

Die naslaanpotensiaal van lemmakandidate vir 'n monofunksionele beknopte Nederlands-Afrikaans-woordeboek

Herman L. Beyer

Departement Taal- en Literatuurstudie, Universiteit van Namibië, Privaatsak 13301, Windhoek, Namibië
E-mail: hbeyer@unam.na

Opsomming

Die beplanning van 'n monofunksionele, beknopte Nederlands-Afrikaans-woordeboek in die gedrukte medium plaas welbekende tradisionele beperkinge op die leksikograaf, waarvan die vernaamste dié op ruimte is. Gegee die konteks, teikengebruiker, gebruikersituasie en funksie van die woordeboek moet die leksikograaf deeglik besin oor die toepaslike lemmatiseringsbeleid. In hierdie artikel word verslag gedoen van 'n empiriese metode wat korpusondersoek kan komplementeer om 'n optimaal funksionele lemmalys vir die onderhawige tipe woordeboek saam te stel. Deur middel van die simulatie van die tipiese gebruikersituasie van die beplande woordeboek word die naslaanpotensiaal van verskillende kategorieë lemmakandidate kwantitatief vasgestel. Sodanige vasstelling lei tot die bepaling van die lemmatiseringsprioriteit van elke kategorie lemmakandidate, wat uiteindelik funksionele lemmatisering en suksesvolle woordeboekgebruik tot gevolg het.

Sleutelwoorde: funksie, gebruikersituasie, leksikografie, lemmakandidate, lemmatisering, naslaanpotensiaal, teikengebruiker, woordeboek.

The consultation potential of lemma candidates for a monofunctional concise Dutch-Afrikaans dictionary

Abstract

The planning of a monofunctional, concise Dutch-Afrikaans dictionary in the printed medium places well-known traditional limitations on the lexicographer, the most important being that of space. Given the context, target user, user situation and function of the dictionary, the lexicographer has to carefully consider the applicable lemmatisation policy. This article reports on an empirical method that could complement corpus study in order to compile an optimally functional lemma list for the planned type of dictionary. By means of a simulation of the typical user situation, the consultation potential of various types of lemma candidates are determined quantitatively. This determination results in the identification of the lemmatisation priority of every category of lemma candidates, which in turn ultimately leads to functional lemmatisation and successful dictionary use.

Keywords: consultation potential, dictionary, function, lemma candidate, lemmatisation, lexicography, target user, user situation.

1. Inleiding

In 2008 is met die beplanning van 'n monofunksionele beknopte Nederlands-Afrikaans-woordeboek vir Afrikaanssprekendes (voortaan NAWA) begin. Hierdie artikel stel 'n prosedure voor om die naslaanpotensiaal van bepaalde kategorieë lemmakandidate na aanleiding van 'n empiriese gebruikersonderzoek te bepaal ten einde 'n lemmatiseringsprioriteitslys in terme van hierdie kategorieë op te stel. Terselfdertyd geld die artikel as 'n oefening in die eksposisie van enkele aspekte van 'n kommunikatiewe metaleksikografie.

2. Metaleksikografiese raamwerk

'n Kommunikatiewe metaleksikografie onderskei nege hoofkomponente (na aanleiding van Beyer (2006) se leksikografiese kommunikasiemodel), naamlik:

- die woordeboek (as manifestasie van leksikografiese kommunikasie: 'n houër van leksikografiese boodskappe met bepaalde inhoude, in bepaalde strukture, met bepaalde funksies);
- die leksikograaf (as leksikografiese kommunikeerder: die sender van leksikografiese boodskappe);
- die leksikografiese proses (wat onder meer die waarneming, seleksie, interpretasie en enkodering van leksikografiese data tot leksikografiese boodskappe verteenwoordig);
- die medium en kanaal (waardeer leksikografiese kommunikasie plaasvind);
- die teikengebruiker (as bestemde ontvanger van leksikografiese boodskappe);
- die gebruikersituasie (wat (die element(e) van) die onmiddellike teikengebruikerspesifieke konteks verteenwoordig wat tot 'n bepaalde funksionele gebruikersdoelwit aanleiding gee, soos 'n teksresepsiesituasie, 'n teksproduksiesituasie, ens.);
- die gebruiksituasie (waarin die waarneming, seleksie, dekodeering en interpretasie van leksikografiese boodskappe deur die teikengebruiker plaasvind);
- terugvoer (wat nie-leksikografiese metaleksikografiese kommunikasie tussen die leksikograaf en onder meer die teikengebruiker verteenwoordig en 'n invloed op toekomstige leksikografiese kommunikasie kan uitoefen);
- die konteks (wat die verskillende fasette van die breër omgewing waarbinne leksikografiese kommunikasie (m.a.w. sowel die leksikografiese proses as woordeboekgebruik) plaasvind, verteenwoordig).

Hierdie hoofkomponente verteenwoordig 'n verfyning op Beyer (2006) se vroeëre elf parameters vir die beskrywing van gebruikersvriendelikheid in woordeboeke. Vir die doel van hierdie artikel sal slegs die vyf mees relevante komponente betrek word: die konteks, die teikengebruiker en die medium; die bespreking van die empiriese ondersoek ressorteer onder die gebruikersituasiekomponent en die lemmatiseringsprioriteitstelling val onder die komponent van die woordeboek.

3. Die konteks

Volgens Beyer (2006:60) omvat die term *konteks* binne 'n kommunikatiewe metaleksikografie daardie versameling omgewingsveranderlikes wat normaalweg buite die beheer van die leksikograaf en die gebruiker in hulle onderskeie kapasiteite as sodanig val,

maar waarbinne die saamstel sowel as die gebruik van die woordeboek (m.a.w. leksikografiese kommunikasie) plaasvind.

