). Geen melding word gemaak of voorbeelde verskaf van saamgestelde eiename wat met ’n koppelteken vas geskryf word nie, maar die afleiding wat gemaak kan word, is dat ’n mens ?*suid-afrikaniseer/suidafrikaniseer* sal moet skryf. Ook multiwoordeiename kom nie ter sprake nie, en ’n mens moet dus aflei dat ons ?*new yorkifiseer/new-yorkifiseer/newyorkifiseer* moet skryf.[4]

Uit hierdie bespreking blyk dit dat daar geen eenduidige reëls in die AWS vir die ortografiese realisering van komposita met en afleidings van multiwoordeiename is nie, terwyl sommige reëls of opmerkings mekaar selfs weerspreek.

In tabel 4 word ’n opsomming gegee van potensiële (teoreties moontlike) skryfwyses van komposita met en afleidings van *Middellandse See* (en *New York* met betrekking tot verbalisering deur middel van suffigering). Gevalle waarvan dit onwaarskynlik is dat hulle

sal voorkom (aangesien daar nie in die AWS voorbeelde of spesifieke reëls is wat dié skryfwyse ondersteun nie), word in grys gemerk. Let daarop dat daar ook in beginsel geen reël is wat **middellandseseegebied* ondersteun nie; dié skryfwyse sou wel moontlik wees in die geval van gewone benamings van plante en diere, argumentsonthalwe ?*middellandseseevlassaad*, of ?*middellandseseemeeu* (in Afrikaans bekend as die swartkopmeeu).

Hierdie moontlikhede is egter lank nie die enigste moontlikhede wat tot die skrywer se beskikking is nie. In die hieropvolgende korpusondersoek word aandag gegee aan alle moontlike opsies waaruit ’n skrywer sou kon kies.

Tabel 4. Moontlike en waarskynlike skryfwyses

	Behou hoofletters Komposita	Slegs beginhoofletter	Beginkleinletter
Los	Middellandse Seegebied†	Middellandse seegebied	middellandse seegebied
Koppelteken vas	Middellandse-Seegebied	Middellandse-seegebied	middellandse-seegebied
Vas	MiddellandseSeegebied	Middellandseseegebied	*middellandseseegebied
	Afleiding: adjektivering (attributiewe posisie) (suffigering)		
Los	Middellandse Sese	Middellandse sese	middellandse sese
Koppelteken vas	Middellandse-Sese	Middellandse-sese	middellandse-sese
Vas	MiddellandseSese	Middellandsesese	middellandsesese
	Afleiding: adjektivering (attributiewe posisie) (prefigering)		
Los	anti-Middellandse Sese	anti-Middellandse sese	antimiddellandse sese
Koppelteken vas	anti-Middellandse-Sese	anti-Middellandse-sese	antimiddellandse-sese
Vas	anti-MiddellandseSese	anti-Middellandsesese	antimiddellandsesese
	Afleiding: verbalisering (prefigering)		
Los	ver-Middellandse See	ver-Middellandse see	vermiddellandse see
Koppelteken vas	ver-Middellandse-See	ver-Middellandse-see	vermiddellandse-see
Vas	ver-MiddellandseSee	ver-Middellandsesee	vermiddellandsesee
	Afleiding: verbalisering (suffigering)		
Los	New Yorkifiseer	New yorkifiseer	new yorkifiseer
Koppelteken vas	New-Yorkifiseer	New-yorkifiseer	new-yorkifiseer
Vas	NewYorkifiseer	Newyorkifiseer	newyorkifiseer

† Ingevolge reël 14.7 in die AWS sou *Middellandse Seegebied* en *Middellandse See-gebied* aanvaarbare wisselvorme wees.

3. Korpusondersoek

3.1 Formulering van ondersoek

3.1.1 Doel en navorsingsvraag

Ten einde die realisering van ortografiese vorme in werklike taalgebruik te ondersoek, is ’n beperkte, verkennende korpusondersoek onderneem. Die basiese navorsingsvraag is: Word die ortografiese integriteit van multiwoordeiename behou in komposita met en afleidings van sulke multiwoordeiename in korpusdata? Oftewel: Is daar ’n positiewe korrelasie tussen die

ortografiese vorm van ’n multiwoordeienaam en die ortografiese vorm van die multiwoordeienaam in komposita daarmee of afleidings daarvan?

Die enigste twee aspekte wat aandag geniet het, is:

1. Hooflettergebruik: Word die hoofletters van multiwoordeiename behou in komposita daarmee en geadjektiveerde vorme daarvan?
2. Los-/vasskryfwyse: Word die spasies in multiwoordeiename behou in komposita daarmee en geadjektiveerde vorme daarvan?

Vir hierdie beperkte ondersoek word slegs aan tweeledige multiwoordeiename aandag gegee (gevalle soos *Departement Kuns en Kultuur* word dus nie ondersoek nie), nominale samestellings met slegs een komponent (komplekse gevalle soos ?*anti-Verenigde State van Amerika-sentimente* word nie ingesluit nie), en gesuffigeerde adjektiverings (verbaliserings soos ?*New Yorkifiseer* word buite rekening gelaat, soos ook gevalle met prefikse, soos ?*anti-Middellandse Sese*).

Hierdie voorlopige ondersoek het ook ten doel om moontlike veranderlikes te identifiseer waaraan in ’n meer uitgebreide korpusondersoek praktiese uitvoering gegee kan word.

3.1.2 Hipoteses en veranderlikes

Geen formele hipoteses wat statisties bewys moes word, is gestel nie. Die doel van die korpusondersoek was slegs om te bepaal watter realiserings in ’n beperkte aantal gevalle die meeste voorkom, en om uit dié frekwensietellings kategoriale afleidings te maak. Deurdat van verskillende tipes korpusmateriaal gebruik gemaak is, is die enigste onafhanklike veranderlike wat indirek verreken word, die potensiële invloed van teksredigering op skryfwyses.

3.2 Operasionalisering van ondersoek

3.2.1 Korpusse

Soektogte is in drie verskillende soorte korpusse op drie platforms gedoen (in die week van 10–17 Junie 2016):

- **Virtuele Instituut vir Afrikaans (VivA) se Korpusportaal: geredigeerde tekste**. Deur van die aanlyn WhiteLab-platform gebruik te maak, kan kragtige soektogte in sewe verskillende korpusse (~85 450 000 woorde) gedoen word (sien tabel 5). Die meerderheid tekste in dié korpusse is afkomstig van bronne wat waarskynlik aan een of ander proses van teksredigering onderworpe was.
- **Afrikaanse Wikipedia: semigeredigeerde tekste**. Deur van die soekfunksie op die webblad af.wikipedia.org gebruik te maak, kan basiese soektogte in alle Afrikaanse Wikipedia-artikels gedoen word. Aangesien die soekalgoritme bekende beperkings het (soos byvoorbeeld dat die koppelteken in soektogte ook as ’n spasie geïnterpreteer word), is alle soekresultate in Notepad++ verwerk om frekwensietellings te bepaal. Gegewe dat Wikipedia-artikels deur die Wikigemeenskap geredigeer word, maar nie deur professionele taalredigeerders nie, word dié tekste as semigeredigeerde tekste

beskou. (Sien ook Kuperman en Bertram 2012 met betrekking tot die gebruik van Wikipedia as ’n korpus in soortgelyke navorsing.)

