

REFERENTIE BESTAND NEDERLANDS

DOCUMENTATIE

**Willy Martin
Isa Maks**

**m.m.v.
Stephan Bopp
Mila Groot**

april 2005

Woord Vooraf

Deze documentatie bestaat uit drie delen. Een eerste deel bevat algemene informatie (ontstaansgeschiedenis, doel, samenstelling, grootte, etc.). Een tweede deel geeft een beschrijving van de datacategorieën weer voor de zelfstandige naamwoorden, bijvoeglijke naamwoorden, werkwoorden, functiewoorden en bijwoorden. Deze laatste categorieën zijn later tot stand gekomen en volgen een eigen, andere presentatiewijze. Hetzelfde geldt voor de geografische namen die deel drie uitmaken. Er is niet naar gestreefd de onderdelen tot één homogeen geheel qua stijl en presentatie te herleiden voor zover dat de inhoudelijk uniformiteit niet hinderde. De enige bedoeling van deze documentatie is dan ook louter informatief.

W. Martin

Amsterdam, april 2005

1.	RBN: ALGEMEEN	7
1.1	MOTIVERING	7
1.2	ALGEMENE KARAKTERISERING	7
1.3	MACROSTRUCTUUR	9
1.3.1	<i>Bronnen</i>	9
1.3.2	<i>Schrappen in en Samenvoegen van Bronnen</i>	10
1.3.3	<i>Toevoegen</i>	12
1.3.4	<i>Omvang van de macro</i>	13
1.3.5	<i>Opmerkingen bij de selectie</i>	13
1.4	BETEKENISSEN IN MICROSTRUCTUUR	15
1.5	RBN EN OMBI	17
1.5.1	<i>Welke gegevens</i>	17
1.5.2	<i>Hoe omgaan met de gegevens</i>	17
2.	NOMINA	19
2.1	IDENTIFICATION	19
2.1.1	<i>FORM</i>	19
2.1.2	<i>CAT</i>	20
2.1.3	<i>SEQUENCE NUMBER en HOMONYM NUMBER</i>	20
2.2	GRAPHEMICS	21
2.2.1	<i>TYPE</i>	21
2.2.2	<i>SPELVAR</i>	21
2.2.3	<i>FORMVAR</i>	21
2.2.4	<i>VERWIJSEENTRY'S</i>	21
2.3	PHONETICS	22
2.4	MORPHOLOGY	22
2.4.1	<i>MORPHOTYPE</i>	22
2.4.2	<i>MORPHOSTRUCTURE</i>	23
2.4.3	<i>PLURAL</i>	23
2.4.4	<i>PRAGMATIC CONSTRAINTS PLURAL</i>	23
2.5	SYNTAX	24
2.5.1	<i>GENDER</i>	24
2.5.2	<i>ARTICLE</i>	25
2.5.3	<i>NUMBERCONCORD</i>	25
2.5.4	<i>NUMBER</i>	25
2.5.5	<i>COMPLEMENTATION</i>	26
2.6	SEMANTICS	27
2.6.1	<i>REFERENCE</i>	27
2.6.2	<i>COUNTABILITY</i>	27
2.6.3	<i>SEMANTIC TYPE</i>	28
2.6.4	<i>SHIFTS</i>	29
2.6.5	<i>RESUME</i>	31
2.6.6	<i>DEFINITION</i>	31
2.6.7	<i>DEFINITION SOURCE</i>	32
2.6.8	<i>SYNONYMS</i>	32
2.7	PRAGMATICS	33
2.7.1	<i>ORIGIN</i>	33
2.7.2	<i>STYLE</i>	33
2.7.3	<i>CONNOTATION</i>	34
2.7.4	<i>SOCIAL GROUP</i>	34
2.7.5	<i>SUBJECT FIELD</i>	35
2.7.6	<i>CHRONOLOGY</i>	37
2.7.7	<i>GEOGRAPHY</i>	38

2.7.8	<i>FREQUENCY</i>	38
2.8	COMBINATORICS	38
2.8.1	<i>IDENTIFICATION</i>	39
2.8.2	<i>SYNTAX</i>	40
2.8.3	<i>SEMANTICS</i>	42
2.8.4	<i>PRAGMATICS</i>	48
2.8.5	<i>COMMENT</i>	48
2.8.6	<i>Collocaties ter illustratie</i>	48
3.	ADJECTIVA	50
3.1	IDENTIFICATION	50
3.1.1	<i>FORM</i>	50
3.1.2	<i>CAT</i>	50
3.1.3	<i>SEQUENCE NUMBER</i>	50
3.2	GRAPHEMICS	51
3.3	PHONETICS	51
3.4	MORPHOLOGY	51
3.4.1	<i>FLECTIONAL TYPE</i>	51
3.4.2	<i>DECLINABILITY</i>	52
3.4.3	<i>COMPAR(ISON)</i>	53
3.4.4	<i>MORPHOTYPE</i>	53
3.4.5	<i>MORPHOSTRUCTURE</i>	54
3.5	SYNTAX	54
3.5.1	<i>POSIT(ION)</i>	54
3.5.2	<i>ADVERBIAL USAGE</i>	54
3.5.3	<i>COMPLEM</i>	55
3.6	SEMANTICS	55
3.6.1	<i>SEMANTIC TYPE</i>	56
3.6.2	<i>SHIFTS</i>	56
3.6.3	<i>RESUME</i>	56
3.6.4	<i>DEFINITION</i>	57
3.6.5	<i>DEFINITION SOURCE</i>	57
3.6.6	<i>SYNONYMS</i>	57
3.7	PRAGMATICS	57
3.8	COMBINATORICS	57
3.8.1	<i>IDENTIFICATION</i>	57
3.8.2	<i>SYNTAX</i>	58
3.8.3	<i>SEMANTICS</i>	59
3.8.4	<i>PRAGMATICS</i>	60
3.8.5	<i>COMMENT</i>	60
4.	VERBA	61
4.1	IDENTIFICATION	61
4.1.1	<i>FORM</i>	61
4.1.2	<i>CAT</i>	61
4.1.3	<i>SEQUENCE NUMBER</i>	61
4.2	GRAPHEMICS	61
4.2.1	<i>TYPE</i>	61
4.2.2	<i>SPELVAR</i>	61
4.2.3	<i>FORMVAR</i>	61
4.2.4	<i>VERWIJSENTRY'S</i>	61
4.3	PHONETICS	62
4.4	MORPHOLOGY AND MORHOSYNTAX	62
4.4.1	<i>MORPHOTYPE</i>	62
4.4.2	<i>SEPARABILITY</i>	62
4.4.3	<i>MORPHOSTRUCTURE</i>	62
4.4.4	<i>CONJUGATION TYPE</i>	62

4.4.5	CONJUGATION-FORM	63
4.4.6	MODE	63
4.4.7	TENSE	63
4.4.8	NUMBER	64
4.4.9	PERSON	64
4.5	SYNTAX	64
4.5.1	CLASS	64
4.5.2	PERAUX	65
4.5.3	VALENCY	65
4.5.4	COMPLEMENTATION	66
4.6.1	PREPCOMPL	69
4.6.2	SUBJECT	70
4.6.3	REFLEXIV(ITY)	70
4.6.4	TRANSITIVITY	71
4.7	SEMANTICS	71
4.7.1	TYPE	71
4.7.2	SUBCLASS/ CASEFRAME/ SR CASEROLES:	71
4.7.3	SHIFTS	74
4.7.4	RESUMES	74
4.7.5	DEFINITION en HYPERONYM	74
4.7.6	DEFINITION SOURCE	75
4.7.7	SYNONYMS	75
4.8	PRAGMATICS	75
4.9	COMBINATORICS	75
5.	ADVERBIA	77
5.1	DE MACROSELECTIE	77
5.2	TYPISCHE EIGENSCHAPPEN VAN DE ADVERBIA	77
5.3	INDELING NAAR FUNCTIE	77
5.4	INDELING NAAR VORM	78
5.4.1	<i>Voorzetselbijwoorden</i>	78
5.4.2	<i>Pronominale of voornaamwoordelijke bijwoorden</i>	78
5.5	INDELING NAAR BETEKENIS	79
5.5.1	<i>Het resumé</i>	79
5.6	OVERZICHT VAN DE VELDEN	80
5.7	EXAMPLE UNITS	81
5.7.1	<i>Aantal EU's</i>	81
5.7.2	<i>Voorzetselbijwoord vs. Pronominaal bijwoord</i>	81
5.7.3	<i>Vaste verbindingen</i>	81
6.	FUNCTIEWOORDEN	82
6.1	INLEIDING	82
6.2	BETEKENISONDERSCHIED	84
6.3	VOORNAAMWOORDEN	85
6.3.1	<i>Niet-zelfstandige voornaamwoorden: determiners</i>	85
6.3.2	<i>Zelfstandige voornaamwoorden: pronouns</i>	87
6.4	LIDWOORDEN	89
6.5	TELWOORDEN	89
6.6	VOORZETSELS	91
6.7	VOEGWOORDEN	92
6.8	DATABASETABEL FUNCTIEWOORDEN	93
6.9	DE TABEL VAN DE COMBINATORICS	96
7.	INTERJECTIES	98
7.1	INDELING MET BETREKKING TOT BETEKENIS	98
7.1.1	<i>Interjecties met voornamelijk communicatieve functie</i>	98
7.1.2	<i>Interjecties met expressieve functie</i>	99

7.2	GEBRUIK EN PLAATS IN DE ZIN	99
7.2.1	<i>Interjecties voor of achter de zin</i>	99
7.2.2	<i>Interjecties binnen de zin</i>	100
7.3	SYNTAXIS	100
7.4	HET RESUMÉ	100
7.5	PRAGMATIEK	101
7.6	EXAMPLE UNITS	101
7.7	DATABASETABEL INTERJECTIES	101
8.	GEOGRAFISCHE NAMEN	103
8.1	BESTANDEN	103
8.2	MACROSTRUCTUUR	103
8.2.1	<i>Bronnen</i>	103
8.2.2	<i>Groot bestand</i>	103
8.2.3	<i>Basisbestand</i>	104
8.3	NOMINA	104
8.3.1	<i>Identification</i>	105
8.3.2	<i>Graphemics</i>	105
8.3.3	<i>Syntax</i>	105
8.3.4	<i>Semantics</i>	106
8.3.5	<i>Combinatorics</i>	106
8.4	ADJECTIVA	106
8.4.1	<i>Identification</i>	107
8.4.2	<i>Graphemics</i>	107
8.4.3	<i>Semantics</i>	107
8.4.4	<i>Combinatorics</i>	107
8.5	OMBI	107
	BIJLAGE 1: SAMENSTELLING SUBCORPUS 38Mi CORPUS INL	108

1. RBN: Algemeen

1.1 Motivering

Het Referentiebestand Nederlands (RBN) is bedoeld als een lexicale databank Nederlands die als moederbestand moet kunnen fungeren voor zowel het begrijpen als het produceren van het Nederlands. Het is gegroeid uit de behoefte die er bestond bij de CLVV¹ (Commissie Lexicografische Vertaalvoorzieningen) aan een exemplarisch lexicaal bestand Nederlands. Bij het uitzetten van vertaalwoordenboekprojecten met het Nederlands als bron- of als doeltaal, werd de CLVV telkens weer geconfronteerd met het ontbreken van een bestand Nederlands van waaruit kon worden vertaald, dan wel waarnaar kon worden verwezen (vanuit de vreemde taal). Beschikte de CLVV niet zelf over een dergelijk bestand dan diende zij bij elk project opnieuw besprekingen te voeren met mogelijke leveranciers van een bestand, in concreto uitgevers.

Een en ander leidt tot:

- het creëren van monopolieposities (alleen bepaalde uitgevers komen nog in aanmerking);
- hoge kosten (voor elk nieuw gebruik moet opnieuw worden betaald);
- beperkte exploitatiemogelijkheden (de CLVV heeft alleen een beperkte licentie, kan niet efficiënt laten omkeren enz.);
- een beperkte onderhandelingspositie bij besprekingen met partners (het ter beschikking stellen van een eigen bestand kan een grote troef zijn in het vaststellen van de financiële participatie van elke partner).

Naast deze directe, lexicografische aanleiding tot het aanmaken van een breed toegankelijk referentiebestand was er voor de CLVV het feit dat, gegeven onze huidige samenleving en haar behoeften, een dergelijk bestand niet los mocht worden gezien van andere, mogelijke, toepassingen. Met name de inzetbaarheid van een RBN bij taaltechnologische toepassingen leek haar van eminent belang. Het ontbreken van een lexicaal bestand voor het Nederlands dat aan de eisen van de informatisering zoals explicietheid en formalisering op het vlak van semantiek, combinatoriek en syntaxis voldeed, was dan ook een tweede reden waarom de CLVV het nodig achtte het RBN te laten entameren. Op die manier zouden niet alleen woordenboeken maar ook elektronische woordenboeken (met andere en met meer informatie dan in de papieren versie), elektronische vertaalhulpmiddelen en ander elektronisch informatie-gereedschap mogelijk kunnen worden gemaakt.

1.2 Algemene karakterisering

Tegen de hierboven geschetste achtergrond zal het duidelijk geworden zijn dat het RBN is

¹ De commissie Lexicale Vertaalvoorzieningen werd per 1 maart 1993 ingesteld door de ministers van Onderwijs van Nederland en Vlaanderen met de bedoeling het beleid mbt. vertaalwoordenboeken met Nederlands als bron- of doeltaal te stroomlijnen en te coördineren; voor meer informatie zie *Woordenboekenbeleid in Nederland en Vlaanderen* (Eindrapport van de Commissie Lexicale Vertaalvoorzieningen, onder redactie van A. Hoorntje en W. Martin, december 2004, Den Haag, 75 pags. + bijlagen).

opgevat als een multifunctioneel lexicon Nederlands met als basisfuncties:

- ✓ productie en
- ✓ receptie

De eerste functie impliceert onder meer dat het lexicon kan dienen als uitgangspunt voor een bestand Nederlands-V (V = vreemde taal), de tweede houdt onder andere in dat, vanuit bijvoorbeeld een vreemde taal kan worden verwezen naar en gekoppeld aan woordbetekenissen uit het RBN. In beide gevallen dient het RBN als een betrouwbaar referentiepunt te fungeren.

In concreto betekent dit dat het RBN voor de CLVV als infrastructuur inzetbaar moet zijn bij haar woordenboekprojecten Nederlands-V en V-Nederlands, inclusief mogelijke multimediale, niet-grafische en computationele voorzieningen, en allerhande derivaten (afhankelijk van de specifieke functie van het woordenboek en/of het specifieke karakter van de gebruiker).

Deze multifunctionaliteit en openheid impliceren uiteraard ook dat, wanneer het RBN bij een concreet woordenboekproject Nederlands – Vreemde Taal wordt ingezet, deze operatie niet kan worden aangezien als een loutere invuloefening d.m.v. de Vreemde Taal. Schrappingen, aanpassingen, eventueel ook toevoegingen kunnen en zullen nodig zijn, afhankelijk van functie en doeltaal. In het gedeelte RBN en OMBI wordt hierop dieper ingegaan. In wat volgt wordt een factuele beschrijving van het RBN gegeven. Deze beschrijving bevat de volgende hoofdstukken: nomina (hoofdstuk 2), adjectieva (hoofdstuk 3), verba (hoofdstuk 4), adverba (hoofdstuk 5), functiewoorden (hoofdstuk 6), interjecties (hoofdstuk 7). Een apart hoofdstuk (8) vormen de geografica. Zoals in het Woord Vooraf vermeld, verschilt de presentatie/uiterlijke vorm/stijl van de diverse onderdelen nogal wat. Dit heeft te maken zowel met inhoudelijke verschillen, verschillen in (oorspronkelijke) auteurs, als met verschillen in ontstaan. De hoofdstukken 1-4 zijn samen ontstaan, evenals de hoofdstukken 5-7. Hoofdstuk 8 is apart ontstaan. Er is niet naar gestreefd deze heterogeniteit weg te werken.

Gegeven het bovenstaande kan het RBN worden gekarakteriseerd als een lexicale databank, waarvan de macro minimaal bestaat uit ongeveer 45.000 trefwoorden (trefwoorden worden hier opgevat als vormeenheden, dus 'lemmata' of woordenboekartikels, i.t.t. lexemen die eenheden van vorm én betekenis zijn). De micro bevat informatie over deze woorden en wel gespreid over zeven linguïstische niveau's, n.l.

1. spelling
2. uitspraak
3. morfologie
4. syntaxis
5. semantiek
6. pragmatiek
7. combinatoriek

In de regel kan de hierboven vermelde informatie op verschillende wijzen gerepresenteerd worden. Voor de CLVV zijn hierbij een aantal zaken belangrijk geweest, te weten:

- de representatie diende de hierboven vermelde inzetbaarheid mogelijk te maken;
- de representatie diende OMBI-compatibel te zijn (d.w.z. compatibel met een editor die de CLVV heeft laten ontwikkelen met het oog op het uitvoeren van redactioneel werk);

- de representatie diende 'haalbaar' te zijn, dat wil zeggen dat met een aantal pragmatische uitgangspunten (b.v. het al dan niet voorhanden zijn van bepaalde data, van een representatieformalisme), een aantal financiële en temporele randvoorwaarden e.d. rekening moest worden gehouden.

Wilde het RBN voor de CLVV-projecten enige relevantie en waarde hebben dan diende het daarenboven:

- a. op korte termijn beschikbaar te zijn
- b. van goede kwaliteit te zijn, en
- c. een voldoende breed draagvlak te hebben (zonder te leiden tot een organisatorische monstrositeit).

Met de aanmaak van het RBN is dan ook begonnen in september 1996. Vier lexicografische teams hebben eraan gewerkt: het INL te Leiden², het Departement Linguïstiek van de KU Leuven³, de vakgroep Nederlands van de RU Utrecht⁴ en de vakgroep Lexicologie van de VU Amsterdam⁵. Bij deze laatste vakgroep berustte ook de algemene leiding en coördinatie. Prof. Dr. W. Martin (VU) was als projectleider verantwoordelijk voor de redactionele instructies en de algemene leiding en Drs. Isa Maks (VU) zorgde voor distributie, coördinatie, voortgang en beheer van de data. Het redactionele werk werd ongeveer gelijkmatig onder de vier teams verdeeld. Na de redactie volgde de controlefase waarbij het leeuwendeel van het werk door de vakgroep Lexicologie van de VU werd verricht. Vanaf 1999 konden delen van het RBN bij lexicografische projecten worden ingezet. Tot op heden (2005) gebeurde dat bij de projecten Arabisch, Deens, Estisch, Fins, Indonesisch, Nieuwgrieks, Portugees, Roemeens en Turks. Daarnaast is het RBN ook gebruikt binnen het project Noors en voor niet-lexicografische toepassingen zoals het Corpus Gesproken Nederlands, de Databank Overheidsterminologie en NL-Translex.

Wat de lexicografische teams betreft, zij kregen een licht aangepaste versie, in de OMBI-editor geïmporteerd. Een toelichting bij deze aanpassing volgt hierna in het gedeelte rbn en ombi (1.5), tevens wordt in de hoofdstukken 2 t.e.m. 4 bij elke datacategorie onder het kopje OMBI, de weergave van de informatie onder OMBI opgegeven.

1.3 Macrostructuur

1.3.1 Bronnen

Zoals bekend wordt met macrostructuur bedoeld: de verzameling lexemen waarover informatie gegeven wordt. Het RBN heeft als kern een zgn. basisvocabulary, i.e. een verzameling Nederlandse lexemen die min of meer centraal zijn voor een gemiddelde moedertaalspreker. Het te beschrijven lexicon van een taal wordt als recursief aangezien waarbij het basisvocabulary

² Met de volgende redacteuren: Ammiranda Abbes, Lineke Oppentocht, Hinke Vissia

³ Met de volgende redacteuren: Geert Ceuppens, Geert Craps

⁴ Met de volgende redacteuren: Peter Ackema, Heleen Hoekstra

⁵ Met de volgende redacteuren en controleurs (bureauredacteuren): Hanny Demeersseman, Hennie van der Vliet, Inger Groesz, Essemie van Dunnee, Mila Groot, Geert Bilsen, Stephan Bopp, Jeroen Redel

als uitbreidbare kern fungeert. Het lexicon (L) wordt daarbij gedefinieerd als⁶:

$$\begin{array}{l} L \text{ --->} L_{\text{kern}} + L_{\text{uitbreiding}} \\ L_{\text{uitbreiding}} \text{ --->} L_{\text{kern}} + L_{\text{uitbreiding}} \end{array}$$

Het lexicon bestaat dus uit een kern en een uitbreiding. Deze uitbreiding bestaat op haar beurt weer uit kern + uitbreiding die weer uit kern + uitbreiding bestaat / kan bestaan enz. Een en ander moet ons in staat stellen het Nederlandse lexicon op een beredeneerde en gestructureerde wijze te laten evolueren, uit te breiden / in te krimpen, en van zichzelf te laten leren (bv. door overerving).

Als uitgangspunt en kern voor de macrostructuur van het RBN werd gekozen voor een eerder aan de VU tot stand gekomen basisvocabularium van ca. 20.000 lemmata. Verder is er vermelding van vakgebieden: of woorden algemeen of vaktalig zijn of weliswaar vaktalig maar toch algemeen (gebruikt buiten de vakttaal/het domein). Op die manier kan het lexicon in verschillende strata of lagen worden opgedeeld (kern (= algemeen/frequent), uitbreiding (= vaktalig/minder frequent) met hierin als kern eerst wat algemener/frequenter is; etc.) en kan uitbreiding/inkrimping op een minder willekeurige wijze verlopen.

Voor de uitbreiding van het basisvocabularium tot de gewenste omvang van het RBN, zo'n 45.000 lemmata, heeft het **5Miljoen-corpus**⁷ (INL) als voornaamste bron gefungeerd (zie ook onder toevoegen). Dit corpus uit 1994 bevat tekstmateriaal uit de volgende Nederlandse bronnen:

- het journaal (NOS 8 uur en jeugdjournaal, '92), de troonredes ('70-'88), het *NRC* ('93-'94), het *Staatsblad* ('91);
- tijdschriften: het *Kwartaalblad voor donateurs van Greenpeace*, het *Waddenbulletin van de Landelijke Vereniging tot Behoud van de Waddenzee* ('89-'93), *Onze Taal* ('91-'92), *Liberaal Reveil* ('92), columns uit de *Playboy* en de *Playboy Stijl* ('92-'93), de *Waterkampioen*, *Orgaan van de ANWB* ('92-'93);
- boeken: *Nieuwe strategieën voor natuurbescherming* ('89), *Productiviteit in Taal en Taalgebruik* ('92), *Sociaal Ontwerpen, werknemers beïnvloeden vormgeving en organisatie van hun werk* ('91), *Vrouwen en nieuwe technologieën in Noord-Nederland* ('90), *Cursus, De Gemeenteraad* ('82).

1.3.2 Schrappen in en Samenvoegen van Bronnen

Zoals hierboven verduidelijkt zijn de basisbronnen voor het RBN:

- het basisvocabulaire Nederlands van de VU
- het INL-corpus

In de eerste bron zijn een beperkt aantal items geschrapt, in de tweede bron is er een selectie

⁶ Martin, W. 1988. Variation in lexical frequency, in: *Distributions spatiales et temporelles*, 1988, ed. P. van Reenen - K. van Reenen - Stein, , Amsterdam: Benjamins 139-152.

⁷ Kruij, T. 1995. Nationale Tekstcorpora in internationaal perspectief.

gemaakt uit het totale aantal items. Daarna zijn de twee bronnen met elkaar vergeleken en samengevoegd, daarbij fungeerde het VU-basisvocabulaire als aan te vullen uitgangspunt.

De selectie uit het INL-corpus verliep aldus: de woorden met een frequentie ≥ 10 zijn zonder meer opgenomen. Ook alle woorden met $5 \leq f \leq 9$ zijn opgenomen, tenzij in geval van twijfel. In dat geval is gecheckt of ze in de Grote Van Dale⁸ (GVD) voorkwamen. Bij woorden met $f=3/4$ was de default 'niet opnemen', toch zijn ook daarvan woorden geselecteerd om redenen van consistentie of op grond van intuïtie. Veel van de zogenaamd zeer frequente woorden zijn toch geschrapt, b.v. omdat ze fout waren ('denemark', 'samenwonend' = adj ipv. nomen) of omdat ze, ondanks hun frequentie, niet in het RBN thuishoren (zie ook 1.3.5). Doorzichtigheid was geen criterium: 'wapenhandelaar' en 'melkhandelaar' zijn even doorzichtig, maar alleen de eerste is (omwille van zijn frequentie) opgenomen (uiteeraard is, in het tweede geval, het frequentere 'melkboer' opgenomen).