Binne 'n bepaalde konteks kan ongereguleerde subkontekste bestaan, bv. die veeltaligheid van 'n gemeenskap, sowel as gereguleerde subkontekste oftewel sisteme, bv. 'n nasionale onderwysstelsel (Beyer 2010:54). In die kommunikasiewetenskap word taal as 'n kodesistiem beskou (vgl. Van Cuilenburg et al. 1992:187, Steinberg 2007:44). Die konteks bevat ook die teikentaal of -tale van 'n woordeboek as konteksafhanklike gedeelde kennis (vgl. Steinberg 2007:44, 50-51); dus word die teikentaal of -tale as (kode)sisteme binne die konteks van die leksikografiese kommunikasie (d.w.s. die woordeboek) beskou. Trouens, dít is die fundamentele verskil tussen leksikografiese kommunikasie en ander soorte kommunikasie: leksikografiese kommunikasie het ten doel om 'n bepaalde kodesistiem ('n leksikon) in terme van sy of 'n ander kodesistiem se tekens (bv. leksikonitems) te kommuniseer. Terwyl ander tipes kommunikasie van verskillende kodesisteme gebruik maak om data oor ander tipes verskynsels oor te dra, maak leksikografiese kommunikasie van 'n kodesistiem gebruik om data oor daardie en/of 'n ander kodesistiem oor te dra. Gevolglik kan die aard van die tale betrokke by NAWA onder die opskrif van die konteks tuisgebring word.

NAWA word ook binne die konteks van 'n bestaande woordeboektipologie gekonseptualiseer; derhalwe moet dié woordeboek se plasing binne hierdie sisteem gekontekstualiseer word.

3.1 Afrikaans en Nederlands

Die Nederlandse basis van Afrikaans (vgl. Ponelis 1993) en die gevolglike ooreenkomste op talle vlakke tussen Afrikaans en Nederlands is welbekend. Hier word kortliks op enkele relevante aspekte van die leksikon ingegaan.

Ponelis (1993:479, 480) meld dat 'n ondersoek van 'n 200-item leksikostatistiese kernleksikon van Afrikaans slegs onbeduidende afwykings van die Nederlandse basis toon, en dat die oorweldigende meerderheid van die mees frekwente Afrikaanse vorme in verskillende leksikale klasse identiese kognate van Nederlands is. In studies wat Afrikaans en Nederlands op vergelykende basis betrek, word dit telkens beklemtoon dat die ooreenkoms tussen die twee tale veral met betrekking tot die leksikon en die sintaksis groot is (vgl. Pheiffer 1989, Raidt 1989, Bosman 2011).

Pheiffer (1989) bied 'n sistematiese uiteensetting van die belangrikste vormverskille wat wel aangetoon kan word, met *vorm* wat verwys na die ortografiese, fonologiese, sintaktiese en morfologiese kenmerke van leksikale items. Wat die wisselwerking tussen ortografie en fonologie betref, wys Pheiffer (1989:73-77) op “klankverskille wat nie in die spelling weerspieël word nie” (bv. Ndl.<Afr. *been*, *dier*, *toer*), “klankverskille wat wel in die spelling weerspieël word” (bv. Ndl. *vooral* < Afr. *veral*, *handje* < *handjie*, *paard* < *perd*) en “verskille in spelling wat geen klankverskille verteenwoordig nie (bv. Ndl. *cilinder* < Afr. *silinder*, *politie* < *polisie*, *synagoge* < *sinagoge*).

In die *Groot Woordeboek Afrikaans en Nederlands* (Martin 2011 – voortaan ANNA) verteenwoordig die kombinasie van vorm en betekenis die basis waarop die lemmata en hulle

bewerkings geklassifiseer word. Die volgende kategorieë word naamlik onderskei (Martin 2011:24-25):

- (1) a. absolute kognate: “woorde in albei tale wat na vorm en betekenis identies is, soos *tafel/tafel*, *kaas/kaas* en *paniek/paniek*”;
- b. absolute kognate met vormverskil: “woorde wat in albei tale dieselfde betekenis het, ’n vormverskil vertoon maar sonder dat hierdie verskil herkenbaarheid belemmer, soos *ontsnappen/ontsnap*, *gel/jel*, *pad/padda*, *kabeljauw/kabeljou*, *zalm/salm* en *spliterwt/splitertjie*”;
- c. gedeeltelike kognate: “woorde wat in teenstelling tot absolute cognates in ten minste een betekenisonderskeiding verskil, soos *robot/robot*”;
- d. nie-kognate: “woorde met dieselfde betekenis maar ’n duidelike vormverskil, soos N. *verkeersdrempel/A. spoedwallekje of spoedhobbel*”;
- e. vals vriende: “woorde wat slegs vormlik (skynbaar) ooreenstem, maar glad nie wat ’n enkele betekenisonderskeiding betref nie, soos N. *limoen* (= klein suurlemoen)/A. *lemoen* (= sinaasappel)”.

’n Verdere beduidende verskil tussen Afrikaans en Nederlands het betrekking op die werkwoorde: Die Nederlandse werkwoordsisteem met sy naamval en onderskeid tussen reëlmatige en onreëlmatige werkwoorde en die gepaardgaande vervoegings volgens persoon en tempus kom nie in Afrikaans voor nie (met die uitsondering van ’n handjievol imperfekta). Op hierdie aspekte, wat gewoonlik uitdagings vir Afrikaanssprekende leerders van Nederlands verteenwoordig, word aanstons teruggekeer.

3.2 Bestaande tweetalige woordeboeke met Afrikaans en Nederlands as taalpaar

Waarskynlik om historiese redes en die waargenome groot ooreenkoms tussen Nederlands en Afrikaans bestaan daar nie ’n uitgebreide tipologie van tweetalige woordeboeke met dié twee tale as taalpaar nie. Vergelyk Swart (2003:241-242) vir opgetekende kommentaar oor hierdie tot onlangs toe merkwaardige hiaat.

Historiese redes sluit die internasionale isolasie van Suid-Afrika tydens die apartheidsera in. Omdat Afrikaans beskou is as die taal wat die apartheidregime verteenwoordig het, is dit verstaanbaar dat die fasilitering van kontak tussen Nederlands en Afrikaans deur die daarstel van Nederlands-Afrikaans-woordeboeke nie destyds aangemoedig sou word nie.