- **Google: gemengde tekste**. Deur van Google se gevorderdesoekfunksie gebruik te maak, kan in Afrikaanse tekste op die internet gesoek word. Alle soektogte is uitgevoer met die “Bladsye in Afrikaans”-opsie aangeskakel, en met uitsondering van tekste op die Afrikaanse Wikipedia (-site:af.wikipedia.org), Maroela Media (-site:maroelamedia.co.za), Radio Sonder Grense (-site:rsg.co.za) en WatKykJy (-site:watkykjy.co.za). Tekste wat ingesluit is in ander korpusse op VivA se Korpusportaal (byvoorbeeld koeranttekste uit die Media24-koerante, of regeringsdokumente), is nie afsonderlik verwyder nie, aangesien dit ’n onbegonne taak was. Aangesien die Google-soekalgoritme bekende beperkings het (soos byvoorbeeld dat die koppelteken in soektogte ook as ’n spasie geïnterpreteer word), en aangesien “[t]he query, search and ranking optimization techniques [search engines] have adopted can either assist or sabotage a scholar’s quest” (Fletcher 2005), is alle soekresultate – soos dit in elke Google-soekresultaatbladsy verskyn – in Notepad++ geplak; daarna kon met reëlmatige uitdrukkings na verskeie vorme gesoek word om frekwensietellings te bepaal. Alle resultate is ook handmatig gekontroleer om seker te maak dat dit wel van Afrikaanse webblaaie kom. Die veronderstelling is dat tekste wat op dié manier verkry is, meestal nie deur professionele teksredigeerders versorg is nie.

Tabel 5. Korpusse op VivA se Korpusportaal

Korpus	Beskrywing	Aantal woorde (~)	Verwysing
NCHLT-Afrikaanse korpus 1.0	'n Versameling regeringsdokumente van die onderskeie webblaaie van die Suid-Afrikaanse regeringswebwerwe en elektroniese publikasies	3 200 000	(Departement Kuns en Kultuur en CTexT 2013)
NWU/Maroela Media-korpus 1.0	Versameling webartikels soos op die Maroela Media-webtuiste gepubliseer	800 000	(Maroela Media en CTexT 2015)
NWU/Lapa-korpus 1.0	Korpus van Afrikaanse boeke deur Lapa Uitgewers gepubliseer	900 000	(Lapa-uitgewers en CTexT 2015)
PUK/Protea Boekhuis-korpus 2.0	Korpus van Afrikaanse boeke deur die uitgewer Protea Boekhuis gepubliseer	12 000 000	(Protea Boekhuis en CTexT 2015)
RSG-nuuskorpus 2.0	Versameling nuusbulletins van 2005–2015 soos op Radio Sonder Grense uitgesaai	20 500 000	(Radio Sonder Grense en CTexT 2015)
Taalkommissie-korpus 1.1	'n Gestratifiseerde korpus soos gebruik deur die Afrikaanse Taalkommissie bestaande uit 'n verskeidenheid genres en domeine, insluitend: wetenskaplike tydskrifte, koerante, letterkundige werke, informele en formele skryfstukke	47 000 000	(Taalkommissie (samest.) 2011)
Watkykjy.co.za-korpus 1.0.	Versameling webartikels soos op watkykjy.co.za gepubliseer	1 043 756	(WatKykJy 2015)

Oor die gebruik van die web as korpus is al baie geskryf (vergelyk byvoorbeeld die jaarlikse kongres van die Web-as-Korpus-bepaaldebelangegroep van die Association for Computational Linguistics),[5] veral ook met betrekking tot die betroubaarheid, akkuraatheid en herhaalbaarheid van Google-resultate vir linguistiese ondersoeke (Lüdeling e.a. 2006). Die gebruik van Google- en Wikipedia-soektogte in hierdie navorsing word gemotiveer teen die agtergrond van wat Fletcher (2005) "web hunting" noem, d.i. om websoektogte te gebruik om rare voorkomste van konstruksies (wat nie noodwendig in standaardkorpusse voorkom nie) te soek. Byvoorbeeld, in Korpusportaal word slegs die ortografiese vorme *Ou Testamenties* en *Ou-Testamenties* aangetref; 'n allesomvattende Google-soektog na alle moontlike ortografiese variante toon egter nog nege ander variante (sien tabel 7). As 'n mens slegs resultate uit Korpusportaal sou gebruik, sou dit 'n eensydige voorstelling van die taalwerklikheid weergee.

3.2.2 Materiaal

Met betrekking tot samestellings is vier gevalle ondersoek, naamlik:

> (19) *Derde Wêreld+land* (in die AWS opgeneem as *Derdewêreldland*, met verwysing na 14.26)
>
> (20) *Middellandse See+gebied* (in die AWS opgeneem as *Middellandse Seegebied*, met verwysing na 14.26.d)
>
> (21) *Dooie See+rolle* (nie in die AWS opgeneem nie, maar *Dooie See* wel, met verwysing na 12.33)
>
> (22) *Konstitusionele Hof+regter* (nie in die AWS opgeneem nie, maar *Konstitusionele Hof* wel, sonder enige reëlverwysing).

Gevalle (19) en (20) is gekies omdat hulle in die AWS opgeneem is, maar met verskillende skryfwyses (om redes soos in die reëlverwysings aangedui). Geval (21) is gekies omdat dit moontlik analogies aan *Middellandse See+gebied* geskryf sou kon word, dit waarskynlik frekwent in die korpusmateriaal sou voorkom, en die samestelling op sigself as 'n multiwoordeienaam geïnterpreteer sou kon word (d.i. as ergoniem, geskryf as *Dooie See Rolle;* in tabel 8 is dit duidelik dat hierdie skryfwyse dan ook die mees waarskynlike skryfwyse in geredigeerde bronne is). Geval (22) is gekies as voorbeeld van 'n multiwoordeienaamkompositum wat waarskynlik as soortnaam geïnterpreteer sal word, en waarskynlik frekwent in korpusse sou voorkom. In alle gevalle is frekwensietellings op lemmas bereken (d.i. *Derde Wêreld+lande* is genormaliseer na *Derde Wêreld+land*, ens.).

Gegewe die fokus van hierdie korpusondersoek, is slegs op variasie ten opsigte van hoof- en kleinletters gekonsentreer, teenoor spasies, koppeltekens en vaste skryfwyse; ander spellingvariasie (soos byvoorbeeld **Middelandse Seegebied*) is buite rekening gelaat. Die potensiële keuseruimte (Wallis, Bowie en Aarts 2012) is 'n matriks van agt (2x2x2; twee opsies – hoof-/kleinletter – in drie posisies), by nege (3x3; drie opsies – spasie/koppelteken/vas – in twee posisies); die totale aantal moontlike opsies is dus 72. In werklikheid het dit uit die ondersoekte korpusdata geblyk dat (ten minste vir die vier gevalle wat ondersoek is) die keuseruimte maksimaal 23 groot is; gevalle soos **DooieSeeRolle* of **dooiesee-Rolle* (wat teoreties moontlik is) kom gewoon nie in die data voor nie.