Ter illustratie volgen drie stukken uit de frequentielijst van het *5 Miljoen Corpus*. Het gaat alleen om nomina. De vetgedrukte woorden zijn in het RBN opgenomen, de andere niet. Sommige ervan zoals saxofoon, scala en scenario waren al in het VU-basisvocabularium aanwezig. Hier werd alleen de frequentie uit het INL-corpus toegevoegd. Andere zoals scenarioschrijver ontbraken, zij werden toegevoegd met frequentieopgave.

Frequentie >10:

item	frequentie
saxofoon	15
scala	45
scalfaro	36
scanshippen	33
scenario	95
scenarioschrijver	11
scenes	51
scepsis	27
scepter	21
schaal	265
schaalvergroting	30
schaamte	30
schaap	90
schaarste	12
schaats	50
schaatser	42
schacht	21
schade	709
schadeclaim	45
schadeloosstelling	11

Frequentie 5-9:

item	frequentie
s-spant	7
s.v.-fonds	6
saaiheid	6
sabb-diesel	5

⁸ Geerts, G. en Heestermans, H. 1992. Van Dale. *Groot woordenboek der Nederlandse taal*. Utrecht: Van Dale Lexicografie (12^e druk).

saboteur	6
sacrament	6
sacre	9
sacrum	5
saga	5
sage	9
saillant	5
sailmate	7
sake	5
salamander	5
salarisverbetering	7
salarisverhoging	7
salman	7
salsa	5
salvo	6

Frequentie 3-4:

item	frequentie
sailmaker	3
sailsimulator	3
sajudis	3
salah	3
salarisachterstand	4
salarisuitgave	4
sallenave	4
salonmotorboot	3
saloon	3
salto	4
salzburger	3
sam	3
sambre	3
samenbundeling	3
samenlevingsvorm	4
samenstellingsanalyse	3
samenwerkingsakkoord	3
samenwerkingsconstructie	3
samenwerkingsproject	4
samenwonend	3

1.3.3 Toevoegen

Als voornaamste bron voor toevoeging aan de macro, tot stand gekomen door samenvoeging van het Basisvocabularium-VU en de (uitgedunde) INL-lijst, gold het HNE⁹. Per letter werd de macrolijst vergeleken met de macro in het HNE. Een woord dat ten onrechte niet in de macrolijst stond, en wel in het HNE voorkwam, werd in een apart toevoeglijstje genoteerd.

Anders dan voor het selecteren uit 1.3.2 waren er geen specifieke ‘regels’ voor het toevoegen uit HNE, maar wel werden principes als ‘consistentie’ en ‘subjectieve frequentie’ gehanteerd. Er is geprobeerd de manier van toevoegen uit HNE zoveel mogelijk te laten aansluiten bij de schrap-

⁹ Hannay, M. 1994 Van Dale *Handwoordenboek Nederlands-Engels*. Utrecht: Van Dale Lexicografie.

en selectiestrategie. Bij twijfel werden andere bronnen geraadpleegd.

1.3.4 Omvang van de macro

Het oorspronkelijke VU-basisvocabularium telde ca. 21.000 woorden, het **RBN**-bestand omvat nu ca. 42.000 znw., bnw. en ww. Hier zijn nog een hoeveelheid functiewoorden en geografische namen bijgekomen. De verdeling is als volgt:

woordcategorie	znw.	ww.	bnw.	totaal	
aanvankelijk in VU-basis		14.132	3.716	3.587	21.435
geschrapt in VU-basis		-945	-89	-101	-1.135
overgenomen uit VU-basis		13.187	3.627	3.486	20.300
toegevoegd uit INL-corpus		+10.217	+1.686	+1.823	+13.726
andere toevoegingen		+5.896	+987	+991	+7.874
subtotaal		29.300	6.300	6.300	41.900
geografische namen					± 2.400
functiewoorden					± 1.450
totaal RBN					± 45.750

1.3.5 Opmerkingen bij de selectie

Bepaalde conceptuele domeinen, grammaticale en morfologische categorieën hebben een aparte behandeling gekregen. Bij het aanpassen van het RBN voor specifieke doelgroepen kan het nuttig zijn hiermee rekening te houden.

a) Scheepstermen. Een van de bronnen waarop het INL-corpus is gebaseerd, is de ‘Waterkampioen’ van de ANWB. Daardoor heeft het een overaanbod aan scheepstermen, vaak met een hoge frequentie. Hierdoor zijn er, vooral in het begin, af en toe ten onrechte woorden toegevoegd die eigenlijk niet in het RBN thuishoren, zoals ‘ankergerei’, ‘bakboordlantaarn’ (f=17), boottrailer (f=42)... Dit is inmiddels enigszins rechtgetrokken.

b) Eigennamen. In de macro zijn in principe geen eigennamen opgenomen. Er zijn wel aardrijkskundige namen in het RBN opgenomen a.d.h.v. het boekje ‘Buitenlandse aardrijkskundige namen in het Nederlands’. Hieruit zijn namen van landen en hun hoofdsteden, inwonersnamen, talen en andere geografische plaatsen behandeld samen met de affectivische afleidingen. Maar ook relevante Belgische en Nederlandse plaatsnamen zijn opgenomen (zoals provincieën en hun hoofdsteden, belangrijke rivieren, streken als ‘de Kempen’, ‘de Veluwe’). Het totale bestand telt ca. 2.400 items waaruit geselecteerd kan worden (zie ook hoofdstuk 8:

Geografische Namen).

c) Kleuren. Van de kleuren zijn alleen ‘rood’, ‘blauw’, ‘geel’, ‘groen’, ‘geel’, ‘wit’, ‘zwart’, ‘oranje’, ‘paars’, ‘purper’, ‘grijs’, ‘bruin’ en ‘roze’ ook als zn. opgenomen. De andere (‘kanariegeel’, ‘flesgroen’, ‘lila’ enz.) zijn alleen als bn. geselecteerd.

d) Vreemde maten en eenheden. Namen van vreemde maten en eenheden (‘pond sterling’, ‘yard’ enz.) zijn geweerd.

e) Vrouwelijke afleidingen. Voor de opname van een vrouwelijke afleiding van een mannelijk zn. waren er verschillende criteria: 1. het woord moet frequent zijn (‘aanstelster’ niet, ‘lerares’, ‘boerin’ wel), of 2. de vrouwelijke afleiding verschilt niet alleen in geslacht van de mannelijke variant (‘voetbalster’, ‘bassiste’ niet, ‘secretaresse’ wel) of 3. de vrouwelijke variant is veel gangbaarder dan de mannelijke of er is geen mannelijke variant (‘caissière’) of 4. de morfologische afleiding is bijzonder (‘ambassadrice’, ‘negerin’, ‘stewardess’, ‘zangeres’). Als productieve vrouwelijke suffixen gelden ‘-ster’ en ‘-e’. Dit laatste punt was ook van belang voor de selectie van aardrijkskundige afleidingen, b.v. Fransman/Française, Belg/Belgische, Duitser/Duitse, Zwitser/Zwitserse, Engelsman/Engelse, Spanjaard/Spaanse.

f) Vlaams/taalvariatie. In het RBN komen ook Vlaamse woorden voor. Hiermee worden die woorden bedoeld die (in een bepaalde betekenis) alleen in Vlaanderen worden gebruikt én die in principe elke Vlaming kent (dus supra-regionaal zijn). Hieronder vallen ook de namen van Belgische/Vlaamse instellingen en andere typisch Belgische/Vlaamse zaken (‘schepen’, ‘mattetaart’). Als Vlaamse woorden zijn nu al opgenomen o.a. ‘panikereren’ (= ‘in paniek raken’), ‘toekomen’ (= ‘aankomen, arriveren’), ‘rechtstaan’ (= ‘opstaan’), ‘microgolfoven’ (‘magnetron’), ‘micro’ (= ‘microfoon’), ‘uitstap’ (= ‘uitstapje’), ‘dampkap’ (= ‘hot’), ‘(kinder)kribbe’ = ‘crèche’.

Zoals bekend, zijn talen onderhevig aan verandering, maar voor het Vlaams geldt dit in hoge mate. Woorden in Nederland gebruikt worden sneller dan vroeger opgepikt en het Vlaams worstelt nog altijd met een Franse erfenis. Daardoor bestaat er onzekerheid over de status van veel Vlaamse woorden. Om die reden is het niet makkelijk om een lijst van Vlaamse woorden te selecteren die nog algemeen gebruikt worden en die niet op korte termijn zullen verdwijnen. Een dergelijke lijst is thans als RBBN (= Referentiebestand Belgisch Nederlands) beschikbaar.

Zolang het RBN niet met het RBBN is gecoördineerd, dient men er echter rekening mee te houden dat de selectie van het Vlaams in het RBN niet consequent is gebeurd, omdat het 5Mi-corpus geen Vlaamse bronnen heeft gebruikt en omdat niet bewust op Vlaamse woorden is gezocht. Overigens gaat het niet alleen om afwijkende Vlaamse woordvormen, maar ook om verbindingen (‘dat zijn vijgen na Pasen’ vs. ‘dat is mosterd na de maaltijd’, ‘je hebt de groeten van’ vs. ‘je moet de groeten hebben van’, ‘uit de lucht vallen’ vs. ‘uit de lucht komen vallen’, ‘lachen met’ vs. ‘lachen om’) die m.n. door de Vlaamse RBN-redacteurs (kunnen) zijn opgevoerd. Het omgekeerde geldt natuurlijk ook voor veel ‘Hollandse’ woorden/uitdrukkingen die in Vlaanderen niet, nauwelijks of minder vaak gebruikt worden (bv. ‘inkomensplaatje’, ‘arriveren’, ‘soortement’, ‘bonbon’, ‘jus’) of niet in een van de ‘Hollandse’ betekenissen (‘gaaf’ als ‘prachtig, geweldig’).

g) Bijvoeglijk naamwoord/deelwoord. Veel voltooide/onvoltooide deelwoorden kunnen als attributieve bnw. gebruikt worden. Is dit voor een bepaalde vorm het geval, dan is het bnw.

opgenomen, op voorwaarde dat het voldoende frequent is. Het HNE fungeerde hierbij als leidraad.

h) Bijvoeglijk naamwoord/bijwoord. De meeste bijvoeglijke naamwoorden kunnen als bijwoord fungeren. Het is meestal weinig zinvol om ze voor beide categorieën op te voeren (b.v. ‘geruisloos’, ‘plechtig’). Dat is alleen gedaan als er sprake is van een significant betekenisverschil (‘praktisch’ als in ‘praktisch onmogelijk’) en/of het bijwoord (frequent) een typische functie vervult: b.v. een intensiverende (‘zwaar’, ‘ernstig’) of modale functie (‘plotseling’, ‘mogelijk’, ‘waarschijnlijk’ i.t.t. ‘onwaarschijnlijk’). Bijvoeglijke naamwoorden die nauwelijks meer als zodanig, maar wel als bijwoord gebruikt worden (zoals ‘dadelijk’), zijn alleen in deze laatste groep opgenomen. Ook hierbij is de selectie van het HNE in grote lijnen gevolgd.

1.4 Betekenenissen in microstructuur

In 1.3 is duidelijk gemaakt wat voor de macrostructuur onze uitgangspunten waren. In de hoofdstukken 2 en volgende zal duidelijk worden dat de empirische basis voor de microstructuur vrij ruim is en dat er op verschillende informatiebronnen een beroep is gedaan.

Voor wat betreft het aantal betekenissen heeft een ‘zuinige’ betekenisstructuur model gestaan, al was het niet altijd makkelijk om dit consequent door te voeren.

De voornaamste leidraad voor de RBN-redacteuren bij het vaststellen van betekenissen was dat woorden/lexemen één (of meer) basisbetekenis(sen) hebben d.w.z. betekenissen van waaruit de andere betekenissen (makkelijk) kunnen worden afgeleid. Bij het naamwoord ‘hel’ b.v. wordt uit de betekenis 1 ‘strafplaats voor zondaars’ de pejoratieve connotatie veralgemeend tot betekenis 2 ‘verschrikkelijke plaats/omstandigheid’. Twee zaken zijn dus aan de orde:

- het aantal betekenissen (en daar wordt zuinig mee omgesprongen)
- de volgorde van die betekenissen

In de regel is er als volgt gewerkt:

--- Bij de woorden uit het VU-basisvocabularium:

De bij deze woorden reeds aanwezige betekenisstructuur is gecheckt op basis van het Corpus¹⁰. B.v. bij ‘honger’ stonden oorspronkelijk twee betekenissen: ‘sterke behoefte om te eten’, en ‘groot verlangen’. Als op basis van het Corpus (b.v. omdat er vaak uitdrukkingen voorkomen als ‘de honger in Afrika’, ‘honger en armoede’ e.d.) het besluit is gevallen dat daarnaast nog een derde betekenis aanwezig is, dan is die derde betekenis ook opgevoerd, b.v. ‘hongersnood’.

--- Bij de nieuwe woorden uit het Corpus:

Hier zijn de betekenissen niet gecheckt e/o veranderd/uitgebreid/gereduceerd, maar wel vastgesteld en aangebracht op basis van het Corpus. De lezingen uit de Hedendaagse Van Dale

¹⁰ Bij het bepalen van de microstructuur is gebruik gemaakt van een selectie van het INL-38miljoen-corpus. Zie ook bijlage 1.

(HVD) en Van Dale Groot Woordenboek Nederlands-Engels (NEVD)¹¹ zijn als hypothesen opgevat.

Overigens is bij het vaststellen van verschil in woordbetekenis geadviseerd de zgn. Leisi-methode te volgen. Leisi gaat er vanuit dat een woord maar één betekenis heeft tot het bewijs van het tegengestelde is geleverd (methodologische monosemie). Om een tweede, derde etc. betekenis te ‘ontdekken’ gaat hij als volgt te werk:

- kies voor een woord W in betekenis x een selector S1 (een woord waarmee W samen in combinatie voorkomt)
- als er een vermoeden is dat W nog een andere betekenis kan hebben, kies dan een typische selector S2 voor deze vermeende betekenis
- combineer W samen met S1 en S2 (b.v. d.m.v. coördinatie) in één uiting
- als de combinatie acceptabel is dan wijst dit erop dat S1 en S2 compatibel zijn, eenheden zijn die tot één verzameling behoren, en wordt W geacht geen twee verschillende betekenissen te realiseren
- als de combinatie een bepaald effect (woordspeling e.d.) sorteert, dan hebben we met twee groepen selectoren te maken die elk met een verschillende betekenis combineren. Bijvoorbeeld ‘slot’: ‘We bekeken twee sloten gisteren: het slot van graaf ten Berge en dat van mijn fiets’. Het is duidelijk dat de vorm ‘slot’ in bovenvermelde zin met twee selectoren uit een andere categorie wordt gecombineerd, een (pseudo-)grappig effect sorteert, en derhalve kan worden beschouwd als zijnde:

slot 1 + S1
slot 2 + S2

i.e. de realisatie van twee betekenissen van ‘slot’.

Al vertoont de hier gevolgde strategie bepaalde moeilijkheden (zoals de keuze van selectoren, het construeren van coördinaties, het beoordelen ervan), toch biedt ze vaak een bruikbaar handvat om in geval van twijfel tot een empirisch-onderbouwde en controleerbare beslissing te komen. Deze methode is echter vrij ‘bewerkelijk’. In de praktijk ging een en ander meestal als volgt. Bij ‘bungalow’ b.v., een entry uit het VU-basisvocabularium, stond één betekenis: ‘losstaand huis met alleen kamers op de begane grond’. A.d.h.v. het Corpus of eigen intuïtie lijkt het dat er naast de ‘fraaie’ bungalow-betekenis, ook nog zoiets is als ‘goedkoop vakantiehuisje’. Werde de tweede ‘betekenis’ bevestigd (vnl. in de NEVD) dan werd die overgenomen. Dit is dus een zeer pragmatische werkwijze die echter meestal tot goede resultaten leidde.

Slotsom:

* Er wordt zuinig om gesprongen met het aantal betekenissen.

* Bij de woorden uit het VU-basisvocabularium is dit basisvocabularium het uitgangspunt; check gebeurd a.d.h.v. het Corpus (zie b.v. honger); toevoeging indien bevestigd door NEVD en/of Leisi-test.

¹¹ Martin, W. en Tops, G. (Eds.) 1991 2^e druk. *Van Dale Groot Woordenboek Nederladns-Engels*. Utrecht: Van Dale Lexicografie.

- Bij de nieuwe woorden fungeerde de NEVD als uitgangspunt en werd er gecheckt a.d.h.v. het Corpus/Leisi.

1.5 RBN en OMBI¹²

Dit hoofdstuk geeft informatie over de integratie van het RBN in OMBI. Dit hoofdstuk is ook van belang voor diegenen die niet met OMBI werken maar wel een tweetalig woordenboek maken met het RBN als referentiebestand.

1.5.1 Welke gegevens

Voor de meeste gegevens die in het RBN voorkomen, geldt dat ze onveranderd zijn overgenomen in OMBI. In sommige gevallen zijn de waarden enigszins aangepast of in het geheel niet overgenomen. In de volgende hoofdstukken wordt per veld de status in het RBN beschreven en steeds vermeld of en hoe/waar dit in OMBI is opgenomen.

De informatie die in het RBN voorkomt is in te delen in 3 categorieën, afhankelijk van het relatieve belang van deze data voor tweetalige woordenboeken.

- Data die relevant zijn voor woordenboekmakers en - gebruikers
Deze gegevens zijn opgenomen in OMBI. Het grootste deel van de RBN-gegevens valt hieronder en zal ook van nut geacht worden voor de toekomstige gebruikers van de bilinguale woordenboeken die met OMBI gemaakt worden.
 - ongemarkeerd
- Data die relevant zijn voor woordenboekmakers
Deze gegevens zijn opgenomen in OMBI . Niet alle woordenboekmakers zullen van deze informatie gebruik maken en de informatie zal vrijwel zeker niet van nut zijn voor de woordenboekgebruiker.
 - gemarkeerd met een *
- Data die vooral voor automatische taalverwerking en/of voor monolinguale bestanden relevant zijn
Deze gegevens zijn meestal niet opgenomen in OMBI
 - gemarkeerd met niet opgenomen

1.5.2 Hoe omgaan met de gegevens

Het staat de gebruikers van het RBN vrij te bepalen welke informatie zij willen handhaven en welke genegeerd of gewijzigd wordt. Bovendien kunnen gegevens worden toegevoegd. Schrappen en negeren zijn weinig tijdrovende ingrepen. Wijzigen vraagt in veel gevallen meer tijd.

¹² OMBI is een editor voor het vervaardigen van bilinguale, omkeerbare bestanden, ontworpen in opdracht van de CLVV.

- Form Units

Lemma's kunnen makkelijk worden verwijderd, uiteraard met het risico een onevenwichtige macro te krijgen. Als de macro aanmerkelijk kleiner moet zijn dan 45.000 lemma's, kan men beter uitgaan van het VU-basisvocabularium met eventuele aanvullingen of gebruik maken van een selectie van het RBN bestemd voor leerwoordenboeken.¹³ Er is tevens een sterk gereduceerde versie van, RBN-klein in de maak met een macro van ongeveer 10.000 items.

- Lexical Units

Wijzigingen in de betekenisstructuur, bijvoorbeeld samenvoegingen, kunnen tijdrovend zijn en dus beter vermeden worden.

- Example en Idiom Units

Wijzigingen in voorbeelden zijn mogelijk. Er moet wel op worden gelet dat de beschrijving van de voorbeelden nog van toepassing is. Wil men b.v. vrije voorbeelden op grote schaal schrappen (of niet vertalen) dan kan dat het best door de waarde van deze voorbeelden (=FREE) ongewijzigd te laten. Er kan dan worden in voorzien dat voorbeelden met waarde FREE niet in het woordenboek worden afgedrukt (en dus niet vertaald hoeven te worden). Wil men bepaalde voorbeelden toch vertalen en opnemen dan verandert men FREE in b.v. FR.

¹³ Gebaseerd op de frequentie van het INL-corpus (5 miljoen) en de frequentielijst van A. Schrooten (Schrooten, A. en Vermeer, A. 1994. *Woorden in het basisonderwijs*. Studies in meertaligheid 6. Tilburg University Press). Het resultaat is een lijst van 20.000 lemma's met een kern van ca. 7.000 items die zeker opgenomen dienen te worden en een uitbreiding van ca. 13.000 items waarover twijfel bestaat.

2. Nomina¹⁴

Voor de beschrijving die volgt is de gebruikte metataal het Engels. De data zijn in acht zgn. 'frames' of informatiecomponenten weergegeven nl.

1. Identification
2. Graphemics
3. Phonetics
4. Morphology
5. Syntax
6. Semantics
7. Pragmatics
8. Combinatorics

De gebruikte basisstructuur voor de descriptie zijn *feature-value*paren. Aangezien kenmerken en waarden per woordsoort kunnen verschillen, is gekozen voor een behandeling per woordsoort.

2.1 IDENTIFICATION

2.1.1 FORM

De Form is de canonieke spellingsvorm (i.e. in de regel geen verbogen/vervoegde vormen). Voor de FORM is als spellingnorm het Groene Boekje 1995 (GB) gehanteerd; komt het item daar niet in voor dan zijn de spellingregels uit het GB toegepast. Bij afkortingen die met en zonder punten kunnen voorkomen, is de vorm zonder punten aangehouden, dus KLM, niet K.L.M..

Bij **FORM** dient nog het volgende vermeld te worden i.v.m. nominale woordgroepen. In de regel worden bij de nomina geen woordgroepen opgenomen; we kunnen echter drie soorten 'probleemgevallen' onderscheiden:

- 'in natura': nergens onder een Nederlands(e) woord(vorm) te hangen; wordt als geheel opgenomen, in dit concrete geval gaat het om een adv.
- 'ten name van': wordt onder 'naam' behandeld, zijn vaak flectierelicten
- 'met argusogen', 'bij de kladden grijpen': bevat items die duidelijk herkenbaar zijn als Nederlandse nouns; zijn ofwel als combinatoriek bij een bestaand entry, ofwel als apart entry opgenomen (dat in de beschrijving gaten bevat, bv. 'argusoog' heeft dan geen semantische beschrijving)

Verkleinwoorden staan als aparte entry in de macro bij verschil in semantiek; als aparte lezing is de betekenis 'kleine N' opgenomen, zie bv. 'uiltje'. (1. kleine uil 2. biljet van fl 100,--)

¹⁴ Tabelnaam RBN versie 1.0 (1998) nomina en voorbeelden
RBN versie 2.0 (2005) zelfstandige naamwoorden en zn-voorbeelden

=OMBI=

veldnaam: spelling
scherm: hoofdscherm

2.1.2 CAT

Cat is de woordklasse. De waarde voor **CAT** is hier **NOUN**. In principe is de CAT-aanduiding in de meeste woordenboeken eenduidig. Soms echter ontbreekt de CAT-aanduiding volledig zoals b.v. bij 'D': Romeins cijfer voor 500 of 'm': <afkorting> mannelijk of 'in dubio' [Latijn]. In het RBN is dan de CAT-waarde op basis van betekenis/functie gekozen. Zo b.v. is bij geschreven afkortingen van het type 'm.' de informatie weergegeven zoals bij de niet-afgekorte vorm.