In Suid-Afrika het wel voor 1994 ’n eerste leksikografiese poging van Dekker en Paardekooper (1990) onder die titel *Nederlands-Afrikaanse Woordeboek* verskyn. Dié 52-bladsy publikasie bevat ongeveer 1 400 Nederlandse lemmata met Afrikaanse vertaalekwivalente. Weens ruimtebeperkings word volstaan met die samestellers se motivering vir die woordeboek, oftewel die woordeboek se werklike doel (Dekker en Paardekooper 1990:1):

Dit wil die Afrikaanssprekende wat nie Nederlands magtig is nie, veral die senior hoërskoolleerling en jong student, help om boeke in daardie taal maklik en met meer begrip te lees. Dit wil die Nederlandse woorde, woordvorme en uitdrukkings wat vir hom verstaansprobleme mag oplewer en wat hy die

waarskynlikste gaan teëkom, in Afrikaans verklaar. Ook word in lastige gevalle 'n benaderde aanduiding van die uitspraak gegee.

Sedert 1994 het kontak tussen Nederlands en Afrikaans toegeneem (vgl. Swart 2003:243), maar dit het nog nie tot 'n uitgebreide woordeboektipologie met Afrikaans en Nederlands as taalpaar aanleiding gegee nie, waarskynlik onder meer weens die persepsie dat die onderlinge verstaanbaarheid van die twee tale nie die leksikografiese moeite regverdig nie. In 2004 verskyn die *Prisma Miniwoordenboek Afrikaans: Afrikaans-Nederlands / Nederlands-Afrikaans* (Demeersseman 2004) met ongeveer 30 000 lemmata, en in 2010 die *Van Dale Miniwoordenboek Afrikaans-Nederlands / Nederlands-Afrikaans* (Beyer 2010a) met ongeveer 20 000 lemmata. Albei hierdie woordeboekies is kennelik op die toeris afgestem en verteenwoordig waarskynlik die primêre aard van post-1994 kontak tussen die Laaglande en Suid-Afrika.

'n Verdere moontlike rede waarom tot onlangs geen omvattender Afrikaans-Nederlands-woordeboek in Suid-Afrika die lig gesien het nie is die beskikbaarheid van omvattender Engels-Nederlands-woordeboeke. Binne die Suid-Afrikaanse konteks is Afrikaanssprekendes meestal ook Engels magtig en kan dus deur dié taal leksikografiese inligting oor Nederlands verkry.

In 2011 verskyn uiteindelik die omvattendste Afrikaans-Nederlands-woordeboek tot nog, naamlik ANNA (waarvan die beplanning in 1999 begin is – vgl. Swart 2003:243). Dié woordeboek bevat sowat 60 000 lemmata in 'n enkele, geamalgameerde makrostruktuur met die volledigste mikrostrukturele bewerking wat tussen Afrikaans en Nederlands bestaan. Weens sy unieke makrostruktuur is ANNA nie bloot 'n tweetalige woordeboek nie, maar ook 'n kontrastiewe woordeboek wat vergelyking tussen Afrikaans en Nederlands maklik maak (Martin 2011:22). 'n Eksemplaar van ANNA kos in Suid-Afrika egter ongeveer R1 000,00 – 'n prys wat die gemiddelde student van Afrikaans en Nederlands waarskynlik nie kan bekostig nie. Daar bestaan dus steeds 'n nis vir 'n bekostigbare Nederlands-Afrikaans-woordeboek van 'n relatief beperkte omvang wat afgestem is op die Afrikaanssprekende student wat vir die eerste keer Nederlandstalige tekste moet lees. Dit is hierdie nis wat uiteindelik deur NAWA gevul moet word.

4. Die teikengebruiker

Soos hierbo vermeld, is die teikengebruiker van NAWA die Afrikaanssprekende student wat vir die eerste keer Nederlandstalige tekste moet lees. Die teikengroep kan egter nader gekarakteriseer word as studente van Afrikaans en Nederlands. Hoërskoolleerders sou hierby ingesluit kon word. Studente in byvoorbeeld die teologie en die regte wat met Nederlandstalige vakliteratuur gekonfronteer word, vorm nie deel van die primêre teikengebruikersgroep nie, soos wat uit die uiteensetting van die teikengebruikersituasie sal blyk.

4.1 Taalvaardigheid

Die teikengebruikersgroep se taalvaardigheid word gestel as dié van die gemiddelde Afrikaanstalige skoolverlater. Hierdie teikengebruiker beskik oor:

- 'n goeie kennis van (die sintaktiese, morfologiese, fonologiese, semantiese en pragmatiese eienskappe van) die Afrikaanse kernwoordeskat;
- 'n gemiddelde kennis van nie-kernwoordeskatitems in Afrikaans;

- geen of 'n baie beperkte kennis van die Nederlandse leksikon (vgl. ook Bosman 2011).

Van belang hier is egter die resente navorsing van Bosman (2011:154) wat die woordeskatontwikkeling van Afrikaanssprekende studente na aanleiding van 'n Nederlandse woordeskattoets indirek bevraagteken.

4.2 Naslaanvaardighede

Beyer en Faul (2010) het in 'n ondersoek onder 500 eerstejaarstudente aan die Universiteit van Namibië vasgestel dat 71,2% van die respondente minstens een maal per week 'n woordeboek gebruik het terwyl hulle op skool was. 60,8% van die respondente het aangedui dat hulle die een of ander vorm van opleiding in woordeboekgebruik op skool ontvang het. Daar is egter bevind dat die respondente wat gereeld woordeboeke gebruik het 'n akkurater persepsie van die inligtingspotensiaal van woordeboeke toon as dié respondente wat opleiding in woordeboekgebruik ontvang het. Vir Suid-Afrika sou waarskynlik soortgelyke statistiek geld, hoewel die Namibiese en Suid-Afrikaanse talige en taalpolitieke kontekste nie sonder meer aan mekaar gelyk gestel kan word nie. Vir die doel van NAWA kan egter aanvaar word dat die teikengebruiker oor die basiese naslaanvaardighede beskik wat nodig is om minstens die gesogte lemma en dié se vertaalekwivalent(e) te vind, hoewel die dekodering van artikels van polisemiese lemmata 'n uitdaging mag wees.