In tabel 6 word aan die hand van *Konstitusionele Hof+regter* getoon hoe die data geënkodeer is. In die eerste kolom verskyn die lemma en in die drie laaste kolomme die frekwensietellings soos bepaal in elk van die drie korpusbronne; die frekwentste voorkoms in ’n bepaalde bron word met vetdruk aangedui. In die tweede kolom word die patroon van die lemma in terme van hoofletters (**H**) of kleinletters (**k**), asook los (byvoorbeeld **k k k**), vas (byvoorbeeld **kkk**), of met koppelteken vas (byvoorbeeld **k-k-k**) aangedui. Kolomme drie tot vyf het slegs betrekking op die multiwoordeienaam se skryfwyse soos dit in die bepaalde kompositum gerealiseer word. In kolom drie word die patroon van die multiwoordeienaam aangedui, in kolom vier of dit los, vas of met ’n koppelteken geskryf word, en in kolom vyf of beide hoofletters van die multiwoordeienaam in die kompositum behoue gebly het.

Tabel 6. Frekwensietellings van Konstitusionele Hof+regter

Lemma	**Patroon-Lemma**	**Patroon-MWE**	**Los/ Vas/ Koppel**	**Hoofletters**	**VivA**	**Wikipedia**	**Google**
Konstitusionele Hofregter	H Hk	H H	Los	Ja	1	0	0
Konstitusionele Hof-regter	H H-k	H H	Los	Ja	**7**	0	2
Konstitusionele Hof Regter	H H H	H H	Los	Ja	0	0	2
Konstitusionele Hof-Regter	H H-H	H H	Los	Ja	0	0	0
Konstitusionele Hof regter	H H k	H H	Los	Ja	0	0	5
Konstitusionele hofregter	H kk	H k	Los	Nee	0	0	1
Konstitusionele hof-regter	H k-k	H k	Los	Nee	2	0	1
Konstitusionele hof regter	H k k	H k	Los	Nee	0	0	0
konstitusionele hofregter	k kk	k k	Los	Nee	0	0	**10**
konstitusionele hof-regter	k k-k	k k	Los	Nee	0	0	3
konstitusionele hof regter	k k k	k k	Los	Nee	0	0	5
Konstitusionelehofregter	Hkk	Hk	Vas	Nee	0	0	1
Konstitusionelehof-regter	Hk-k	Hk	Vas	Nee	0	0	0
Konstitusionelehof regter	Hk k	Hk	Vas	Nee	0	0	1
Konstitusionelehof Regter	Hk H	Hk	Vas	Nee	0	0	0
konstitusionelehofregter	kkk	kk	Vas	Nee	0	0	2
konstitusionelehof-regter	kk-k	kk	Vas	Nee	2	0	2
konstitusionelehof regter	kk k	kk	Vas	Nee	0	0	0
Konstitusionele-Hofregter	H-Hk	H-H	Koppel	Ja	0	0	0
Konstitusionele-Hof-regter	H-H-k	H-H	Koppel	Ja	0	0	0
Konstitusionele-hofregter	H-kk	H-k	Koppel	Nee	0	0	0
konstitusionele-hof-regter	k-k-k	k-k	Koppel	Nee	1	0	1
konstitusionele-hof regter	k-k k	k-k	Koppel	Nee	0	0	0

Met betrekking tot geadjektiveerde vorme is vyf gevalle ondersoek, naamlik:

(23) *Nabye Ooste·ADJZ* (in die AWS opgeneem as *Nabye-Oosters*, met verwysing na 12.33.c)

(24) *Verre Ooste·ADJZ* (in die AWS opgeneem as *Verre-Oosters*, met verwysing na 12.33.c)

(25) *Ou Testament·ADJZ* (in die AWS opgeneem as *Ou-Testamenties*, met verwysing na 14.25)

(26) *Nuwe Testament·ADJZ* (in die AWS opgeneem as *Nuwe-Testamenties of Nieu-Testamenties*, sonder reëlverwysing; *Nieu-Testamenties* is nie verder ondersoek nie)

(27) *Derde Wêreld·ADJZ* (in die AWS opgeneem as *Derdewêrelds*, met verwysing na 14.25).

Al vyf gevalle is gekies omdat hulle in die AWS opgeneem is, met *Derdewêrelds* wat in skryfwyse verskil van die ander vier gevalle. *Nabye Ooste* en *Verre Ooste* is semanties vergelykbaar met *Middellandse See+gebied* (geografiese gebied), terwyl *Ou Testament* en *Nuwe Testament* semanties naby aan *Dooie See+rolle* is. In alle gevalle is frekwensietellings ook op lemmas (*Derdewêreldse* is dus genormaliseer na *Derdewêrelds*, ens.) bereken. Vir gesuffigeerde vorme is die potensiële keuseruimte twaalf groot (d.i. (2x2)x3); in werklike taalgebruik blyk dit dat slegs vorme soos **derdeWêrelds* (d.i. **kH**) nie voorkom nie, en die werklike keuseruimte is dus elf groot. Dieselfde konvensie as vir komposita is in die enkodering van data gevolg, soos in tabel 7 aan die hand van *Ou Testament·ADJZ* geïllustreer word.

Tabel 7. Frekwensietellings van *Ou Testament·ADJZ*

Lemma	Patroon-MWE	Los/ Vas/ Koppel	Hoofletters	VivA	Wikipedia	Google
Ou Testamenties	H H	Los	Ja	63	**17**	**151**
ou testamenties	k k	Los	Nee	0	2	11
ou Testamenties	k H	Los	Nee	0	0	2
Ou testamenties	H k	Los	Nee	0	0	4
Outestamenties	Hk	Vas	Nee	0	0	23
outestamenties	kk	Vas	Nee	0	0	25
OuTestamenties	HH	Vas	Ja	0	0	18
Ou-Testamenties	H-K	Koppel	Ja	**550**	4	131
ou-testamenties	k-k	Koppel	Nee	0	0	12
Ou-testamenties	H-k	Koppel	Nee	0	1	4
ou-Testamenties	k-H	Koppel	Nee	0	0	2

