

Bachelorscriptie Aimée Talen 3472892

Taal- en Cultuurstudies

Universiteit Utrecht

Begeleider: Martin Everaert

Tweede lezer: Joost Zwarts

22 oktober 2014



*Een woord is uit, een woord is in  
wordt het gebruikt, komt het erin  
Willem III z'n selfies zijn nog steeds meer kop dan munt  
en de Van Dale die is weer een stukkie dikker*

*de radio en de TV  
spouwen ze uit  
en dat werkt mee  
ookal heeft je moeder geen idee wat shinen is  
is de Van Dale toch weer mooi een stukkie dikker*

*een woord erin, een woord eruit*

*'t Geheugenkoor zingt bij Holland's Got talent  
de mijlpaaljury heeft beslist dit wordt het helemaal  
en de Van Dale is alweer een stukkie dikker*

*een woord erin, een woord eruit*

*En sommigen die mogen dan wel hip zijn  
maar sukkelseks dat heeft een man pas als 'tie dronken is  
en de Van Dale is alweer een stukkie dikker*

*een woord is uit, een woord is in  
wordt het gebruikt, komt het erin  
Geishaballen gooien heeft de kermis niet gehaald  
maar de Van Dale is weer mooi een stukkie dikker*

*een woord erin, een woord eruit*

*of Zwarte Piet een vrouw een man of trans is  
maakt eigenlijk niet uit  
want Piet is flexiseksueel*

*en De Van Dale die is weer een stukkie dikker*

(Liedje dat The Kik ten gehore bracht tijdens de theatervoorstelling Hotel van Dale (28 augustus 2014) ter ere van het 150-jarig jubileum van de woordenboeken van Van Dale)



## Inhoud

1. Inleiding .....	7
2. Van Dale woordenboeken .....	8
3. Theoretisch kader .....	11
3.1 Semantic primes en primitives .....	11
3.2 NSM .....	12
3.3 NSM in woordenboeken.....	14
3.3.1 Semantic invariant.....	14
3.3.2 Verbeteringen voor woordenboeken .....	15
3.3 Verwachtingen van woordenboeken .....	20
4. Methode.....	21
5. Resultaten .....	26
5.1 Papieren woordenboeken chronologisch.....	27
5.2 Papieren woordenboeken versus Basiswoordenboek.....	30
5.3 Betaald onlinewoordenboek versus gratis onlinewoordenboek .....	32
5.4 Papieren woordenboeken versus betaald onlinewoordenboek .....	34
5.5 Subbetekenissen .....	35
6. Conclusie.....	40
<b>Bibliografie.....</b>	<b>43</b>
Bijlage I Korte Swadeshlijst (35 woorden) .....	45
Bijlage II Semantic primes .....	46



## 1. Inleiding

De dikke Van Dale is al 150 jaar een begrip in het Nederlandse taalgebied. Het woordenboek, tegenwoordig uitgegeven door Van Dale Uitgevers, wordt voor het Nederlandse taalgebied als het meest gezaghebbende woordenboek beschouwd. Het woordenboek omschrijft de taal zoals die nu gesproken wordt. Bij elke editie worden er nieuwe woorden aan het woordenboek toegevoegd. Het lijkt dan ook niet uit te kunnen blijven dat het woordenboek elke editie een beetje groeit. De eerste van Dale, uit 1864 was één boekwerk van ongeveer 1400 pagina's. In 1970 paste het niet meer in één boek, en verscheen het in twee delen met totaal 2800 pagina's. De meest recente Van Dale, uit 2005, is een driedelig werk met ongeveer 4500 pagina's. Welke keuzes maken lexicografen om tot een hanteerbaar woordenboek te komen? De hoeveelheid lemma's is in deze scriptie buiten beschouwing gelaten, maar ik heb mij geconcentreerd op de omvang van de lemma's. Is de opbouw van die lemma's veranderd over de edities? Daarbij zal gekeken worden naar de verschillende betekenissen die een woord kan hebben, het aantal voorbeelden dat per lemma gegeven wordt, en het aantal verwijzingen dat een lemma bevat. Juist die elementen hebben veel invloed op de lengte van een lemma, zoals later zal blijken. Ik heb gekeken naar de omvang van de betekenisomschrijvingen. Anna Wierzbicka heeft daar een theorie over – de betekenisomschrijvingen zouden vooral niet te omslachtig moeten zijn – die in het Theoretisch Kader wordt verhelderd.

Van Dale is inmiddels ook toegankelijk via het internet. Het is gedaan met ruimtegebrek: digitale lemma's zouden in principe oneindig lang kunnen zijn. Dat bracht mij tot de vraag hoe de digitale betekenisomschrijvingen zich verhouden tot de betekenisomschrijvingen in de papieren woordenboeken. Zijn de betekenisomschrijvingen in de digitale versies inderdaad uitgebreider dan de papieren versies, en op welk vlak dan precies? Verandert het lexicografische werk nu het medium geen fysieke beperkingen meer oplegt?

## 2. Van Dale woordenboeken

Van Dale woordenboeken gaan al geruime tijd mee: in 1864 verscheen de eerste editie, en in 2005 verscheen de recentste papieren uitgave. Intussen is ook van Dale gemoderniseerd en heeft het een online woordenboek, dat steeds up-to-date wordt gehouden, en zelfs een applicatie voor tablet en smartphone. In de 150 jaar dat Van Dale bestaat zijn er 14 papieren uitgaven gemaakt (de 15<sup>e</sup> is onderweg: die moet in 2015 uitkomen), elk een verbeterde uitgave van de voorgaande, maar “[e]en nieuwe uitgave is geen nieuw woordenboek” (Van Dale 2005:XIV). Ook niet alle nieuwe woorden die erin staan zijn daadwerkelijk nieuwe woorden, want “[w]erkelijk nieuwe woorden zijn zeer zeldzaam; het overgrote deel van het materiaal bestaat uit nieuwe toepassingen, nieuwe betekenissen van bestaande woorden, uit nieuwe samenstellingen en uit ontleningen.” (Van Dale 1970:V)

Het eerste woordenboek dat als Van Dale wordt beschouwd, is strikt genomen geen Van Dale. Het is samengesteld door I. M. Calisch en N. S. Calisch. De eerste echte Van Dale, die uitkwam in 1872, is daar een bewerking op, gemaakt door Johan Hendrik Van Dale. Helaas heeft Van Dale niet meer meegemaakt dat zijn woordenboek daadwerkelijk uitkwam: hij overleed toen het woordenboek pas tot de letter Y voltooid was. Zijn medewerker J. Manhave voltooide dit “reuzenwerk” en besloot dat de naam van Van Dale “steeds met eere aan dit Woordenboek verbonden zal blijven” (J. Manhave 1884:IX).

Van Dale streeft al vanaf het begin naar volledigheid, maar “[e]en volstrekt volledig woordenboek is een ideaal, dat nimmer te bereiken is” (J. Manhave 1884:XII). Het is erg moeilijk om te besluiten dat een lemma ‘af’ is: “... niets is gemakkelijker dan een woordenboek in omvang te doen uitdijen” (J. Manhave 1884:XI).

Van Dale wil volledig zijn, maar daarbij wel een wóordenboek blijven: “Was er van een of ander dier sprake, dan luidde het: zekere vogel, zeker insect enz. Ik heb getracht,



zonder daarom van dit Woordenboek eene encyclopedie te willen maken, menigeen dergelijke teleurstelling te besparen” (J. Manhave 1884:VIII). Jarenlang wilde Van Dale ook niets weten van uitgebreide etymologieën, tot ze in 1992 - om onbekende redenen – voor het eerst toch in de dikke Van Dale verschenen.

Samenstellingen werden in de woordenboeken ook niet gevonden, want “[d]e hoofdschotel vormt, als gezegd, de algemene Nieuwnederlandse gesproken en geschreven taal. Daarbij is naar volledigheid gestreefd, maar dit wil niet zeggen dat nu ook alle bestaande Nederlandse woorden - om nog te zwijgen van de bestaانبare - er in staan. Wat de grondwoorden betreft is deze volledigheid wel benaderd, maar voor het overige bedenke men dat het Nederlands een taal is die op ongelimiteerde wijze nieuwe woorden kan vormen door samenstelling en op grote schaal ook door afleiding. Alle samenstellingen te vermelden was dus praktisch onmogelijk en ook overbodig, daar verreweg de meeste te begrijpen zijn als men de samenstellende delen kent. Niettemin zijn gangbare samenstellingen zo veel mogelijk opgenomen, soms zonder omschrijving van de betekenis, als deze duidelijk genoeg was, dus alleen volledigheidshalve. De afleidingen van werkwoorden op -aar, -er, -ster en -ing zijn niet altijd opgenomen, daar zeer vele van deze bij elk werkwoord denkbare formaties in de praktijk nooit gebruikt worden” (Van Dale 1950:XI). In 1970 leek het absoluut ondenkbaar om samenstellingen op te nemen in een Van Dale, want “[n]iet alleen is het aantal mogelijkheden op dit terrein onoverzienbaar, maar er wordt ook op schromelijke wijze misbruik van gemaakt. Het is niet te veel gezegd als men dit misbruik als de ernstigste ziekte aanwijst waaraan de Nederlandse taal thans lijdt, een kwaal die men verbale lintwormziekte zou kunnen noemen, of ook horror praepositionis.” (1970:VI)

In 1970 werd het woordenboek toch zo volledig dat het niet meer in één boek paste, en daarom is het in 1970 uitgegeven in twee delen. In 1984, slechts twee uitgaven later, was het al nodig om er drie delen van te maken.

In 1984 is besloten tot een andere manier van het opschrijven van de lemma's: “[z]ij zijn niet in kapitalen, maar in kleine vette letters gedrukt, waardoor het mogelijk is direct te zien of een woord met een hoofdletter geschreven moet worden of niet (...)” (1984:XIX).

De Van Dale woordenboeken zijn in de afgelopen 150 jaar steeds vollediger geworden, of in ieder geval steeds dikker. Waar is die volledigheid voornamelijk in terug te vinden?