4.3 Verwysingsraamwerk

Soos reeds aangetoon, beskik die tipiese teikengebruiker oor 'n goeie konsep van wat 'n woordeboek is en behoort die gebruik van NAWA nie ernstige probleme op te lewer nie. In terme van struktuur en inhoud wyk NAWA nie noemenswaardig af van die tipes woordeboeke waarmee die teikengebruiker waarskynlik al te doen gekry het nie. Daar kan wel voorsien word dat NAWA as 'n nuwigheid in terme van sy monofunksionaliteit en skopus binne die Suider-Afrikaanse konteks ondervind mag word, maar ook hierdie kenmerke behoort nie leksikografiese kommunikasie te belemmer nie.

4.4 Inligtingsbehoefte

In die moderne leksikografie dryf die teikengebruiker se inligtingsbehoefte die leksikograaf se besluite tydens die leksikografiese proses (vgl. Gouws en Prinsloo 2005). Spesifieke inligtingsbehoefte ontstaan in die tipiese gebruikersituasie (vgl. 6) en sal aldaar bespreek word.

4.5 Bekostigbaarheid

Daar is reeds in die vorige afdeling gewys op die vereiste dat 'n woordeboek wat op studente afgestem is ook deur studente bekostig behoort te kan word.

5. Die medium

Alhoewel die snelle tegnologiese vooruitgang die elektroniese medium as die aangewese medium vir NAWA voorhou, bestaan die vraag steeds of hierdie medium tans vir die teikengebruiker en in die tipiese gebruikersituasie vanselfsprekend of bloot geredelik toeganklik is.

Daarbenewens word studente gedurende toetse en eksamens (ook potensiële gebruikersituasies) gewoonlik nie toegelaat om enige elektroniese toerusting op hulle persoon te hê nie. Indien studente wel toegelaat word om 'n woordeboek soos NAWA gedurende 'n toets of eksamen te raadpleeg en dié woordeboek is slegs in die elektroniese medium beskikbaar, sal dit uiteraard nie benut kan word nie.

Derhalwe word NAWA (minstens aanvanklik) as 'n gedrukte woordeboek beplan. Die keuse van die gedrukte medium gaan met mediumsiesifieke beperkinge gepaard, in die eerste plek ten einde die eindproduk se bekostigbaarheid vir die teikengebruiker te verseker. Die vernaamste beperking is dié aan ruimte, wat die beperking van die makrostruktuur tot maksimaal 40 000 lemmata noop. Om hoofsaaklik hierdie rede moet 'n spesifieke lemmatiseringsbeleid na aanleiding van aanvanklike lemmatiseringshipoteses vasgestel word.

6. Die tipiese gebruikersituasie – 'n empiriese studie

Die tipiese gebruikersituasie waarop NAWA afgestem is, kom ooreen met dit wat vir die *Nederlands-Afrikaanse Woordeboek* (Dekker en Paardekooper 1990) voorsien is. Dit is naamlik die situasie waarin die tipiese teikengebruikers (die eerste maal) Nederlandse tekste moet lees as deel van hulle studie van Afrikaans en Nederlands. Wanneer die (potensiële) teikengebruiker 'n woord of frase in die Nederlandse teks teëkom wat hy nie verstaan nie, kan hy hom tot NAWA wend om 'n Afrikaanse vertaalekwivalent vir die Nederlandse probleemitem te vind. Gegee die vlakke van leksikale ooreenkoms en verskil tussen Afrikaans en Nederlands (vgl. 3.1) en die ruimtespesifieke beperkinge op die woordeboek (vgl. 5), is dit egter nie vanselfsprekend watter Nederlandse items deur die teikengebruiker as problematies ondervind sou word nie. Om insig hierin te bekom, is 'n empiriese studie onder potensiële teikengebruikers uitgevoer om die naslaanpotensiaal van bepaalde tipes Nederlandse leksikale items vas te stel.

6.1 Doel van die empiriese studie

Die doel van die studie was om die naslaanpotensiaal van 'n versameling Nederlandse leksikale items vas te stel. Met die *naslaanpotensiaal* van 'n (versameling) leksikale items word verwys na die mate waartoe die gemiddelde teikengebruiker in die tipiese gebruikersituasie geneig sou wees om daardie (versameling) leksikale items in 'n gegewe woordeboek na te slaan. 'n Aanduiding hiervan kan bepaalde lemmatiseringshipoteses bewys, lemmatiseringsprioriteite identifiseer en die vasstelling van 'n optimale lemmatiseringsbeleid bemiddel.

6.2 Die populasie

Die populasie het bestaan uit 269 eerstejaarstudente wat in 2010 met hulle tweede semester in die vak Afrikaans en Nederlands aan die Universiteit Stellenbosch besig was. Van hierdie getal het 142 studente (52,8%) 'n module in Nederlandse leesvaardigheid gedurende die eerste semester voltooi en 121 (45,0%) 'n module in Afrikaanse leksikale semantiek. Ses van die studente (2,2%) het nie aangedui watter module hulle gedurende die eerste semester gevolg het nie; gevolglik is hulle response nie by die verwerking van die data betrek nie. Die eersgenoemde groep het dus reeds oor 'n basiskennis van Nederlands beskik terwyl die tweede groep nog nie met Nederlands kennis gemaak het nie.

Die primêre teikengroep in die populasie is die 121 studente wat nie voorkennis van Nederlands gehad het nie – hulle is die lede van Groep 1. Die 142 studente wat die module in Nederlandse leesvaardigheid gedurende die eerste semester voltooi het, is die lede van Groep 2, wat as 'n soort kontrolegroep sou kon dien.

6.3 Metodologie

Die studente is gedurende een van hulle normale lesings 'n Nederlandse teks gegee om te lees. Die teks het bestaan uit die eerste 413 woorde van *Het Fregatschip Johanna Maria* van Arthur van Schendel, 'n populêre teks in Nederlands (vergelyk die Addendum – voortaan “die toetstekst”). Die studente is gevra om die toetstekst te lees met die doel om dit in die geheel te verstaan en om die woorde in die teks te omkring wat hulle sou naslaan indien hulle oor 'n Nederlands-Afrikaans-woordeboek sou beskik. Hierdeur is die tipiese gebruikersituasie gesimuleer. Studente is nie gevra om hulle identiteit te verstrek nie en anonimiteit is verseker.