3.3 Resultate

Aangesien hierdie ondersoek oor drie verskillende soorte korpusmateriaal strek, en die totale populasie (d.i. aantal woorde) van die Wikipedia- en Google-materiaal nie bepaalbaar is nie,

word daar in navolging van Wallis, Bowie en Aarts (2012) met ’n keusemodel gewerk waar elke keuseopsie vry is om te varieer, d.i. “a genuine choice exists and all cases could theoretically be of one type or the other” (Wallis e.a. 2012:4). Die waarskynlikheid *P* dat ’n bepaalde vorm n_1 uit die totale aantal werklike keuseopsies (23 in die geval van komposita en elf in die geval van afleidings) gekies sal word, is:

$$P\ (n_1 \mid \{n_1, n_2, n_3 \ldots n_23\}) = F\ (n_1)/F(\{n_1, n_2, n_3 \ldots n_23\})\ ,$$

waar $F(n_1)$ die totale aantal gevalle (ongenormaliseerde frekwensie) van die eerste ortografiese variant is, ens. (Wallis e.a. 2012:4; Baayen 2003). Die waarskynlikheid *P* word uitgedruk as ’n waarde tussen 0 (onwaarskynlik) en 1 (volkome waarskynlik). Aangesien daar nou met waarskynlikhede gewerk word, kan ook met die gemiddeld van waarskynlikhede oor die drie korpusse heen gewerk word. In tabel 8 word die waarskynlikhede in elke korpus (P_{VivA}, P_{Wiki} en P_{Goog}) vir *Dooie See+rolle* geïllustreer, met die gemiddelde waarskynlikheid $P_{\text{Gemiddeld}}$ in die laaste kolom. In alle hieropvolgende tabelle word met sodanige waarskynlikhede gewerk.

Tabel 8. Waarskynlikhede van *Dooie See+rolle*

Lemma	P_{VivA}	P_{Wiki}	P_{Goog}	$P_{\text{Gemiddeld}}$
Dooie Seerolle	0,01	0,09	0,25	0,12
Dooie See-rolle	0,31	**0,91**	**0,30**	**0,51**
Dooie See Rolle	**0,63**	0,00	0,09	0,24
Dooie See-Rolle	0,00	0,00	0,02	0,01
Dooie See rolle	0,00	0,00	0,11	0,04
Dooie seerolle	0,00	0,00	0,00	0,00
Dooie see-rolle	0,00	0,00	0,00	0,00
Dooie see rolle	0,00	0,00	0,00	0,00
dooie seerolle	0,00	0,00	0,02	0,01
dooie see-rolle	0,00	0,00	0,00	0,00
dooie see rolle	0,00	0,00	0,00	0,00
Dooieseerolle	0,02	0,00	0,07	0,03
Dooiesee-rolle	0,02	0,00	0,05	0,02
Dooiesee rolle	0,00	0,00	0,02	0,01
Dooiesee Rolle	0,00	0,00	0,02	0,01
dooieseerolle	0,00	0,00	0,01	0,00
dooiesee-rolle	0,00	0,00	0,01	0,00
dooiesee rolle	0,00	0,00	0,02	0,01
Dooie-Seerolle	0,00	0,00	0,02	0,01
Dooie-See-rolle	0,00	0,00	0,00	0,00
Dooie-seerolle	0,00	0,00	0,01	0,00
dooie-see-rolle	0,00	0,00	0,00	0,00
dooie-see rolle	0,00	0,00	0,00	0,00

Die eerste vraag wat in hierdie korpusondersoek gevra word, is of die hoofletters van multiwoordeiename behoue bly in komposita daarmee en geadjektiveerde vorme daarvan. Overgeset synde: Wat is die waarskynlikheid dat die hoofletters van ’n multiwoordeienaam behou sal word in ’n kompositum daarmee en geadjektiveerde vorm daarvan?

In tabel 9 word die waarskynlikhede van die verskillende komposita opgesom in terme van of die hoofletter van die multiwoordeienaam behou word (“Ja”) of nie (“Nee”). In die laaste ry word die gemiddelde waarskynlikheid van die drie korpusse tesame gegee. (In die geval van *Konstitusionele Hof+regter* word die waarskynlikhede van die Wikipedia-data buite rekening gelaat by die berekening van die gemiddeldes.)

Tabel 9. Hoof-/kleinletters: Waarskynlikhede van verskillende komposita

Komp.	*Middellandse See+gebied*		*Dooie See+rolle*		*Konstitusionele Hof+regter*		*Derde Wêreld+land*	
hl.?	Ja	Nee	Ja	Nee	Ja	Nee	Ja	Nee
P_{VivA}	**1,00**	0,00	**0,95**	0,05	**0,62**	0,38	0,20	**0,80**
P_{Wiki}	**1,00**	0,00	**1,00**	0,00	0,00	0,00	**0,60**	0,40
P_{Goog}	**0,82**	0,18	**0,79**	0,21	0,25	**0,75**	0,13	**0,87**
$P_{\text{gemiddeld}}$	**0,94**	0,06	**0,91**	0,09	0,43	**0,57**	0,31	**0,69**

In die gevalle van *Middellandse See+gebied* en *Dooie See+rolle* is dit duidelik dat die waarskynlikheid baie groot is dat die hoofletters van die multiwoordeienaam behoue bly in komposita. By *Konstitusionele Hof+regter* en *Derde Wêreld+land* sien ons ’n meer eweredige verspreiding van waarskynlikhede, met ’n verskuiwing dat dit meer waarskynlik is dat die hoofletters nie behoue sal bly nie. Die waarskynlikheid is egter nie so groot as in die eersgenoemde twee gevalle nie; dit vra dus vir noukeuriger statistiese ontleding in uitgebreider korpusnavorsing.

Uit hierdie data kan ook reeds vier veranderlikes geïdentifiseer word wat in toekomstige navorsing verreken moet word, te wete:

- Die woordsoort van die resulterende woordvorm: Uit tabel 8 kan ons sien dat *Dooie See+rolle* in sy geheel as multiwoordeienaam geïnterpreteer kan word (sien die relatief hoë waarskynlikheid van *Dooie See Rolle* in geredigeerde teks).
- Die voorkoms van die resulterende woordvorm in normerende bronne en woordeboeke: Let op die sterk waarskynlikheid van *Derdewêreldland* in die VivA-materiaal (meestal geredigeerde bronne), moontlik omdat dié samestelling op hierdie wyse in die AWS opgeneem is.
- Die mate van vestiging (Van Huyssteen 2005) van multiwoordeiename: Uit die data wil dit lyk asof *Middellandse See* en *Dooie See* moontlik sterker as multiwoordeiename geïnstitusionaliseerd is as *Konstitusionele Hof* en *Derde Wêreld*; vir laasgenoemde twee is gewone woordgroepe soos *rondgaande hof*, *militêre hof* of *siviele hof*, en *ontwikkelende wêreld*, *veranderende wêreld* of *Westerse wêreld* dalk bekende analogieë.
- Die verskil tussen geredigeerde teks en ongeredigeerde teks: Let byvoorbeeld op die verskil in waarskynlikhede tussen die VivA- en Google-materiaal vir *Konstitusionele Hof+regter*.

Die waarskynlikhede van die verskillende geadjektiveerde vorme (tabel 10) bevestig hierdie waarnemings: Die hoofletters van die multiwoordeiename word meestal behou, behalwe in die geval van *Derdewêrelds* wat sonder hoofletters in die AWS opgeneem is. Geen ooglopende verskil word egter tussen geredigeerde (VivA), semigeredigeerde (Wikipedia) en ongeredigeerde (Google) korpusmateriaal opgemerk nie.