### 3. Theoretisch kader

Anna Wierzbicka denkt dat metataal een handig hulpmiddel kan zijn bij het schrijven van woordenboeken. Metataal is geschikt om op een heel rigide manier woordbetekenissen te omschrijven. Het zoekt de absolute betekeniskern van een woord en verwoordt dat door gebruik te maken van *semantic primes*, ook wel *semantic primitives* genoemd.

#### 3.1 Semantic primes en primitives

In het volgende citaat legt Wierzbicka (1996:10) uit wat die semantic primes zijn:

“The elements which can be used to define the meaning of words (or any other meanings) cannot be defined themselves; rather, they must be accepted as “indefinibilia”, that is, as semantic primes, in terms of which all complex meanings can be coherently represented. A definition which attempts to explain the simple word *if* via the complex word *implication* flies in the face of the basic principle of sound semantic analysis put forward more than two millennia ago by Aristotle (1937: 141a) (...)”

Semantic primes zijn dus woorden die ondefinieerbaar zijn. De concepten die met die woorden samenhangen zijn aangeboren (Wierzbicka 1996:14, 16). Enkele semantic primes zijn bijvoorbeeld “IK”, “ZIEN”, en “HIER”. Ze worden weergegeven in hoofdletters, omdat niet per se het Nederlandse woord (of het woord uit een andere taal) een semantic prime is, maar het concept dat erachter schuilgaat. Wel hebben alle (tot nog toe onderzochte) talen een woord of gebonden morfeem voor alle semantic primes (1992:10). Een recente lijst met die ondefinieerbare woorden is te vinden in het artikel van Goddard en Wierzbicka (2007) en in bijlage II. Door hun eigen ondefinieerbaarheid kunnen die ondefinieerbare semantic primes juist goed gebruikt worden voor het definiëren van andere woorden. Wierzbicka (1996:10) stelt het nog sterker: zonder *semantic primes* kunnen er geen betekenissen omschreven worden: dat is “translating the unknowns into the unknowns.”

Die simpele semantic primes zijn nodig om andere woorden te kunnen begrijpen. Wierzbicka (1996:10) spreekt dan ook van een “absolute order of understanding” die afhangt van semantische complexiteit. Het is onmogelijk om een concept als ‘beloven’ te begrijpen, zonder eerst het concept ‘zeggen’ te begrijpen (Wierzbicka 1996:10). In de absolute order of understanding moet eerst ‘zeggen’ begrepen worden, alvorens ‘beloven’ begrepen kan worden.

Semantic primes zijn er niet zomaar. Ze kunnen worden gevonden door een diepte-analyse uit te voeren op willekeurige talen (Wierzbicka 1996:13). De sets met vermoedelijke semantic primes uit al die talen moeten dan wel overeenkomen, want anders zijn er alleen maar taalspecifieke concepten vastgelegd (Wierzbicka 1992:10, 13, 1996:13).

Semantic primes zijn dus universele concepten en de woorden die daarbij horen. Ze komen voor in iedere natuurlijke taal. Zelf zijn ze ondefinieerbaar, maar daardoor zijn ze handig bij het verklaren van andere, wel definieerbare woorden.

### **3.2 NSM**

Naast universele woorden en concepten (de semantic primes) bestaat er ook een universele grammatica. Die universele grammatica bevat elementen die universeel voorkomen in natuurlijke talen. Syntax (de universele grammatica) en semantiek (de semantic primes) kunnen niet los van elkaar gezien worden, omdat betekenis pas tot stand komt als grammaticale constructies en semantiek samenkomen (Wierzbicka 1988:1, 8, 293).

Het primaire doel van Wierzbicka is om de gedeelde kern (dus semantic primes in combinatie met de universele grammatica) van alle natuurlijke talen vast te leggen en om daar een “natural semantic metalanguage” (NSM) van te maken, of “Lingua Mentalis” zoals zij het zelf noemt (Wierzbicka 1980). Met NSM is het mogelijk om betekenissen te omschrijven die in welke natuurlijke taal dan ook kunnen worden overgedragen. Vaak worden glossen

gebruikt voor vertalingen, maar daarbij gaat vrijwel altijd betekenis verloren. Het gebruik van NSM is veel nauwkeuriger, omdat daarbij geen betekenis verloren gaat (Wierzbicka 1996:23).

Het idee dat alle talen een gedeelde kern hebben is niet nieuw, maar het idee dat daaruit een metataal gedistilleerd kan worden wel (Wierzbicka 1996:23). NSM onderscheidt zich van andere betekenisomschrijvende kunstmatige talen doordat NSM gedistilleerd is uit natuurlijke taal, en begrepen kan worden zonder verdere uitleg (Wierzbicka 1996:23).

De theorie van NSM is al jarenlang dezelfde: het is een zoektocht naar universele semantic primes, die geïmplementeerd worden in een universele grammatica. Het is een “self-explanatory” systeem: het maakt geen gebruik van features, markers, of logica, maar steunt volledig op natuurlijke taal (Wierzbicka 1996:31). De optimale metataal is zo minimaal mogelijk. Het bestaat uit slechts die elementen die nodig zijn om semantische relaties aan te geven, en een minimale set van universele grammaticale constructies (Wierzbicka 1988:11-12).

De omschrijving van een woord, in dit voorbeeld het Engelse “courageous” komt er in NSM zo uit te zien:

“X is courageous. =

X can do very good things when other people can't because when other people think something like this:

I don't want bad things to happen to me

X thinks something like this:

it is good if I do this

it is bad if I don't do it

I want to do it because for this

this is good”

(Wierzbicka 1996:241)

### **3.3 NSM in woordenboeken**

Wierzbicka (1996:278) ziet in haar lijst met semantic primes potentie voor lexicografie, en daarmee woordenboeken. De lijst heeft zich in de lexicale semantiek en in NSM al bewezen en zou voor een- en tweetalige woordenboeken ook uitkomst bieden (Wierzbicka 1996:281-282). Er kleeft immers een groot praktisch voordeel aan: een taalleerder kan woordenboekverklaringen in NSM veel makkelijker begrijpen dan verklaringen waarbij gebruik gemaakt wordt van gloss, omdat er gebruik gemaakt wordt van woorden en concepten die ook in zijn moedertaal voorkomen (Wierzbicka 1996:281-282).

Toch worden verklaringen in reguliere woordenboeken meestal niet in een taal als NSM weergegeven. Een dergelijke metataal is handig om nauwkeurige en rigide definities te geven, maar niet voor het gebruik van bijvoorbeeld tweedetaalonderwijs, beschrijvende grammatica, of lexicografie. De formules die de metataal oplevert zijn daar veel te lang en te moeilijk voor (Wierzbicka 1988:12-13). In die gevallen is het veel handiger om een soort uitgebreide versie van de metataal te hebben, die meer woorden bevat, maar die dan vanzelfsprekend niet allemaal ondefinieerbaar zijn. Tevens willen woordenboekgebruikers niet enkel definities opzoeken, maar ook gezegdes en synoniemen. Metataal is goed voor de definities, maar zal niet snel zijn intrede doen in een woordenboek als Van Dale, omdat metataal niet bedoeld is voor gezegdes en dergelijke.

#### **3.3.1 Semantic invariant**

Alle bestaande woorden hebben een “semantic invariant”: een harde kern waarin de betekenis van een woord vastligt (Wierzbicka 1996:240). Dat is een deel van de impliciete kennis van de moedertaal (1996:264.) Het zoeken naar een semantic invariant die alle gebruiken van een woord omvat is een zeer tijdrovende klus, waar lexicografen vaak niet in slagen (Wierzbicka 1996:240-241). Doordat de semantic invariant niet gevonden wordt, worden begrippen als *of* en *in het bijzonder* gebruikt in de definities, en worden vele glossen gebruikt, en er wordt verkeerdt gebruik gemaakt van polysemie (Wierzbicka 1996:267). Op deze manier worden

woordenboeken veel onhandiger dan ze zouden kunnen zijn (Wierzbicka 1996:267). Een nauwgezette, consistente lexicografische theorie, “with a firmly established principle of determinacy of meaning”, zou de woordenboeken veel beter kunnen maken (Wierzbicka 1996:267). Met een theorie is het veel makkelijker om betekenissen te definiëren en polysemie duidelijk te houden (Wierzbicka 1996:267). De letterlijke betekenissen kunnen makkelijker worden gescheiden van de niet-letterlijke (Wierzbicka 1996:267). “It can offer lexicographic criteria on the basis of which meanings can be firmly separated from one another, clearly identified, and intelligible stated” (Wierzbicka 1996:267).

### **3.3.2 Verbeteringen voor woordenboeken**

Wierzbicka (1996:286) stelt dat NSM misschien niet voor alle woordenboeken handig is. Het zou wel handig zijn als lexicografen van het bestaan van NSM af weten, dan kunnen ze zelf beslissen in hoeverre ze een dergelijke metataal handig vinden voor hun woordenboek (Wierzbicka 1996:286). Ook zonder dat nsm gebruikt wordt in woordenboeken kan er al veel aan woordenboeken verbeterd worden (Wierzbicka 1996:241-242). Er zijn vier punten waarop het vaak misgaat in woordenboeken (Wierzbicka 1996:241-242):

1. Het gebrek aan een lijst met indefinables

Woordenboeken zouden een gebrek hebben aan lijsten met indefinables. Door NSM – die maakt gebruik van een lijst met indefinables, de lijst met semantic primes – zou hieraan een einde moeten komen.

2. Het gebrek aan een goede theorie over de verscheidenheid van verschillende betekenissen

Homografie is het verschijnsel dat er woorden bestaan die op dezelfde manier gespeld of uitgesproken worden, maar die toch niet aan elkaar verwant zijn. Soms is het verschil

tussen homografie en polysemie niet zo helder. Een woord als “vis” krijgt in Van Dale 2005 bijvoorbeeld vier verschillende betekenissen:

<sup>1</sup>**vis** (chem.) *visible* (symbool voor zichtbaar licht) (...)