Vir die doeleindes van analise is die toetstekst gedigitaliseer en met behulp van die konkordansierekenaarprogram AntConc 3.2 is 'n woordelys daaruit onttrek wat bestaan uit 235 tipes. Tipes wat eiename (bv. *Amsterdam* en *Johanna*) verteenwoordig en 'n nie-Nederlandse vreemdetaalkonstruksie (*Nildesp-erandum*) is gediskwalifiseer, wat 228 Nederlandse tipes gelaat het.

Die tipes is in 'n Microsoft Excel[®]-spreiblad ingevoer en die spesifieke tipes wat elke respondent aangedui het hy/sy sou naslaan, is teenoor die betrokke teken en respondent aangeteken. Hierdie metode maak dit moontlik om 'n oorsig te kry van die aantal male wat 'n spesifieke tipe volgens die respondente potensieel nageslaan sou word en om maklik eenvoudige berekeninge te doen. Die 228 tipes verteenwoordig die leksikale items waaruit lemmata vir NAWA geselekteer sou word indien die teks die volledige woordeboekbasis sou verteenwoordig, en daar sal voortaan na hulle as die *lemmakandidate* verwys word.

Na aanleiding van die vergelyking tussen Afrikaans en Nederlands in 3.1 en met inagneming van die eienskappe van die teikengebruikersgroep en die tipiese gebruikersituasie is die lemmakandidate in ses verskillende kategorieë verdeel, naamlik:

- (2) a. absolute kognate (AK): lemmakandidate, werkwoorde uitgesluit, wat na vorm identies is aan hulle Afrikaanse vertaalekwivalente (vgl. (1)a), bv. Ndl./Afr. *anker*;
- b. absolute kognate met vormverskil (AK+V): lemmakandidate, werkwoorde uitgesluit, wat 'n geringe verskil in ortografie met hulle Afrikaanse vertaalekwivalente vertoon, maar die verskil belemmer nie vormlike identifikasie met die Afrikaanse vertaalekwivalent nie (vgl. (1)b), bv. Ndl. *bedenkelyk* x Afr. *bedenklik*;
- c. gedeeltelike kognate (GK): lemmakandidate, werkwoorde uitgesluit, wat in teenstelling met absolute kognate en kognate met vormverskil, met minstens een betekenisonderskeiding en/of in terme van woordsoort verskil van die vormlik (gedeeltelik) identiese Afrikaanse leksikale items (vgl. (1)c), bv. Ndl./Afr. *dan*;
- d. nie-kognate (NK): lemmakandidate, werkwoorde uitgesluit, waarmee die Afrikaanse vertaalekwivalente vir die teikengebruiker geen vormlike identiteit vertoon nie (vgl. (1)d), bv. Ndl. *door* x Afr. *deur*, Ndl. *het* x Afr. *die*;
- e. reëlmatig verboë werkwoorde (Ww.(r)), bv. *piepten*;
- f. onreëlmatig verboë werkwoorde (Ww.(o)), bv. *onderzocht*.

Nie-kognate soos Ndl. *door* x Afr. *deur* sou in ANNA se klassifikasie as absolute kognate met vormverskil gereken word, maar vir die doel van hierdie studie is die vormverskil as voldoende geag om te aanvaar dat die teikengebruiker nie die verband tussen die leksikale items sou agterhaal nie. Die objektiewe kriterium hiervoor is 'n verskil in die vokale of diftonge binne die stamme van die betrokke leksikale items (bv. [o:] in Ndl. *door* x [ø] in Afr. *deur*).

Die feit dat werkwoorde twee aparte kategorieë lemmakandidate vorm, erken die afwykende werkwoordsisteem van Nederlands in vergelyking met Afrikaans en die gepaardgaande uitdagings met betrekking tot teksresepsie vir die teikengebruiker van NAWA.

Lemmakandidate is gekategoriseer op grond van hulle werklike gebruik in die toetsteks (d.w.s. as woorde) en nie op grond van hulle infinitiefvorme (of as lekseme) nie. Dus is die lemmakandidaat *wimpels* as absolute kognaat gekategoriseer aangesien die meervoudsvorm in Nederlands en Afrikaans identies is, terwyl die lemmakandidaat *engelsen* as absolute kognaat met vormverskil gekategoriseer is aangesien die meervoudsvorme van mekaar afwyk (Ndl. *engelsen* x Afr. *Engelse*).

Geen valse vriende is in die toetsteks verteenwoordig nie; derhalwe geld dié kategorie nie in die onderhawige studie nie.

6.4 Bevindinge

Daar is 'n beduidende verskil in die aantal naslaanprosedures wat respondente met geen voorkennis van Nederlands nie (Groep 1) aangedui het hulle sou uitvoer teenoor die aantal deur respondente wat reeds met Nederlands kennis gemaak het (Groep 2) aangedui: Groep 1 het meer as twee maal soveel potensiële naslaanprosedures as Groep 2 gerapporteer, soos wat Tabel 1 hieronder toon.

Tabel 1: Aantal potensiële naslaanprosedures per respondentegroep

Groep	Aantal potensiële naslaanprosedures gerapporteer
Groep 1: Geen voorkennis van Ndl. Nie	3 374 (70%)
Groep 2: Basiese lees kennis van Ndl.	1 449 (30%)
Totaal	4 823 (100%)

Hierdie resultaat is uiteraard nie verrassend nie. Trouens, dit bewys minstens dat die respondente in Groep 2 minder behoefte aan die tersaaklike leksikale inligting ondervind, hoogs waarskynlik as gevolg van die kennis wat hulle tydens die eerste semester in die module in Nederlandse leesvaardigheid opgedoen het.