=OMBI=

veldnaam: word category
waarde: noun
scherm: morphology

2.1.3 SEQUENCE NUMBER en HOMONYM NUMBER

Is in feite een volgnummer (=cijfer). Het **VOLGNUMMER** scheidt de lezingen binnen een polyseem. Binnen één woordklasse werden oorspronkelijk geen homoniemen onderscheiden, tenzij het ging om woorden met verschillende uitspraak (b.v. vóorkomen en voorkómen beide als ww.). Gevallen als vorst (koning) en vorst (het vriezen) die op etymologische gronden onderscheiden kunnen worden, werden in het RBN dus niet van polysemen onderscheiden. Recentelijk (2005) is hierin verandering gekomen en is er een veld aan het RBN toegevoegd, het zgn. homoniemnummer (**H-NR**). Hierin staat het homoniemnummer waartoe een woord behoort, opgegeven. De homoniemen worden onderscheiden op basis van woordcategoriale, subcategoriale morfo-syntactische, en, eventueel oor de wijze waarop dit gerealiseerd is, zie onder OMBI.

=OMBI=

In OMBI worden binnen één woordklasse wel homoniemen onderscheiden. Het volgnummer is op 2 manieren weergegeven: als sequence number ter onderscheiding van verschillende Form Units met dezelfde vorm (zgn. homoniemen); en als betekenisonderscheiding op Lexical Unit niveau (zgn. polysemen).

veldnaam: sequence number
scherm: hoofdscherm
veldnaam: order
scherm: lexical unit

NB.: de betekenissen zijn , op het scherm, niet expliciet genummerd, maar wel duidelijk van elkaar onderscheiden.

◆ Onderscheid tussen homoniemen en polysemen in OMBI/RBN

⇒ Ombi-homoniemen (dwz. verschillende Form Units) worden onderscheiden op basis van woordcategoriale, subcategoriale morfo-syntactische, en, eventueel, pragmatische verschillen.

Nomina: verschil in lidwoord, meervoud, oorsprong¹⁵

Verba: verschil in transitiviteit, reflexiviteit, scheidbaarheid, werkwoordparadigma (= hoofdtijden), hulpwerkwoord

Adjectiva: hier worden geen homoniemen onderscheiden

Op die wijze wordt er wel een onderscheid gemaakt tussen vorst (het vriezen, geen mv.) en vorst (koning, wel mv.), maar eveneens tussen 'een landbouwbedrijf' (= 'agent') en 'het

¹⁵ In het RBN wordt de taal van herkomst gegeven wanneer dat tot uitspraakverschillen leidt.

landbouwbedrijf (= 'actie').

2.2 GRAPHEMICS

Hier worden drie features onderscheiden (2.2.1 t/m 2.2.3).

2.2.1 TYPE

Het Type verwijst naar de lengte van het te beschrijven item; waarden:

SHORTENING: reductie van woord tot een gedeelte van zijn oorspronkelijke vorm, zoals b.v. demo, info.

ORIGSHORT: oorspronkelijke verkorting: b.v. bus, auto

ABBREVIATION: afkorting, zoals: b.v., enz., inclusief letterwoorden als afzonderlijke strings van letters uitgesproken, b.v. NOS

ACRONYM: i.t.t. afkorting (zie NOS) gaat het nu om letterwoorden die als een woord, niet als een reeks letters worden uitgesproken, bv. NATO.

CONTR(action): samentrekking: zoals 'ten' of 't'

FULL: volledige vorm (onverkort), default

=OMBI=

veldnaam: spelling type

waarde: *zie boven*

scherm: spelling

2.2.2 SPELVAR

--- default is SPELVARNIL = geen variant; als andere waarden worden hier de spellingsvarianten opgegeven. Dit is alles wat van de officiële spelling afwijkt en voldoende frequent is. Bijvoorbeeld i.v. niveau: 'nivo', i.v. eega: 'ega'

--- de norm voor de FORM is het GB, in SPELVAR zal een van deze norm afwijkende vorm staan (hiervoor was Corpus vaak de bron).

=OMBI=

niet opgenomen

2.2.3 FORMVAR

default is FORMVARNIL = geen variant; vormvarianten vertonen naast verschillende spellingvorm ook nog verschillen in uitspraak. B.v. druppel, droppel; onweer, onweder (bron hiervoor: Corpus).

=OMBI=

niet opgenomen

2.2.4 VERWIJSEENTRY'S

Spellings- en vormvarianten (dus ook abbreviations naast hun voluitvormen) zijn ook als zgn. verwijseentry opgenomen. Ze zijn als gewoon behandelentry aangemaakt, waarbij REFERS TO de verwijzing bevat naar een ander behandelentry (het hoofdbehandelentry) en de informatie die

afwijkt tov. dat hoofdentry. Alle niet opgegeven info wordt van het hoofdbehandelentry overgeërfd. Plurale vormen kunnen, zo nodig, bij het verwijssentry worden opgenomen, evenals combinaties die voor het verwijssentry uniek zijn. In REFERS TO worden het TYPE, de status (officieel of niet) en de (relatieve) pragmatische beperkingen van het verwijssentry ingevuld, pragmatische beperkingen volgens de features van de PRAGMATICS, zie 2.7. Voorbeelden van verwijssentry's zijn *nivo*, *ega*, maar ook afkortingen als *km* (*kilometer*), *p.* (*pagina*).

Paarlemoer als variant van *parelmoer* krijgt als verwijssentry de volgende informatie mee:

FORM	<i>paarlemoer</i>
CAT	noun
NUMBER	1
TYPE	FULL
PLURFORM	nil
REFERS TO	<i>parelmoer</i>
PRAGM CONSTRAINTS	LESS FREQUENT OFFICIAL FORMAL OLDFASH

=OMBI=

niet opgenomen

SHORTENING, ORIGSHORT, ABBREVIATION en ACRONYM:

Als de vormen niet worden opgevangen door verwijssentry's, bv. omdat het geen varianten van woorden, maar van woordgroepen zijn, dan is de 'voluitvorm' in COMMENT gezet. Bv. 'KLM': ABBREVIATION en COMMENT 'Koninklijke Luchtvaart Maatschappij'.

=OMBI=

veldnaam:	explanation
waarde:	string
scherm:	spelling

N.B. In de huidige fase van het RBN is geen informatie mbt. woordafbreking opgenomen. Een en ander is wel door OMBI voorzien.

2.3 PHONETICS

In de huidige fase is geen uitspraakinformatie opgenomen. Een en ander is wel door OMBI voorzien.

2.4 MORPHOLOGY

Vier features worden onderscheiden waarvan het tweede, de MORPHOSTRUCTURE, alleen bij de woorden uit het VU-basisvocabularium voorkomt. De features zijn (2.4.1 t/m 2.4.4):

2.4.1 MORPHOTYPE

SIMPMORPH	ongeleed woord
COMPOUND	samenstelling
XCOMPOUND	idiomatische samenstelling

DERIVATION	afleiding
WORDGROUP	bv. <i>ten einde, in natura</i>
ZERO-DERIVATION	bv. verbaalsubstantieven als <i>loop</i>

=OMBI=

*

veldnaam:	morphotype
waarde:	<i>zie boven</i> XCOMPOUND is opgenomen als COMPOUND ZERO-DERIVATION is opgenomen als DERIVATION
scherm:	morphology

2.4.2 MORPHOSTRUCTURE

Al naar gelang van het morfologische type wordt een bepaalde stringwaarde ingevuld als waarde. Zo b.v.:

WOORD	MORPHOTYPE	MORPHOSTRUCTURE
<i>pluspunt</i>	compound	'[plus]punt'

de waarde van een compound is het item opgedeeld in een head en een modifier (= tussen vierkante haakjes). Samenstellingen worden niet recursief gedefinieerd.

Nog enkele voorbeelden:

<i>poging</i>	DERIVATION	'pogen [*ing]'
<i>herschikken</i>	DERIVATION	'[her*]schikken'
<i>zelfmoordpoging</i>	COMPOUND	'[zelfmoord]poging'
<i>rozengeur</i>	COMPOUND	'[roos]<en>geur'

=OMBI=

niet opgenomen

2.4.3 PLURAL

-- waarden: string (= de meervoudsvorm) of NIL (indien er geen meervoudsvorm bestaat), voornaamste bron: GB

-- het feature kan eenmaal herhaald worden; twee mogelijke vormen kunnen worden opgegeven

=OMBI=

veldnaam:	flections-plural1, plural2
waarde:	<i>zie boven</i> NIL is opgenomen als -
scherm:	morphology
bijzonderheden:	alle meervouden die niet gevormd worden door -en, -s of -'s achter het woord te zetten worden als onregelmatig gemarkeerd

2.4.4 PRAGMATIC CONSTRAINTS PLURAL

De beperkingen kunnen liggen op het vlak van :

- STYLE
- SUBJECT FIELD

- CHRONOLOGY
- GEOGRAPHY
- FREQUENCY

Voor waarden: zie PRAGMATICS 2.7

=OMBI=

Niet opgenomen

2.5 SYNTAX

Bij de nomina gaat het om de volgende vijf features:

- GENDER
- ARTICLE
- NUMBER
- NUMBERCONCORD
- COMPLEMENTATION

De features GENDER, ARTICLE en COMPLEMENTATION zijn bij alle woorden ingevuld, de features NUMBER en NUMBERCO zijn alleen ingevuld bij de woorden uit het VU-basisvocabularium en niet ingevuld bij de nieuwe woorden.

2.5.1 GENDER

Staat hier voor pronominale referentie of voornaamwoordelijke aanduiding, er worden thans acht waarden onderscheiden n.l.:

M	referentie d.m.v. hij/hem/zijn
F	zij/haar
MF	hij/hem/zijn of zij/haar (b.v. `schaar')
N	het/zijn
MFN	hij/hem/zijn of zij/haar of het/zijn (b.v. `kind')
MN	hij/hem/zijn of het/zijn (b.v. `joch')
FN	het/zijn of zij/haar (b.v. `meisje')
M/F	hij/hem/zijn of zij/haar (b.v. `dokter'), alle levende wezens, tenzij evident of M of F
n.a.	not applicable, bv. bij pluralia tantum

Bij M/F betreft het de zgn. `gemeenkunnige namen': hier beslist de sekse. Verder moge het duidelijk zijn dat de features ARTICLE en GENDER wel gerelateerd zijn maar elkaar niet volledig overlappen.

Tenslotte, dit feature wijkt soms af van de meeste lexicografische bronnen. Zo b.v.:

	GVD-12	GB	RBN
burenhulp	v (m)	de ¹⁶	MF

¹⁶ Wordt gebruikt ter aanduiding van wat vroeger ml. en vrwl. kreeg met alle nuancerings van dien, b.v.

burgemeester m de (m.) M/F

MF in RBN betekent voor de meerderheid van de Nederlandstaligen M (in feite praktisch alleen M in Nederland, in Vlaanderen F met hoe langer hoe meer M (door het verdwijnen van de dialecten waar het genusonderscheid nog heel sterk aanwezig is). Er zijn gevallen waarbij het GB in afwijking van Van Dale bij de-woorden M geeft en niet MF. In deze gevallen was het GB de richtlijn.

=OMBI=

veldnaam: flexions-gender
waarde: *zie boven*
scherm: morphology

2.5.2 ARTICLE

bepaalt keuze van het lidwoord:

DE (= de-woorden)

HET (= het-woorden)

DE/HET (zowel de- als het-woord)

NARTICLE (= geen lidwoord)

Bij twijfel werd GB e/o Corpus geraadpleegd.

=OMBI=

veldnaam: flexions-article
waarde: *zie boven*
 NARTICLE is opgenomen als -
scherm: morphology

2.5.3 NUMBERCONCORD

Bepaalt de overeenkomst in getal tussen het naamwoord als onderwerp en de pv. Waarden: NORMCONC (i.e. indien Nounsubject = sg dan pv = sg, indien Nounsubject = plur dan pv = plur); SINGPLUR (b.v. bij 'aantal', 'hoop' gevolgd door postmodnoun in het mv. kan pv. in enk. of mv. staan).

=OMBI=

niet opgenomen

2.5.4 NUMBER

Getal; in feite een morfologische categorie van het trefwoord met syntactische implicaties; waarden: LEXSING b.v. 'ziekte'; LEXPLUR b.v. 'mazelen'.

=OMBI=

niet opgenomen

m(v), v en m, v(m)

2.5.5 COMPLEMENTATION

De volgende complementeerwaarden die elkaar onderling niet uitsluiten (tenzij NCOMPLEM), zijn mogelijk bij nomina:

- NCOMPLEM (= de defaultwaarde)

geen complementatie

- PSMODNOUN

duidt aan of een noun al of niet een nabepaling d.m.v. een ander noun kan krijgen. Zo b.v. 'aantal', 'groep', 'ruiker', 'zwerm', 'fles' (die in feite allemaal partitieve nomina zijn)

- FACTIVE

duidt nouns aan die een zinscomplement kunnen krijgen ingeleid door de dat-complementeerder (niet te verwarren met de relatief-dat, cf. de ambiguïteit van 'het feit dat hij bekend heeft, komt mij vreemd voor'). Zo heeft 'probleem' als waarde FACTIVE vanwege 'het probleem dat we niet genoeg koffie in huis hebben', niet echter op grond van 'het probleem dat je signaleerde' of 'het is een probleem dat ...'.

- OFCLAUSE

zoals in 'de vraag of ...'

- WHCLAUSE

wh staat hier voor vraagwoord als complementeerder, zo b.v.: 'de vraag hoe/wanneer/waarom' enz.

- TOINF

complement met 'te'-infinitief (met de exclusie van 'om'; zie hierna), b.v. 'de bewering te zullen komen'

- OMTOINF

complement met 'om te' + infinitief waarbij de 'om' een optionele alternantie is.

'Bewering' zal op die manier onder to + inf-complementering vallen ('de bewering te komen' laat de alternantie 'de bewering om te komen' niet toe), 'poging' daarentegen wordt bij OMTOINF gerangschikt (cf. 'de poging (om) te ontsnappen').

OMTOINF-complementering is niet te verwarren met constructies als: 'een boek om te lezen', 'een vrouw om van te houden' e.d., waar 'om te' niet met 'te' kan alterneren. In deze gevallen is het noun tevens object van de inf op een lager niveau (een zgn. doel-clause).

- FIXPREP

het gaat om de preps als inleiders van vaste (hoewel niet noodzakelijke) voorzetselbepalingen bij nomina, de noun zelf functioneert als head van de NP waarvan de PP deel uitmaakt. Het gaat bij fixed preposition om vaste (i.e. niet te voorspellen) voorzetsels en niet b.v. om attributieve bepalingen (dus wel 'een gevoel van angst', niet 'de fiets van vader', wel 'de grens tussen B. en N.', niet 'de grenzen van B.'. Als de 'van'-PP het eerste argument (agent, bezitter etc) introduceert, wordt hij niet opgenomen. Discontinue constructies zoals b.v. 'de grens tussen ... en ...' worden als 'tussen' aangeduid (het tweede lid kan wegvallen: de grens tussen twee buurlanden).

Dit is een belangrijk feature waarbij zoveel mogelijk ook het Corpus werd gebruikt. Uiteraard is het zo dat wat niet in het Corpus voorkomt, niet impliceert dat het niet zou kunnen voorkomen.

=OMBI=

niet opgenomen

2.6 SEMANTICS

2.6.1 REFERENCE

Er wordt onderscheid gemaakt tussen nouns met een unieke referentie die de waarden PROPER (= eigennaam) krijgen, en deze met een niet-unieke referentie, zij krijgen de waarde COMMON.

Eigennamen hebben i.t.t. soortnamen, een unieke referentie: ze identificeren één entiteit. In de regel worden ze met een hoofdletter geschreven. Zo zijn namen voor politieke partijen of andere groeperingen, namen van personen, landen, steden, rivieren e.d. eigennamen. Ook namen van feestdagen i.t.t. dagen van de week (b.v. Witte Donderdag tgo. donderdag) en van talen beschouwen we als eigennamen.

=OMBI=

*

veldnaam:	reference
waarde:	<i>zie boven</i>
scherm:	semantics

2.6.2 COUNTABILITY

Zes waarden zijn mogelijk t.w.:

- COUNT

nouns die in het meervoud kunnen voorkomen; voorafgegaan door numerieke determinatoren krijgen zij een meervoudsuitgang; in het enkelvoud worden zij in de regel door een determinator voorafgegaan, (b.v. 'thema')

- UNCOUNT

nouns die niet in het meervoud kunnen voorkomen (b.v. 'haat'); het onderscheid COUNT/UNCOUNT kan betekenisgebonden zijn en wordt dan ook in dergelijke gevallen per betekenis anders gespecificeerd (bv. 'liefde' COUNT en 'liefde' UNCOUNT) (zie ook verder bij COLL)

- COUNT/UNCOUNT

nouns die in één lezing zowel COUNT als UNCOUNT zijn (bv. 'onweer')

- MASS

UNCOUNT nouns die, wanneer ze naar een bepaald type/soort/merk/hoeveelheid e.d. verwijzen, in het meervoud kunnen voorkomen b.v. 'thee', 'koffie', 'bier', 'olie'.

- PLURTANT

pluralia tanta. Dit zijn nouns die qua morfologische vorm meervoudig zijn (b.v. 'thermen', 'graffiti') met alle implicaties m.b.t. NUMBERCO en pronominale referentie vandien.

- COLL

collectiva of verzamelnamen. Hiermee worden aangeduid de nouns die

a) UNCOUNT zijn en

b) een groep of verzameling aanduiden.

'Groep', 'bende', 'kudde' (vaak collectiva genoemd) hebben dus de waarde COUNT. 'Vee' heeft de waarde COLL, 'thematiek' UNCOUNT en 'politie' COLL. Het semantische criterium is hierbij dus ondergeschikt aan het morfo-syntactische. Het semantische type (= de selectie-restrictie) bij COLL wordt bepaald door de elementen van de verzameling (dus 'vee' =

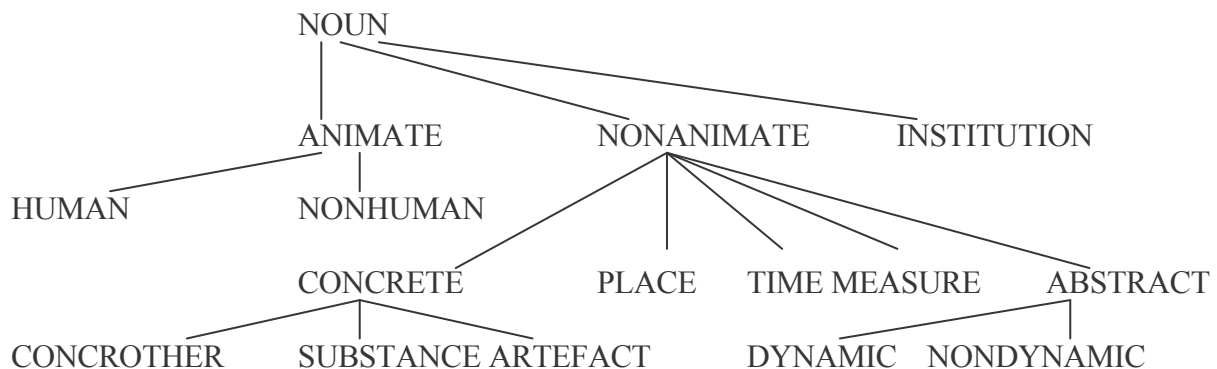
NONHUMAN).

=OMBI=

veldnaam: syntactic subcategory
waarde: *zie boven*
scherm: syntax

2.6.3 SEMANTIC TYPE

Er is van een beperkte verzameling markers (in feite waarden voor selectierestricties) uitgegaan zoals uit de hieronderstaande boomstructuur blijkt.



De set van markers is beperkt gehouden omdat a) met deze beperkte verzameling al heel wat kan worden bereikt (b.v. (in)compatibiliteit vaststellen tussen de nounwaarden en de waarden die het caseframe van het werkwoord oproept) en b) een en ander toch in een semantische representatie verder uitgewerkt en gespecificeerd moet worden (cf. ook infra). Een verklaring van eventueel minder bekende en/of problematische waarden volgt:

- ANIMATE

verwijst naar het al of niet levend zijn, b.v. 'advocaat' ANIMATE; wordt, zoals in het geval 'advocaat', ook een beroep aangeduid, dan wordt dat betekenisaspect in de definitie e/o het commentaar opgevangen.

- INSTIT

duidt instellingen en groepen aan die als Agent kunnen optreden b.v. 'EEG', 'de Belgische staat', 'het Rijk', 'de vakgroep' enz. Noteer dat b.v. plaatsnamen als 'België', 'Den Haag' enz. kunnen verschuiven van PLACE naar INSTIT. Een dergelijke verschuiving wordt in principe door metalexicaal regels opgevangen of wordt reeds in het SHIFT-veld aangegeven (zie verderop).

- MEASURE

duidt eenheden van lengte, gewicht, afstand, geld enz. aan (bv. 'kilometer', 'euro')

- CONCRETE: <--> ABSTRACT

nouns die niet de waarden ANIMATE, INSTIT, PLACE, TIME, MEASURE konden krijgen en als filler van het slot 'dit/deze ... is/zijn kleurloos/rood/blauw enz.' voorkomen, kregen de waarde CONCRETE, de overige de waarde ABSTRACT. 'Water' en 'borrel' zijn dus concreet, 'idee' en 'enthousiasme' abstract.

---binnen de concreta onderscheiden we nouns die een substantie aanduiden ('water/zeep/goud/glas/papier') tgo. nouns die (menselijke) artefacten (= kunstmatig vervaardigde objecten) aanduiden zoals 'fiets', een 'glas', 'stoel', 'paperclip' enz. Artefacten zijn dus altijd COUNT, bij SUBSTANCE maken we echter geen onderscheid tussen zgn. stofnamen (b.v. 'water'), 'natural kinds', en door de mens gemaakte substanties (b.v. 'bier'). 'Koekje', 'borrel' echter zijn door de mens vervaardigde objecten en dus artefacten. Hiermee zijn de concrete nouns zeker niet uitputtend behandeld. 'Sinaasappel' b.v. valt noch onder SUBSTANCE, noch onder ARTEFACT; voorlopig krijgen dergelijke nomina de waarde CONCROTHER. Zo wordt CONCROTHER b.v. ook gebruikt voor lichaamsdelen.

---binnen de abstracta onderscheiden we nouns die kunnen optreden in constituenten ingeleid door 'tijdens/na/gedurende/voor' enz. Die krijgen de waarde DYNAMIC (ze veronderstellen een zeker gebeuren of moment), de andere krijgen de waarde NONDYNAMIC. Concreta kunnen ook vaak in het slot van de dynamische abstracta optreden maar moeten dan als verkortingen worden aangezien: b.v. 'na de tweede borrel' = na het drinken van de tweede borrel. Voorbeelden: DYNAMIC: 'collecte', 'oratie', 'blikseminslag', NONDYNAMIC: 'haat', 'eigenschap', 'schoonheid'.

Vooralsnog zijn alleen de terminale waarden in het lexicon opgenomen; CONCRETE, ABSTRACT, ANIMATE en NONANIMATE kunnen overgeërfd worden.

=OMBI

*

veldnaam:	semantic type
waarde:	<i>zie boven</i>
scherm	semantics

Voor het Nederlands zijn er geen woordenboeken voorhanden die waarden voor het semantische type en voor de syntactische subcategorisatie (complementatie) bevatten. Met de opgave van deze informatie beoogt het RBN dan ook een reële meerwaarde te creëren. Bij het vaststellen van de waarden voor deze categorieën heeft het Corpus en de competentie van de redacteuren een belangrijke rol gespeeld.

Inmiddels is voor RBN-klein (zie 1.5.2) de semantische typologie verder uitgebreid. Binnenkort zal deze uitbreiding ook in het RBN geïmporteerd worden. De categorie HUMAN b.v. bestaat dan uit 17 subcategorieën, o.m. beroepsnaam, familienaam, verwantschapsnaam, naam voor inwoners uit bepaald land/stad/gebied/behoorend tot bepaald volk etc.

2.6.4 SHIFTS

Bij de nomina is er bij voorkeur een keuze gemaakt uit 29 shifts. Neutrale default is NOSHIFT. Een shift voorspelt een verschuiving in betekenisaspect van het opgegeven SEMANTIC TYPE naar een of meer van de andere TYPES, bv. DYNAMIC (proces) --> NONDYNAMIC (resultaat). Zo bv. 'spreiding' is zowel DYNAMIC als NONDYNAMIC. Als het hier om twee betekenisaspecten gaat, niet om twee verschillende betekenissen, dan wordt als SEMANTIC TYPE de basiscategorie ingevuld en als SHIFT de waarde 'Shift 5'. SHIFTS gaan dus uit van een basisbetekenis. Bij het vaststellen van deze basisbetekenis spelen o.m. factoren als frequentie, psychologische relevantie (de eerste betekenis die door proefpersonen out-of-context wordt gegeven), systeemdwang (bv. er kan wel van Artefact --> Place, niet van Place --> Artefact geshift worden), volgorde-regels (bv. wat geïmpliceerd wordt, gaat vooraf aan wat impliceert) e.d. een rol.