Tabel 10. Hoof-/kleinletters: Waarskynlikhede van verskillende geadjektiveerde vorme

·ADJZ	*Nabye Ooste*		*Verre Ooste*		*Ou Testament*		*Nuwe Testament*		*Derde Wêreld*	
hl.?	Ja	Nee	Ja	Nee	Ja	Nee	Ja	Nee	Ja	Nee
P_{VivA}	**1,00**	0,00	**1,00**	0,00	**1,00**	0,00	**1,00**	0,00	0,10	**0,90**
P_{Wiki}	**1,00**	0,00	**0,78**	0,22	**0,88**	0,13	**0,96**	0,04	0,00	**1,00**
P_{Goog}	**0,95**	0,05	**0,90**	0,10	**0,78**	0,22	**0,83**	0,17	0,13	**0,87**
$P_{gemiddeld}$	**0,98**	0,02	**0,89**	0,11	**0,89**	0,11	**0,93**	0,07	0,08	**0,92**

Die tweede vraag wat in hierdie korpusondersoek gevra word, is of die spasies in multiwoordeiename behou bly in komposita daarmee en geadjektiveerde vorme daarvan. Anders gestel: Wat is die waarskynlikheid dat die spasie in ’n multiwoordeienaam behou sal word in ’n kompositum daarmee en geadjektiveerde vorm daarvan?

In Tabel 11 word die waarskynlikhede van die verskillende komposita opgesom in terme van skryfwyse (los, vas of koppelteken (“kop.”)). In die laaste ry word die gemiddelde waarskynlikheid van die drie korpusse tesame gegee. (In die geval van *Konstitusionele Hof+regter* word die waarskynlikhede van die Wikipedia-data weer eens buite rekening gelaat by die berekening van die gemiddeldes.)

Tabel 11. Los/vas/koppelteken: Waarskynlikhede van verskillende komposita

Komp.	*Middellandse See+gebied*			*Dooie See+rolle*			*Konstitusionele Hof+regter*			*Derde Wêreld+land*		
los/vas/ kop.	Los	Vas	Kop.	Los	Vas	Kop.	Los	Vas	Kop.	Los	Vas	Kop.
P_{VivA}	**0,94**	0,00	0,06	**0,95**	0,05	0,00	**0,77**	0,15	0,08	0,23	**0,76**	0,02
P_{Wiki}	**1,00**	0,00	0,00	**1,00**	0,00	0,00	0,00	0,00	0,00	**0,60**	0,30	0,10
P_{Goog}	**0,95**	0,04	0,01	**0,79**	0,19	0,03	**0,81**	0,17	0,03	**0,53**	0,39	0,08
$P_{gemiddeld}$	**0,96**	0,01	0,02	**0,91**	0,08	0,01	**0,79**	0,16	0,05	0,45	**0,48**	0,06

In vergelyking met die keuse tussen hoof- en kleinletters is dit opvallend dat daar in hierdie data nie so ’n duidelike onderskeid tussen los, vas en met koppelteken vas geskrewe vorme is nie; slegs *Derdewêrelds* (wat só in die AWS opgeneem is) word in die VivA-materiaal (geredigeerd) vas geskryf, terwyl die ortografiese integriteit van die multiwoordeienaam in al die ander gevalle meestal behoue blyk te word. Dit bevestig ook die moontlike rol van normerende bronne, soos hier bo uitgelig.

Ook in die data vir geadjektiveerde vorme (tabel 12) staan *Derdewêrelds* uit: In al die verskillende korpusmateriaal het dit ’n relatief hoë waarskynlikheid dat dit vas geskryf sal word (soos in die AWS), terwyl die ander gevalle in die verskillende korpusmateriaal wisselend los of met ’n koppelteken vas geskryf word. In den brede bevestig hierdie data ook

die voorlopige waarnemings wat ten opsigte van hoof- en kleinletters oor veranderlikes gemaak is.

Tabel 12. Los/vas/koppelteken: Waarskynlikhede van verskillende geadjektiveerde vorme

·ADJZ	*Nabye Ooste*			*Verre Ooste*			*Ou Testament*			*Nuwe Testament*			*Derde Wêreld*		
los/vas/kop.	Los	Vas	Kop.	Los	Vas	Kop.	Los	Vas	Kop.	Los	Vas	Kop.	Los	Vas	Kop.
P_{VivA}	***0,50***	0,00	***0,50***	0,00	0,00	**1,00**	0,10	0,00	**0,90**	0,11	0,00	**0,89**	0,10	**0,81**	0,10
P_{Wiki}	0,00	0,00	**1,00**	0,22	0,00	**0,78**	**0,79**	0,00	0,21	**0,71**	0,04	0,25	0,00	**1,00**	0,00
P_{Goog}	**0,76**	0,00	0,24	**0,52**	0,00	0,48	**0,44**	0,17	0,39	**0,44**	0,14	0,43	0,26	**0,60**	0,14
$P_{\text{gemiddeld}}$	0,42	0,00	**0,58**	0,25	0,00	**0,75**	0,44	0,06	**0,50**	0,42	0,06	**0,52**	0,12	**0,80**	0,08

As ons veralgemeen, kan ons samevattend tot die gevolgtrekking kom dat die ortografiese integriteit van multiwoordeiename oor die algemeen in korpusmateriaal behou word, veral as die multiwoordeienaam sterk as eienaam gekonseptualiseer word (byvoorbeeld *Middellandse See*, teenoor *Derde Wêreld*). Die rol van normerende bronne en woordeboeke in die institusionalisering van ’n bepaalde skryfwyse in veral geredigeerde bronne kan nie onderskat word nie, maar moet verder empiries ondersoek word.

4. Meningspeiling

Teen die agtergrond van die potensiële impak wat normerende bronne op die skryfwyse van komposita met en afleidings van multiwoordeiename kan hê, word daar hier ook oor die resultate van ’n meningspeiling wat die TK in 2012 onder taalpraktisyns onderneem het, gerapporteer. Ten einde reëls oor hierdie twee konstruksies in ’n toekomstige uitgawe van die AWS in te sluit, het die TK twee stelle konsepreëls (konsepreëls A & B en C & D) ontwikkel, soos uiteengesit in tabel 13. (Let op dat die TK die term *meerledige eiename* gebruik om na multiwoordeiename te verwys.)