Die resultate in Tabel 1 gee egter geen aanduiding van hoeveel lemmakandidate per kategorie potensiël deur hoeveel van die respondente in elke groep nageslaan sou word nie. Om hierdie verhouding uit te druk, moet 'n bepaalde indeks ontwerp word. Die indeks wat vir hierdie doel ontwikkel is, word die *naslaanpotensiaalindeks* (NPI) of die *werklike naslaanindeks* (WNI) genoem, na gelang van of die gerapporteerde naslaanprosedures potensiële naslaanprosedures (soos in hierdie studie) of werklike naslaanprosedures (soos

tydens die waarneming van werklike woordeboekgebruik) verteenwoordig. Die NPI/WNI word uitgedruk as 'n waarde relatief tot 1,0. 'n NPI/WNI van 1,0 beteken dat elke lid van die populasie van teikengebruikers elke lemmakandidaat in 'n bepaalde kategorie of in 'n bepaalde teks nageslaan het of sou naslaan; 'n NPI/WNI van 0,5 sou beteken dat een uit twee lede van die populasie elke lemmakandidaat in 'n bepaalde kategorie of teks nageslaan het of sou naslaan, of dat elke lid van die populasie een uit twee lemmakandidate in 'n bepaalde kategorie of teks nageslaan het of sou naslaan, ensovoorts. Die NPI/WNI word bereken aan die hand van die volgende vergelyking:

$$\text{NPI of WNI} = \frac{\text{aantal naslaanprosedures per kategorie of teks gerapporteer (} p \text{)}}{\text{aantal potensiële lemmata per kategorie of teks (} l \text{)} \times \text{populasie (} n \text{)}}$$

oftewel

$$\text{NPI of WNI} = \frac{p}{ln}$$

Die toepassing van hierdie vergelyking lewer die resultate in Tabel 2:

Tabel 2: Naslaanpotensiaalindekse (NPI's) vir die onderskeie kategorieë lemmakandidate per respondentegroep

Kategorie	Groep	<i>n</i>	<i>p</i>	<i>l</i>	NPI
AK	1	121	313	77	0,03
	2	142	117		0,01
AK+V	1	121	1 201	65	0,15
	2	142	474		0,05
GK	1	121	46	13	0,03
	2	142	20		0,01
NK	1	121	612	22	0,23
	2	142	371		0,12
Ww. (r.)	1	121	560	25	0,19
	2	142	202		0,06
Ww. (o.)	1	121	642	26	0,20
	2	142	265		0,07
Totaal	1	121	3 374	228	0,12
	2	142	1 449		0,04

Die onderskeie NPI's kan meer oorsigtelik in die staafgrafiek in Figuur 1 hieronder voorgestel word.

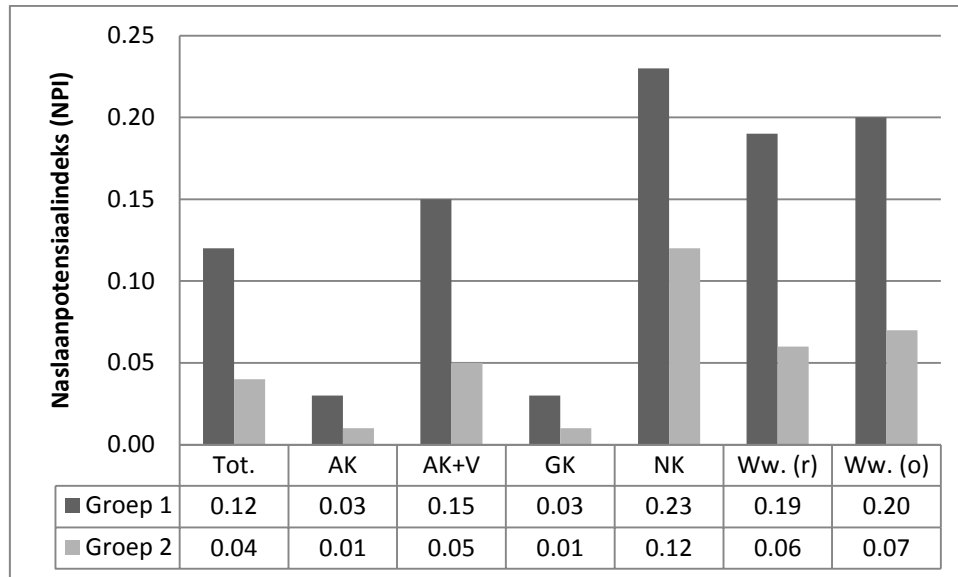


Fig. 1: NPI's vir die onderskeie kategorieë lemmakandidate per respondentegroep¹

Van belang is die NPI's vir die onderskeie kategorieë lemmakandidate met betrekking tot Groep 1, dié groep respondente met geen voorkennis van Nederlands nie. Die NPI's met betrekking tot Groep 2 word interessantheidshalwe betrek.

Duidelik vertoon nie-kognate die hoogste NPI, waarna onreëlmatig verboë werkwoorde en reëlmatig verboë werkwoorde volg. Daarna vertoon absolute kognate met vormverskil steeds 'n NPI wat hoër is as die totale NPI, met gedeeltelike kognate en absolute kognate wat die laagste NPI's het (wat ook laer as die totale NPI is).

Interessant is die beduidende verskil in die NPI vir absolute kognate met vormverskil (0,15) teenoor dié vir gedeeltelike kognate (0,03). Die moontlike rede hiervoor is dat die kategorie gedeeltelike kognate meestal lemmakandidate insluit wat vormlik identies aan hulle Afrikaanse vertaalekwivalente is met betrekking tot dié betekeniswaardes wat deur die lees van die teks geaktiveer word, afgesien daarvan dat 'n ekwivalentverhouding van semantiese divergensie of polidivergensie telkens geld. Hulle sou binne dié teks dus vir absolute kognate aangesien word. Dít verklaar ook die identiese NPI's vir gedeeltelike kognate en absolute kognate (0,03).

Die NPI's vir Groep 2 vertoon 'n ongeveer identiese patroon aan dié vir Groep 1, behalwe dat die indekswaardes (uiteraard) proporsioneel laer is ('n totaal van 0,12 vir Groep 1 en 0,04 vir Groep 2).