Overzicht van semantische shift met voorbeelden

1	Artefact	--> Place		<i>fabriek</i>
2	Artefact	--> Place	--> Institution	<i>school</i>
3	Institution	--> Place		<i>secretariaat</i>
4	Institution	--> Place	--> Artefact	<i>ambassade</i>
5	Dynamic	--> Nondynamic		<i>spreiding</i>
6	Nondynamic	--> Dynamic		<i>etappe</i>
7a	Nondynamic	--> Artefact		<i>contract</i>
7b	Nondynamic	--> Concrother		<i>gunstkoopje</i>
8a	Artefact	--> Nondynamic		<i>machtiging</i>
8b	Concrother	--> Nondynamic		<i>geraamte</i>
8c	Substance	--> Nondynamic		<i>melange</i>
9	Measure	--> Artefact		<i>dirham</i>
10	Measure	--> Time		<i>lichtjaar</i>
11a	Nonhuman	--> Substance		<i>anijs</i>
11b	Nonhuman	--> Concrother		<i>biet</i>
11c	Nonhuman	--> Human		<i>mormel</i>
12a	Human	--> Nonhuman		<i>scharminkel</i>
12b	Human	--> Artefact		<i>schoonheid (2)</i>
12c	Human	--> Place		<i>kruidenier</i>
12d	Human	--> Nondynamic		<i>gigant</i>
12e	Human	--> Institution		<i>commando</i>
13a	Dynamic	--> Human		<i>bijslaap</i>
13b	Dynamic	--> Nondyn	--> Artefact	<i>reconstructie</i>
13c	Dynamic	--> Nondyn	--> Concroth	<i>slijtage</i>
13d	Dynamic	--> Place		<i>start</i>
13e	Dynamic	--> Human	--> Institution	<i>verpleging</i>
14	Nondynamic	--> Human		<i>beta</i>
15	Institution	--> Human		<i>marechaussee</i>
16	Artefact	--> Human		<i>marionet</i>

=OMBI=

*

veldnaam:

shift

waarde:

zie boven

scherm

semantics

2.6.5 RESUME

Waarde = string die een verkorte definitie weergeeft. B.v. onweer RESUME = `weer met bliksem en donder'. Resumés worden in eerste instantie ter betekenisonderscheiding bij polysemen gebruikt en zijn heterogeen van aard. Dit betekenisonderscheid kan nl. op de volgende wijze worden weergegeven:

- synoniemen: fiets = rijwiel
- korte betekenisomschrijving: bungalow 1 = vrijstaand huis
 bungalow 2 = zomerhuisje
- selectier restrictie: bundeling 1 = van krachten
 bundeling 2 = van geschriften
- intralexematisch label: sleutel 1 = werktuig m.b.t. een slot
 sleutel 2 = figuurlijk gebruik van sleutel 1
- domeinlabel: sleutel 1 = werktuig m.b.t. een slot
 sleutel 3 = muziek

Wat de weergave d.m.v. selectier restricties betreft: subject restricties worden weergegeven d.m.v. `van ...!', object restricties d.m.v. `m.b.t. ...!'. Ook bij monosemen moet een betekenisresumé opgegeven worden. De maximale lengte van het resumé is 40 karakters.

=OMBI=

veldnaam: resume
waarde: zie boven
scherm: semantics

2.6.6 DEFINITION

Waarde = string die definitie bevat waarvan resumé kan worden afgeleid; vorm (voor nomina) is een analytische definitie met het hyperoniem (genus proximum) en differentiae specificae, die gescheiden worden opgevoerd.

Aldus:

onweer	GENUS	PROX.	weer
	DIFF	SPEC	dat gepaard gaat met donder en bliksem

In tegenstelling tot resumés kunnen definities

---wel letterlijk uit een bron zijn overgenomen (met bronvermelding) of door een van de redacteuren zijn aangepast; in dat geval staan er initialen als bron

---niet door synoniemen worden weergegeven. Voor nomina gelden alleen analytische definities.

Nominalisering op -ing e.d., zoals b.v. `bundeling', zijn weergegeven

als:	GENUS PROXIMUM	het bundelen
	DIFF. SPECIF.	van krachten

waarde:	<i>zie boven</i>
scherm:	semantics
bijzonderheden:	OMBI biedt de mogelijkheid synoniemen met verwijzingen op te nemen in het scherm "related lexical units", ook andere semantische relaties zijn mogelijk

2.7 PRAGMATICS

In het RBN worden acht pragmatische features onderscheiden: in feite gaat het om zeven kenmerken plus frequentie die vaak het resultaat is van (een aantal van) deze kenmerken.

2.7.1 ORIGIN

Het gaat hierbij om

- a) een aanduiding van nog duidelijk herkenbare leenwoorden; zij krijgen als waarde de brontaal waaruit zij ontleend werden, b.v. `label' ORIGIN: ENGLISH, en
- b) een aanduiding van ontleningen aan andere talen die als zodanig in strijd zijn met het Nederlandse taaleigen b.v. `zich interesseren aan' (Gallicisme). Origin krijgt dus de structuur: PRAGM ORIGIN ENG(LISH), FR(ENCH), GER(MAN), enz., NORIGIN (= inheems, de default-waarde), GALL(ICISM), ANGL(ICISM), GERM(ANISM) enz.

In dit feature gaat het NIET om etymologische/diachrone informatie, ook al zijn GREEK en LATIN in de waarden opgenomen. Voorbeelden van synchroon herkenbare (fonetisch/orthografisch) vreemde woorden: `jazz' ENG, `feta' GREEK, `belladonna' ITALIAN.

=OMBI=
niet opgenomen

2.7.2 STYLE

Met `style' wordt bedoeld de graad van formaliteit die een woord(betekenis) heeft. Er worden 5 waarden op deze formaliteitsschaal geplaatst, t.w.

FORMAL --- NSTYLE --- INFORMAL --- SLANG --- VULGAR

Deze waarden corresponderen met de 5 stijllagen zoals die in de Van Dale-vertaalwoordenboeken en de HVD (= Hedendaagse Van Dale) worden onderscheiden. Aldus:

- FORMAL
schrijftalig, (zeer) formeel, literair, zou in dagelijkse conversatie erg opvallen c.q. potsierlijk aandoen, b.v. `sponde', `ega'.
- NSTYLE
ongemarkeerd, de default-waarde, gewone spreek- en schrijftaal, b.v. `bed', `vrouw', `stelen'.
- INFORMAL
informele spreektaal; gebruik hiervan in geschreven taal, die niet uitdrukkelijk de spreektaal wil weergeven, zou opvallen, b.v. `jatten'.
- SLANG
slang, altijd informeel, vaak aan sociale context of groep gebonden, b.v. `gaaf'.
dit is in het RBN: STYLE: SLANG en SOCIAL: CHILD
- VULGAR
uitgesproken vulgair of grof; werkt vaak choquerend, b.v. `zeiken', `kut'.

=OMBI=

veldnaam: pragmatics- style
waarde: *zie boven*
de neutrale waarde NSTYLE is niet opgenomen
scherm: semantics

2.7.3 CONNOTATION

Met dit feature worden gefixeerde emotionele waarden aangeduid. Dit emotionele spectrum beweegt zich tussen een positieve en een negatieve pool. De volgende gevoelsnuanceringen zijn gebruikt:

- OFFENS

beledigend, alleen voor personen gebruikt
(cf. scheldnamen), b.v. 'nikker'.

- PEJOR

pejoratief, met negatieve gevoelsnuance, b.v. 'vertrossen'.

- EUPHEM

eufemistisch, b.v. 'zwangerschapsonderbreking' voor 'abortus', '65-plusser'.

- JOCULAR

schertsend, grappig, b.v. 'gemeentepils' voor 'leidingwater'.

De default-waarde is 'unmarked' (= NCONNOT) en komt uiteraard bij de meeste woorden/betekenisissen voor.

Bij OFFENS (= beledigend) geldt, net zoals overigens voor EUPHEM (= eufemistisch) en JOCULAR (=schertsend), dat voor deze items telkens ook neutrale denotaties bestaan. Naast 'nikker' is er 'neger', naast 'club' 'bordeel', naast 'kankerstokje' 'sigaret'. Bij PEJOR (= pejoratief) ligt dat wat anders: cf. BVD 'muzak' = lichte achtergrondmuziek <gezegd door mensen die daar niet van houden>. Komt een dergelijk subjectief emotioneel negatief gebruik voor dan gebruiken wij pejoratief. Het gaat hier dus om een (subjectief) negatieve evaluatie van een objectief gegeven.

Beledigend (OFFENS) is alleen voor personen gebruikt en komt ipso facto bij items die als scheldnamen gebruikt worden voor. Dat negatieve waarden als OFFENS en PEJOR wel gemarkeerd zijn en positieve waarden b.v. minder, heeft verder te maken met het feit dat het hier in de eerste plaats om restrictieve labels gaat: iemand een 'crack' noemen schaadt niemand, iemand een 'krenge' noemen daarentegen wel ...

=OMBI=

veldnaam: pragmatics- connotation
waarde: *zie boven*
de neutrale waarde NCONNOT is niet opgenomen
scherm: semantics

2.7.4 SOCIAL GROUP

Dit feature duidt de beperking aan van taalgebruik tot bepaalde groepen van taalgebruikers;

voorkomende waarden (o.m.):

- GENSOC = ongemarkeerd, defaultwaarde
- CHILD = kindertaal, jongerentaal
- SOLD = soldatentaal
- STUD = studententaal

Ook woorden/betekenis die gerestringeerd zijn tot religieuze of kerkelijke groeperingen ressorteren hieronder, aldus: 'eucharistieviering' ROMCATH.

=OMBI=

veldnaam: pragmatics- soc. group

waarde: *zie boven*

de neutrale waarde GENSOC is niet opgenomen

scherm: semantics

2.7.5 SUBJECT FIELD

Dit feature is gebruikt om het vakinterne gebruik van woorden/betekenis weer te geven (in principe gaat het dus om items die alleen door vaklui gebruikt worden; 'roos' b.v. krijgt de waarde PLANTK GEN, 'belladonna' krijgt de waarde PLANTK(UNDE)).

Naast de defaultwaarde (= GENSUBJF) worden de namen van vakgebieden als waarden gebruikt. GEN is aangekruist als het woord tussen vakintern (= zonder GEN) en algemeen gebruik in ligt. N.B.: het gaat hier om een indeling naar pragmatische, niet naar semantische criteria.

Samengevat zijn er dus drie waarden mogelijk:

▼ GENSUBJ:

lexemen die niet tot een specifiek vakgebied gerekend kunnen worden, tot de algemene taal behoren

▼ NAAM VAKGEBIED + GEN:

vakexterne lexemen; deze behoren tot een vakgebied maar zijn doorgedrongen in de algemene taal

▼ NAAM VAKGEBIED:

vakinterne lexemen; worden alleen gebruikt in de communicatie waarbij (een of meer experts betrokken zijn (expert - expert van hetzelfde of verwant vakgebied; expert -- leek: deze laatste situatie kan ook aanleiding geven tot vakextern gebruik)

De volgende waarden/gebieden/namen voor vakgebieden komen voor. Deze lijst was voor de redacteurs in principe uitbreidbaar, er kunnen dus ook andere namen voorkomen.

RBN – waarden voor Subjectfield

Beeld(Ende) K(Unsten
Bouwk(Unst
Dansk(Unst
Dram(Atiek, -Aturgie, Theater
Kunst
Druk(Kunst
Film(Kunde
Grafi(Ische Kunst)
Lit(Eratuur)
Muz(Iek)
Voetbal
Wielrennen
Paardensp(Ort)
Zwemsport
Tennis
Sport
Spel
Computer
Elek(Triciteit)
Foto(Grafie)
Landmeetk(Unde)
Lucht(Vaart)
Mijnw(Ezen)
Ruim(Tevaart)
Ind(Ustrie)
Landb(Ouw)
Veet(Eelt)
Scheep(Vaart/Bouw)
Spoorwegen
Typografie
Weg (-) Enwaterb(Ouw)
Anatomie
Aardrijkskunde
Alch(Emie)
Antropologie
Bioch(Emie)
Dierk(Unde)
Farmacie
Fil(Osofie)
Geol(Ologie)
Logica
Med(Iejinen, Geneeskunde)
Meteo(Rologie)
Nat(Uurkunde)
Plantk(Unde)
Psych(Ologie)
Schei(Kunde)
Sociol(Ologie)
Ster(Rekunde)
Taalk(Unde)
Theol(Ogie)

Wetenschap
Wisk(Unde)
Adm(Inistratie)
Ambacht(Elijk)
Astr(Ologie)
Bijb(El)
Com(Municatie, Media)
Conf(Ectie)
Cul(Inaria)
Dipl(Omatie)
Econ(Onomie)
Folk(Lore)
Geneal(Ogie)
Geldw(Ezen)
Gesch(Iedenis)
Arch(Eologie)
Handel
Herald(Iek)
Journal(Istiek)
Jur(Idisch, Recht)
Pol(Itiek)
School(Wezen, Onderwijs)
Techn(Iek)
Textiel
Verzek(Eringen)
Visserij
Mil(Itair, Leger)
Verkeer(Swezen)

=OMBI=

veldnaam:

pragmatics- subject

waarde:

zie boven

de neutrale waarde GENSUBJ is niet opgenomen

VAKGEBIED + GEN (markering van de vakexterne lexemen) is niet opgenomen

scherm:

semantics

2.7.6 CHRONOLOGY

Dit feature situeert elk item op een tijdsas. Waarden:

- SYNCHRON (= de defaultwaarde)
- OLDFASH verouderd, b.v. `diender' voor `politieagent'.
- NEOLOG voor neologismen, zo b.v. modieus aandoend taalgebruik als: 'bloedgroep' in politieke betekenis (dus niet zozeer nieuwe begrippen, conceptuele neologismen, als wel stilistische neologismen)
- OLDFASH wordt gegeven als het woord verouderd is; of het concept oud is of niet, is niet relevant, bv. 'diender' OLDFASH, maar 'mammoet' SYNCHRON

Neologismen worden slechts sporadisch opgegeven.

=OMBI=

veldnaam:

pragmatics- chronology

waarde: *zie boven*
de neutrale waarde SYNCHRON is niet opgenomen
scherm: semantics

2.7.7 GEOGRAPHY

Dit feature situeert items binnen een regio. Drie waarden zijn nu toegepast:

- GENGEO (Algemeen Nederlands; defaultwaarde)
- BELG vnl. namen van officiële instellingen, beroepen ed. die alleen in België gebruikt worden, b.v. `schepen' in België tgo. `wethouder' in Nederland
- DIALECT dialectaal taalgebruik, voor zover opgenomen

Voor de opname van Belgisch Nederlands en voor de afstemming hierbij met het RBBN (Referentiebestand Belgisch Nederlands) zie 1.3.5 f.

=OMBI=

veldnaam: pragmatics-geography
waarde: *zie boven*
de neutrale waarde GENGEO is niet opgenomen
scherm: semantics

2.7.8 FREQUENCY

Bij de woorden uit het oorspronkelijke VU-basisvocabulary zijn de waarden BASIC (neutraal), BASICF (hoogfrequent) en RARE ingevuld. Bij de nieuwe woorden waren de waarden NFREQ/NEUTRAL (neutraal; default) en RARE beschikbaar. Nu ingevulde waarden geven slechts een indicatie, geen precieze aanduiding van de frequentie.

=OMBI=

veldnaam: pragmatics- frequency
waarde: *zie boven*
de neutrale waarde NFREQ/NEUTRAL is niet opgenomen
scherm: semantics

2.8 COMBINATORICS

Hieronder vallen zowel vrije als vaste verbindingen. Wat de eerste betreft zijn alleen frequente vrije verbindingen opgenomen, wat de laatste aangaat: het gaat hier zowel om idiomatische als om niet-idiomatische of minder-idiomatische verbindingen. Elke verbinding is als een afzonderlijke eenheid opgenomen en is in de database verbonden aan het entry van het woord waar de verbinding bijhoort. De vaste verbindingen die een nomen bevatten (Adj + Nomen, Verbum + Nomen e.d.) worden in de regel onder het Nomen behandeld. Zo zal 'lange vingers hebben' niet onder 'lang' te vinden zijn, wel onder 'vinger'. Wel kunnen verwijzingen voorkomen (b.v. van 'lang' naar 'vinger'). Ook kunnen, ter illustratie, vrije voorbeelden onder het woord waarvan het semantische/syntactische gedrag geïllustreerd wordt, worden opgevoerd (b.v. 'aanhoudende droogte' onder 'aanhouden'). Zie ook 2.8.2.2 onder lexicale collocaties. Een entry kan meerdere verbindingen/voorbeelden dragen of onder zich hebben, vaak zijn het er

meer dan één. Het RBN heeft ernaar gestreefd een rijk bestand te zijn met veel clichés en contextuele voorbeelden gebaseerd op een Corpus (38Mi Corpus INL)(zie Kruijt, o.c., p. 8) Het combinatorics-gedeelte is bedoeld zowel ter illustratie van syntactisch, collocatieel, als van semantisch gedrag. Er is naar gestreefd een en ander op een zo systematisch mogelijke wijze weer te geven dit in tegenstelling tot de meeste bestaande mono-en bilinguale woordenboeken.

Met name de aanpak waarbij de relaties tussen de delen van een collocatie op een systematische wijze worden weergegeven, maken het RBN niet alleen uiterst geschikt voor lexicografische doeleinden (verklarend, vertalend, collocatieel woordenboek), maar tevens voor taaltechnologische zowel bij zoek- als bij productiefuncties. Zo b.v. is 'boeken' als collocator gebruikelijk bij 'charter' en is de relatie met dit laatste OPER1, bij 'overwinning' daarentegen is 'boeken' = REAL. Op die manier wordt 'boeken' zowel gedesambigueerd als correct gebruikt.

2.8.1 IDENTIFICATION

- CAN(onical) FORM:
waarde = string, de standaardvorm waarin de verbinding voorkomt

- TEXT(ual) FORM:
ter illustratie een gecontextualiseerde versie van de canform

- CATEGORY:
waarden zijn S (voor zin) en constituenten van S als NP, VP e.d.

- NIL
NIL wordt aangekruist als de opgegeven verbinding niet bij de betekenis van het entry hoort, maar nergens anders past.

Bij vrije verbindingen wordt altijd minstens één vorm opgegeven en in de CANFORM-balk ingevuld ook al is het een gecontextualiseerd voorbeeld. Bij spreekwoorden en pragmatische formules (cf. infra) zijn de canonieke vormen in de regel gecontextualiseerd: werkwoordsvormen staan in de finiete vorm bv. Denk aan spreekwoorden als: 'de appel valt niet ver van de stam' of pragmatische formules als 'zo zie je maar', 'noem maar op' etc. Canonieke en gecontextualiseerde vormen vallen hier a.h.w. samen. De vuistregel is dus:

-- CAN(onical) FORM: is altijd aanwezig; bij idioms, , spreekwoorden, pragmatische formules en vrije voorbeelden is dat in de regel een gecontextualiseerde vorm; bij collocaties gaat het wel degelijk om een canonieke vorm

-- TEXT(ual) FORM: komt alleen voor als het gaat om een gecontextualiseerde vorm, afwijkend van wat in de canform-balk staat; dit komt meestal voor bij collocaties en is ook mogelijk bij idioms, en vrije voorbeelden

Alle verbindingen staan zonder hoofdletters en punten, tenzij eigennamen e.d.. Met / (slash) kunnen varianten zijn opgevoerd; met [] en () kunnen er in de CANFORM variabelen zijn aangegeven:

[] = invulling op deze plaats is verplicht, maar variabel
() = invulling op deze plaats is NIET verplicht en variabel

=OMBI=
veldnaam:

canonical form

waarde:	<i>zie boven</i>
scherm:	hoofdscherm en example-syntax
veldnaam:	textual form
waarde:	<i>zie boven</i>
scherm:	example-syntax
veldnaam:	category
waarde:	<i>zie boven</i>
scherm:	hoofdscherm en example-syntax
bijzonderheden:	

- als alleen een textual form is opgegeven is die in OMBI als canonical form weergegeven; de category is dan niet opgegeven
- [] haakjes zijn verwijderd; het example type is dan *Free* (zie onder)
- NIL: idiom is in OMBI weergegeven als IDIOM op het hoofdscherm

2.8.2 SYNTAX

2.8.2.1 TYPE

= het type verbinding: FREE of FIXED.

=OMBI=

*

veldnaam:	example type
waarde:	<i>free</i>
scherm:	example-syntax
bijzonderheden:	De combinaties van het type <u>free</u> zijn met name opgenomen in OMBI als hulpmiddel voor de woordenboekmaker. De voorbeelden zijn over het algemeen afkomstig uit een corpus en bieden informatie over het gebruik van het woord in context. Het gaat om zgn. <u>pregnante</u> voorbeelden: voorbeelden die rijk zijn aan conceptuele en/of grammaticale informatie. In veel gevallen zullen zij niet van belang zijn voor de <u>gebruiker</u> van een <u>bilinguaal</u> woordenboek.

2.8.2.2 SUBTYPE

bij FIXED is een van de volgende SUBTYPEs ingevuld:

- grammatical collocation **GRACOL** = verbinding tussen een lexicaal woord (het betreffend entry) en een bepaald functiewoord (b.v. prepositie, conjunctie) dat ofwel verplicht (b.v. 'van' in 'houden van') dan wel optioneel voorkomt ('op' in 'wachten op'). Bv. 'angst voor..', 'in paniek', 'aansluiting op ..'
- lexical collocation **LEXCOL**: a collocation A + B is a phrase S such that to express the meaning of S with the lexeme B the choice of A is lexically determined by B (definitie Melcuk) (b.v. indien lexeem B (base) = vrijgezel, en meaning S = 'in hoge mate' 'vrijgezel', dan lexeem A (collocator) = verstokt). Deze LEXCOL's zijn opgenomen onder de basis van de phrase, NN 'zwerm bijen', AN 'verstokte vrijgezel', VN/NV 'honden blaffen', 'medicijnen innemen' en - buiten de nomina - AdvV 'zich groen en geel ergeren' en AdvA 'diep ongelukkig'. Het gaat om een gerestringeerde verbinding (the choice of A is lexically determined by B).
- **IDIOM**: uitdrukking die niet-letterlijk en niet compositioneel is, de betekenis van het geheel is niet de som van de betekenis van de delen (b.v. het hoekje omgaan). Wordt bij het eerste lexicale woord van de uitdrukking opgenomen, in volgorde eerst nomen, dan adjectief, dan verb.

- **PRAGMA**: een vaste formule die niet idioom is, of een lexicale of grammaticale collocatie; de verbinding heeft een routine-achtige vorm en een pragmatische functie in de context. De volgende communicatieve functies zijn typisch voor PRAGM: opvulling van hiaten, stoplappen 'weet je wel', een inleiding van een conclusie 'goed', 'alles in acht genomen', een waarschuwing 'voor je het weet', een aansporing 'toe nou', een soort van instemming 'daar zeg je wat', een wens 'beterschap', 'mazzel' etc. Het gaat steeds om zg. 'stock phrases', stereotiepe zinswendingen.
- **PROVERB**: spreekwoorden hebben een vaste, syntagmatische vorm en geven qua inhoud een algemene 'volkse wijsheid' weer. Bv. 'zoals het klokje thuis tikt, tikt het nergens', 'oost, west, thuis best' etc.