<sup>2</sup>**vis** (...) ingezonken deurhengsel (...)

<sup>3</sup>**vis** (...) naam van in het water levende gewervelde dieren met een gestrekt, meestal spoelvormig lichaam, rood, koudbloedig, met vinnen als ledematen en kieuwen als ademhalingsorganen (...)

<sup>4</sup>**vis** (...) vits (zie ald.)”

Geen van bovenstaande de betekenissen lijkt verwant aan een andere, daarom moeten ze allemaal apart worden opgenomen. Het zou echter wel kunnen dat, bijvoorbeeld, een etymologisch verband is tussen <sup>2</sup>vis en <sup>4</sup>vis zodat hier sprake zou zijn van polysemie. Polysemie in woordenboeken lijkt soms volstrekt willekeurig te zijn: soms worden er teveel aparte ingangen gemaakt, en soms te weinig (Wierzbicka 1996:242-244). Polysemie komt veelvuldig voor in natuurlijke taal, maar soms is moeilijk te achterhalen of een woord echt polysemisch is, of dat het gebruik van het woord een beetje vaag is (Wierzbicka 1996:270).

### 3. Verwarring van lexicale betekenis met retorica zoals

metaforen, ironie, sarcasme, hyperbool, en dergelijke

Een woord kan alleen adequaat gedefinieerd worden als de letterlijke betekenis wordt gescheiden van de metaforische, ironische, speelse, en eufemistische gebruiken (Wierzbicka 1996:244). Woordenboeken gaan hier weleens de mist in door bijvoorbeeld een metafoor te letterlijk te nemen (Wierzbicka 1996:244). Dan wordt het inderdaad erg lastig om de betekenis van een woord te omschrijven, maar dat is niet de fout van het woord (Wierzbicka 1996:244)!



Omdat het zo moeilijk is om een sluitende definitie te maken met behulp van een semantic invariant, nemen lexicografen hun toevlucht in “hedges” en dergelijke, waardoor de generalisatie uit het oog wordt verloren (Wierzbicka 1996:271). Een hedge is een uiting waaruit blijkt dat de lexicograaf de definitie niet sluitend kan krijgen of de zoektocht naar de semantic invariant staakt, denk daarbij aan *met name, in het bijzonder* en *en dergelijke*. Zoals in de Van Dale uit 2005 bij het lemma “oog” en hedge voorkomt: “**oog** (...) **10** (...) het sluitwerk van een jupon e.d., als alternatief voor knopen (...)”. Een oog is niet alleen een sluitwerk voor een jupon, maar ook voor andere dingen die hier niet nader gespecificeerd worden. In tegenstelling tot die betekenis van “oog” wordt een van de betekenissen van “zon” in diezelfde Van Dale uit 2005 wel uitvoerig beschreven: “**zon** (...) **3** (g.mv) de straling, de warmte of het licht van de zon (1) (...)”.

4. De verleidelijkheid van de theorie van “family resemblances”, familiekenmerken  
Er wordt beweerd dat woorden geen afgebakende betekenis hebben, en dat ze hun betekenissen krijgen doordat ze aan andere woorden zijn gerelateerd: ze zijn daar familie van (Wierzbicka 1992:24-25, 1996:245). De betekenis van woorden zou te “fuzzy” zijn om te kunnen definiëren (Wierzbicka 1992:23). Deze theorie van “family resemblances” lijkt op de theorie van “prototypes”: men heeft een prototype van iets in zijn hoofd, en baseert andere betekenissen daarop (Wierzbicka 1996:245). Volgens Wierzbicka (1996:245) zitten beide benaderingen ernaast, omdat woorden wel degelijk gedefinieerd kunnen worden. Ieder woord heeft immers een semantic invariant (Wierzbicka 1992:24).

Daarnaast zijn vrijwel alle woordenboeken doorspekt van circulariteit (Wierzbicka 1996:274). Woorden verwijzen naar elkaar, zonder dat ze duidelijk gedefinieerd worden. Zonder het gebruik van semantic primitives zijn eigenlijk alle omschrijvingen circulair (Wierzbicka 1996:11). Het probleem is in de loop der jaren niet opgelost, maar juist verergerd (Wierzbicka

1996:277-278). Ton den Boon, hoofdredacteur van Van Dale, verdedigt de circulariteit in zijn woordenboek, als Van Nieuwkerk in het televisieprogramma De Wereld Draait Door<sup>1</sup> vraagt waarom “lamsragout” in 2014 in de Van Dale is opgenomen:

“Er is een moeilijk woord, een Frans woord, daarvoor ingeburgerd geraakt, navarin. En je wilt een woordenboek natuurlijk ook raadplegen om, als je dat moeilijke woord even niet kent, maar je weet wel dat het lamsragout betekent, dan ga je dus bij lamsragout zoeken en dan vind je navarin. Daarom, om die reden, hebben we dus nu lamsragout opgenomen, omdat het woord navarin ingeburgerd is.”

Die verwijzingen staan er dus niet in om van het verklaren af te zijn, maar omdat woordenboekgebruikers synoniemen willen opzoeken voor woorden die ze wel kennen.

Volgens Wierzbicka (1992:23) is er eigenlijk niemand die woordenboekdefinities goed vindt, maar niemand lijkt er in geïnteresseerd om ze te verbeteren of om daar methodes voor op te stellen. Dat wordt, in de tweede helft van de twintigste eeuw, gezien als onnodig en onmogelijk: theoretici zouden zich daar in ieder geval niet mee bezig moeten houden (Wierzbicka 1992:23). Het serieuze lexicografische onderzoek gaat pas net beginnen (Wierzbicka 1992:25). Het succes daarvan hangt (onder andere) af van de semantic primes, waarmee alle concepten kunnen worden gedefinieerd (Wierzbicka 1992:25).

Desondanks is Wierzbicka (1996:249) van mening dat definities in woordenboeken zo slecht nog niet zijn, omdat er immers miljoenen van worden verkocht. Ze stelt wel dat het vertrouwen in woordenboeken gebaseerd is op een illusie. De omschrijvingen zijn niet al te best, maar dat weten de lezers niet. Zij weten niet wat de beste definitie van een woord is, en hoeveel informatie ze uit een woordenboek zouden kunnen halen als de verklaringen zo goed zouden zijn als ze zouden kunnen zijn (Wierzbicka 1996:250). Woordenboeken geven wel wat houvast, maar hun definities zijn verre van adequaat en zouden makkelijk verbeterd

---

<sup>1</sup> De Wereld Draait Door, 15 mei 2014, Nederland 2, 20:00 uur

kunnen worden als praktische lexicografie en lexicale semantiek zouden samenwerken (Wierzbicka 1996:251).

Woordenboeken hebben beperkte ruimte, daar is Ton den Boon, hoofdredacteur van Van Dale zich van bewust. Hij vertelde in het televisieprogramma De Wereld draait Door<sup>2</sup> dat het papieren woordenboek een beetje te vol wordt. Toch moeten er nieuwe woorden toegevoegd, omdat ze steeds algemener worden. Vroeger werden die woorden alleen uit “deftige kranten” gehaald, nu ook bijvoorbeeld van internet en sociale media. Over de vraag wat het nieuw-toegevoegde woord “twerken” betekent, begint het antwoord met: “het is een hele lange definitie”. In verband met ruimtegebrek zijn eind jaren ’80, begin jaren ‘90 veel doorzichtige samenstellingen uit het woordenboek gehaald, het woord “fietsbel” werd bijvoorbeeld ondergebracht bij “bel”. Nu de Van Dale op internet staat, is onder andere “fietsbel” weer terug opgenomen in de online Van Dale.

Ondanks het ruimtegebrek komt er vaak overbodige technische of wetenschappelijke informatie in een woordenboek terecht, waardoor het meer een encyclopedie lijkt (Wierzbicka 1996:261-262). Dergelijke omslachtige verklaringen zijn niet per se fout qua definitie, maar de implicaties zijn fout omdat er belangrijke componenten weggelaten worden (Wierzbicka 1996:263). Het komt ook voor dat de definities juist te klein zijn, maar dat gebeurt veel minder vaak (Wierzbicka 1996:264).

Daarnaast worden de gegeven definities veel te indiscreet weergegeven (Wierzbicka 1996:267). Lexicografen geven vaak op omdat ze niet zeker weten óf ze wel tot een eenduidige definitie kunnen komen (Wierzbicka 1996:267-268). Wierzbicka (1996:269) zegt dat het niet makkelijk is om tot een eenduidige definitie te komen. Betekenissen zijn begrensd en er zijn geen subbetekenissen (“shades of meaning”) en “hedges” (*en, of, in het bijzonder*) zijn onnodig en onacceptabel. Een definitie moet altijd op zichzelf kunnen staan (Wierzbicka

---

<sup>2</sup> De Wereld Draait Door, 15 mei 2014 Nederland 2, 20:00 uur

1996: 270). Als een woord echt polysemisch is dan moeten beide betekenissen omschreven worden, maar ook die moeten op zichzelf kunnen staan (Wierzbicka 1996:270). Betekenissen hangen namelijk niet van hun buurwoorden af, zoals eerder omschreven.

Woordenboeken zijn niet zo goed als ze zouden kunnen zijn volgens Wierzbicka. Woordenboeken beginnen overvol te raken. Toch wordt er onnodige informatie in opgenomen, waardoor menig woordenboek op een encyclopedie lijkt. Daarnaast wordt in woordenboeken verkeerd gebruik gemaakt van polysemie, er zitten te veel hedges in en ook te veel voorbeelden. Daarnaast is er sprake van circulariteit. Dat komt allemaal doordat er niet hard genoeg gezocht wordt naar de semantic invariant.

### **3.3 Verwachtingen van woordenboeken**

Van woordenboeken wordt verwacht dat ze min of meer compleet zijn, ten minste op een bepaald domein of aspect van de taal (Wierzbicka 1996:264-265). Daarnaast moeten ze ook praktisch zijn, commercieel haalbaar, en hanteerbaar (Wierzbicka 1996:258). Hoe valt dat alles met elkaar te rijmen? Volgens Wierzbicka (1996:259) moeten praktisch lexicografen niet langer steunen op “experience and common sense” maar de hulp inschakelen van theoretisch lexicografen.