Veral die NPI-waardes vir Groep 1 kan intuïtief (in die afwesigheid van WNI-waardes) as relatief hoog beskou word. Hiervoor kan twee redes aangevoer word: In die eerste plek is die teks wat respondente moes lees nie baie lank nie; in die tweede plek het werklike woordeboekgebruik nie plaasgevind nie – slegs potensiele woordeboekgebruik is gerapporteer. In hoe 'n mate 'n NPI sou afwyk van 'n WNI kan indirek aangetoon word deur die berekening van *potensiele naslaanfrekwensie* (PNF), waarvolgens die aantal potensiele

¹ Sleutel: Tot. = totaal; AK = absolute kognate; AK+V = absolute kognate met vormverskil; GK = gedeeltelike kognate; NK = nie-kognate; Ww. (r) = reëlmatig verboë werkwoorde; Ww. (o) = onreëlmatig verboë werkwoorde.

naslaanprosedures per stuk gelese teks uitgedruk kan word. 'n PNF staan in teenstelling tot 'n *werklike naslaanfrekwensie* (WNF), waarvolgens die werklike aantal naslaanprosedures per stuk gelese teks uitgedruk word. Die vergelyking wat vir hierdie doel voorgestel word, is:

PNF of WNF

$$= \frac{\text{aantal naslaanprosedures per teks gerapporteer } (p) \times \text{maateenheid in woorde } (w)}{\text{aantal tekens in die teks } (t) \times \text{populasie } (n)}$$

oftewel

$$\text{PNF of WNF} = \frac{pw}{tn}$$

Om die PNF vir hierdie studie as die aantal naslaanprosedures per 100 tekens/woorde (nié tipes nie) lopende teks uit te druk, word die waarde 100 aan die maateenheid in woorde (w) toegeken. Die resultate vir die twee groepe is soos volg:

Tabel 3: Potensiële naslaanfrekwensie (PNF) in naslaanprosedures per 100 woorde lopende teks

Groep	p	W	t	n	PNF/100 woorde
1	3 374	100	413	121	6,8
2	1 449			142	2,5

Groep 1 toon 'n PNF van 6,8 naslaanprosedures per 100 woorde lopende teks. Dít veronderstel een naslaanprosedure na elke 14,7 gelese woorde. Volgens MS Word[®] is die gemiddelde sinslengte in die toetsteks 34,5 woorde. Die PNF dui dus op gemiddeld twee potensiële naslaanprosedures per sin, wat 'n enigszins onrealistiese naslaanfrekwensie verteenwoordig: Woordeboekgebruiknavorsing in die tweede- en vreemdetaalonderrigkonteks dui daarop dat leerders wat woordeboeke gebruik tot twee maal so lank neem om 'n teks te lees as leerders wat nie woordeboeke gebruik nie (Nation 2001:283). Boonop onderbreek elke naslaanprosedure die leesproses en kan teksbegrip ironies genoeg daaronder ly. Dit is dus onwaarskynlik dat 'n WNF ten opsigte van dieselfde teks so hoog soos die PNF sou wees.

Volgens Cook (2008:59) is die raai van woordbetekenis uit die konteks 'n strategie wat baie deur vreemdetaalleerders aangewend word wanneer hulle onbekende woorde tydens teksresepsie in die vreemde taal teëkom. Trouens, Qian (2004) het bevind dat dit dié populêrste strategie onder Koreaanse leerders van Engels is, met woordeboekgebruik wat in die tweede plek staan. Indien leerders wat 'n genealogies vreemde taal aanleer die meeste van die woordbetekenis uit die konteks aflei om tekste te verstaan, kan aanvaar word dat dit waarskynlik ook die voorkeurstrategie sal wees van Afrikaanssprekende studente wat 'n teks in die genealogies nou verwante Nederlands probeer verstaan. Hier sou dit veral oor die kognitiewe koppeling van kognate gaan, soos wat die tersaaklike relatief lae NPI's suggereer. Cook (2008:60) het egter bevind dat leerders gereeld sodanige koppeling vermy, moontlik uit vrees vir valse vriende. Daarom kan selfs nabye of absolute kognate ook nageslaan word, soos

wat die tersaaklike NPI's eweneens suggereer. Die punt bly egter staan dat die PNF onrealisties hoog is, en dat 'n WNF waarskynlik beduidend laer sou wees. Dit beteken egter nie dat die NPI's per respondentegroep nie steeds relatief tot mekaar akkuraat is nie, soos wat die duidelike verband tussen die NPI's van Groep 1 en dié van Groep 2 suggereer.

Daar kan dus aangeneem word dat die WNI's beduidend laer sou wees as die NPI's, maar dat dieselfde relatiewe tendense met betrekking tot die onderskeie kategorieë lemmakandidate vertoon sou word.

7. Gevolgtrekking: Die woordeboek

Die vasstelling van die naslaanpotensiaal van verskillende kategorieë potensiële lemmata bemiddel die identifisering van 'n lemmatiseringsrangorde in terme van dié kategorieë. Die lemmatiseringsprioriteitslys vir NAWA kan streng hiervolgens soos volg gestel word:

1. Nie-kognate
2. Onreëlmatige werkwoorde gelemmatiseer in hulle infinitief- én verboë vorme
3. Reëlmatige werkwoorde gelemmatiseer in hulle infinitief- én verboë vorme
4. Absolute kognate met vormverskil
5. Gedeeltelike kognate
6. Absolute kognate

8. Slot

In hierdie artikel is gedemonstreer hoe die empiriese vasstelling van die naslaanpotensiaal van lemmakandidate van hulp kan wees in die ontwikkeling van 'n lemmatiseringsbeleid, wat tot woordeboeke aanleiding sal gee waarvan minstens die makrostruktuur aan die inligtingsbehoefte van die teikengebruiker in die tipiese gebruikersituasie sal voldoen. Die bespreking is binne die raamwerk van 'n kommunikatiewe metaleksikografie gevoer.

Die data wat deur die empiriese ondersoek gegenerer is, is hoegenaamd nie uitgeput nie. Verskeie onderlinge verbande tussen tot nog toe ongeïdentifiseerde relatiewe kategorieë lemmakandidate mag steeds lê en wag om verken te word.