=OMBI=

*

veldnaam: example type
 waarde: free, lex. collocation, gramm. collocation, pragm, prover
 scherm: example-syntax
 veldnaam: textual form
 waarde: zie boven
 scherm: example-syntax
 veldnaam: category
 waarde: zie boven
 scherm: hoofdscherm en example-semantics
 bijzonderheden:

- als alleen een textual form is opgegeven is die in OMBI als canonical form weergegeven; de category is dan niet opgegeven
- [] haakjes zijn verwijderd; het example type is dan *Free* (zie boven)
- de volgende relatie bestaat tussen *lexicalisation status* en *example type*:

Free example	non-lexic
Lexicale en grammaticale collocatie	semi-lexic
Idioms, pragma, prover	fully-lexic

2.8.2.3 COMBINATIONWORD

= stringwaarde, het gaat om een woord waarmee het entrywoord gecombineerd wordt; twee woorden konden hier worden ingevuld. Het veld wordt bij elke verbinding, vrij of vast, ingevuld, waardoor er dwarsverbanden door het lexicon heen zijn ontstaan. De voorkomende combinatiewoorden per TYPE:

GRACOL: grammaticale woord uit de verbinding, veelal voorzetsel
 LEXCOL: collocator

IDIOM: syntactische structuur van verbinding is bepalend: nomina, adjectiva en verba gaan in die volgorde vóór op functiewoorden; hier worden meestal 2 combi-woorden ingevuld

FREE: meest saillante woord(en) waarmee het entry geïllustreerd wordt

PRAGM: zie IDIOM (vaak zijn echter alle woorden belangrijk)

PROVERB: zie IDIOM

Voorbeeld: IDIOM '(niet) over één nacht ijs gaan'

entry: 'nacht'

combinationwords : 'ijs', 'gaan'

=OMBI=

*

veldnaam: const1
waarde: *zie boven*
scherm: example-constituents
bijzonderheden: alleen opgenomen bij *lex. collocation* en *gramm. collocation*

2.8.2.4 CAT COMBINATIONWORD

de wordcats NOUN, ADJ, PREP, VERB, ADV(erbium), PRON (voornaamwoord), NUM(eral) (ook soms TELW), DET(erminer), en CONJ(unctie) zijn gebruikt.

=OMBI=

niet opgenomen

2.8.3 SEMANTICS

De semantische descriptie van combinaties is als volgt:

2.8.3.1 MEANING DESCRIPTION

Bij idioms, proverbs en pragmatic expressions wordt een definitie gegeven, de zgn. **MEANING DESCRIPTION**. Bij de andere example types komt deze alleen voor als de betekenis ambigu is.

=OMBI=

*

veldnaam: definition
waarde: *string*
scherm: example-semantics

2.8.3.2 MEANING COMPLEMENT

Bij **GRACOL met complementeerder** wordt een **MEANING COMPLEMENT** gegeven, = waarde uit lijst SEMANTIC TYPEs (zie 2.6.3), maar ook hogere knopen in de boom komen voor, evenals meer types tegelijk met slashes ertussen; soms is er niets ingevuld, soms staat er de code ALL (= neutrale, ongespecificeerde waarde)

=OMBI=

*

veldnaam: func1
waarde: *zie boven*
scherm: example-constituents

2.8.3.3 MEANING GRAMWORD

Bij **GRACOL zonder** complementeerder wordt een **MEANING GRAMWORD** (= waarde uit lijst waarden functiewoorden: Time, Location, Circumstance, Manner, Means, Cause, Reason, Origin, Concession, Purpose, Result, Comparison, Restriction en Modality); bijv. 'over twee

weken' Time, 'aan het park' Location, 'in ernst' Manner, 'tot ieders opluchting' Result, 'uit medelijden' Reason, 'onder applaus' Circumstance

=OMBI=

*

veldnaam: func1
waarde: zie boven
scherm: example-constituents

2.8.3.4 MEANING COLLOCATOR

Bij LEXCOL komt er een **MEANING COLLOCATOR**. De gebruikte waarden voor lexicale functies zijn geïnspireerd op Melcuk¹⁷. De prototypische gevallen, zoals de hieronder genoemde voorbeelden, hebben een collocator gekregen. Andere lexicale collocaties die moeilijker interpreteerbaar waren, zijn soms wel, soms niet benoemd. De beschrijving van de lexicale collocaties is een van de meest innoverende kenmerken van het RBN. Het is een rijk potentieel niet alleen voor lexicografische, maar ook voor taaltechnologische toepassingen.

-- Adjectivale relaties/Qualifiers (Adj-Nomen)

-Bonus (een 'goede' x)

een sterke marketing

gezond verstand

een wijs besluit

een gelukkige keuze

een sterke prestatie

-Verus (een 'echte' x zoals die hoort of verwacht wordt te zijn)

een geboren en getogen Amsterdammer

ware doodsverachting

-Magnus (een intensificatie van x)

een daverend applaus

een pril begin

-Qualifier (neutraal, niet verder gespecificeerd)

een grijs kenteken

-Anti (aspectueel: drukt in combinatie met een andere LF het tegenovergestelde uit)

een zwak bod AntiBonus

een vals gebit AntiVerus

- Nominale relaties (Nomen-Nomen)

-Mult (een verzameling van x)

een doos sigaren

een slof sigaretten

-Sing (één instantiëring van x)

een fles wijn

een glas wijn

¹⁷Melcuk, I., e.a. Dictionnaire explicatif et combinatorie du français contemporain, vol 1. 1984.

Gevallen als *Kamer van Koophandel* zijn gecodeerd als FIXED en verder vooralsnog ongespecificeerd gelaten. De COMBIWORDS worden wel opgegeven.

-- Verbale relaties (Nomen-Verb/Verb-Nomen)

Hier maken we onderscheid tussen relaties die

- een **Action** uitdrukken (+ dynamisch, + controle; + dynamisch, impliceert een verandering; test: combinatie met b.v. 'langzaam', 'geleidelijk', 'aan het ..' is mogelijk; + controle veronderstelt een actor die bepalend is voor het al of niet plaatsvinden van de actie en de imperatief is mogelijk)

en

- **geen Action** uitdrukken maar een **State** (- dynamic, +/- control) of **Process** (+ dynamic, - control)

Bij de **Action-verba** maken we gebruik van de volgende LF's (lexical functions):

- Oper1

het verrichten, het doen van of met x; het werkwoord is een louter 'supportverb', semantisch leeg, vaak hebben de naamwoorden een verbale counterpart: *prestatie --> presteren* etc.; ze zijn te parafaseren met 'doen/maken/geven (van/met)'

een prestatie leveren

koffie zetten (koffie drinken beschouwen we als vrije combinatie)

een wens doen

stage lopen

verraad plegen

een schatting maken

de was doen

een verklaring afleggen

weerstand bieden

- Oper2

als Oper1, maar Gram. Subject verb = logisch Object situatie (Patiens of Beneficiary)

een face-lift ondergaan (tgo. een face-lift geven)

aandacht krijgen (tgo. aandacht geven)

op weerstand stuiten (tgo weerstand bieden)

- Real

te vergelijken met Oper-ww, alleen, nu verwijst het werkwoord naar het vervullen van een bepaalde eis of voorwaarde die in het entry geïmpliceerd is, dus parafraseerbaar met het realiseren van de voorwaarde vervat in het entry, verdergaand dan maken/doen/geven

een wens vervullen (doen/uiten Oper1)

een tentamen halen (doen Oper1)

een belofte houden (doen Oper1)

een overwinning behalen

een hypothese bevestigen

een droom realiseren

een diploma halen

de trein halen (nemen Oper1)

-**Liqui** het opheffen/eliminieren (van het functioneren, het bestaan) van x

een brand blussen

dorst lessen

honger stillen
een hypothese verwerpen
een huwelijk ontbinden
weerstand breken

Als tegenhanger voor Liqui kunnen we gebruiken:

-CauseIncepFunct

doen beginnen functioneren (waar Liqui in feite CauseFinFunct betekent);
een commissie instellen, zie verder onder Cause.

Net zoals we bij de adjectivale relaties een niet-verder gespecificeerde Qualifier aantreffen, zullen we ook bij de verbale relaties gebruik maken van een dergelijke neutrale waarde. Hier is dat de waarde **Action**. **Action** is dus voor lexicale functies met een actie-ww gebruikt die niet onder Oper1, Oper2, Real, Liqui of Cause vallen en niet verder gespecificeerd worden.

een wedstrijd spelen Oper1 tgo. *een wedstrijd winnen* Real tgo *een wedstrijd organiseren* Action.

Andere voorbeelden: *je tanden poetsen* Action
je hond africhten Action
een uurwerk opwinden Action

Bij **niet-Action-werkwoorden** (toestanden, gebeurtenissen of processen) zijn de volgende LFs gebruikt:

-Funct

duidt het (proto)typische functioneren of de typische actie van iets aan, vaak dus bij artefacten die in een bepaalde toestand verkeren of een bepaald proces ondergaan of bij natuurfenomenen

een bom explodeert

het hart klopt

de kachel brandt kan hieronder ressorteren of anders als frequente, vrije verbinding worden opgenomen.

-Excess

is het overmatig functioneren van iets

mijn hart bonst

de storm raast/woedt

-Son

min of meer aan Funct verwant, is het typische geluid dat x maakt

met je tanden knarsen

de deur piept

honden blaffen

bijen zoemen

het klokje tikt

de bel rinkelt

paarden hinniken

-Manif

duidt op een nieuwe toestand en betekent het zich manifesteren van deze nieuwe toestand

tranen wellen op

-Possess

wijst op een bezitten van, een hebben, een lijden aan; is dus ook in algemene zin een supportverb; Melcuk gebruikt hiervoor trouwens ook Oper; het lijkt goed hier te differentiëren.

Ziektes vallen hieronder.

de bof/griep hebben

aan kanker lijden

de griep krijgen = IncepPossess; wordt op deze manier onderscheiden van *een cadeau krijgen*
Oper2 en *een nederlaag lijden* Oper2.

raakvlakken vertonen

gelijkenis vertonen

een ambt bekleden

een fonds beheren

een wens koesteren

een kater hebben

-Exist

wijst op het loutere bestaan/zijn van x

er heerst griep

het antwoord luidt

-Degrad/Upgrad

wijst op negatief aspect, slechter worden, slecht aflopen van x

melk wordt zuur

de liefde bekoelt

tegenhanger is Upgrad, het verbeteren van x

De neutrale waarden bij de niet-gespecificeerde processen en states zijn **Process en State**, vergelijkbaar met Action bij de Action-werkwoorden.

dankbaarheid verschuldigd zijn State

de kansen keren Process

Bij al deze verbale relaties zijn **aspecten (aspectuele relaties)** mogelijk. Wij gebruiken de volgende:

-Incep (het beginnen van x)

-Cont (het voortduren van x)

-Fin (het eindigen van x)

-Plus/Minus (het toenemen/afnemen van x)

-Cause (het veroorzaken van x)

In de regel hebben we aspecten als Incep, Cont en Fin alleen gebruikt als er duidelijke fasen in een actie of proces te onderscheiden zijn. Bv.

het onweer steekt op IncepFunct

het onweer drijft over FinFunct

het onweer houdt aan ContFunct

het vuur aanwakkeren zou in feite moeten worden gerepresenteerd als CausePlusFunct; vaak is echter bij de Funct+aspect-relaties Funct weggelaten, dus Incep bij *het onweer steekt op* is correct. Ook CausePlus bij *het vuur aanwakkeren* is correct. Causatieven horen uiteraard bij de actie-werkwoorden thuis.

Overzicht verbale relaties:

Action-ww:	Oper1	Niet-action-ww:	Funct
	Oper2		Excess
	Real		Son

Liqui
Cause ...
Action

Possess
Exist
Degrad
Upgrad
Process
State

Cause kan alle LF's (actief en niet-actief) actief maken. Bij sommige LF's kunnen combinaties met aspecten optreden: Incep, Cont, Fin, Cause.

=OMBI=

*

veldnaam: collocator
waarde: *zie boven*
scherm: example-constituents
bijzonderheden: alleen opgenomen bij *lex. collocation*

2.8.3.5 SEMANTIC SPECIFICATION COLLOCATOR

N.B. deze datacategorie is nog niet volledig doorgevoerd en gecontroleerd.

Een groot aantal lexicale collocaties is verder gesubcategoriseerd (zie veld SEM-SPEC-COLLOCATOR):

- de Oper1-categorie en de ongespecificeerde Action-categorie zijn verder gespecificeerd in sub-klassen als:
 - give (= het geven (in brede zin) van x, waarbij x ook abstract kan zijn (bv. een akkoord)). Er moet een gever en een ontvanger bij betrokken zijn; voorbeeld: 'een akkoord aanbieden'
 - andere waarden zijn: consume (tgo. give = aannemen), engage, consult, create, prepare, introduce, use

N.B. De subcategorisatie is (nog) niet volledig doorgevoerd.

=OMBI =

niet opgenomen

2.8.3.6 SEMANTIC SUBTYPE ARGUMENT

N.B. deze datacategorie is nog niet volledig doorgevoerd en gecontroleerd.

De basis van een groot aantal lexicale collocatie is gesubcategoriseerd in semantische subklassen (zie veld SEM-SUBTYPE-ARGUMENT). Zo heeft bijvoorbeeld in het voorbeeld 'een akkoord aanbieden' 'akkoord' niet alleen het semantische type = 'abstract' maar ook het subtype = 'agreement'. Andere waarden zijn: profession, fammember, plant, animal, nutrition, musintr, machine, building, clothing, written (voorwerpen bedoeld om gelezen te worden of om info in schrift vast te leggen) etc. In toto: 22 subklassen die weliswaar nog verder dienen gecontroleerd te worden maar die een zeer goede aanzet kunnen geven voor de totstandkoming van een semantisch lexicon van het Nederlands.

=OMBI =

niet opgenomen

2.8.4 PRAGMATICS

Zie 2.7 voor de features en waarden. Waarden die bij het entry zelf zijn opgegeven, bijv. SUBJECTFIELD, worden hier **NIET** herhaald. Alleen bij afwijkingen tov. het entry is er iets ingevuld.

=OMBI =

zie 2.7

2.8.5 COMMENT

Soms staan hier bijzonderheden over de opgegeven verbinding.

=OMBI =

niet opgenomen

2.8.6 Collocaties ter illustratie

PRAGM:

de mazzel!

laat me niet lachen

blijft u even aan de lijn

hierover is het laatste woord nog niet gezegd

mooi zo

alles kits?

we zullen zien

om erger te voorkomen

GRACOL:

in het kader van: in, Location

op de kaft prijken: op, Location

op de kalender staan: op, Location

een boom van een vent: (bij 'boom') van + comparison; comment 'collocatie bij 'vent'

een boom van een vent: (bij 'vent') FIXED: LEXCOL Magnus

een kast van een huis: idemdito

in de commissie zitten: in, Location

door de middenberm schieten: door, Location of FREE

verdenken van medeplichtigheid: FREE, (bij 'verdenken' GRACOL, van + COMPLEMENT)

met meerderheid van stemmen: met, Manner

in overeenstemming met: in, Manner

LEXCOL:

een teken van medeleven: Sing

een kop koffie: Sing bij 'koffie'

een echte verrassing: Verus

een aangename verrassing: Bonus

het kabinet is afgetreden: FinFunct

een kabinet afzetten: Liqui

de regering is gevallen: FinExist

het kabinet treedt aan: IncepFunct

een commissie installeren: CauseIncep(Funct)
een wet afschaffen: Liqui
een wet stemmen: Action of CauseFunct
een wet goedkeuren: Action of CauseFunct
de wet eerbiedigen: Real
zijn kandidatuur intrekken: Liqui
tegen een rode kaart aan/oplopen Oper2
de score opdrijven: Upgrad
een doelpunt scoren: Oper1

kans zien/hebben om: Possess
de kans krijgen om: Oper2
de kans nemen/grijpen/wagen om: Real
de kans geven om: Oper1
de kans doet zich voor: Manif

de bel gaat/rinkelt/luidt: Son
de bel rinkelen/luiden: Oper1
de radio aanzetten: IncepFunct
de radio uitzetten: FinFunct
de radio staat aan: Funct

mijnen vegen/opruimen: Liqui
mijnen leggen: Oper1
oorlog voeren: Oper1
er woedt oorlog in..: Exist
een oorlog ontketenen: CauseExist
de oorlog barstte uit/los: IncepExist

een talent ontwikkelen: CauseExist
gezag uitoefenen: Oper1
dankbaarheid verschuldigd zijn: State
een gerucht verspreiden: Oper1
zijn/haar leven vergallen: CauseDegrad
een steekproef nemen: Oper1
heersen over een gebied: Possess
een keizerssnode uitvoeren: Oper1
een standpunt innemen/bepalen: Oper1
de wind luwt: MinusFunct
er rust een taboe op..: Exist
talent hebben voor..: Possess
een vlieger oplaten: Oper1

3. ADJECTIVA

Ook bij deze woordsoort volgt een overzicht van de informatiecomponenten samen met de desbetreffende feature/value-paren. Overigens zal in vele gevallen kunnen worden volstaan met een verwijzing naar de behandeling bij de zelfstandige naamwoorden.¹⁸

3.1 IDENTIFICATION

Zie ook naamwoorden onder 2.1

3.1.1 FORM

Canonieke spellingsvorm, i.e. onverbogen vorm van het adjectief volgens de spellingnorm van het Groene Boekje, b.v. *productief*

=OMBI=

veldnaam:	spelling
scherm:	hoofdscherm

3.1.2 CAT

woordklasse, i.c. ADJ

=OMBI=

veldnaam:	word category
waarde:	adj
scherm:	morphology

3.1.3 SEQUENCE NUMBER

= cijfer (= nummer betekenis)

Wat betreft zgn. dubbelcategorieën (ADJ + ADV): alleen het adjectief is als entry opgenomen. De grammatica laat zo nodig toe te verschuiven van adjectief naar adverbium. Als het adverbium een andere betekenis heeft, is ofwel dit adverbium als 'functiewoord' in het RBN opgenomen, ofwel is er een extra lezing onder het adjectief opgenomen, die in vele gevallen een intensifierbetekenis heeft, bv. 'erg'1 naar, vervelend en 'erg'2 heel, heel veel (intensifier).

Homoniemen zoals *blauw* (noun) en *blauw* (adj) worden door de categorie-aanduiding gescheiden, hun betekenisnummer is in beide gevallen '1'.

Verbale vormen zoals oorspronkelijke voltooide en onvoltooide deelwoorden, bv. *gezocht*,

¹⁸ Tabelnaam RBN versie 1.0 (1998) adjectiva en voorbeelden
RBN versie 2.0 (2005) bijvoeglijke naamwoorden en bn-voorbeelden

gloeiend en *golvend*, met een eigen adjectivische betekenis, zijn als adjectiefentry opgenomen. Bronnen als de VDL-woordenboeken golden vaak als informatiebron in deze, ook werd getest op intensifieerbaarheid (zeer, erg etc.).

=OMBI=

zie onder nomina 2.1

3.2 GRAPHEMICS

Zie voor de features TYPE, SPELVA, en FORMVAR nomina 2.2

Voorbeeld van een **VERWIJSEENTRY**:

FORM	m.
CAT	adj
NUMBER	1
TYPE	abbreviation
REFERS-TO	mannelijk
PRAGM	(nil)

=OMBI=

veldnaam: explanation

waarde: string

scherm: spelling

bijzonderheden: alleen verwijseentries van het type *abbreviation*, *shortening*, *acronym* zijn opgenomen

3.3 PHONETICS

Zie nomina 2.3

3.4 MORPHOLOGY

De volgende features komen voor:

FLECTIONAL TYPE,

DECLINABILITY,

COMPARISON,

MORPHOLOGICAL TYPE (zie ook nomina),

MORPHOLOGICAL STRUCTURE (zie ook nomina)

3.4.1 **FLECTIONAL TYPE**

De volgende waarden komen voor:

PRESPA

PAPA

PSEUDOPRESPA

PSEUDOPAPA
CO(mp)
SU(p)
NOFLEC (= default)

Bij PRESPA en PAPA gaat het om adjectieven die morfologisch gelijk zijn aan een tegenwoordig dan wel een voltooid deelwoord, b.v.: *doorslaand* en *gedragen*. Deze oorspronkelijk verbale vormen leiden een eigen adjectivaal leven, semantisch e/o syntactisch (b.v. attributief gebruik, typische combinaties).

Bij elk van de waarden komt een feature BASE voor. Bij (oorspronkelijk) tegenwoordige en voltooidde deelwoorden gaat het om de infinitief van het werkwoord. Deze is als stringwaarde opgegeven. Ook woordgroepen zijn hier soms ingevuld.

Aldus:

<i>denderend</i>	FLECTIONAL TYPE BASE	PRESPA <i>denderen</i>
<i>haatdragend</i>	FLECTIONAL TYPE BASE	PRESPA <i>haat dragen</i>

PSEUDOPRESPA en PSEUDOPAPA worden gebruikt voor die vormen waar geen BASE voor te vinden is maar waarin toch de vorm van een PRESPA dan wel PAPA herkend kan worden. Voorbeelden: *aangeboren*, *gelaarsd*, *gemustst*.

CO(mp) en SU(p) zijn gebruikt als het entrywoord als FORM al een comparatief of superlatief is, dit is zeer zelden het geval.

=OMBI=

veldnaam:	flections-flec
waarde:	<i>zie boven</i>
	noflec (defaultwaarde) is niet opgenomen
scherm:	morphology

3.4.2 DECLINABILITY

Adjectieven die verbogen kunnen worden krijgen het feature DECLIN (= Default). NONDECLIN geldt voor morfologisch onveranderlijke adjectieven, zoals *lila*, *rose* of *gouden* maar ook participia op -en die attributief gebruikt worden.

Een woord als *begane* dat altijd geflecteerd is, is wel DECLIN genoemd. Het gegeven dat het altijd geflecteerd is, is dan in COMMENT weergegeven.

'Versteende' adjectieven die slechts in (idiomatische) uitdrukkingen met een werkwoord voorkomen (b.v. *kond* in *kond doen*) zijn behandeld zoals nomina in vaste uitdrukkingen (cf. *bij de lurven vatten/grijpen*) (zie nomina 2.1.1).

=OMBI=

veldnaam:	flections-declin
waarde:	<i>zie boven</i>
	declin (defaultwaarde) is niet opgenomen
scherm:	morphology

3.4.3 COMPAR(ISON)

Dit feature komt alleen bij de adjectieven voor (= trappen van vergelijking). Hier zijn 5 waarden aanwezig, namelijk:

	Adjectief	vergotende trap	overtreffende trap
REGULAR (default)	Snel	-(d)er	-st
IRREGULAR	Goed	<i>beter</i>	<i>best</i>
MIXED	Achterbaks	-er	meest
PERIPHRAS(TIC)	Moe	meer	meest
NCOMPAR(ISON)	Dood	--	--

De betekenis van de waarde is de volgende:

irregular:	trappen van vergelijking zijn onregelmatig gevormd
mixed:	vergotende trap op -er en overtreffende trap omschreven met <i>meest</i> of vergotende trap omschreven met <i>meer</i> en overtreffende trap op -st
ncompar:	geen trappen van vergelijking
periphras:	vergotende trap omschreven met <i>meer</i> en overtreffende trap omschreven met <i>meest</i>
regular:	vergotende trap op -er en overtreffende trap op -st

Frequentie is hier het criterium. Zeldzame trappen zoals bij *besneeuwd* zijn toch REGULAR genoemd; iets als *breedgeschouderd* wordt NCOMPAR.

Wat betreft adjectieven met de waarde REGULAR dient opgemerkt te worden dat deze zeer wel in gemarkeerde situaties met een omschreven vergotende en overtreffende trap kunnen voorkomen. Vergelijk 'dit idee is het belachelijkste dat ik ken' met 'dit is (wel) het meest belachelijke idee dat ik ken'. Deze eigenschap is niet apart gecodeerd. Bij de adjectieven met de waarde PERIPHRAS is sprake van adjectieven die slechts op deze manier compareerbaar zijn.

De features **COMP** en **SUP** geven – indien relevant - de volledige vormen van de vergotende en overtrappende trap.

=OMBI=

veldnaam:	comparison
waarde:	-er/-st (voor regular) <i>irregular</i> <i>mixed</i> <i>periphrastic</i> <i>ncompar</i>
scherm:	morphology

3.4.4 MORPHOTYPE

De volgende waarden gelden (zie ook 2.4 nomina):

SIMPMORPH
COMPOUND
XCOMPOUND

DERIVATION
WORDGROUP

COMPDERIV - voor samenstellende afleidingen als *baanbrekend*

=OMBI=

veldnaam: morphotype
waarde: *zie boven*
scherm: morphology

3.4.5 MORPHOSTRUCTURE

Zie nomina 2.4.2. Ook alleen in het VU-basisvocabularium toegepast.