Ook als woordenboeken worden behouden in hun traditionele vorm kan er al veel worden verbeterd, te beginnen met taalgebruik (Wierzbicka 1996:283). Er worden vaak veel te moeilijke woorden gebruikt, waardoor de definities veel te ingewikkeld worden (Wierzbicka 1996:283). De definities moeten worden versimpeld, de “hedges” moeten verdwijnen, en daar moeten semantic invariants voor in de plaats komen (Wierzbicka 1996:283).

#### 4. Methode

Om antwoord te kunnen geven op de vraag of en hoe woordenboeken steeds groter worden, worden in dit onderzoek lemma's in een tiental woordenboeken geteld. Het is niet nodig om alle woordenboeken te onderzoeken. Als een trend te merken is, is die ook te merken zonder dat er alle woordenboeken voor nagekeken zijn. Er worden zeven van de veertien papieren dikke Van Dales onderzocht. Uiteraard de eerste en de laatste uitgave, en de daar tussenin liggende woordenboeken met een interval van rond de twintig jaar.

Daarnaast wordt ook het Basiswoordenboek onderzocht. Dat is gericht op een ander publiek: namelijk op kinderen vanaf tien jaar oud. Dat heeft consequenties voor welke woorden en betekenissen worden opgenomen (het basiswoordenboek is licht gecensureerd: er komen bijvoorbeeld maar weinig vulgaire woorden in voor), en dus voor de omvang van de lemma's.

Tevens heeft Van Dale heeft een gratis onlinewoordenboek en een betaald onlinewoordenboek die allebei zijn te raadplegen via de Van Dale-website. Ten opzichte van de papieren woordenboeken zijn de online woordenboeken interessant omdat er op internet geen ruimtegebrek is: de definities kunnen in principe oneindig lang zijn. Onderling zijn de onlinewoordenboeken interessant omdat er voor de een betaald moet worden en voor de ander niet. Welke informatie krijgt men als men betaalt, maar niet als men dat niet doet?

In totaal worden de volgende tien woordenboeken onderzocht:

- *Nieuw Woordenboek der Nederlandsche Taal*, 1864
- *Nieuw Woordenboek der Nederlandsche Taal*, 1884
- *Van Dale's Groot Woordenboek der Nederlandsche Taal*, 1914
- *Van Dale's Nieuw Groot Woordenboek der Nederlandsche Taal*, 1950
- *Van Dale Groot Woordenboek der Nederlandse Taal*, 1970
- *Van Dale Groot Woordenboek der Nederlandse Taal*, 1984

- *Van Dale Groot Woordenboek van de Nederlandse Taal*, 2005
- *Van Dale Basiswoordenboek Nederlands*, 2009
- *Van Dale onlinewoordenboek* - betaald
- *Van Dale onlinewoordenboek* - gratis

In alle woordenboeken worden dezelfde 35 lemma's geteld, namelijk de woorden zoals beschreven op de korte Swadeshlijst (zie bijlage I). Het is een lijst met daarop woorden die een kernwoordenschat omschrijven, met daarop onder andere werkwoorden, zelfstandig naamwoorden en bijvoeglijk naamwoorden. Op de Swadeshlijst komen geen neologismen voor: de woorden komen in de oudste en in het nieuwste woordenboek voor. Het zijn ook geen specifieke of technische woorden dus ook in het Basiswoordenboek komen ze voor. De woordenboeken worden onderzocht op de volgende zeven punten, zoals eerder vermeld door Wierzbicka:

1. Polysemie: Onder polysemie worden lemma's verstaan die meerde ingangen nodig hebben, omdat hun spelling wel overeenkomt, maar hun betekenis of woordsoort niet. De twee lemma's hieronder uit de Van Dale 2005 hebben wel hun spelling gemeenschappelijk en ook de betekenissen lijken wel verwant te zijn, maar de woordsoort en het gebruik is anders:

“<sup>1</sup>**weten** (overg.; wist, h. geweten) kennis hebben van of omtrent, bekend zijn met-  
(...)

<sup>2</sup>**weten** (het; g.mv.) **1** kennis, wetenschap omtrent iets (...) **2** [betekenisontlening van Du. *Wissen*] kennis, wetenschap: *het westerse weten*”

Deze visie op polysemie wijkt iets af van het van wat gangbaar is binnen de taalkunde/lexicografie. Wat in die vakgebieden ‘polysemie’ genoemd wordt, is in deze scriptie ‘subbetekenis’ genoemd.

2. Het aantal woorden: alle woorden die nodig zijn om een betekenis te omschrijven worden bij elkaar opgeteld. Hierboven is het lemma “weten” opgedeeld in “<sup>1</sup>weten” en “<sup>2</sup>weten”. Bij het aantal woorden worden de woorden van zowel “<sup>1</sup>weten” als “<sup>2</sup>weten” bij elkaar opgeteld.

3. Het aantal subbetekenissen: het woord “hand” bijvoorbeeld heeft verschillende betekenisnuances. Anders dan bij polysemie hangen de betekenissen van hand wel allemaal met elkaar samen, en komen ook de woordsoorten overeen:

“**hand** (de;-en) **1** elk van de beide lichaamsdelen aan de uiteinden van de armen, zich uitstrekkend van de pols tot het uiteinde van de vingers, tot grijpen en vasthouden geschikt (...) **2** in verbinding met ‘aan’ (...) **3** (meton.) breedte, dikte of holte van een hand van een volwassen persoon als maat (...) **4** (meton.) manier, trant van schrijven (...) **5** (fig.) zijde, kant (...) **6** (vand.) partij (...) **7** (bij apen; soms ook wel in verband met andere dieren) het met de menselijke hand (1) in gedaante overeenkomende lichaamsdeel **8** als onderdeel van werktuigen (...) **9** (gew.) kruk aan gereedschappen (...) **10** (meton., kaartsp.) kaartverdeling in één hand (...) **11** (meton.) afstand die een man met de volle kruitwagen in steeds dezelfde tijd kruit (...) **12** (op behangselrukkerijen) drukvorm (...)”

4. Het aantal subsubbetekenissen: soms is een subbetekenis niet genoeg, en krijgt ook de subbetekenis een subbetekenis. In het voorbeeld van het lemma “bloed” uit Van Dale 2005 hieronder een voorbeeld. Er worden twee subbetekenissen aangeduid met **I** en **II**, maar die subbetekenissen hebben óók weer subbetekenissen (aangeduid met **1**, **2**, **3**, et cetera):

“**bloed I** (het; g.mv.) (stofn.) **1** de in het levend organisme circulerende vloeistof die de stofwisseling draagt (...) **2** (gewelddadig) vergoten bloed (...) **3** bloed (I, 1) beschouwd als drager van aandoeningen en eigenschappen (...) **4** bloed (I, 1)

beschouwd als drager der verwantschap, als het essentiële van het geslacht (...) **5** (veet.) ras (...) **6** in een bijz. toepassing: (r.-k.) *het bloed van Christus* (...) **7** (...) *het bloed der druiven*, het sap **II** (de (m.); -en) **1** *arme bloed, onnozele bloed*, arm, beklagenswaardig of sukkelachtig mens **2** (...) *bloeien van kinderen*, bloedjes (...)"

5. Het aantal hedges: de gevallen waaruit blijkt dat de lexicograaf er zelf ook de vinger niet kan opleggen, woorden als *in het bijzonder*, *enzovoort*, en *en dergelijke*. In het voorbeeld hieronder uit Van Dale 2005 wordt niet uitvoerig uitgelegd:

“<sup>1</sup>**vuur** (...) **5** (...) in toepassing op hetgeen met (een) brand vergeleken wordt, m.n. op twist en strijd en op liefde en hartstocht (...).” Het woordje “m.n.” (*met name*) duidt erop dat de omschreven betekenis de meest gebruikte betekenis is. het impliceert ook dat er meer gebruiken zijn, maar die worden niet vermeld.

6. Het aantal voorbeelden: alle voorbeelden die in een lemma worden gebruikt, ook in ook in de (sub)subbeteknissen. Ook varianten op voorbeelden tellen mee. Een gezegde als “*zo ziek als een hond*” telt mee als één voorbeeld, waar “*groeien, gedijen als haren op een hond*” als twee telt, omdat “groeien” en “gedijen” varianten van het voorbeeld aanduiden.

7. Het aantal verwijzingen: het aantal keren dat er naar een ander woord verwezen wordt, dat kan expliciet zijn, maar ook een als er een synoniem of antoniem gegeven wordt, of als de uitleg van een bepaald gezegde elders gegeven wordt.

“<sup>1</sup>**hond** (...) *de hond in de pot vinden*, zie bij *pot*<sup>1</sup>”. Hier wordt een spreekwoord gegeven, maar wat het spreekwoord betekent is niet te vinden bij “hond”, maar bij “pot”.

“<sup>1</sup>**wat** (...) **1** (...) uitroep van verbazing, verontwaardiging of misnoegen, vgl. *wat*<sup>3</sup>.” De betekenis van deze “wat<sup>1</sup>” komt overeen met de betekenis van “wat<sup>3</sup>”, vandaar dat men wordt doorverwezen.



Als alle data bekend is, wordt gekeken op welke van bovenstaande punten de woordenboeken het meest zijn gegroeid en met welk percentage. De verwachting is dat de papieren dikke Van Dales bij iedere editie groei doormaken. Het Basiswoordenboek is met 630 bladzijden kleiner dan de kleinste dikke Van Dale. De verwachting is dat dat ligt aan het aantal voorbeelden en subbetekenissen. De onlinewoordenboeken hebben geen ruimtegebrek. Het is te verwachten dat het betaalde onlinewoordenboek groter zal zijn dan de laatste papieren dikke Van Dale, uit 2005. Het gratis onlinewoordenboek daarentegen, zal waarschijnlijk minder groot zijn dan de papieren Van Dales en het betaalde onlinewoordenboek. Het gratis onlinewoordenboek is meer een voorproefje op de betaalde variant dan een woordenboek an sich, en zal om die reden waarschijnlijk een minder uitvoerige uitleg bevatten en minder voorbeelden.