Erkenning

Die empiriese komponent van die navorsing is in die tweede semester van 2010 aan die Universiteit Stellenbosch uitgevoer. Die outeur bedank prof. Rufus Gouws, Voorsitter van die Departement Afrikaans en Nederlands, vir toestemming om die opname onder studente in die Departement te doen, en dr. Amanda Lourens, wat die opname tydens een van haar eerstejaarlesings gefasiliteer het. Die navorsing is finansiële ondersteun deur die Kaapse Forum vir Neerlandistiek en die Van Ewijck Stigting. Die verantwoordelikheid vir die inhoud van hierdie artikel bly uiteraard dié van die outeur.

Verwysings

Beyer, H.L. 2006. 'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelikheid in woordeboeke. *Lexikos* 16:46-66.

Beyer, H.L. 2010. Kontekstuele bepalers by die beplanning van skoolwoordeboeke. *Lexikos* 20:52-72.

Beyer, H.L. 2010a. *Van Dale Miniwoordenboek Afrikaans-Nederlands / Nederlands-Afrikaans*. Utrecht/Antwerpen: Van Dale.

Beyer, H.L. en A.K. Faul. 2010. Namibian University Entrants' Concepts of 'a Dictionary'. *Lexikos* 20:644-679.

Bosman, N. 2011. Die diagnostiese waarde van 'n Nederlandse woordeskattoets vir Afrikaanssprekende aanleerders. *LitNet Akademies* 8(3):137-166.

Cook, V. 2008. *Second Language Learning and Language Teaching*. Vierde uitgawe. Londen: Hodder Education.

Dekker, L. en P.C. Paardekooper. 1990. *Nederlands-Afrikaanse Woordeboek*. Pretoria: J.L. van Schaik.

Demeersseman, H.J. (red.). 2004. *Prisma Miniwoordenboek Afrikaans: Afrikaans-Nederlands / Nederlands-Afrikaans*. Utrecht: Het Spectrum.

Gouws, R.H. en D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PRéSS.

Martin, W. (red.). 2011. *Groot Woordeboek Afrikaans en Nederlands*. Kaapstad: Pharos.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Pheiffer, R.H. 1989. Afrikaans en Standaard-Nederlands. In T.J.R. Botha (red.). *Inleiding tot die Afrikaanse taalkunde*. Tweede, hersiene uitgawe. Pretoria/Kaapstad: Academica. 72-95.

Ponelis, F. 1993. *The Development of Afrikaans*. Frankfurt am Main: Peter Lang.

Qian, D.D. 2004. Second language lexical inferencing: Preferences, perceptions, and practices. In P. Bogaards en B. Laufer (reds.). *Vocabulary in a Second Language. Selection, acquisition and testing*. Amsterdam/Philadelphia: John Benjamins. 155-169.

Raidt, E.H. 1989. Ontwikkeling van Vroeë Afrikaans. In T.J.R. Botha (red.). *Inleiding tot die Afrikaanse taalkunde*. Tweede, hersiene uitgawe. Pretoria/Kaapstad: Academica. 96-126.

Steinberg, S. 2007. *An Introduction to Communication Studies*. Kaapstad: Juta.

Swart, C. 2003. Uitdagings vir die Afrikaanse leksikograaf ten opsigte van Nederlandse lemmata in 'n vertalende Nederlands-Afrikaanse woordeboek. *Lexikos* 13:240-249.

Van Cuilenburg, J.J., O. Scholten en G.W. Noomen. 1992. *Communicatiewetenschap*. Bussum: Couthino.

Van Schendel, A. 1930. *Het fregatschip Johanna Maria*. Amsterdam: Meulenhof.

ADDENDUM: TOETSTEKS

Het driemast volschip Johanna Maria werd op een dag in februari 1865 van de werf op Oostenburg in Amsterdam te water gelaten. Het was windstil en vochtig weer, maar de zon scheen wit door de nevel, glanzend op het verse rood van de kielt en van de vlag. Toen een meisjeshand hem had bevrijd begon de romp onder het geroep van mannen langzaam te glijden, dan stortte hij zich voorwaarts en hij plonsde onstuimig in het opspattend water dat over de steven sloeg, tot de kabels die hem hielden kraakten en piepten. Nu lag daar een schip dat deinde op de kabbeling met de kleuren van zijn vlag weerspiegeld in het nat.

Een maand later, nadat de masten waren opgesteld, het binnenwerk gedaan en deugdelijk onderzocht, werd het naar het Nieuwediep getrokken voor de voltooiing van het tuig. Daar kwam aan de kade ook een schuit langszij met werklieden die het schegbeeld brachten, een vergulde vrouwenfiguur, de Hoop voorstellende, met een anker in de linkerhand en de rechter op de borst gedrukt; van de hals kronkelde een band, aan de scheg zelf geslagen, waarop in zeven letters ter ene en zeven ter andere, de zinspreuk stond: Nildesp-erandum. Van koper waren band en letters opdat, of het verguldsel ook vergaan mocht, de zinspreuk vast zou blijven aan het vaartuig.

Mannen met ringbaarden stonden morgen en middag op de wal te kijken; soms knipoogde er een voor hij verder ging, soms nam er een de pijp uit de mond om een enkel woord te zeggen; een hoofdknik was voor de bewondering, maar geen die niet bedenkelijk de ogen kneep als hij de hoogte van masten en stengen mat. De Johanna Maria, hoewel kloek gebouwd, rijzig in de boeg, vol in de ribben, droeg inderdaad onder de wimpels een vermetele tuigage. Toen zij voltooid lag en de reders, de heren ten Hope, haar kwamen beschouwen met de bouwmeester en de kapitein, meenden zij zelf dat de bovenste stengen gekort moesten worden; de kapitein echter, die een gelukkig man was, vertelde lachend van juist zulk tuig waarmede de Engelsen wonderen van snelheid haalden, en hij gaf zijn hand er op hetzelfde te zullen doen als de heren maar voor de goede lading zorgden. Na dit woord bezichtigden zij ieder onderdeel voor en achter, in ruimen, kombuis en bak, en teruggekeerd in de kajuit dan, voldaan over de hechtheid en de keurigheid van het werk, beklonken zij met Rijnse wijn hun vertrouwen op de zinspreuk van hun eigendom.