=OMBI=

niet opgenomen

3.5 SYNTAX

Features zijn POSITION, ADVERBIAL USAGE en COMPLEMENTATION.

3.5.1 POSIT(ION)

Er zijn adjectieven die predicatief en attributief gebruikt kunnen worden, zoals 'slecht' of 'vervelend' (het voorbeeld is slecht vs. het slechte voorbeeld), maar er zijn er ook, die maar op één van beide posities kunnen voorkomen zoals 'ijzeren' (*deze staaf is ijzeren) en 'jammer' (*een jammere gebeurtenis). Bij elk adjectief is opgesomd welke eigenschap het heeft. Dit levert de volgende waarden op:

AP (attributief, predicatief: default)
ATTR (alleen attributief)
PRED (alleen predicatief)

=OMBI=

veldnaam: flexions-pos
waarde: *zie boven*
scherm: AP (default) is niet opgenomen
morphology

3.5.2 ADVERBIAL USAGE

ADV (Default)
NONADV
SENT

ADV: het adj. kan ook als bijwoord worden gebruikt. Testje hierbij was: 'op een ... wijze'. Als die constructie mogelijk was, dan ADV.

NONADV: het Adj kan niet als bijwoord gebruikt worden.

De waarde SENT betreft adjectieven die bij adverbiaal gebruik een hele zin als bereik kunnen hebben zoals `waarschijnlijk' in `Waarschijnlijk is zij ziek' = Het is waarschijnlijk dat zij ziek is (zgn. Sentential Adverbs).

=OMBI=

veldnaam: flections-usage
waarde: *zie boven*
ADV (default) is niet opgenomen
scherm: morphology

3.5.3 COMPLEM

De complementatiewaarden van de nouns zijn bekend (zie 2.5.5). Van toepassing voor de adjectieven zijn:

- DATCLAUSE
- WHCLAUSE
- OFCLAUSE
- TOINF
- OMTOINF
- FIXPREP "over", "op", etc.

Speciaal voor adjectieven worden verder ingevoerd:

- OBLOBJ Met verplicht object: het beu zijn. Ook bij 'het benauwd krijgen'; dan in COMMENT 'bij adverbiaal gebruik'.
- QUANT Met facultatieve bepaling van kwantiteit: drie meter hoog
- DANCOMP Met facultatief dan bij `ander' en `anders'
- OBLPREP Het verschil met FIXPREP is het volgende: adjectieven met FIXPREP hebben weliswaar een vast voorzetsel, ze kunnen echter ook zonder dit voorzetsel gebruikt worden, b.v. `schuldig (aan)'. Adjectieven met OBLPREP hebben ook een vast voorzetsel, maar kunnen slechts in combinatie hiermee voorkomen: `onderhevig aan'.

Ook bij de adjectieven kan in dit veld meer dan een feature en meer dan één waarde worden toegekend.

=OMBI=

veldnaam: adj-comp1,2,3
waarde: *zie boven*
scherm: LU-complementation

3.6 SEMANTICS

Hier onderscheiden we de volgende zes features:

3.6.1 SEMANTIC TYPE

De volgende waarden komen voor:

- PHYPER (= Physical/Perceptive Attributes). Hierbij moet voornamelijk gedacht worden aan adjectieven die betrekking hebben op als concreet gedefinieerde 'nouns', zoals 'rond', 'vierkant', 'zwart' etc. Dit wil overigens niet zeggen dat een adjectief niet een tweede of derde lezing kan hebben die een andere waarde heeft: 'vierkant' in 'een vierkante weigering' heeft als waarde ABSTRACT.
- STUFF Stofadjectieven: b.v. 'plastic'
- COLOUR Kleuradjectieven: b.v. 'groen'
- EMOMEN Emotional/Mental Attributes, hiertoe worden adjectieven gerekend die direct dan wel indirect horen bij nomina die aan mensen refereren, bv. *blij*, *woedend* of *gek*.
- PLACE Adjectieven van plaatsnamen afgeleid: b.v. 'Nederlands'
- TEMP Tijdsaanduidende adjectieven: b.v. 'toenmalig'
- ABSTRACT Hiertoe behoren de andere adjectieven.

Deskundig b.v. in de betekenis *vakbekwaam* is van het type EMOMEN. Een shift maakt vervolgens duidelijk dat *deskundig* niet alleen naar mentale wezens/personen verwijst maar ook en vooral naar hun daden: *een deskundig onderzoek*, *deskundig advies/oordeel/antwoord* = advies/oordeel/onderzoek/antwoord van/door deskundige(n). Zie hieronder SHIFTS.

=OMBI=

veldnaam: semantic type
waarde: *zie boven*
scherm: LU-semantics

3.6.2 SHIFTS

Desolaat heeft twee betekenissen:

1. zoals in: de stad zag er desolaat en verlaten uit --> PHYPER
2. zoals in: een desolaat gevoel --> EMOMEN

Beide kunnen worden uitgebreid naar ABSTRACT: ze kunnen ook verwijzen naar abstracta, bv. 1 'een desolate aanblik' en 2 'een desolate situatie'. Deze uitbreiding is met een SHIFT weergegeven, in de vorm van '--> ABSTRACT' of met alleen de eindwaarde.

=OMBI=

veldnaam: shift
waarde: *zie boven*
scherm: LU-semantics

3.6.3 RESUME

Zie nomina 2.6.5. Waarde = string die o.a. een verkorte definitie weergeeft. Heeft zelfde functie als bij nomina. Soms zijn i.p.v. korte betekenisomschrijvingen synoniemen, dan wel antoniemen gebruikt: *zwijgzaam* = niet spraakzaam; of een morfo-semantische relatie (b.v. in de 1e betekenis

van houten: 'van hout').

=OMBI=

veldnaam: resume
waarde: *zie boven*
scherm: hoofdscherm en LU-semantics

3.6.4 DEFINITION

Daar waar resumé's vooral van synoniemen en antoniemen gebruik zullen maken is een definitie in de regel uitgebreider.

gelovig	RESUME	godsdienstig (= SYN)
gelovig	DEFINITION	gezegd van iemand die gelooft in God

Definities van adjectieven zijn in tegenstelling tot definities van nomina niet opgedeeld in genus proximum en differentiae specifica.

=OMBI=

veldnaam: definition genus
waarde: *zie boven*
scherm: LU-definition

3.6.5 DEFINITION SOURCE

Zie nomina 2.6.7

3.6.6 SYNONYMS

Zie nomina 2.6.8

3.7 PRAGMATICS

De features en waarden zijn dezelfde als voor de nomina, zie 2.7

3.8 COMBINATORICS

Hoewel er veel gelijkenis is met de combinatorics voor nomina blijven er anderzijds toch ook een aantal verschillen. Daarom worden alle features en waarden hieronder genoemd.

=OMBI=

Niet opgenomen zijn alle vaste verbindingen die een verwijzing hebben naar een zelfstandig naamwoord. Er zullen dus over het algemeen alleen FREE combinations bij de adjectiva voorkomen. In het RBN zelf staan deze verbindingen hier wel, in de regel, met verwijzingen naar het nomen in het feature entry word reference (zie hieronder 3.8.3.2)

3.8.1 IDENTIFICATION

- CANFORM

- TEXTFORM
- CATEGORY
- NIL wel/niet aangekruist

Waarden en conventies zoals bij de nomina, zie 2.8.1

Voorbeeld	'mesjogge'	
	CANFORM	mesjogge zijn
	CAT	VP
	TEXTFORM	ben je nu helemaal mesjogge!
	CAT	S
	NIL	niet aangekruist

3.8.2 SYNTAX

- TYPE
- SUBTYPE
- COMBINATIONWORD
- CAT COMBINATIONWORD

3.8.2.1 TYPE

Keuze tussen FREE en FIXED.

3.8.2.2 SUBTYPE

Wordt FIXED gekozen dan zijn de volgende subtypes mogelijk:

LEXCOL	Lexical Collocation
GRACOL	Grammatical Collocation
IDIOM	
PRAGMA	Pragmatic Formula
PROVERB	

Voor definities zie naamwoorden 2.8.2

Voorbeeld:

CANFORM:	<i>er bedenkelijk uitzien</i>
TEXTFORM:	<i>dat ziet er bedenkelijk uit</i>

Bedenkelijk is een adjectief dat ook als adv. gebruikt kan worden. Bij adjectiva kunnen dus naast AdjN en AdvA ook Adv V-verbindingen voorkomen; deze worden in de regel bij het V behandeld (= basis van de phrase).

Indien bij de adjectieven opgevoerd, zijn ze als volgt te behandelen (pragmatics achterwege gelaten):

<i>bedenkelijk</i>	
CANFORM	<i>er bedenkelijk uitzien</i>
CAT	VP
TEXTFORM	<i>dat ziet er bedenkelijk uit</i>
CAT	S
Nil	-

ENTRYWORDREF eruitzien
CAT verb

N.B.1 LEXCOL's als 'centraal staan', 'chic staan', 'bankroet gaan' (AdjV of AdvV) zijn wel bij de adjectieven opgenomen en semantisch beschreven. Dit omdat het werkwoord in deze gevallen een support werkwoord is en slechts een lege, neutrale betekenis heeft. De MEANING COLLOCATOR heet in dit geval 'SUPPORT'. Soms is ACTION gebruikt als naam van de collocator.

N.B.2 De Adj-Nomen-constructies van de LEXCOL's zijn al opgenomen bij de nomina aangezien de lexicale collocaties bij de Basis van de woordgroep dienen te worden opgenomen. B.v. `verstokt vrijgezel' bij `vrijgezel'. Zie ook 2.8.2.2 onder Lexicol waar de onderstreepte categorieën in voorbeelden als NN zwerm bijen, AN verstokt vrijgezel, Adv A bar koud, Adv V diep teleurgesteld, VN (Subj) de motor sputtert, VN (Obj) zelfmoord plegen, duidelijk maken wat de basis/het behandel(t)refwoord van de collocatie is. Omdat adjectieven echter bij uitstek collocationale woorden zijn, (ze krijgen hun volle betekenis - zie b.v. verstokt - in combinatie met andere woorden) zijn hier de AN collocaties weer opgevoerd, zonder hun semantische descriptie te herhalen (zie hieronder **ENTRYWORDREF 3.8.3**).

GRACOL

Adj in combinatie met een functiewoord.

IDIOMS, PRAGMA en PROVERB: voor zover nomina in deze verbindingen voorkomen zijn ze aldaar behandeld en wordt in het semantische gedeelte met een verwijzing volstaan (zie hieronder 3.8.3). Voorbeelden waarvoor dit niet is opgegaan, omdat er geen nomina in voorkomen:

IDIOM:	`klein maar dapper'
PRAGMA:	`mooi zo'
PROVERB:	`wie niet sterk is, moet slim zijn'

3.8.2.3 COMBINATION WORD

In veel gevallen zal dat een naamwoord zijn; string staat in de canonieke vorm.

3.8.2.4 CAT COMBINATION WORD

Default is 'noun'.

Zie b.v. een blanco stem onder 'blanco' opgenomen met Entrywordref stem, CAT = noun.

3.8.3 SEMANTICS

De semantische descriptie van combinaties met adjectieven is afhankelijk van het type combinatie.

3.8.3.1 MEANING DESCRIPTION

bij **IDIOMS, PROVERBS en PRAGMA's** waarvoor niet geldt dat ze onder nomina behandeld kunnen worden, wordt een **MEANING DESCRIPTION** gegeven.

=OMBI=
veldnaam:

definition

waarde: *zie string*
scherm: example-semantics

3.8.3.2 ENTRY WORD REFERENCE

-bij FIXED verbindingen met een zelfstandig naamwoord komt er alleen een verwijzing naar het zelfstandig naamwoord onder:

ENTRY WORD REFERENCE waarde = string
CAT noun

Dit kan zich voordoen bij lexicale collocaties van het type Adj-Nomen, bij IDIOMS, PROVERBS en PRAGMAS, waarbij het eerste noun bij ENTRYWORDREF wordt ingevuld.
B.v.

verstokt vrijgezel	(lexcol)	ENTRYWORDREF	vrijgezel
zich (niet) op glad ijs begeven	(idiom)	ENTRYWORDREF	ijs
beter een half ei dan een lege dop (proverb)		ENTRYWORDREF	ei

=OMBI=
niet opgenomen

3.8.3.3 MEANING COLLOCATOR

bij LEXCOL's van het type Adv + Adj komt er een **MEANING COLLOCATOR: Magnus of Modif(ier)** (b.v. erg beperkt, erg = Magnus; vrij beperkt, vrij = modifier);

=OMBI=
veldnaam: func
waarde: *zie boven*
scherm: example-constituents
bijzonderheden: alleen opgenomen bij *lex. collocation* en *gramm. collocation*

3.8.3.4 MEANING COMPLEMENT

bij GRACOL's komt er alleen een **MEANING COMPLEMENT** (i.t.t. grammaticale collocaties bij nomina waar ook meaning gram word mogelijk is);

=OMBI=
veldnaam: func1
waarde: *zie nomina 2.8.3.2 (=waarde uit lijst semantische types nomina)*
scherm: example-constituents

Bij (frequente) vrije verbindingen komt er geen meaning description

3.8.4 PRAGMATICS

Zie nomina 2.8.4 en 2.7

3.8.5 COMMENT

Zie nomina 2.8.5

4. Verba¹⁹

4.1 IDENTIFICATION

Zie ook nomina 2.1 en adjectiva. 3.1

Ook hier zijn drie features opgegeven, nl.:

4.1.1 FORM

Canonieke spellingsvorm, i.c. de infinitief, volgens de spellingsnorm van GB, b.v. koketteren.

4.1.2 CAT

woordklasse, i.c. VERB

4.1.3 SEQUENCE NUMBER

cijfer (= nummer betekenis)

=OMBI=

zie nomina 2.1

4.2 GRAPHEMICS

Dezelfde features gelden voor de andere woordcategorieën:

4.2.1 TYPE

4.2.2 SPELVAR

4.2.3 FORMVAR

4.2.4 VERWIJSEENTRY'S

Zie nomina 2.2.4

Kunnen voorkomen bij spellings- en vormvarianten. Voorbeeld:

FORM	coquetteren
CAT	verb
NUMBER	1
TYPE	FULL
REFERS_TO	koketteren
PRAGM	n(on-)offic(ial)

¹⁹ Tabelnaam RBN versie 1.0 (1998) verba en voorbeelden
RBN versie 2.0 (2005) werkwoorden en ww-voorbeelden

=OMBI=

niet opgenomen

4.3 PHONETICS

Zie nomina 3

4.4 MORPHOLOGY AND MORPHOSYNTAX

Hier onderscheiden we negen features.

4.4.1 MORPHOTYPE

SIMPMORPH	bv. lachen
COMPOUND	bv. ademhalen
DERIVATION	bv. dineren
ZERO-DERIVATION	
PHRASAL	bv. omkomen
WORDGROUP	bv. houden van

PHRASAL wordt gebruikt voor werkwoorden met een scheidbaar partikel. Een partikel is een onverbuigbaar woord(deel) zoals voorzetsel of bijwoord. Niet alle scheidbare werkwoorden zijn ook meteen PHRASAL; daar waar het scheidbare deel geen partikel is, is het werkwoord een **COMPOUND**. Bijv. 'samenkomen' en 'onderuitgaan' zijn PHRASALS, 'schoonmaken' en 'scheefgroeien' zijn COMPOUNDS.

=OMBI=

veldnaam:	morphotype
waarde:	<i>zie boven</i>
scherm:	morphology

4.4.2 SEPARABILITY

In dit aparte veld wordt expliciet aangegeven of het werkwoord scheidbaar (SCH) is of niet (ONSCH). Dit als toevoeging op het MORPHOTYPE. (vgl. ademhalen met stofzuigen)

=OMBI=

veldnaam:	flections-scheidbaar
waarde:	<i>sch</i>
	default <i>onsch</i> is niet opgenomen
scherm:	morphology

4.4.3 MORPHOSTRUCTURE

Zie nomina 2.4.2

4.4.4 CONJUGATION TYPE

Het type vervoeging waartoe het werkwoord behoort:

- REGULAR= stam + te(n)/de(n) (ge) + stam + d/t
- STRONG= klinkerwisseling -en (+/- kl. wisseling)
- MIXEDA= klinkerwisseling (ge) + stam + d/t
- MIXEDB= stam + te(n)/de(n) -en (+/- kl. wisseling)
- OTHER= overig

=OMBI=
 veldnaam: flections-flection
 waarde: *zie boven*
 default *regular* is niet opgenomen
 scherm: morphology

4.4.5 CONJUGATION-FORM

In de velden OVT (onvoltooid verleden tijd) en VOLTOOIDDEELW (voltooid deelwoord) zijn in de niet-REGULAR-gevallen de vormen van de onvoltooid verleden tijd en het voltooid deelwoord expliciet opgegeven. Dus bij de typen strong, mixed en other.

=OMBI=
 veldnaam: flections-ovt
 waarde: *string*
 scherm: morphology
 veldnaam: flections-voltdeelw
 waarde: *string*
 scherm: morphology

4.4.6 MODE

- INF(INITIVE)
- IMP(ERATIVE)
- CONJ(UNCTIVE)
- PART(ICIPLE)
- INDIC(ATIVE)

Default is INF; in het lexicon komen de andere waarden slechts zeer uitzonderlijk voor. Bv. dunkt INDIC.

=OMBI=
 veldnaam: flections-mode
 waarde: *zie boven*
 inf (default) is niet opgenomen
 scherm: morphology

4.4.7 TENSE

In de regel is de mode INF en wordt geen tijd (tense) opgegeven. Slechts uitzonderlijk, zie b.v. dunkt hierboven, is dit wel het geval. Mogelijke waarden:

- NTENSE
- PRES(ent)

- PAST

=OMBI=

veldnaam: flexions-tense
 waarde: *zie boven*
ntense (default) is niet opgenomen
 scherm: morphology

4.4.8 NUMBER

Zie opmerkingen hierboven i.v.m. TENSE. Mogelijke waarden:

- NNUMBER
- LEXSING
- LEXPLUR

=OMBI=

niet opgenomen

4.4.9 PERSON

Idem als hierboven. Waarden:

- NPERSON
- ONE
- TWO
- THREE

In de regel wordt vanaf 4.6 het frame, bij default, als volgt ingevuld:

MODE: Infinitive
 TENSE: N Tense
 NUMBER: N Number
 PERSON: N Person

=OMBI=

niet opgenomen

4.5 SYNTAX

Syntax is een belangrijk subframe o.m. omdat hier in vergelijking met de bestaande woordenboeken/lexicale bestanden nieuwe informatie wordt gegeven (vnl. valentie en complementatie). Er zijn zeven f/v-paren:

4.5.1 CLASS

- MAIN (zelfstandig werkwoord)
- AUX (hulpwerkwoord)
- COPULA (koppelwerkwoord)

=OMBI=

veldnaam: flexions-class
 waarde: *zie boven*
main (default) is niet opgenomen
 scherm: morphology

4.5.2 PERAUX

Waarden voor het perfectieve hulpwerkwoord 'zijn' en/of 'hebben':

- H(EBBEN)
- Z(IJN)
- H(EBBEN)Z(IJN)
- H(EBBEN)/Z(IJN)

Het onderscheid tussen hz en h/z bestaat hierin dat hz duidt op een vrije alternantie, (b.v. *achtenarijden*) daar waar h/z staat vooreen aspectuele alternantie (ik heb gefietst = Perfectief; ik ben van Leden naar Den Haag gefietst = Duratief)

=OMBI=

veldnaam:	flections-aux
waarde:	<i>h,z,hz</i>
scherm:	morphology

De waarden voor deze categorie zijn meestal ontleend aan de VDL-woordenboeken.

4.5.3 VALENCY

Met valentie wordt bedoeld het aantal semantische slots (argumenten, rollen, cases, participants) dat main verbs opnemen. De voorkomende waarden zijn:

- MONO *slapen*
- DI *lezen*
- TRI *geven*

(één-, twee- of drieplaatsige predicaten)

wonen is ook DI en maakt meteen duidelijk dat een en ander niet samenvalt met de traditionele subcategorieën transitief en intransitief. Syntactisch gezien kunnen de slots door de volgende functies gevuld worden:

MONO	:	Subject
DI	:	Subject + DO/IO/Prep O/Compl
TRI	:	Subject + (IO) + DO/Compl/Prep O

Aangezien het IO niet altijd gerealiseerd wordt bij trivalente werkwoorden is het tussen haakjes geplaatst. Als het voorkomt kan het zowel als NP (en komt dan vóór het DO) dan wel als PP (met *aan* of *voor*, en komt dan voor of na het DO) voorkomen: *zij gaf hem geld. Zij gaf geld aan Piet. Zij gaf aan Piet geld.*

=OMBI=

veldnaam:	syntactic subcategory
waarde:	<i>zie boven</i>
scherm:	LU-syntax

Bij de vaststelling van valentie is vnl. van het corpus en van voorbeeldzinnen uit lexicografische bronnen gebruik gemaakt.

4.5.4 COMPLEMENTATION

Met COMPLEMENTATION bedoelen we de patronen die werkwoorden kunnen vormen met objecten en/of complementen (vgl. met de Engelse verb patterns uit de leerwoordenboeken zoals Langmans en OALD). In principe geldt dat alleen voor di- en trivalente werkwoorden. Monovalente krijgen in de regel geen complementering behalve in gevallen als *slapen --> de slaap der onschuldigen slapen*. Om die gevallen op te vangen is er een slot IN(HERO) (= inherent object) met als waarde de woordgroep in kwestie.

In wat volgt geven we de gebruikte complementeringspatronen. Vooraf zijn er nog enkele verklarende opmerkingen te maken:

De hieronder opgesomde complementatiemogelijkheden zijn in beginsel slechts voor zgn. "zelfstandige werkwoorden" bedoeld. Het voorgestelde systeem zou alle complementatiemogelijkheden moeten kunnen opnemen, behalve dan subjectcomplementatie.

Om de complementatiepatronen weer te geven wordt een eigen code gebruikt. Hieronder volgt een verklaring van deze code. Wat het complementatie-feature betreft, het RBN opteert hier voor duidelijk, eigen, expliciete beschrijving, dit i.t.t. de meeste andere bestaande bronnen.

De letters in de tweede kolom hebben betrekking op de basiscategorieën voor complementatie, namelijk N(IL), D(O), I(O), P(O), S(PECC), C(OMPL) en IN(HERO); de E staat voor "expansie".

De volgende "functional slots" worden aangenomen:

NIL, DO, IO, PO, OC (Object Complement), SC (Subject Complement), Sx (verschillende soorten 'Specifying Complement'), COMPL en INHERO.

Wanneer in de kolom onder "constituent filler" achter een bepaald complementatienummer meerdere "fillers" staan, gescheiden door een schuine streep, dan dient dit als volgt te worden gelezen: "filler" is bij voorkeur de eerste categorie. Indien niet aanwezig dan (één van) de andere. Wanneer in de kolom 'constituent filler' bij een PP geen verdere specificatie volgt, dan is er sprake van een PP als plaatsbepaling: specificatie van een uniek voorzetsel is dan dus onmogelijk. Vanzelfsprekend wordt er dan ook geen prepositie onder 'functioneel slot' gespecificeerd.

Valentie	Complementatie-code	Functioneel slot	Constituent filler	Mnemonische code
1	N	(DO)	NIL	INTRAN(SITIEF)
2	D1	DO	NP	TRAN(SITIEF)
3	D2a	IO DO	NP NP	DITRAN(SITIEF)
3	D2b	IO _{<aan>} DO	PP _{<aan>} NP	NPADITRAN
3	D3	PO _{<.>} DO	PP _{<.>} NP	PPTRAN
3	D4a	OC DO	AP NP	CMPXTRAN
3	D4b	OC DO	AP/NP NP	COPTRN
3	D4c	OC DO	PP _{<als>} NP	ALSTRAN
3	D4d	OC DO	NP PP _{<res>}	

2	I1a	IO	NP	IOCOMP
2	I1b	IO _{<aan>}	PP _{<aan>}	AIOCOMP
3	I2a	PO _{<>} IO	PP _{<>} NP	IOPP
3	I2b	PO _{<>} IO _{<aan>}	PP _{<aan>} NP	AIOPP
2	P1	PO _{<>}	PP _{<>}	PPCOMP
2	SQ1	SPECC	NP/AP	QUANTCOMP
3	SQ2a	IO SPECC	NP NP/AP	QUANTCOMP
3	SQ3	PO _{<>} SPECC	PP NP/AP	QUANTCOMP
2	SA1	SPECC	PP/AP	ADVCOMP
2	SA2a	SPECC IO	PP/AP NP	ADVCOMP
2	SC	SPECC	PP _{<als>}	ALSCOMP

De E-combinaties komen vrijwel uitsluitend voor als expansies van D-en/of P-categorieën, zie verder beneden.