## 5. Resultaten

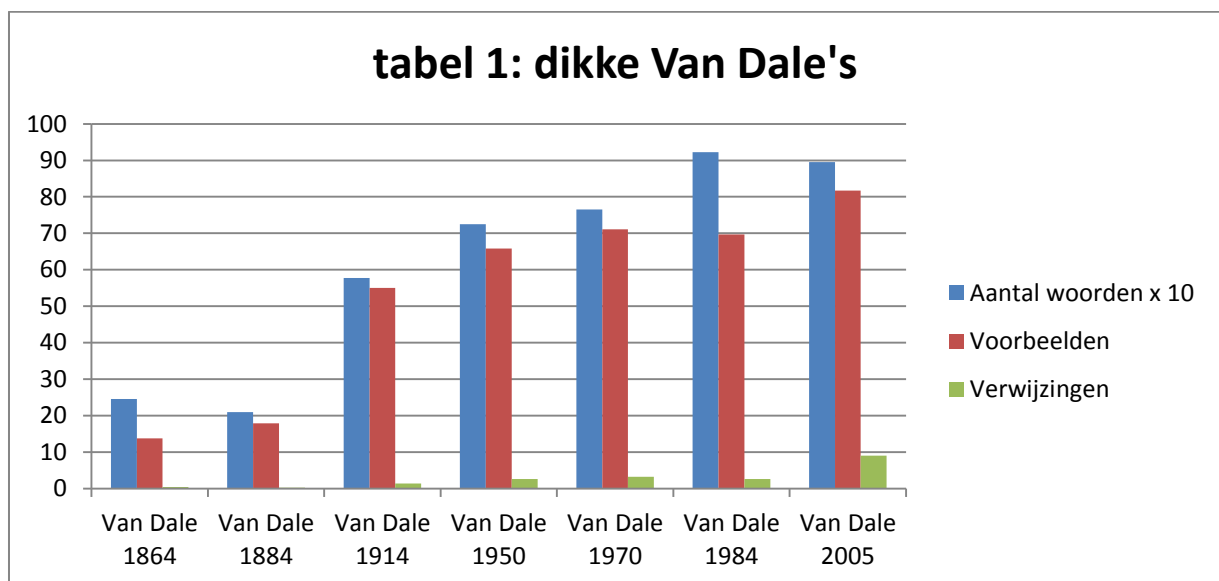
De resultaten zijn opgedeeld in vier subresultaten: *Papieren woordenboeken chronologisch*, *Papieren woordenboeken versus Basiswoordenboek*, *Betaald onlinewoordenboek versus gratis onlinewoordenboek*, en *Papieren woordenboeken versus onlinewoordenboeken*.

In die resultaten zijn enkel het aantal woorden, het aantal voorbeelden, en het aantal verwijzingen opgenomen. De resultaten voor subbetekenissen hebben een eigen hoofdstuk gekregen.

Hedges laten wel een mooie correlatie zien ( $R^2$  0,8264), maar de verschillen tussen de woordenboeken zijn zo klein dat deze ook buiten beschouwing gelaten worden. Het aantal hedges stijgt en daalt over de verschillende woordenboeken: er is eigenlijk geen peil op te trekken, ondanks een goede correlatie. Het kleine Basiswoordenboek en het gratis onlinewoordenboeken hadden weliswaar het kleinste aantal hedges, maar dat komt waarschijnlijk doordat hun omschrijvingen heel kort zijn en daardoor is er geen plek voor veel hedges. In de papieren dikke Van Dales en het betaalde onlinewoordenboek worden wel meer hedges gevonden dan in het Basiswoordenboek en gratis onlinewoordenboek. Het zijn er echter niet zoveel meer dat daar conclusies uit getrokken kunnen worden. Ook is er geen duidelijke stijgende of dalende lijn in te vinden.

## 5.1 Papieren woordenboeken chronologisch

In tabel 1 hieronder zijn de resultaten te zien van het gemiddeld aantal woorden per lemma, het gemiddeld aantal voorbeelden per lemma, en het gemiddeld aantal verwijzingen per lemma. Voor het overzicht is het aantal woorden gedeeld door 10, anders werden de woordenstaafjes zo lang dat de andere bijna niet meer zichtbaar waren. De dikke Van Dale uit 2005 bijvoorbeeld heeft in werkelijkheid per lemma gemiddeld dus geen 89,5 woorden, maar 895.



Het aantal woorden neemt lineair toe over de jaren ( $y = 127,23x + 111,07$ ;  $R^2 = 0,8893$ ). Er vindt in die tijd een stijging plaats van gemiddeld 245 woorden in 1864, naar 895 in 2005: een stijging van 265,3%. Er is desondanks niet in alle gevallen sprake van een stijging: de editie van 1884 is kleiner dan zijn voorganger uit 1864: een daling van 14,7%. Ook de Van Dale uit 2005 is kleiner dan zijn voorganger uit 1984: er vindt een kleine krimp van 3,1% plaats. De overige edities groeien allemaal ten opzichte van elkaar, en ook allemaal significant:

1864 – 1884: daling van 14,7%.

1884 – 1914: stijging van 175,7%.

1914 – 1950: stijging van 25,5%.

1950 – 1970: stijging van 5,5%.

1970 – 1984: stijging van 20,6%.

1984 – 2005: daling van 3,1%.

Over het algemeen groeien de betekenisomschrijvingen van de lemma's, maar de recentste uitgave omvat niet de meeste woorden.

Ook het gemiddelde aantal voorbeelden per lemma neemt lineair toe ( $y = 11,569x + 7,2735$ ;  $R^2 = 0,8559$ ). In 1864 had een lemma gemiddeld 13,7 voorbeelden, in 2005 zijn dat er al 81,7: een stijging van maar liefst 496,4%. Ook bij de voorbeelden vindt er niet per editie een stijging plaats, in de editie van 1984 staan 2% minder voorbeelden dan in zijn voorganger uit 1970.

1864 – 1884: stijging van 30,0%.

1884 – 1914: stijging van 208,3%.

1914 – 1950: stijging van 19,7%.

1950 – 1970: stijging van 18,1%.

1970 – 1984: daling van 2%.

1984 – 2005: stijging van 17,3%.

Hier is iets opmerkelijks aan de hand: het aantal woorden stijgt tussen 1970 en 1984 wel, maar blijkbaar ligt dat niet aan het aantal voorbeelden, want die dalen. Het tegenovergestelde is geldig bij de Van Dale uit 2005: die gebruikt minder woorden om een lemma te omschrijven dan zijn voorganger uit 1984, maar gebruikt daarbij wel meer voorbeelden. Waarom dat zo is, wordt in deze scriptie niet aan de orde gesteld.

Het gemiddeld aantal verwijzingen per lemma is enorm gestegen: van 0,4 in 1864 naar 9 in 2005. Dat is een stijging van 2323,3% ( $y = 1,1561x - 1,8612$ ;  $R^2 = 0,7053$ ). Het gemiddeld aantal verwijzingen fluctueert nogal. Er zijn zeer sterke stijgingen, zoals tussen 1884 en 1914 en tussen 1985 en 2005. Er zijn ook dalingen, tussen 1864 en 1884 en tussen 1970 en 1984:

1864 – 1884: daling van 30,8%.

1884 – 1914: stijging van 433,4%.

1914 – 1950: stijging van 87,5%.

1950 – 1970: stijging van 23,3%.

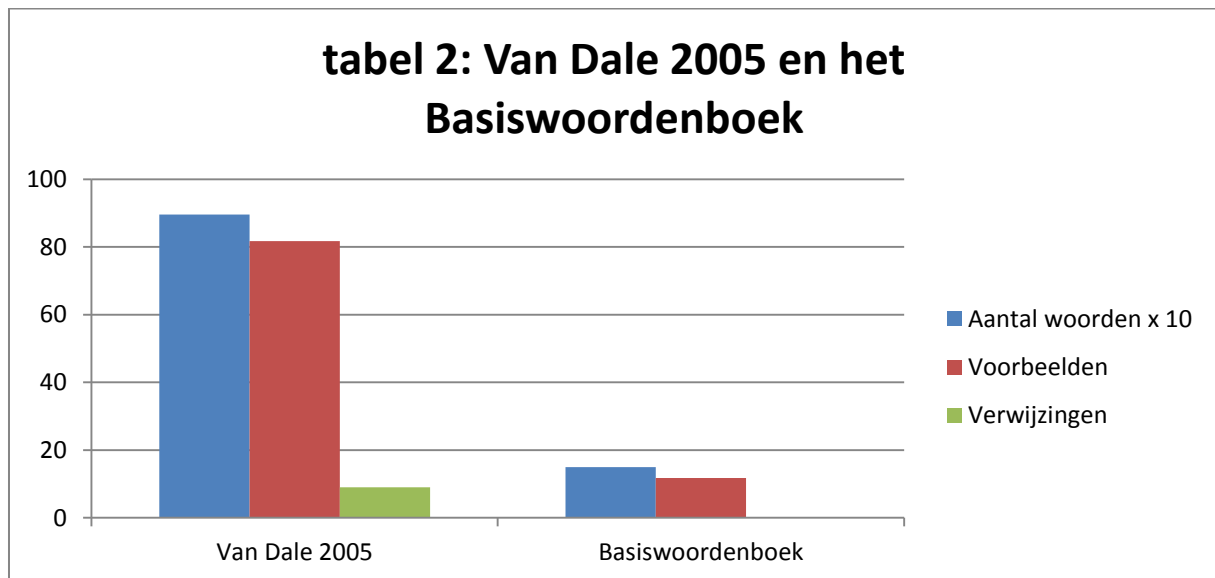
1970 – 1984: daling van 15,3%.