Tabel 13. Konsepreëls A & B en C & D

REËL A: Afleidings van meerledige eiename word met ’n koppelteken tussen die dele van die meerledige eienaam geskryf, met die laaste gedeelte vas aan die laaste deel van die meerledige eienaam. **Addis-Abebitiese** **Ou-Testamentiese** **Dooie-Sese** **Stille-Oseaniese**

REËL B: Samestellings met meerledige eiename word met ’n koppelteken tussen die dele van die meerledige eienaam geskryf, met die laaste gedeelte vas of met ’n koppelteken vas aan die laaste deel van die meerledige eienaam. **Addis-Abeba-beraad** of **Addis-Abebaberaad** **Ou-Testament-vertaling** of **Ou-Testamentvertaling** **Dooie-See-rolle** of **Dooie-Seerolle** **Koning-Shaka-Lughawe** of **Koning-Shakalughawe**
REËL C: Afleidings van meerledige eiename word sonder ’n koppelteken tussen die dele van die meerledige eienaam geskryf, met die laaste gedeelte vas aan die laaste deel van die meerledige eienaam. **Addis Abebitiese** **Ou Testamenties** **Dooie Sese** **Stille Oseaniese**
REËL D: Samestellings met meerledige eiename word sonder ’n koppelteken tussen die dele van die meerledige eienaam geskryf, met die laaste gedeelte vas of met ’n koppelteken vas aan die laaste deel van die meerledige eienaam. **Addis Abeba-beraad** of **Addis Abebaberaad** **Ou Testament-vertaling** of **Ou Testamentvertaling** **Dooie See-rolle** of **Dooie Seerolle** **Koning Shaka-Lughawe** of **Koning Shakalughawe**

Konsepreëls A & B hanteer multiwoordeiename as woordgroepe (reël 14.24), en die skryfwyse van samestellende afleidings en samestellende samestellings ingevolge reëls 14.25 en 14.26 in die AWS word in beginsel gevolg; die nuwe woordvorme word met koppeltekens vas geskryf, met behoud van die hoofletters van die multiwoordeienaam. Konsepreëls C & D hanteer dit egter nie as woordgroepe nie, maar as multiwoord leksikale items: die ortografiese integriteit van multiwoordeiename word dus behou, met inbegrip van spasies en hoofletters.

Met betrekking tot komposita geld reël 14.6 in die AWS (Taalkommissie 2009:120) sonder meer: In eienaamverbindings wat as eiename optree, kan die eienaamgedeelte los van, of met of sonder ’n koppelteken vas aan die res van die kompositum geskryf word. Die tweede gedeelte van so ’n eienaamverbinding kan met ’n hoofletter of kleinletter geskryf word wanneer dit los of met ’n koppelteken vas geskryf word; vergelyk (28). Soos vroeër bespreek, sou *Dooie See+rolle* in sy geheel gesien kon word as ’n multiwoordeienaam, soos ook *Koning Shaka+lughawe;* dié twee voorbeelde is dus nie noodwendig gepaste/prototipiese voorbeelde in hierdie konsepreëls nie, aangesien *Dooie See Rolle* en *Koning Shaka Lughawe* ewe “korrekte” skryfwyses sou kon wees.

(28) *Bloemfontein Museum*
Bloemfontein-Museum
Bloemfontein-museum
Bloemfonteinmuseum.

4.1 Vraelys en respondente

Die TK het bostaande konsepreëls as voorbeelde in ’n informele vraelys gebruik (sien bylae). Respondente kon kies tussen ’n aantal kombinasies, naamlik: A & B, C & D, A & D en B & C, asook ’n opsie dat die reëls soos wat dit tans in die AWS verskyn, voldoende is. Verder kon respondente ook enige kommentaar byvoeg, indien hulle sou voel dat dit nodig is. Die vraelys is deur Tom McLachlan namens die TK via e-pos aan ongeveer 80 taalpraktisyns gestuur, waaronder vertalers, teksredakteurs, kopieskrywers, navorsers, dosente, onderwysers, en dies meer; 62 respondente het die vraelys uiteindelik voltooi.

4.2 Resultate

Die resultate van die ondersoek word in tabel 14 opgesom. 76% (47) van al die respondente het vir C & D gekies (d.i. waar die ortografiese integriteit van die multiwoordeienaam behou word). Dit is dus duidelik dat taalpraktisyns ook ’n voorkeur vir ’n skryfwyse het wat grotendeels ooreenstem met die resultate uit die korpusdata, naamlik *Middellandse Seegebied*, *Dooie Seerolle*, *Konstitusionele Hofregter* en *Derde Wêreldland*; en *Nabye Oosters*, *Verre Oosters*, *Ou Testamenties*, *Nuwe Testamenties* en *Derde Wêrelds*.

Tabel 14. Resultaat van meningspeiling

	Ja	Nee
Het die bestaande reëls hoegenaamd wysiging nodig?	0	1
Lewer die bestaande reëls/stelsel ooit vir u probleme?	0	0
My voorkeur is vir A en B	4	0
My voorkeur is vir C en D	**47**	0
My voorkeur is vir A en D	9	0
My voorkeur is vir B en C	0	0
Slegs C	1	0
Totaal	**61**	**1**

5. Aanbevelings en implikasies

In hierdie artikel is ’n multiwoordeienaam gedefinieer as ’n frasekonstruksie wat na ’n benoemde entiteit verwys en wat uit twee of meer ortografiese woorde bestaan; dié ortografiese woorde word deur spasies geskei en gewoonlik met hoofletters geskryf (uitgesonderd funksiewoorde soos lidwoorde, voorsetsels en partikels). Daar is in afdeling 2.2.1 aangetoon dat daar nie eenduidige reëls in die AWS vir die skryfwyse van komposita met en afleidings van sodanige multiwoordeenhede is nie.

Die basiese navorsingsvraag van die korpusondersoek was of daar ’n positiewe korrelasie tussen die ortografiese vorm van ’n multiwoordeienaam en die ortografiese vorm van die multiwoordeienaam in komposita daarmee of afleidings daarvan is. Uit dié kleinskaalse ondersoek wat geloods is, blyk dit dat die ortografiese integriteit van multiwoordeiename oorwegend behoue bly in komposita daarmee en gesuffigeerde afleidings daarvan.

Op sterkte van hierdie bevindings sou die konsepreëls van die TK soos volg aangepas kon word as ’n enkele reël X:

REËL X: In samestellings met en afleidings van ’n multiwoordeienaam word die hoofletters en spasies van die multiwoordeienaam behou.

- Addis Abebitiese
  Middellandse Sese
  Ou Testamenties
  Stille Oseaniese
- Addis Abeba-populasie *of* Addis Abebapopulasie
  Ou Testament-vertaling *of* Ou Testament-vertaling
  Middellandse Seebootreis *of* Middellandse See-bootreis
  Stille Oseaanverskynsel *of* Stille Oseaan-verskynsel

’n Opmerking sou bygevoeg kon word om voorsiening te maak vir komposita wat as eiename funksioneer, met verwysing na reël 14.6 en voorbeelde soos dié in (29).

> (29) *Dooie Seerolle* of *Dooie See-rolle* of *Dooie See-Rolle* of *Dooie See Rolle*
> *Nelson Mandelabaaisakekamer* of *Nelson Mandelabaai-sakekamer* of *Nelson Mandelabaai-Sakekamer* of *Nelson Mandelabaai Sakekamer*
> *Solomon Mahlangurylaan* of *Solomon Mahlangu-rylaan* of *Solomon Mahlangu-Rylaan* of *Solomon Mahlangu Rylaan*.