2	E1		NP _(subj) INF	ACI
2	E2		NP _(subj) TOINF	ACTI
2	E3		OF	OFTRAN
2	E4		DAT	DATTRAN
2	E5		WH	WHTRAN
2	E6		HOE	WHTRAN
2	E7		UTTERANCE	UTTRAN
2	E8		OMTOINF	OMTETRAN
2	E9		TOINF	TETRAN
2	E10		INF	INFTRAN
2	C1	COMPL	INF	INFCOMP
2	C2	COMPL	OMTOINF	OMTECOMP
2	C3	COMPL	TOINF	TECOMP
2	C4	COMPL _{<>}	<> INF	PPINFCOMP
2	IN	INHERO	IN	INHERO

Aan de volgorde van de fillers e.d. worden geen specifieke claims ontleend. Criterium voor de beslissing of een bepaalde constituent DO dan wel IO is, vormt de mogelijkheid van een DO in een corresponderende passieve zin als subject te fungeren:

Jan scheept hem op met de klus --> *Hij wordt ... (DO)*
Jan draagt hem de klus op --> *Hem wordt .. (IO)*

Voorbeelden:

N	lezen, slapen
D1	lezen, verwezenlijken
D2a	geven, toevertrouwen
D2b	geven, aan iem. iets afdragen
D3	iem. met iets opscheppen
D4a	de deur groen verven
D4b	iem. een ei noemen, iets leuk vinden
D4c	iem. beschouwen als
D4d	iets tot een bal kneden, ook: mensen naar de overkant roeien
I1a	iem. gehoorzamen
I1b	aan iem. gehoorzamen, aan iets toegeven
I2a	iem. vertellen over
I2b	aan iem. vertellen over
P1	wachten op
SQ1	drie kilo wegen
SQ2a	veel/een hele hoop schelen, iem. veel berekenen
SQ3	veel van iemand opsteken
SA1	rustig wonen, zich goed houden, wonen in
SA2a	dat doet me prettig aan
SC	zich als een nar gedragen

De categorieën E1-E10 zijn, zoals gezegd, als expansies van objecten te beschouwen: zij kunnen dus in combinatie met elk van de categorieën D1 en/of P1 voorkomen. Dit wil overigens niet zeggen dat deze combinaties ook werkelijk allemaal bestaan.

Voorbeelden van bestaande combinaties:

D1	E4	aankondigen, dat
D2a	E8	iem. aanbieden, om te
D2b	E8	aan iem. aanraden, om te
D3	E8	iem. (ertoe) aanzetten, om te
D2a	E10	iem. leren lezen
D2a	E9	iem. gebieden/gelasten iets te doen

Bij de onderstaande voorbeelden zijn voor het gemak steeds werkwoorden opgenomen die D1 dan wel D2 als complementatiepatroon (vanaf hier: CP) hebben:

E1	iemand zien fietsen
E2	iemand weten te wonen
E3	vragen of
E4	zeggen dat
E5	vragen hoe, wie, waar, wat
E6	horen hoe iem. langzaam: 'dat' als 'hoe'
E7	beweren: '....': directe rede

E8	weigeren, proberen
E9	hopen te overleven
E1	Karel hoort lachen

Bij de categorieën C1-C4 is van complementatie in de vorm van een werkwoord sprake:

C1	komen
C2	dienen?
C3	lopen te
C4	op ... staan
IN	een slag slaan

In navolging van M. Gross wordt er dus bij werkwoorden op basis van complementatiepatronen niet echt gesubcategoriseerd: alle mogelijke patterns bij een betekenis worden expliciet opgesomd, er wordt niet van voorafbepaalde subcategorieën uitgegaan.

Bij het invullen van de complementatiepatronen is grofweg de volgende procedure gevolgd:

1. Stel de valentie van het werkwoord (in een bepaalde betekenis) vast
2. Ga na welke patterns volgens eigen inzicht gerealiseerd kunnen worden; gebruik de lijst als checklist, n.g.v. de valentie zijn verschillende waarden mogelijk, nl.:

valentie	1:1 mog.
valentie	2:8 mog. + Inhero + Ww. Compl (4 mog.) + Expansie
valentie	3:11 mog. + Expansie

3. Confronteer met de bronnen en het Corpus. Het is niet uitgesloten dat deze laatste minder info opleveren, maar het tegendeel kan ook.

N.B. In Access is in totaal plaats voor zes verschillende patronen. Extra informatie is in het COMMENT-veld weergegeven.

OMBI	
veldnaam:	compl,2,3,.., ext1,2,3
waarde:	<i>zie boven</i>
scherm:	LU-complementation
4.6 bijzonderheden:	Tijdens het redigeren werden codes met een expansie E, b.v. D2aE4, ingevoerd door beide kolommen in het veld complementatie op 1 regel in te vullen, in dit geval links D2a en rechts E4

4.6.1 PREPCOMPL

Dit veld is een vrij tekstveld dat speciaal voor het woordenboekgebruik is opgenomen. Voor het grootste deel bevat het info die in het complementatieveld voorkomt, alleen geeft het deze informatie weer in een vorm die dichter staat bij de Nederlandse lexicografische traditie. Het gaat om het gebruik van voorzetsels die als complementeerdere van argumenten bij werkwoorden optreden. De meeste hiervan worden door de complementatiepatronen D2b, D3, D4d, I1b, I2a, I2b en P1 ondervangen. Ze worden in dit veld als tekststring vermeld (zodat ze letterlijk kunnen worden overgenomen in woordenboeken) samen met een indicatie voor een selectierestrictie op het argument.

Bv. bij *verkopen* staat in dit veld PREPCOMPL: *iets aan iemand verkopen voor iets*. (Vergelijk

in deze met A. Abeling Handwoordenboek Nederlands.)

=OMBI=

veldnaam: pattern
waarde: *zie boven*
scherm: LU-semantics

4.6.2 SUBJECT

Twee waarden, nl. PERS(ONAL) en IMPERS(ONAL), het laatste voor gevallen als 'regenen'

=OMBI=

veldnaam: tr/intr/imp (cf. infra verba 4.6.4)
waarde: *impers*
pers (standaardwaarde) wordt niet opgenomen
scherm: morphology
bijzonderheden: de waarde is samen met de transiviteit in 1 veld opgenomen
NB de waarde 'imp' gaat voor 'tr' en 'intr'.

4.6.3 REFLEXIV(ITY)

Hiermee wordt de reflexiviteit van werkwoorden aangeduid. Waarden

- NREFL niet
- OBLREFL verplicht
- OPTREFL optioneel

Deze worden gedefinieerd als volgt:

Reflexiva zijn werkwoorden die vaak voorkomen met het reflexief pronomen ZICH of, in mindere mate, ZICHZELF.

Obligatoire reflexiva moeten met ZICH voorkomen dat noch door ZICHZELF noch door een DO (direct object) vervangen kan worden.

Optionele reflexiva (in de ANS: toevallig wederkerende werkwoorden) zijn werkwoorden die gecombineerd kunnen worden met een DO dat dezelfde persoon of zaak aanduidt als het S (subject) door middel van het reflexief pronomen, in de regel ZICH – dat ook door ZICHZELF kan worden vervangen.

Zij kunnen tevens met een DO dat niet naar het S verwijst gecombineerd worden.

Voorbeelden:

zich herinneren OBLREFL *zich* kan niet vervangen worden door *zichzelf*

zich wassen OPTREFL ook *zichzelf wassen*

NB De reflexivity-waarde kan per betekenis verschillen:

zich ergeren (over) OBLREFL vs.

iemand ergeren NREFL

=OMBI=

veldnaam: flexion-reflex
waarde: *zie boven*
nrefl (defaultwaarde) wordt niet vermeld
scherm: morphology

4.6.4 TRANSITIVITY

De waarden TRAN (transitief) en INTR (intransitief) zijn vermeld. Dit is de traditionele tweedeling tussen werkwoorden met, en die zonder mogelijk DO. Is bedoeld voor gebruikelijke opgave van transitiviteit in woordenboeken.

=OMBI=

veldnaam: tr/intr/imp
waarde: tr,intr
scherm: morphology
zie verba 4.6.2

4.7 SEMANTICS

Op dit ogenblik zijn er 8 subframes. Een negende subframe, subclass + caseframes en caseroles is ten dele ingevuld. Zie hiervoor bijlage 4.

4.7.1 TYPE

Klassen ACTION, STATE en PROCESS die tot stand komen op basis van de features [dynamic] en [control]:

- ACTION +dynamic,+control
- PROCESS +dynamic, -control
- STATE - dynamic, +/-control

Het feature [dynamic] kan o.m. met de syntactische test *aan het V...en* vastgesteld worden, indien mogelijk dan [+dynamic]. Anders gezegd: [dynamic] impliceert verandering, STATE niet. (Andere test is bijv. de combinatiemogelijkheid van Non State-verbs met Adverbs zoals: *geleidelijk aan, langzamerhand*). De control-parameter is positief wanneer het Subject in de combinatie S + V actief het gebeuren bepaalt, m.a.w. bepaalt of V plaats vindt of niet.

Bijv.

<i>kijken</i>	ACTION
<i>zien</i>	STATE
<i>groeien</i>	PROCESS

=OMBI=

*

veldnaam: semantic type
aarde: zie boven
scherm: LU-semantics

4.7.2 SUBCLASS/ CASEFRAME/ SR CASEROLES:

N.B. deze datacategorie is nog niet volledig doorgevoerd en gecontroleerd.

De verdere onderverdeling van Action-, State- en Process-verbs is samen met een indicatie van case frames en selectierestricties bij ongeveer de helft van de werkwoorden (± 5000) uitgevoerd.

Deze onderverdeling gebeurde op basis van 2 criteria:

1. de valentie van het predikaat (de n-plaatsigheid: 1,2 of 3)
2. het gebruiken van attributieve, spatiale of cognitieve primitieve features:
 - het attributieve feature drukt een bezitsrelatie uit: hebben, verwerven, verkiezen, overdragen van bezit e.d. bij één of meer semantische actanten van het werkwoord;
 - het spatiale feature verwijst naar een ruimtelijke relatie: de ruimte waarin de semantische actant(en) zich bevinden, zich naar toe bewegen, etc.;
 - het cognitieve feature duidt op emotionele, perceptuele of mentale eigenschappen die bij de semantische actanten, gegeven de stand-van-zaken door het werkwoord opgeroepen, aanwezig (moeten) zijn.

Het geheel is gebaseerd op het proefschrift van G. Deville (Modelization of task-oriented utterances in a man-machine dialogue system, Antwerp, 1989). Deville onderscheidt 23 subklassen types elk met een eigen caseframe (dat dus is af te leiden uit het type):

‘Kopen’ bijv. is een 3-plaatsig Actie-werkwoord, waarbij een (transfer-)relatie m.b.t. het bezit van iets/iem. wordt uitgedrukt. Op basis van deze kenmerken zal ‘kopen’ in de SUBCLASS EXCH3 vallen (iem. koopt iets/iem. van iem.). Sommige Actie/Proces/Toestand-werkwoorden worden niet door één van de opgegeven parameters gespecificeerd ze blijven dan ook ongespecificeerd zoals blijkt uit het onderstaande overzicht:

TYPE/SUBCLASS	FEATURES	VOORBEELDEN
ACTION1	C D	applaudiseren
ACTION2	C D T	afgieten
ACTION3	C D Di	combineren
EXCPROD2	C D A T	inkopen
EXCPROD3	C D A Di	afnemen
MVMT1	C D S	duikelen
MVMT2	C D S T	patrouilleren
MVMT3	C D S Di	deponeren
COGNT1	C D C o	huichelen
COGNT2	C D C o T	bagatelliseren
COGNT3	C D C o Di	inpeperen
STATE1		sneuvelen
STATE2	T	leunen
STATE3	Di	noodzaken
POSSESS2	A T	grossieren
LOCATION1	S	schoolblijven
LOCATION2	S T	stranden
LOCATION3	C S Di	vasthouden
STCOGNT1	C o	wanhopen
STCOGNT2	C o T	aanmatigen
STCOGNT3	C o Di	ontlenen
PROCESS1	D	achteruitgaan
PROCESS2	D T	schaden
PRMVMT1	D S	stuiteren
PRMVMT2	D S T	struikelen
PRMVMT3	---	---
PROCOGNT1	D C o	hallucineren

PROCOGNT2	D C o T	pochen
PROCOGNT3	D C o Di	inspireren
EXCOBT1	---	---
EXCOBT2	D A T	bijdragen
EXCOBT3	D A Di	reserveren

Verklaring:

- slechts de positieve waarden werden opgenomen, m.a.w. State 1 betekent: -C, -D, -A, -S, -Co, -T, -Di; D, zie PROCESS1, impliceert: -C+D-A-Co-S-T-Di
- afkortingen: D (dynamic); C (control), A (attributive); S (spatial); Co (cognitive); T (transitive); Di (ditransitive)

Deville zelf onderscheidt 23 subklassen, maar opent de mogelijkheid voor meer (in principe 36 verschillende). Voor het RBN zijn meerdere mogelijkheden voorzien.

Ad CASEFRAME

Gekoppeld aan een TYPE/SUBCLASS is een CASEFRAME, bijv. ‘kopen’ krijgt het CASEFRAME van de EXCPROD3 werkwoorden, i.e.: AGENT THEME BENEFICIARY. Het CASEFRAME wordt afgeleid uit de SUBCLASS (zie onderstaande tabel).

	TYPE/SUBCLASS	CASEFRAME	
A	Action 1	Agent	
	Action 2	Agent Theme	
	Action 3	Agent Theme Reference	
	Exchprod 2	Agent Theme	
	Exchprod 3	Agent Theme Beneficiary	
	Mvmt 1	Agent	
	Mvmt 2	Agent Source/Direction/Path	
	Mvmt 3	Agent Theme Source/Direction/Path	
	Cognt 1	Agent	
	Cognt 2	Agent Theme	
	Cognt 3	Agent Theme Cognizant	
	S	State 1	Theme
		State 2	Theme Referent
State 3		Effector Theme Referent	
Possess 2		Beneficiary Theme	
Loc 1		Theme	
Loc 2		Theme Locative	
Loc 3		Positioner Theme Locative	
StateCogn 1		Cognizant	
StateCogn 2		Cognizant Theme	
StateCogn 3		Cognizant Theme Cognitive Source	
P		Process 1	Processed
		Process 2	Effector Theme
		Prmvmt 1	Processed
	Prmvmt 2	Processed Source/Direction/Path	
	Prmvmt 3	Effector Theme Source/Direction/Path	
	Prcognt 1	Cognizant	
	Prcognt 2	Cognizant Theme	
	Prcognt 3	Effector Cognizant Referent	
	Excobj 2	Beneficiary Theme	
	Excobj 3	Beneficiary Theme Referent	

Soms moet wel (zoals in het geval 'kopen') de polariteit worden aangeduid. Bij kopen is die negatief (van Beneficiary naar Agent: ik koop een boek van Jan).

Ad Selectional RESTRICTIONS en CASEROLES

Voor de rollen in het CASEFRAME kunnen bepaalde restricties gelden. Zo is AGENT vrijwel altijd ANIMATE. Het THEME van 'vermorzelen' is CONCRETE enz. AGENT krijgt automatisch de restrictie ANIMATE, de andere "cases" zijn op dit moment nog niet van default-restricties voorzien. Alleen die rollen die een specifieke restrictie hebben, worden opgegeven, de andere krijgen de waarde NSELRES.

Het geheel is zoals gezegd nog niet volledig en zou nog verder aangevuld en gecontroleerd moeten worden. Het kan echter een mooie basis vormen voor een semantisch lexicon.

Voorbeeld:

aandikken (bet. mooier, beter etc. maken) heeft de volgende datacategorieën:

valency	2
subclass	cognt 2
caseframe	agent, theme (= roles)

selectional restrictions		
roles	agent	animate
	theme	inanimate

4.7.3 SHIFTS

Alle SHIFTS tussen de drie semantic types zijn gebruikt.

Een geval als *opschieten* (= vorderen) en *opschieten* (= vorderen met) is ook d.m.v. een shift beschreven, nl. Process --> Action.

=OMBI=

*

veldnaam:	shift
waarde:	<i>zie boven</i>
scherm:	LU-semantics

4.7.4 RESUMES

Zie nomina 2.6.5.

4.7.5 DEFINITION en HYPERONYM

De definitie bestaat uit twee delen:

HYPERONYM	werkwoord dat het semantisch direct 'bovenliggende' is
DEFINITION	volledige definitie

bijv. *laden*:

HYPERONYM	(iets) ergens in doen
-----------	-----------------------

DEFINITION er een vracht in doen

bijv. *lachen*:

HYPERONYM een geluid maken

DEFINITION wanneer je iets leuk vindt een geluid maken dat klinkt als 'hahaha'

=OMBI=

*

veldnaam: definition
waarde: *zie boven: string*
scherm: LU-definition

veldnaam: hyperonym
waarde: *zie boven: string*
scherm: LU-definition

4.7.6 DEFINITION SOURCE

Zie nomina 2.6.7

4.7.7 SYNONYMS

Zie nomina 2.6.8

4.8 PRAGMATICS

Zie nomina 2.7

4.9 COMBINATORICS

Zie nomina 2.8 en adjectieven 3.8

Typische LEXCOL's met werkwoorden zijn van het type

Adv V

VN

NV

- De laatste twee zijn in principe bij het nomen opgenomen en behandeld. Ook nu geldt dat ze opnieuw opgevoerd zijn bij het werkwoord maar zonder semantische descriptie. Dus onder 'sputteren' staat:

CANFORM *de motor sputtert*
CAT S
TEXTFORM *de motor begon te sputteren*
CAT S
TYPE FIXED
ENTRYWORDREF *motor*
CAT noun

De **COLLOCATOREN** van **LEXCOLS** die **wel** bij de verba worden behandeld, zijn
MAGNUS
MODIF(ier)
INTENS(ifier)
QUALIFIER
AdvV-gevallen van het type 'categorisch weigeren' zijn bij het verb opgenomen en van een semantische descriptie voorzien; bij het adjectief, in dit geval 'categorisch', is de combinatie ook opgenomen maar dan zonder semantische descriptie, nl. met een verwijzing d.m.v. ENTRYWORDREF naar 'weigeren'.

AdvV-gevallen van het type 'chic staan', 'centraal staan', 'bankroet gaan' waarbij het werkwoord alleen de rol heeft van supportverb en geen volle betekenis heeft, zijn bij het adjectief opgenomen en daar semantisch gecodeerd. De verwijzing is dan vanuit het werkwoord met ENTRYWORDREF naar dit adjectief.

Uiteraard zijn bij verba ook nog andere verbindingen dan lexicale collocaties mogelijk zo bv. pragmatische formules, grammaticale collocaties, idioms, proverbs en vrije verbindingen (zie nomina 2.8 en adjectiva 3.8).

5. Adverbia²⁰

5.1 De macroselectie

Voor de samenstelling van de macroselectie van de adverbia geldt de volgende richtlijn uit de *Documentatie Referentiebestand Nederlands* (1998):

De meeste bijvoeglijke naamwoorden kunnen als bijwoord fungeren. Het is meestal weinig zinvol om ze voor beide categorieën op te voeren (b.v. *geruisloos*, *plechtig*). Dat is alleen gedaan als er sprake is van een significant betekenisverschil (*praktisch* als in “praktisch onmogelijk”) en/of het bijwoord (frequent) een typische functie vervult: b.v. een intensiverende (*zwaar*, *ernstig*) of modale functie (*plotseling*, *mogelijk*, *waarschijnlijk* i.t.t. *onwaarschijnlijk*). Bijvoeglijke naamwoorden die nauwelijks meer als zodanig, maar wel als bijwoord gebruikt worden (zoals *dadelijk*), zijn alleen in deze laatste groep opgenomen.

Wanneer een woordvorm naast zijn voorkomen als adjectief toch ook een bijwoordlezing heeft (bijvoorbeeld *praktisch*), worden bij het bijwoord niet de adjectieflezingen herhaald, maar wordt alleen het bijwoordelijk gebruik beschreven en geïllustreerd.

In totaal zijn er ca. 1.000 adverbia in het bestand. Indien men dit aantal te groot acht, kunnen er naar eigen inzicht adverbia geschrapt worden.

5.2 Typische eigenschappen van de adverbia

Volgens de ANS zijn er drie manieren waarop de adverbia in groepen ingedeeld kunnen worden:

- naar functie,
- naar vorm, en
- naar betekenis.

Deze indelingen bestaan naast elkaar; dat wil zeggen dat de indeling naar functie onafhankelijk is van de indeling naar vorm en betekenis. In de database van het RBN is voor elk van deze indelingen een apart veld gebruikt om een zo volledig mogelijke beschrijving van de adverbia te geven. Omdat deze velden en hun waarden zoals ze hieronder worden beschreven niet allemaal even relevant zijn voor gebruik in een bilinguaal woordenboek, worden ze niet allemaal en/of niet in hun volledigheid in OMBI opgenomen. Per indeling wordt aangegeven welke waarden in OMBI gebruikt worden.

5.3 Indeling naar functie

De indeling naar functie houdt in dat er een onderverdeling wordt gemaakt tussen noemende

²⁰ Tabelnaam RBN versie 1.0 (1998) functiewoorden en voorbeelden
RBN versie 2.0 (2005) bijwoorden+functiewoorden en fw-voorbeelden

en verwijzende adverbia. Noemende adverbia zijn bijvoorbeeld *vanmorgen* en *buiten*: zij noemen resp. een bepaalde tijd of plaats. Verwijzende adverbia zijn bijvoorbeeld *dan* en *daar*: zij verwijzen resp. naar een bepaalde tijd of plaats (zonder die specifiek te noemen). Deze adverbia kunnen vervolgens weer in vier categorieën verdeeld worden. Hiervoor wordt het veld [sub-type] gebruikt. De noemende adverbia vormen een grote groep waarin geen verder onderscheid wordt gemaakt; zij worden in de database voor deze indeling gecodeerd met de waarde “other”. In de onderstaande tabel staan de vier waarden voor de verwijzende adverbia.

WAARDE	VERKLARING	VOORBEELDEN
definite	verwijst naar een bepaalde plaats of tijdstip	<i>daarheen, hierna, meteen, nu, toen</i>
indefinite	verwijst naar een onbepaalde plaats of tijdstip	<i>ergens, overal, eens, dikwijls, nooit</i>
interrogative	vraagt naar een plaats, tijdstip of hoedanigheid	<i>wanneer, waar, hoe, hoelang</i>
relative	verwijst naar een betrekkelijke plaats of tijdstip	<i>het huis waar, de dag waarop, de plaats alwaar</i>
other	noemende adverbia	<i>gisteren, tenslotte, absoluut, bijna</i>

=OMBI=

VELDNAAM	WAARDEN	AFKORTING IN OMBI	veldnaam in OMBI
subtype	definite	<i>niet overgenomen</i>	subtype
	indefinite	<i>niet overgenomen</i>	scherm: morphology
	interrogative	interrogat	
	relative	relative	
	other	<i>niet overgenomen</i>	

5.4 Indeling naar vorm

Met betrekking tot hun vorm kunnen adverbia in drie groepen worden onderverdeeld: de voorzetselbijwoorden, de voornaamwoordelijke bijwoorden en die bijwoorden die in geen van deze twee groepen horen.