1984 – 2005: stijging van 235,1%.

De stijgingen zijn sterker dan de dalingen. In nieuwe edities kregen meer woorden verwijzingen. In 1864 hadden slechts 9 van de 35 onderzochte lemma's verwijzingen, maar in 1950 zijn dat er 24 van de 35, en in 2005 zelfs 34 van de 35. Vooral in het lemma “hand” is de stijging van het aantal verwijzingen goed te zien. In 1864 was er maar 1 verwijzing bij “hand”, in 1950 al 17, en in 2005 zijn dat er 55. Dat is een overweldigende groei van 5400%. Het aantal woorden in de betekenisomschrijving van het lemma stijgt in diezelfde periode met ‘slechts’ 594,6% en het aantal voorbeelden met ‘slechts’ 645,9%. Dat is dus ook een flinke stijging, maar niet zo enorm als het aantal verwijzingen. De grote stijging in het aantal verwijzingen wordt in deze scriptie niet aan de orde gesteld.

## 5.2 Papieren woordenboeken versus Basiswoordenboek

Het Basiswoordenboek (uit 2009) is vergeleken met de meeste recente dikke Van Dale: die uit 2005. Ook hier, in tabel 2, geldt dat het aantal woorden nog dient vermenigvuldigd te worden met 10.



Het gemiddeld aantal woorden dat de beide woordenboeken gebruiken om lemma's te omschrijven verschilt nogal: het dikke Van Dale woordenboek heeft er gemiddeld 895 per lemma, het Basiswoordenboek maar 150. Met andere woorden: de Dikke van Dale heeft bijna zesmaal zoveel woorden nodig om een lemma te omschrijven als het Basiswoordenboek.

Het Basiswoordenboek heeft ook gemiddeld minder voorbeelden per lemma. De dikke Van Dale uit 2005 geeft gemiddeld 82 voorbeelden, het Basiswoordenboek maar 12. Bijna zevenmaal zo weinig. De dikke Van Dale geeft bijvoorbeeld verouderde spreekwoorden weer, maar het Basiswoordenboek laat die weg. Het Basiswoordenboek gebruikt alleen voorbeelden die relevant zijn voor het jonge publiek, en oude spreekwoorden zijn dat bijvoorbeeld niet.

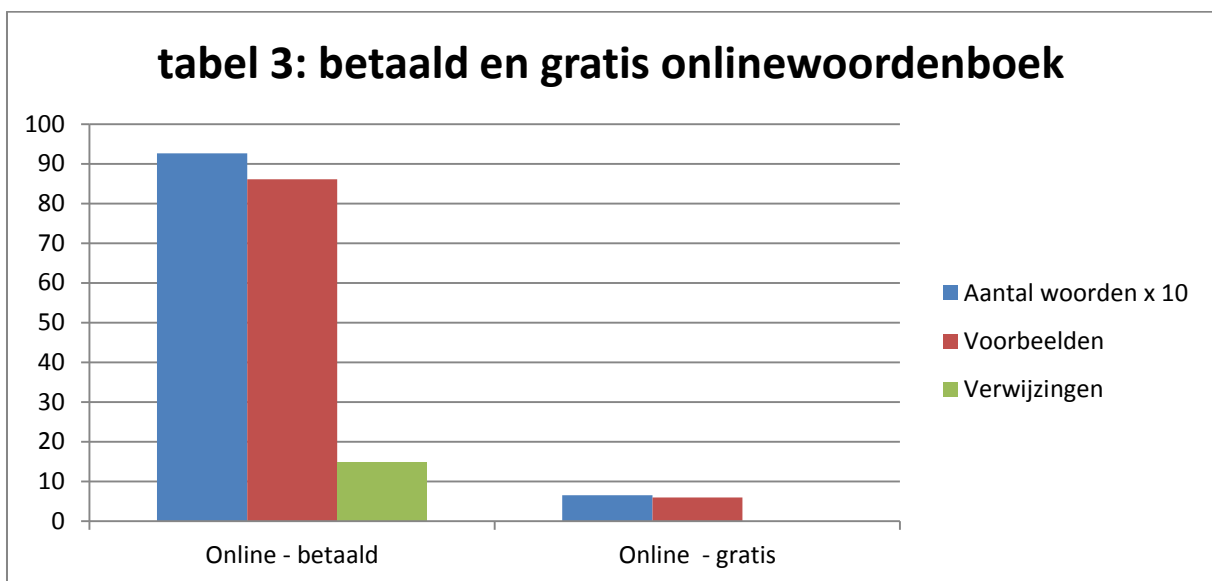
In het basiswoordenboek wordt ook lang niet zoveel verwezen als in de dikke Van Dale uit 2005. In de dikke Van Dale staan 9 verwijzingen, in het Basiswoordenboek maar 0,26. De Dikke Van Dale gebruikt 30 keer zoveel verwijzingen als het Basiswoordenboek. Het Basiswoordenboek heeft maar bij 8 lemma's verwijzingen; de dikke Van Dale bij 34 van

de 35. Het lijkt erop dat het basiswoordenboek de definitie bij het betreffende lemma wil afhandelen. De dikke Van Dale wil juist additieve informatie verschaffen, die te zien is bij een ander lemma.

Omdat het Basiswoordenboek op een jong publiek is gericht en uit is op de belangrijkste, frequentste betekenissen, is te verwachten dat het Basiswoordenboek minder gevallen van polysemie, subbetekenissen, en subsubbetekenissen behelst. Dat blijkt ook zo te zijn. Een lemma in de dikke Van Dale uit 2005 heeft gemiddeld 1,8 betekenissen, in het Basiswoordenboek zijn dat er 1,3. De dikke Van Dale uit 2005 heeft daarbij nog gemiddeld 8,8 subbetekenissen per lemma, het Basiswoordenboek slechts 1,9. Ook heeft Van Dale uit 2005 gemiddeld 1,1 subsubbetekenissen per lemma, maar het Basiswoordenboek gebruikt geen subsubbetekenissen. Het basiswoordenboek veegt meer subtiele betekenisnuances op één hoop.

### 5.3 Betaald onlinewoordenboek versus gratis onlinewoordenboek

Het gratis woordenboek is het kleinste woordenboek uit dit onderzoek – het gebruikt het minste aantal woorden om een lemma te omschrijven – en het betaalde onlinewoordenboek is het grootste. De verschillen zijn dan ook groot. Het onlinewoordenboek is niet echt een opzichzelfstaand woordenboek, maar meer een lokkertje voor het betaalde onlinewoordenboek. Tabel 3 laat goed zien dat het betaalde onlinewoordenboek veel groter is dan het gratis onlinewoordenboek. Ook hier, in tabel 3, geldt dat het aantal woorden nog dient vermenigvuldigd te worden met 10.



Een lemma in het betaalde onlinewoordenboek bevat gemiddeld 927 woorden, een lemma in het gratis onlinewoordenboek slechts 86. Het betaalde onlinewoordenboek is hiermee 1308,7% groter dan het gratis onlinewoordenboek.

Het betaalde onlinewoordenboek omvat ook aanzienlijk meer voorbeelden dan het gratis onlinewoordenboek: respectievelijk 86 om 6. Het betaalde onlinewoordenboek is hiermee 1434,8% groter dan het gratis onlinewoordenboek. Het gratis onlinewoordenboek toont alleen frequent gebruikte uitdrukkingen, maar het betaalde onlinewoordenboek toont ook minder frequente.



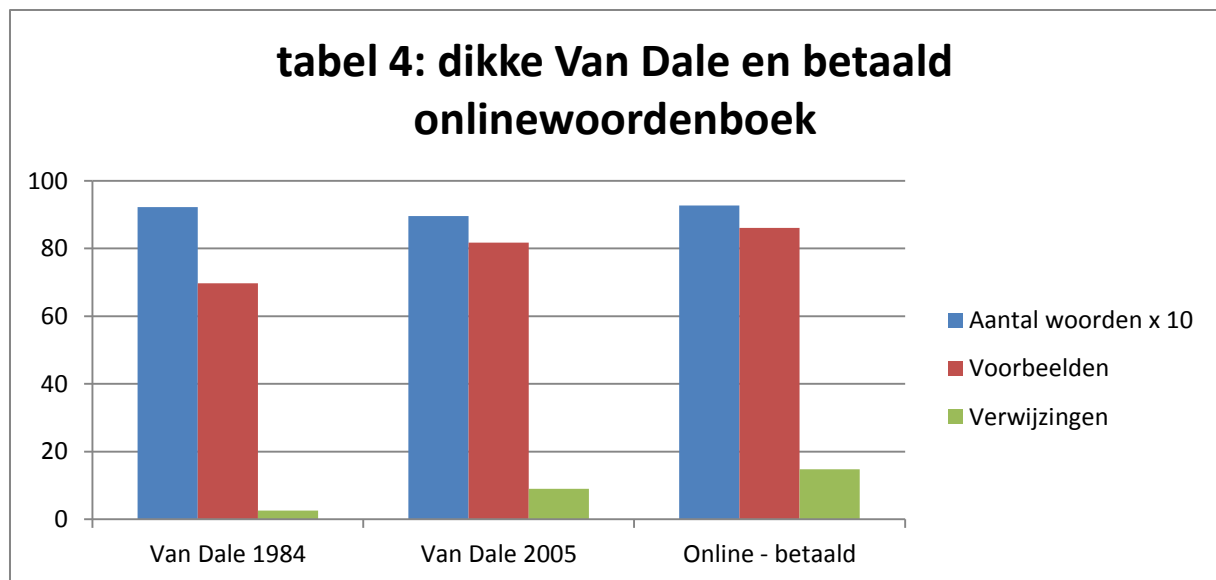
Het aantal verwijzingen is snel besproken: het betaalde onlinewoordenboek gebruikt gemiddeld 14,7 verwijzingen per lemma, maar het gratis onlinewoordenboek gebruikt geen verwijzingen.