Vier veranderlikes waaraan in toekomstige navorsing praktiese uitvoering gegee moet word, is geïdentifiseer, te wete:

- Die woordsoort van die resulterende woordvorm.
- Die voorkoms van die resulterende woordvorm in normerende bronne en woordeboeke.
- Die mate van vestiging van multiwoordeiename.
- Die verskil tussen geredigeerde tekste en ongeredigeerde tekste.

In toekomstige (korpus)navorsing moet bogenoemde veranderlikes verreken word en aandag moet ook gegee word aan multiwoordeiename wat uit meer as twee dele bestaan (byvoorbeeld *Departement Kuns en Kultuur*), komplekser konstruksies (byvoorbeeld ?*anti-Verenigde State van Amerika-sentimente*), verbaliserings (soos ?*New Yorkifiseer* of ?*ver-Suid-Afrikaans*) en ander geprefigeerde gevalle (soos ?*anti-Middellandse Sese*).

Erkenning

Hierdie navorsing is deel van die Automatic Compound Processing-projek (gerhard.pro/aucopro) en is gefinansier deur die Nederlandse Taalunie, Departement Kuns en Kultuur van die Suid-Afrikaanse regering, die Nasionale Navorsingstigting (toelaagnommer: 81794) en die European Network on Word Structure (NetWordS) van die European Science Foundation (toelaagnommer: 5570). Geen van die opinies in hierdie artikel kan aan enige van die organisasies of hulle werknemers toegedig word nie.

Nanette van den Berg het ’n waardevolle rol gespeel in die verkenningsondersoek wat vir hierdie navorsing gedoen is; haar insette en hulp met aanvanklike data-insameling word met dank erken. Dank ook aan Roald Eiselen, Martin Puttkammer en Bertus van Rooy vir hulle kommentaar op en leiding met betrekking tot die korpusondersoek. Die TK, insonderheid Ernst Kotzé, Frikkie Lombard, Tom McLachlan en Suléne Pilon, se insette en stimulerende gesprekke word ook met dank erken. Alle eventuele denk- en berekeningsfoute kan egter slegs voor my deur gelê word.

Bibliografie

Baayen, R.H. 2003. Probabilistic approaches to morphology. In Bod e.a. (reds.) 2003.

Bergman, M. 2016. How fine grained can entity types get? http://www.mkbergman.com/1927/how-fine-grained-can-entity-types-get (6 Julie 2016 geraadpleeg).

Bizer, C., J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak en S. Hellmann. 2009. DBpedia – A crystallization point for the web of data. *Web semantics: science, services and agents on the world wide web*, 7(3):154–65.

Bod, R., J. Hay en S. Jannedy (reds.). 2003. *Probabilistic linguistics*. Cambridge: MIT Press.

Bold, C. 2013. *The frontier club: popular westerns and cultural power: 1880–1924*. New York: Oxford University Press.

Booij, G. 2002. Separable complex verbs in Dutch: a case of periphrastic word formation. In Dehé e.a. (reds.) 2002.

Brown, K. (red.). 2006. *Encyclopedia of language & linguistics*. 2de uitgawe. Oxford: Elsevier.

De Haas, W. en M. Trommelen. 1993. *Morfologisch handboek van het Nederlands. Een overzicht van de woordvorming*. ’s-Gravenhage: SDU Uitgeverij.

Dehé, N., R. Jackendoff, A. Macintyre en S. Urban (reds.). 2002. *Verb-particle explorations*. Berlyn: Mouton de Gruyter.

Departement Kuns en Kultuur en CTexT. 2013. Afrikaans NCHLT Text Corpora. Potchefstroom: Noordwes-Universiteit. http://rma.nwu.ac.za (6 Julie 2016 geraadpleeg).

Desmet, B. en V. Hoste. 2014. Fine-grained Dutch named entity recognition. *Language Resources and Evaluation*, 48(2):307–43.

Fletcher, W.H. 2005. Concordancing the web: promise and problems, tools and techniques. *Corpus Linguistics and the Web*, 1–22.

Gillick, D., N. Lazic, K. Ganchev, J. Kirchner en D. Huynh. 2014. Context-dependent fine-grained entity type tagging. *arXiv preprint arXiv*, 1412.1820. http://www.aclweb.org/anthology/W10-3702 (6 Julie 2016 geraadpleeg).

Gouws, R. 1989. *Leksikografie*. Pretoria: Academica.

Gralinski, F., A. Savary, M. Czerepowicka en F. Makowiecki. 2010. Computational lexicography of multi-word units: how efficient can it be? In *Proceedings of the workshop on multiword expressions: From theory to applications*, Beijing.

Hanks, P. 2006. Proper names: linguistic status. In Brown (red.) 2006.

Kirk, J.M. 2006. Word frequency: use or misuse? In *Word frequency and keyword extraction: AHRC ICT methods network expert seminar on linguistics*, AHRC ICT Methods Network, Londen.

Krstev, C., D. Vitas, D. Maurel en M. Tran. 2005. Multilingual ontology of proper names. In *Proceedings of the second language & technology conference: Human language technologies as a challenge for computer science and linguistics*, Poznan.

Kuperman, V. en R. Bertram. 2012. Moving spaces: spelling alternation in English noun-noun compounds. *Language and Cognitive Processes*, 0965 (Junie), ble. 1–28.

Lapa-uitgewers en CTexT. 2015. NWU/Lapa-korpus 1.0. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Lehrer, A. 2006. Proper names: semantic aspects. In Brown (red.) 2006.

Ling, X. en D.S. Weld. 2012. Fine-grained entity recognition. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. http://xiaoling.github.io/pubs/ling-aaai12.pdf (6 Julie 2016 geraadpleeg).

Louw, P. 2006. Inclusion strategies for multi-word units in monolingual dictionaries. *Lexikos*, 16:95–103.

Lüdeling, A., S. Evert en M. Baroni. 2006. Using web data for linguistic purposes. *Language and computers*, 59:7–24.

Maroela Media en CTexT. 2015. NWU-Maroela Media-korpus 1.0. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Maurel, D. 2008. Prolexbase: a multilingual relational lexical database of proper names. In *6th Language resources and evaluation conference LREC 2008*, Marrakesj.

McArthur, T. 1992. What is a word? In McArthur (red.) 1992.

McArthur, T. (red.). 1992. *Oxford companion to the English language*. Oxford: Oxford University Press.

Newman, M. 2015. *New York City English*. Berlyn: Mouton de Gruyter.

Philip, G. 2005. Identifying multi-word units in context. *AMS Acta*. http://amsacta.cib.unibo.it/archive/00001133/01/MultiWordUnits.pdf (6 Julie 2016 geraadpleeg).

Protea Boekhuis en CTexT. 2015. PUK/Protea Boekhuis-korpus 2.0. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Puttkammer, M. 2005. Outomatiese Afrikaanse tekseenheididentifisering. Ongepubliseerde MA-verhandeling, Noordwes-Universiteit.

Radio Sonder Grense en CTexT. 2015. RSG Afrikaanse korpus 2.0. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Sekine, S. 2010. Sekine's extended named entity hierarchy. http://nlp.cs.nyu.edu/ene (6 Julie 2016 geraadpleeg).