Het gratis onlinewoordenboek is meer een voorproefje op het betaalde onlinewoordenboek dan een woordenboek op zich. Onder ieder lemma wordt vermeld dat het gratis onlinewoordenboek “een beperkt aantal woorden en mogelijkheden” bevat. Men kan via een linkje dan direct het betaalde onlinewoordenboek voor een beperkte tijd gratis uitproberen.

## 5.4 Papieren woordenboeken versus betaald onlinewoordenboek

In dit hoofdstukje worden de dikke van Dales uit 1984 en 2005 vergeleken met het betaalde onlinewoordenboek. Het betaalde onlinewoordenboek kan gezien worden als een volwaardig woordenboek, maar dan online. Het gratis onlinewoordenboek is geen opzichzelfstaand woordenboek en dus niet vergelijkbaar met de papieren woordenboeken. In dit hoofdstuk wordt het gratis onlinewoordenboek dan ook buiten beschouwing gelaten.

Ook hier, in tabel 4, geldt dat het aantal woorden nog dient vermenigvuldigd te worden met 10.



Qua woorden is het onlinewoordenboek het grootse woordenboek met 927 woorden. Het scheidt echter niet veel met de dikke Van Dale uit 1984 die 922 woorden gebruikt. Het verschil is nog geen half procent, en daarom niet significant te noemen. De dikke Van Dale uit 2005 heeft 895 woorden; het onlinewoordenboek is daarmee slechts 3,5% groter. De verschillen tussen de woordenboeken is zo klein, dat met niet echt van groei kan spreken.

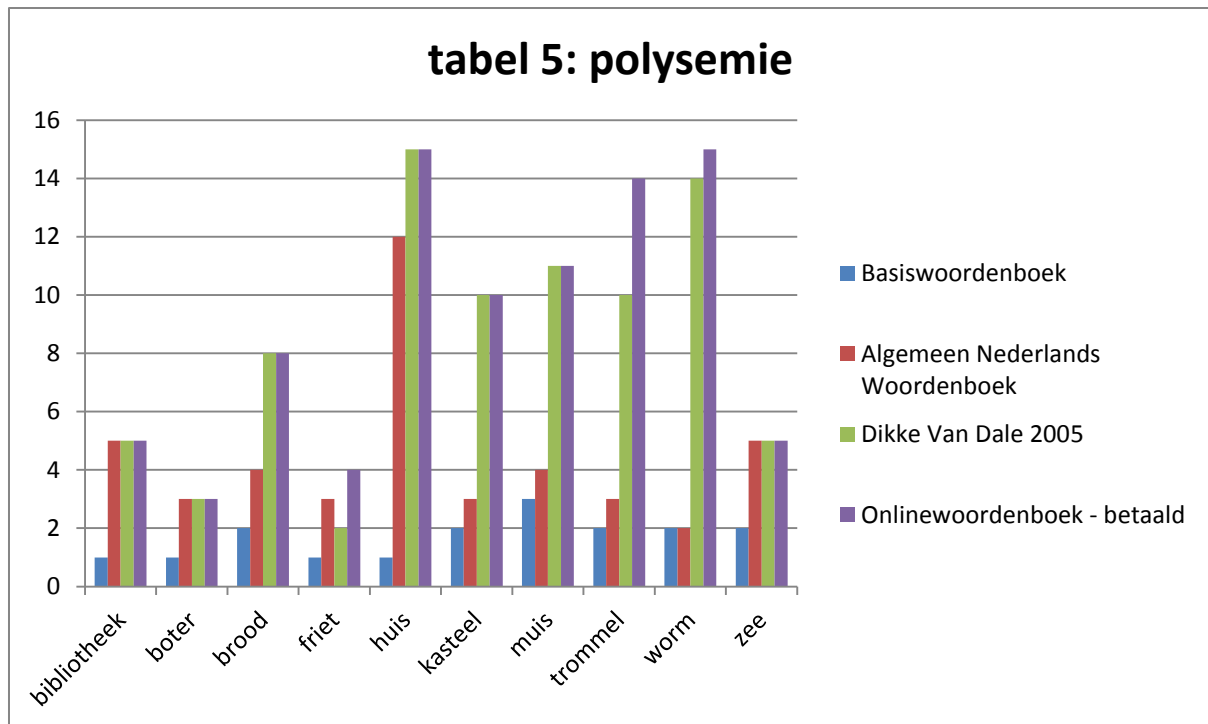
Ondanks dat de woorden niet noemenswaardig gestegen zijn, zijn de voorbeelden dat wel. De dikke Van Dale 1984 heeft gemiddeld 70 voorbeelden, die van 2005 heeft er 82, en het onlinewoordenboek 86. Het aantal voorbeelden is tussen 1984 en 2005 met 17,3% gestegen, tussen 2005 en onlinewoordenboek 5,3%. Over het geheel met 23,5%.

Het aantal verwijzingen is het hardst gestegen. Tussen 1984 en 2005 met 146,2%, tussen 2005 en online met 83,3%. Over de hele linie met 467,0%. Dat is te verklaren. Het onlinewoordenboek geeft vaak per voorbeeld een verzwijging. Het gebeurt dan dat er per lemma meermalen naar een zelfde woord verzworen wordt. In de dikke Van Dales uit 1984 en 2005 gebeurt dat veel minder.

Vooralsnog lijkt het erop dat het onlinewoordenboek een gedigitaliseerde versie is van de papieren dikke van Dales.

## **5.5 Subbetekenissen**

De subbetekenissen verdienen een eigen hoofdstuk. Wat ik in deze scriptie ‘subbetekenissen’ noem, zou normaal ‘polysemie’ genoemd worden. In dit hoofdstuk zijn iets andere woordenboeken onderzocht dan in de rest van de scriptie. Het betreft het Basiswoordenboek, de dikke Van Dale uit 2005, het betaalde onlinewoordenboek van Van Dale, en het algemeen Nederlands Woordenboek (ANW). Het ANW is een woordenboek dat louter op internet bestaat. Het is geenszins verbonden aan van Dale en dat maakt het interessant om te zien hoe het ANW omgaat met verschillende betekenissen ten opzichte van de Van Dale. Ook maakt dit hoofdstuk gebruik van andere lemma’s. Het ANW is nog niet helemaal af: veel lemma’s van de korte Swadeshlijst zijn er (nog) niet in opgenomen. Daarom was het noodzakelijk andere lemma’s te gebruiken, die wel in het ANW voorkomen en uiteraard ook in de drie andere woordenboeken. In tabel 5 (op de volgende pagina) is te zien hoeveel betekenissen een bepaald lemma krijgt.



Uit het onderzoek blijkt dat het Basiswoordenboek verreweg de minste betekenissen onderscheidt. Het Basiswoordenboek neemt betekenissen op een hoop die de andere woordenboeken apart opnemen. Ook worden bepaalde betekenissen helemaal niet opgenomen. Bij ‘zee’ geeft het Basiswoordenboek maar twee betekenissen: “grote hoeveelheid zout water die een groot oppervlak van de aarde bedekt” en “grote hoeveelheid”. Het ANW, de Van Dale uit 2005, en het onlinewoordenboek onderscheiden allemaal vijf betekenissen, maar niet allemaal dezelfde vijf.

Het betaalde onlinewoordenboek van Van Dale lijkt op een digitale versie van de Van Dale uit 2005. Bij veel lemma’s worden evenveel betekenissen onderscheiden, met allemaal dezelfde betekenissen. Bij enkele woorden geeft het onlinewoordenboek meer betekenissen dan de Van Dale uit 2005, bij het woord ‘friet’ bijvoorbeeld. Van Dale 2005 vermeldt maar twee betekenissen: “in kokend vet gefrituurde reepjes aardappel” en “stukje friet”. Het onlinewoordenboek vermeldt vier betekenissen: “in kokend vet gefrituurde reepjes aardappel”, “stukje friet”, “gerecht van in dunne reepjes gesneden en vervolgens gefrituurde (of knapperig gebakken) groente, vis, vlees” en “portie frieten”. De eerste twee betekenissen

komen overeen met die uit Van Dale 2005, de laatste twee zijn daaraan toegevoegd. Het onlinewoordenboek van Van Dale is min of meer een gedigitaliseerde versie van Van Dale 2005, waaraan bij enkele lemma's informatie is toegevoegd. Het lijkt erop dat het onlinewoordenboek welbetekenissen toevoegt, maar geen betekenissen weghaalt. Het onlinewoordenboek heeft of evenveel betekenissen als Van Dale 2005 of meer, maar nooit minder.

Het ANW zit tussen het Basiswoordenboek en de Van Dale uit 2005 in. Het onderscheidt meer betekenissen dan het Basiswoordenboek, maar minder dan de Van Dale 2005. Het woord 'brood' leent zich uitstekend om dit te verhelderen. Op de volgende pagina een tabel:

brood		
Basiswoordenboek	ANW	Van Dale 2005
1. “baksel van gekneede deeg, dat voor veel mensen dagelijks voedsel is”	1. “voedsel dat bestaat uit gebakken brooddeeg”	1. “voedsel uit gekneede, gerezen en gebakken deeg (van meel met water of melk) bestaande”
2. “blok samengeperste stof”	2. “levensonderhoud”	2. “levensonderhoud”
	3. “hoeveelheid brood”	3. “avondmaalsbrood”
	4. “broodvormig stuk”	4. “een in een bep. vorm gekneede en gebakken hoeveelheid brooddeeg”
		5. “de hoeveelheid brood die mens als maaltijd eet”
		6. “broodmaaltijd”
		7. “in een bep. vorm gegoten of geperste hoeveelheid van sommige stoffen, als zeep, goud, salpeter, zout en suiker”
		8. “zeker gewicht”

Het Basiswoordenboek geeft slechts twee betekenissen bij 'brood', maar merk hierbij wel op dat er veel informatie in een betekenis wordt gestopt. Wat er bij de eerste betekenis in het Basiswoordenboek uitgelegd wordt, is in het ANW en in de Van Dale 2005 opgedeeld in twee betekenissen, in beide gevallen 1. en 2. Daar zou zelfs betekenis 4. van Van Dale 2005 ook nog in passen.