Sekine, S., K. Sudo en C. Nobata. 2002. Extended named entity hierarchy. In *3rd Language resources and evaluation conference LREC 2002*, Las Palmas.

Taalkommissie (samest.). 2009. *Afrikaanse woordelys en spelreëls*. 10de uitgawe. Kaapstad: Pharos.

—. 2011. Taalkommissiekorpus 1.1. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Trask, R.L. 2004. What is a word? Working papers 11, Department of Linguistics and English Language, University of Sussex.

Van Huyssteen, G.B. 2005. 'n Kognitiewe gebruiksgebaseerde beskrywingsmodel vir die Afrikaanse grammatika. *Southern African Linguistics and Applied Language Studies*, 23(2):125–37.

Vitas, D., C. Krstev en D. Maurel. 2007. A note on the semantic and morphological properties of proper names in the Prolex project. *Lingvisticae Investigationes*, 30(1):115–33.

Wallis, S., J. Bowie en B. Aarts. 2012. That vexed problem of choice. Some reflections on experimental design and statistics with corpora. In *International Conference on Advances in Mechanics Engineering (ICAME 33)*, Leuven. http://www.ucl.ac.uk/english-usage/staff/sean/resources/vexedchoice.pdf (6 Julie 2016 geraadpleeg).

WatKykJy. 2015. Watkykjy.co.za-korpus. Potchefstroom: Noordwes-Universiteit. http://viva-afrikaans.org (6 Julie 2016 geraadpleeg).

Bylae: Vraelys van die Taalkommissie

Geagte Vriend

Die Taalkommissie van die SA Akademie vir Wetenskap en Kuns, wat die opsteller van die *Afrikaanse Woordelys en Spelreëls* is, het reeds voor die publikasie van die 9de uitgawe van die *AWS* besluit om taalpraktisyns te raadpleeg voordat sekere reëls ingevoer word. Hierdie brief is so ’n poging om die menings van kundige gebruikers in te win. Mag ek dus namens die Taalkommissie asseblief u mening oor die volgende saak vra?

Tradisioneel skryf ons in Afrikaans ***Ou Testament***, maar wanneer ’n afleiding daarvan gevorm word, word dit vas geskryf: ***Ou-Testamentiese*** (reël 14.25 van die jongste *AWS*). Insgelyks skryf ons ***Derde Wêreld***, maar indien dit verder verbind, word dit vas geskryf, byvoorbeeld ***Derdewêreldland*** (bv. reël 14.26).

Hierdie stelsel word egter nie konsekwent toegepas nie, want in gevalle soos ***New Yorkse*** of ***Sri Lankaan*** word die dele van die eienaam nie aan mekaar gekoppel soos by ***Ou-Testamentiese*** nie (reël 14.25, opm. (d)). Die praktyk toon ook dat ’n naam soos ***Kaap Verde*** nie altyd gekoppel (of vas geskryf) word in byvoorbeeld ***Kaap Verde-landingstrook*** nie. Daar is ook groot onsekerheid of ’n mens ***Dooieseerolle*** of ***Dooie-Seerolle*** of ***Dooie See-rolle*** moet skryf.

Die Taalkommissie is tans besig om groter eenvormigheid in gevalle soos bostaande te probeer bewerkstellig. Ons wil dus probeer om alle meerledige eiename waarvan die dele los van mekaar geskryf word, dieselfde te behandel wanneer ’n mens afleidings daarvan of samestellings daarmee vorm.

Indien moontlik wil ons nie verval in ’n klomp subreëls nie. ’n Mens sou byvoorbeeld kon sê dat ’n eienaam soos *Derde Wêreld* verskil van een soos *New York* omdat eersgenoemde gewone Afrikaans is terwyl laasgenoemde duidelik nie Afrikaans is nie, en dat daardie verskil ook jou hantering van afleidings en samestellings moet bepaal. Daarrondom ontstaan egter ander vrae, soos die volgende:

- In watter mate is ’n leenwoord (al is dit ’n eienaam) Afrikaans of nie? Ons voeg immers Afrikaanse voor- en agtervoegsels daaraan toe.
- Is name wat uit ander alfabette getranslitereer is, Afrikaans of vreemdtalig?
- Wat is ’n naam wat deels verafrikaans is, soos *Kaap Verde*?
- Is ’n persoonseienaam soos *Nelson Mandela* vreemdtalig?
- Moet ’n persoonseienaam anders behandel word as ander eiename?

En so meer.

Nog ’n vraag is of afleidings en samestellings noodwendig op dieselfde manier gehanteer moet word. Moet koppeltekens byvoorbeeld by afleidings EN samestellings in meerledige eiename ingevoeg word, of kan slegs afleidings koppeltekens kry?

Die Taalkommissie het gevolglik twee stelle konsepreëls geformuleer. Konsepreëls 14.A en 14.B berus op die beginsel dat meerledige eiename net soos ander woordgroepe in Afrikaans behandel word, terwyl konsepreëls 14.C en 14.D op die beginsel berus dat meerledige eiename onveranderlike eenhede is waarvan die integriteit grotendeels ongeskonde behou moet word. Die gevolge van die reëls word in die voorbeelde by elk geïllustreer.

Wat miskien nie so opvallend is nie, is dat sulke sambreelreëls noodwendig ’n uitwerking op sekere bekende skryfwyses gaan hê, na watter kant toe ook al. Ná die konsepreëls (op p. 5) is daar ’n ruimte waarin u u voorkeure kan aandui. Ons sal ook enige kommentaar wat u wil lewer, verwelkom.

Stuur asseblief u antwoorde aan my by [e-posadres].

Baie dankie vir u tyd en aandag hieraan.

Die uwe

Tom McLachlan

Voorsitter: Taalkommissie van die SA Akademie vir Wetenskap en Kuns

23 Oktober 2012

Eindnotas

[1] Samestellings word aangedui met ’n plusteken tussen die woorde, terwyl affiksering en morfeemgrense met ’n gesentreerde punt (·) aangedui word. Die afkorting *ADJZ* verwys na ’n adjektiveerder.

[2] Let byvoorbeeld daarop dat reël 14.25 voorsiening maak vir die adjektivering van werkwoordfrases deur middel van die deelwoordvormer ·***end***, byvoorbeeld *alles omvat* > *allesomvattend* of *swak sien* > *swaksiende* (AWS 2009:129). Waar *voor die hand liggend* as woordgroep aangedui word by reël 12.22, lei dit m.i. tot ’n verkeerde ontleding, aangesien die basiswoordgroep *voor die hand lê* is. Dit is ook nie duidelik waarom die deelwoordvormer ·***end*** by 14.25 geld, maar nie by 12.33 nie; hiervolgens is *swak sien* > *swaksiende* aanvaarbaar, maar *op sigself staan* > **opsigselfstaande* nie.

[3] Sien onder andere Newman (2015:120) vir ’n voorbeeld van *New Yorkism* in Engels.

[4] Sien onder andere Bold (2013:68) vir ’n voorbeeld van *New Yorkify* in Engels.

[5] https://www.sigwac.org.uk.