Het ANW en het Basiswoordenboek trekken betekenissen bij elkaar, maar laten ook betekenissen weg die van Dale 2005 wel opneemt. Daarnaast trekt Van Dale 2005 vrijwel alle betekenissen uit elkaar. Daarom geeft Van Dale 2005 meer betekenissen dan het ANW en het Basiswoordenboek.

## 6. Conclusie

Woordenboeken zijn niet zo goed als ze kunnen zijn, denkt Anna Wierzbicka. Zij stelt dat dat is omdat er veel te veel informatie wordt gegeven en dat dat ook nog eens steeds meer lijkt te worden. De definities zijn te omslachtig en missen een semantic invariant. De oplossing die zij hiervoor geeft is Natural Semantic Metalanguage (NSM). Met NSM kunnen definities heel rigide omschreven worden. NSM is goed om definities mee te omschrijven, maar voor Van Dale schiet het toch tekort, omdat woordenboekgebruikers niet alleen definities willen opzoeken, maar ook weleens bijvoorbeeld spreekwoorden of synoniemen.

Volgens Wierzbicka maken woordenboeken ook nog eens verkeerd gebruik van polysemie. Ook (sub)subbetekenissen zouden niet moeten voorkomen in woordenboeken, er bestaan volgens haar immers geen “shades of meaning”. Als het woordenboek erin slaagt om de semantic invariant te omschrijven, dan kan het woordenboek toe zonder voorbeelden. Ook verwijzingen zijn dan overbodig, omdat alle informatie in de semantic invariant zit en niet meer elders gezocht hoeft te worden.

In dit onderzoek zijn Van Dale woordenboeken onderzocht op polysemie, sub- en subsubbetekenissen, het aantal hedges, het aantal verwijzingen, het aantal voorbeelden, en het aantal woorden. Het onderzoek is uitgevoerd met zeven dikke Van Dales, het van Dale basiswoordenboek, en twee Van Dale onlinewoordenboeken.

Uit het onderzoek blijkt dat de lemma's in Van Dale inderdaad gestaag groeien. Vooral het aantal woorden en het aantal voorbeelden groeien snel. Het enorme aantal voorbeelden is te verklaren doordat er ook varianten op voorbeelden worden gegeven (dit voorbeeld uit Van Dale 1970 telt al mee voor zes voorbeelden: “*een fraaie, mooie, losse, lopende, stijve, kriebelige hand schrijven*”) en omdat er in de latere edities samenstellingen gegeven worden (bij het lemma “vol” worden in van Dale 2005 twaalf samenstellingen



gegeven: “*volbouwen, volgieten, volgooien, volkiepen, volpennen, volpompen, volpraten, volschenken, volschrijven, volstoppen, volstorten, volwerpen, volzuigen*”).

Ook zijn er veel (sub)subbetekenissen in Van Dale, die allemaal niet al te veel van elkaar verschillen. Sommige daarvan kunnen wellicht beter onder dezelfde (sub)subbetekenis geschaard worden. In het Basiswoordenboek is dat inderdaad gedaan: daarin zijn minder gevallen van polysemie, subbetekenissen, en subsubbetekenissen te vinden.

Verwijzingen zijn nog altijd present in van Dale, en het aantal verwijzingen groeit per editie. In een woordenboek als Van Dale zijn die verwijzingen onvermijdbaar, omdat het ook synoniemen en antoniemen aanbiedt.

Vooraf in woordaantallen zijn grote verschillen te zien. Die verschillen zijn chronologisch: het eerste woordenboek had aanzienlijk minder woorden dan het meest recente deel. Van Dale groeit de laatste twee edities niet meer zo veel, en ook het onlinewoordenboek is niet veel groter dan de twee laatste papieren edities. In 2015 komt een nieuwe editie van het papieren woordenboek uit. In een vervolgonderzoek kan met dat woordenboek gekeken worden of de Van Dale nog steeds bij elke editie groeit, of dat dat stagneert.

Het Basiswoordenboek is gericht op een jonger publiek dan de Dikke Van Dale. Dat laat zich voornamelijk merken door het aantal woorden en ook het aantal voorbeelden en verwijzingen: dat is veel minder. Ook maakt het Basiswoordenboek geen gebruik van subsubbetekenissen, waar de andere papieren edities (op 1884 en 1914 na) dat wel doen. Er is ook een gratis onlinewoordenboek, dat doet denken aan het basiswoordenboek. Het is zeer beknopt en ook op woorden, voorbeelden en verwijzingen kleiner dan het betaalde onlinewoordenboek en de papieren versies.

Van Dale streeft naar volledigheid. Omdat het woordenboek steeds vollediger wordt, zijn er meer woorden nodig om de betekenissen allemaal te omschrijven. De laatste onderzochte edities en het betaalde onlinewoordenboek zijn niet zo hard gegroeid als hun

voorgangers. Het lijkt erop dat de Van Dale woordenboeken bijna volledig genoeg zijn. Om te weten of de groei echt stagneert zou er een recentere Van Dale onderzocht moeten worden. In 2015 komt er een nieuwe editie uit, die naar alle waarschijnlijkheid niet veel groter zal zijn dan de editie uit 2005.

## **Bibliografie**

- Algemeen Nederlands Woordenboek (ANW)*. INL de schatkamer van de Nederlandse taal, juli 2014. Web.
- Boon, Ton den, Dirk Geeraerts, en Nicoline van der Sijs. *Van Dale Groot Woordenboek van de Nederlandse Taal*. 14<sup>e</sup> druk. Utrecht/Antwerpen: Van Dale Lexicografie, 2005. Print. 3 vol.
- Calisch, I. M. en N. S. Calisch. *Nieuw Woordenboek der Nederlandsche Taal*. Tiel: H. C. A. Campagne, 1864. Web.
- Geerts, G., H. Heestermans, en C. Kruyskamp. *Van Dale Groot Woordenboek der Nederlandse Taal*. 2<sup>e</sup> oplage. Utrecht/Antwerpen: Van Dale Lexicografie, 1984. Print. 3 vol.
- Geerts, G., en H. Heestermans. *Van Dale Groot Woordenboek der Nederlandse Taal*. Utrecht/Antwerpen: Van Dale Lexicografie, 1992. Print.
- Kruyskamp, C. en F. de Tollenaere. *Nieuw Groot Woordenboek der Nederlandsche Taal*. 's-Gravenhage: Martinus Nijhoff, 1950. Print.
- Kruyskamp, C. *Van Dale Groot Woordenboek der Nederlandse Taal*. 1970. 's-Gravenhage: Martinus Nijhoff, 1973. Print. 2 vol.
- Malssen, P. J. van. *Van Dale's Groot Woordenboek der Nederlandsche Taal*. 's-Gravenhage/Leiden: Martinus Nijhoff/A. W. Sijthoff's Uitgevers-M<sup>ll</sup>, 1914. Print.
- Manhave, J. *Nieuw Woordenboek der Nederlandsche Taal*. 's-Gravenhage/Leiden: Martinus Nijhoff/A. W. Sijthoff, 1884. Print.
- Palmer, Gary, en Farzad Sharifan (eds.) *Applied Cultural Linguistics: Implications fro second language learning and intercultural communication*. "Semantic primes and cultural scripts in language learning and intercultural communication." Door Cliff Goddard en Anna Wierzbicka. Amsterdam: John Benjamins, 2007. Web.

*Van Dale Betaald Woordenboek*. Van Dale uitgevers, 2014. Web.

*Van Dale Gratis Woordenboek*. Van Dale uitgevers, 2014. Web.

Verburg, Marja. *Van Dale Basiswoordenboek Nederlands: Vanaf 10 jaar*. Van Dale/Noordhoff Uitgevers: Utrecht/Antwerpen, 2009. Print.

Wierzbicka, Anna. *Lingua Mentalis: The Semantics of Natural Language*. Sydney/New York/Londen/Toronto/San Fransisco: Academic press Australia, 1980. Print.

Wierzbicka, Anna. *Semantics, Culture, and Cognition: Human Concepts in Culture-Specific Configurations*. New York/Oxford: Oxford University Press, 1992. Print.

Wierzbicka, Anna. *Semantics: Primes and universals*. New York/Oxford: Oxford University Press, 1996. Print.

Wierzbicka, Anna. *The Semantics of Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1988. Print. Studies in Language Companion, series 8.

### Bijlage I Korte Swadeshlijst (35 woorden)

	Polysemie	Aantal woorden	Aantal subbetekenissen	Aantal subsubbetekenissen	Aantal hedges	Aantal voorbeelden	Aantal verwijzingen
Ik							
U							
Dit							
Wie							
Wat							
Een							
Twee							
Vis							
Hond							
Luis							
Bloed							
Been							
Ei							
Hoorn							
Staart							
Oor							
Oog							
Neus							
Tand							
Tong							
Hand							
Weten							
Sterven							
Geven							
Zon							
Maan							
Water							
Zout							
Steen							
Wind							
Vuur							
Jaar							
Vol							
Nieuw							
Naam							

## **Bijlage II Semantic primes**

Substantives:	I, YOU, SOMEONE/PERSON, PEOPLE, SOMETHING/THING, BODY
Relational substantives:	KIND, PART
Determiners:	THIS, THE SAME, OTHER
Quantifiers:	ONE, TWO, SOME, ALL, MANY/MUCH
Evaluators:	GOOD, BAD
Descriptors:	BIG, SMALL
Mental/experiential predicates:	THINK, KNOW, WANT, FEEL, SEE, HEAR
Speech:	SAY, WORDS, TRUE
Actions and events:	DO, HAPPEN, MOVE
Existence and possession:	THERE IS/EXIST, HAVE
Life and death:	LIVE, DIE
Time:	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
Space:	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH (CONTACT)
Logical concepts:	NOT, MAYBE, CAN, BECAUSE, IF
Intensifier, augmentor:	VERY, MORE
Similarity:	LIKE/WAY