5.4.1 Voorzetselbijwoorden

Een voorzetselbijwoord is homoniem met een voorzetsel: het heeft dezelfde vorm en vaak ook (een) vergelijkbare betekenis(sen). Zo kan *om* een voorzetsel zijn in “ze liepen om de kerk” en een adverbium in “de tijd is om”. Twee opmerkingen zijn hierbij van belang:

1. Het voorkomen van *om* in de zin “ze liepen het hoekje om” wordt tot de voorzetsels gerekend, en wel één in postpositie.
2. Het voorkomen van *aan* in de zin “zij kleepte zich aan” wordt niet als aparte lezing van het voorzetsel noch van het adverbium gerekend, maar wordt beschouwd als partikel van *aankleden*. Bij *aan* 1 [tegen het lichaam] vindt men dus wel het voorbeeld “hij heeft een trui aan” maar niet “zij kleepte zich aan”.

5.4.2 Pronominale of voornaamwoordelijke bijwoorden

Pronominale bijwoorden zijn samenstellingen van een voorzetsel(bijwoord) met één van de volgende plaatsaanduidende bijwoorden: *er, daar, hier, waar, ergens, nergens* en *overal*.

Voor de indeling naar vorm bestaat in de RBN-database voor de functiewoorden het veld [morpho-form].

WAARDE	VERKLARING	VOORBEELDEN
prepositional pronominal other	voorzetselbijwoorden voornaamwoordelijke bijwoorden	<i>buiten, middenin, onder, voorbij ertegenop, daarvandaan, hiertegenover, waaruit achterstevoren, dakpansgewijs, ergens</i>

=OMBI=

niet overgenomen

5.5 Indeling naar betekenis

Ook met betrekking tot betekenis zijn er bepaalde groepen te onderscheiden binnen de adverbia. Veel adverbia hebben een lokatieve of temporele betekenis, maar er zijn ook categorieën voor de modale of herhalende adverbia. Het veld voor de betekenisindeling is [sem-type]. De waarden hiervoor zijn gebaseerd op de informatie uit de ANS.

WAARDE	VERKLARING	VOORBEELDEN
Cause circumstance conjunctive contrast degree duration effect loc_dir manner modality negation nuance purpose quantifier reason repetition state time other	oorzaak omstandigheid voegwoordelijk beperkende tegenstelling aanduiding van graad tijdsduur gevolg of resultaat plaats of richting wijze modaliteit ontkenning nuancering doel kwantificerend reden herhaling hoedanigheid tijdstip overig	<i>daardoor, namelijk, zodoende beroepshalve, halverwege, en plein public bovendien, enfin, daarnaast desalniettemin, nochtans, echter zeer, drommels, beestachtig aldoor, mettertijd, zolang dus, ergo, bijgevolg binnen, ervoorbij, heuvelaf a capella, kalmpjes, ruitsgewijs misschien, waarschijnlijk, nota bene niet, geensinzs, nooit nu, eens, toch, wel daarvoor, waartoe absoluut, lichtjes, bijna, praktisch beleefdheidshalve, immers, waarvoor andermaal, telkens, vaak blootshoofds, incognito, lichterlaaie gisterenmorgen, binnenkort, hoelang desnoods, achterwege, gegarandeerd</i>

=OMBI=

niet opgenomen

5.5.1 Het resumé

In tegenstelling tot de functiewoorden hebben de adverbia vaak voldoende semantiek om het resumé met een betekenisomschrijving te vullen. In een enkel geval waarin dat toch moeilijk is, zoals bijvoorbeeld *er*, wordt tussen haakjes een functionele beschrijving van het adverbium gegeven: “(de existentie van iets vastlegend)”.

=OMBI=

Dit veld wordt onveranderd in OMBI opgenomen.

5.6 Overzicht van de velden

In de onderstaande tabel staan alle velden die voor de adverbia in het RBN worden gebruikt. De velden die onveranderd naar OMBI worden geëxporteerd zijn aangegeven in de tweede kolom met een grote stip. De kleinere stip geeft aan dat voor dit veld aangepaste waarden worden gebruikt.

VELDNAAM	OMBI	WAARDEN	BETEKENIS VAN DE WAARDEN
<i>id-form</i>	●	<het adverbium>	Het adverbium.
<i>id-number</i>	●	<nummer>	Betekenisnummer.
<i>cat</i>	●	adv – bijwoord adv comb – bijwoordgroep	Woordklasse.
<i>sub-type</i>	•	[zie paragraaf 5.3]	Het sub-type.
<i>Graph-status</i>	●	full – volle vorm clitic – clitische vorm weak – zwakke vorm emph – benadrukte vorm	Status van deze woordvorm; altijd full.
<i>formvar</i>	●	<vormvarianten>	Vormvarianten.
<i>morpho-type</i>	●	simplex – ongeleed woord compderiv – samenstellende afleiding compound – samenstelling derivation – afleiding xcompound – ondoorzichtige samenstelling wordgroup – woordgroep	Morfologisch type.
<i>morpho-form</i>		[zie paragraaf 5.4]	Morfologische structuur van bijwoorden.
<i>morpho-base</i>		<basisvorm>	Vorm waarvan deze is afgeleid.

<i>sem-type</i>		[zie paragraaf 5.5]	Semantisch type.
<i>Sem-resume</i>	●	<resumé >	Betekenisonderscheidende kenmerken.
<i>Sem-crossref</i>		<kruisverwijzingen>	Synoniemen.
<i>prag-*</i>	●	<pragmatiek>	Pragmatiek zoals voor de nomina etc.

5.7 Example Units

De volgende richtlijnen zijn aangehouden bij de redactie van de Example Units van de adverbia.

5.7.1 Aantal EU's

Omwille van de uniformiteit en de eenduidigheid is er getracht alle lezingen van de adverbia van minimaal één EU te voorzien; in de praktijk zijn het er meestal meer. De EU's worden vaak in de vorm van een volzin opgegeven en laten ook vaak de scheidbaarheid van voornaamwoordelijke adverbia zien. Tegen een bilingaal-contrastieve achtergrond is dat niet altijd nodig. Louter illustratieve voorbeelden mogen dan ook geschrappt en/of ingekort worden.

5.7.2 Voorzetselbijwoord vs. Pronominaal bijwoord

Bij de voorzetselbijwoorden mogen geen EU's worden opgenomen die eigenlijk bij een pronominaal bijwoord horen. Zo is de voorbeeldzin "ze sprong er met gemak bovenop" een illustratie van het gescheiden gebruik van *erbovenop* en staat bij het adverbium *bovenop* het voorbeeld "leg het maar bovenop".

5.7.3 Vaste verbindingen

Ook bij de voorbeeldzinnen van de adverbia is er onderscheid gemaakt tussen vaste en vrije verbindingen. Voor de vaste verbindingen bestaan dezelfde mogelijkheden als voor de overige woordsoorten: lexicale en grammaticale collocaties, pragmatische formules, zegswijzen en (nil-)idroom. Indien het een vaste verbinding betreft waarbij het adverbium niet zelf het hoofd is (bijvoorbeeld "iets dunnetjes overdoen" in de tabel hieronder), dan wordt er verwezen naar het hoofd van de verbinding. Een vaste verbinding waarbij het adverbium wel het hoofd is wordt op normale wijze beschreven. Vrije verbindingen worden altijd als "free" gemarkeerd en niet verder beschreven.

VERBINDIN G	free EU	fixed EU met verwijzing	fixed EU met beschrijving
vrij	" <u>misschien</u> gaat het morgen regenen" "achterover leunen"		
vast		"iets <u>dunnetjes</u> overdoen" "cum laude afstuderen"	"iemand <u>au sérieux</u> nemen" "erbovenop komen"

6. Funcitiewoorden²¹

6.1 Inleiding

In dit deel van de RBN-manual zullen de macro- en microstructuur van de functiewoorden (ca. 350) worden beschreven. De beschrijving gaat uit van de tabellen die door het RBN-team in Leuven zijn samengesteld en aan de VU verder zijn bewerkt. Naast de oorspronkelijke data uit Leuven zijn ook andere bronnen gebruikt om de descriptiemodellen op te stellen: de *Algemene Nederlandse Spraakkunst* [ANS, 1997] en *Part-of-Speech Tagging* van Frank Van Eynde [Van Eynde, 1999].

De specifieke eigenschappen van de verschillende soorten functiewoorden lopen dusdanig uiteen, dat het nodig is voor elke soort een eigen descriptiemodel op te stellen. Toch zijn er ondanks de verschillen ook velden die voor alle functiewoorden hetzelfde zijn. Hieronder volgt het overzicht.

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
idnr-functiewrd	teller		Teller.
id-form	<woordvorm>		Woordvorm.
id-number	<nummer van de lezing>		Nummer van de betekenis.
cat	art conj conj comb num prep prep comb pron	lidwoord voegwoord voegwoord multiwordgroup telwoord voorzetsel voorzetsel multiwordgroup voornaamwoord	Woordsoort.
morpho-type	compderiv compound derivation simplex word group xcompound	samenstellende afleiding samenstelling afleiding ongeleed woord woordgroep idiomatische samenstelling	Morfologisch type.
syn-disc	Yes/No		Syntactische discontinuïteit: scheidbaarheid.
sem-résumé	<résumé>		Résumé.

²¹ Tabelnaam RBN versie 1.0 (1998) functiewoorden en voorbeelden
RBN versie 2.0 (2005) bjwoorden+functiewoorden en fw-voorbeelden

sem-crossref	<verwijzing>		Kruisverwijzingen naar synoniemen.
prag-origin	<standaard RBN-waarden>		Origin.
prag-style	<standaard RBN-waarden>		Stijl.
prag-connot	<standaard RBN-waarden>		Connotatie.
prag-soc-group	<standaard RBN-waarden>		Sociale groep.
prag-chronol	<standaard RBN-waarden>		Chronologie.
prag-geogr	<standaard RBN-waarden>		Geografische beperking.
prag-subjf	<standaard RBN-waarden>		Vakgebied.
prag-freq	Rare		Frequentie.

Tabel 1 Algemeen tabeloverzicht

OMBI			
VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
id-form	<woordvorm>		
cat	art conj conj comb num prep prep comb pron	<i>onveranderd</i>	word category scherm: morphology
morpho-type	compderiv compound derivation simplex word group xcompound	<i>onveranderd</i>	morphotype scherm: morphology
syn-disc	Yes/No	<i>niet opgenomen</i>	
sem-résumé	<résumé>	<i>onveranderd</i>	résumé scherm: semantics (lexical unit)
sem-crossref		<i>niet opgenomen</i>	
pragmatics		<i>onveranderd</i>	pragmatics scherm: semantics (lexical unit)

De descriptiemodellen zijn allemaal op dezelfde manier opgebouwd:

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	VOORBEELDEN/UITLEG
----------	---------	--------------------------	--------------------

veldnaam	<de waarden>	uitleg van de waarde op deze regel	een voorbeeld of nadere uitleg
----------	--------------	------------------------------------	--------------------------------

In de eerste kolom staat altijd in cursief schrift de naam van het veld. De kolom met waarden kan op twee manieren worden gevuld: een eindige opsomming van alle mogelijke waarden waarvan er één gekozen moet worden, of een herhaling van de veldnaam tussen de hoekhaakjes < en >. Dit laatste staat voor een door de editor te bepalen waarde; deze invulling wordt voornamelijk gebruikt voor het résumé en (verbogen vormen van) het lemma. Alleen in dit geval mag het veld ook leeg blijven.

Sommige velden kunnen de waarde “unmarked” krijgen indien dit veld niet van toepassing is (*person* voor een conjunctie), of “all” als alle waarden mogelijk zijn ofwel de woordvorm niet voor één van de waarden in het bijzonder gespecificeerd is (*number* voor het persoonlijk voornaamwoord *u*).

Aangezien de mogelijke waarden van de diverse velden vaak afkortingen, wordt de derde kolom gebruikt om een volledige of meer beschrijvende vorm op te geven. In de laatste kolom kan een beschrijving van het hele veld volgen, of een voorbeeld voor de specifieke waarde van dit veld op de betreffende regel.

Voor zover de velden en/of waarden gebaseerd zijn op Van Eynde, 1999 (POS) of de *Algemene Nederlandse Spraakkunst*, 1997 (ANS), wordt dit vermeld achter de veldnaam in de overzichtstabel. De overige velden en waarden worden niet per veld worden verantwoord, aangezien de keuze en/of invulling wordt gestuurd door algemeen gebruik in de lexicografie. Elk veld zal maar één maal worden voorzien van een dergelijk label, met uitzondering van de velden *sub-type* en *sem-type*, die voor elke woordklasse verschillende waarden kunnen bevatten.

Na iedere tabel volgt een tweede tabel genaamd **OMBI** waarin beschreven staat welke velden en waarden zijn overgenomen in het RBN dat geladen is in OMBI. De tabel ziet er als volgt uit:

OMBI

VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
veldnaam RBN	waarden RBN	afkorting van waarde in RBN-OMBI	naam v. het betreffende veld op OMBI-scherm

6.2 Betekenisonderscheid

Het résumé heeft de functie betekenissen van elkaar te onderscheiden. In het geval van de adverbia, conjuncties, preposities en telwoorden wordt een korte betekenisomschrijving gegeven. In het geval van de telwoorden, lidwoorden en voornaamwoorden wordt het meest onderscheidende kenmerk van de functionele beschrijving als résumé gegeven. De functionele résumé's worden onderscheiden van de inhoudelijke door het gebruik van ronde haken (..).

In OMBI is dit onderscheid niet opgenomen.

6.3 Voornaamwoorden

Binnen de groep voornaamwoorden wordt een extra onderscheid gemaakt met betrekking tot het gebruik van het voornaamwoord: zelfstandig of niet-zelfstandig. Om deze twee groepen beter van elkaar en van het bovenliggende begrip te onderscheiden, worden hiervoor de Engelse termen gebruikt: *determiner* (niet-zelfstandig of bijvoeglijk gebruikt) en *pronoun* (zelfstandig gebruikt). De term *voornaamwoord* is de overkoepelende term en wordt als hoofdwoordklasse in RBN worden opgegeven. Behalve de persoonlijke en de wederkerende voornaamwoorden kunnen alle voornaamwoorden zowel zelfstandig als bijvoeglijk worden gebruikt. Andere woordsoorten met een determinerfunctie zijn telwoorden en lidwoorden.

Als een voornaamwoord zowel een determiner is als een pronoun, krijgt de woordvorm voor elk gebruik een aparte betekenis. De twee betekenissen worden van elkaar onderscheiden door het veld *determiner*²² voor de één op “Ja” en voor de ander op “Nee” te zetten. Omdat met dit onderscheid ook de invulling van een aantal andere velden samenhangt, worden de determiners en de pronouns apart behandeld.

6.3.1 Niet-zelfstandige voornaamwoorden: *determiners*

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het pronomen>		
<i>sub-type</i>	Demonstrative Exclamative Indefinite Interrogative Possessive Relative	aanwijzend uitroepend onbepaald vragend bezittelijk betrekkelijk	De subsoort; <i>hetzelfde</i> <i>zo 'n</i> <i>al</i> <i>wiens</i> <i>mijn</i> <i>welk</i>
<i>status</i> ^{POS}	Full Clitic Weak	vol clitisch zwak	<i>jouw</i> <i>'r</i> <i>je</i>
<i>declined form</i>	<verbogen vorm>		Verbogen vorm op <i>-e</i> : <i>onze</i> .
<i>person</i>	Unmarked First Second Third	niet van toepassing eerste tweede derde	Persoon (bezittelijk vnw). <i>mijn</i> <i>jouw</i> <i>hun</i>
<i>number</i>	Unmarked Sing Plur All	niet van toepassing enkelvoud meervoud zowel enkel- als meervoud	Getal (idem). <i>haar</i> <i>ons</i> <i>uwe Heiligheid</i>
<i>gender</i>	Unmarked Male	niet van toepassing mannelijk	Natuurlijk geslacht (idem). <i>zijn</i>

²² In de modellen in deze paragraaf wordt dit veld niet genoemd; het verschijnt wel in de tabel in paragraaf **Error! Reference source not found.**

	Female all	vrouwelijk zowel mannelijk als vrouwelijk	<i>haar</i> <i>ons</i>
<i>base-form</i>	<onverbogen vorm>		<i>haar</i> voor <i>hare Majesteit</i>
<i>concord</i>	sing-all sing-mf-plur sing-neut quantity plur all	alle enkelvoudsvormen niet-onzijdig enkelvoud en meervoud onzijdig enkelvoud ontelbaar enkelvoud en telbaar meervoud alle meervoudsvormen alle enkelvouds- en meervoudsvormen	<i>elk, ieder</i> <i>die</i> <i>dat, dit</i> <i>voldoende</i> <i>sommige</i> <i>dergelijke</i>
<i>usage</i>	Prenominal prepost	voor het nomen zowel voor als achter het nomen	<i>dat, die</i> <i>deze</i> (namelijk met <i>dezes</i>)

OMBI

VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
<i>wordform</i>	<het pronomen>		
<i>sub-type</i>	demonstrative exclamative indefinite interrogative possessive relative	dem exclam indef interrogat poss rel	subtype ²³ scherm: morphology
<i>status</i> ^{POS}	Full Clitic Weak	<i>onveranderd</i> <i>onveramderd</i> <i>onveranderd</i>	spelling status scherm.: spelling
<i>declined form</i>	<verbogen vorm>	<i>onveranderd</i>	declined scherm: morphology, flections
<i>person</i>	Unmarked First Second Third	niet van toepassing <i>onveranderd</i> <i>onveranderd</i> <i>onveranderd</i>	person scherm: morphology, flections
<i>number</i>	Unmarked Sing Plur All	niet van toepassing <i>onveranderd</i> <i>onveranderd</i> sg/pl	numb scherm: morphology, flections
<i>gender</i>	Unmarked Male Female All	niet van toepassing m. f. m.,f.	gender scherm: morphology, flections
<i>base-form</i>		<i>niet opgenomen</i>	
<i>concord</i>		<i>niet opgenomen</i> <i>niet opgenomen</i> <i>niet opgenomen</i> <i>niet opgenomen</i> <i>niet opgenomen</i>	

²³ veld heet in de meeste RBN-OMBI versies op dit moment *tr/intr/pers*

		<i>niet opgenomen</i>	
<i>usage</i>	Prenominal Prepost	<i>niet opgenomen</i> <i>niet opgenomen</i>	

6.3.2 Zelfstandige voornaamwoorden: *pronouns*

De voornaamwoorden *ieder*, *elk*, *alle(n)*, *allemaal* en *alles* hebben ook een predicatieve betekenis (“ze kregen ieder een glas”). Ondanks dat dit noch een determiner- noch een pronounfunctie is, wordt deze betekenis toch bij de pronouns ondergebracht. In het veld *function* kiest men daarvoor de waarde “predicative”.

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het pronomen>		
<i>sub-type</i>	Demonstrative indefinite interrogative personal possessive reciprocal reflexive relative	aanwijzend onbepaald vragend persoonlijk bezittelijk wederkerig wederkerend betrekkelijk	De subsoort; <i>diegene</i> <i>niemand</i> <i>wie</i> <i>zij</i> <i>zijne</i> <i>elkaar</i> <i>zich</i> <i>hetgeen</i>
<i>status</i>	full clitic weak emph	vol clitisch zwak nadruk	<i>hem</i> ' <i>k</i> <i>je</i> <i>ikke</i>
<i>person</i>	unmarked first second third all	niet van toepassing eerste tweede derde zowel eerste, tweede als derde persoon	Persoon. <i>wij</i> <i>jullie</i> <i>hem</i> <i>elkaar</i>
<i>number</i>	unmarked sing plur all	niet van toepassing enkelvoud meervoud zowel enkel- als meervoud	Getal. <i>u</i> <i>zij</i> <i>elkaar</i>
<i>gender</i>	unmarked male female neutre all	niet van toepassing mannelijk vrouwelijk onzijdig zowel mannelijk als vrouwelijk	Natuurlijk geslacht. <i>hij</i> <i>zij</i> <i>het</i> <i>ons</i>
<i>concord</i>	unmarked plur all	niet van toepassing alle meervoudsvormen alle enkelvouds- en meervoudsvormen	Concordantie met het nomen. <i>geen</i> <i>jouwe</i>
<i>nominalisati</i>	<nominalisatie>		Vorm op <i>-n</i> waarin

<i>on</i>			een determinator verschijnt als deze zelfstandig gebruikt wordt: <i>dezelfden</i> .
<i>base-form</i>	<onverbogen vorm>		<i>ik</i> voor <i>ikke</i>
<i>function</i> ^{POS}	subject object indirect object subj-obj predicative all_roles	alleen als subject als object van een werkwoord of voorzetsel alleen als indirect object alles behalve indirect object predicatieve functie zowel subject als alle soorten object	<i>ik</i> <i>hem</i> <i>hun</i> <i>ze</i> <i>allemaal</i> <i>elkeen</i>

VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
<i>wordform</i>	<het pronomen>		
<i>sub-type</i>	demonstrative indefinite interrogative personal possessive reciprocal reflexive relative	dem indef interrogat pers poss recip reflex rel	subtype scherm: morphology
<i>status</i>	full clitic weak emph	<i>onveranderd</i> <i>onveranderd</i> <i>onveranderd</i> <i>onveranderd</i>	spelling status scherm: spelling
<i>person</i>	unmarked first second third all	niet van toepassing <i>onveranderd</i> <i>onveranderd</i> <i>onveranderd</i> 1/2/3	person scherm: morphology, flections
<i>number</i>	unmarked sing plur all	niet van toepassing <i>onveranderd</i> <i>onveranderd</i> sg/pl	numb scherm: morphology, flections
<i>gender</i>	unmarked male female neutre all	niet van toepassing m. f. n. m./f.	gender scherm: morphology, flections
<i>concord</i>		<i>niet opgenomen</i>	

<i>nominalisati on</i>		<i>niet opgenomen</i>	
<i>base-form</i>		<i>niet opgenomen</i>	
<i>function</i> ^{POS}		<i>niet opgenomen</i>	

6.4 Lidwoorden

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het lidwoord>		
<i>sub-type</i>	definite indefinite	bepaald onbepaald	<i>de, het, 's een, geen</i>
<i>status</i>	full clitic	vol clitisch	Zie ook de Voornaamwoorden.
<i>determiner</i>	Yes		Alle lidwoorden zijn determiners.
<i>concord</i>	sing-all sing-neut sing-mf-plur all	alle enkelvoud onzijdig enkelvoud niet-onzijdig enkelvoud en meervoud alle enkelvouds en meervoudsvormen	<i>een het de geen, 's</i>
<i>usage</i>	prenominal	prenominaal	Allen komen voor het nomen.

OMBI

VELDNAAM	WAARDEN	AFKORTING IN OMBI	VELDNAAM IN OMBI
<i>wordform</i>	<het lidwoord>		
<i>sub-type</i>	definite indefinite	def indef	subtype scherm: morphology
<i>status</i>	full clitic	<i>onveranderd onveranderd</i>	spelling status scherm: spelling
<i>determiner</i>	Yes	<i>niet opgenomen</i>	
<i>concord</i>		<i>niet opgenomen</i>	
<i>usage</i>		<i>niet opgenomen</i>	

6.5 Telwoorden

Net zoals de voornaamwoorden kunnen de telwoorden zowel de determiner- als de pronounfunctie hebben. Toch wordt er slechts één betekenis gegeven voor beide functies--- ondanks dat de dubbele functie voor alle telwoorden geldt---omdat dat overeenkomt met de gebruikelijke lexicografische praktijk. De interpretatie van het veld *determiner* is dan ook iets

anders dan voor de voornaamwoorden: “No” betekent dat de woordvorm altijd zelfstandig wordt gebruikt en “Yes” houdt in dat het zowel zelfstandig als adjectivistisch kan worden gebruikt. In tegenstelling tot de voornaamwoorden, moet dus voor alle telwoorden (behalve *elfendertigst*) zowel het veld *function* als het usageveld worden ingevuld.

Naast de verbogen vormen *beider*, *elfder* en *veler* (die als verwijzingen behandeld worden) hebben de telwoorden nog meer afgeleide vormen: de verkleinvorm en de genominaliseerde vorm. Deze vormen worden bij de beschrijving van de basisvorm opgegeven (*nominalisation* en *diminutive*).

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het telwoord>		
<i>sub-type</i> ^{ANS}	cardinal-def cardinal-indef ordinal-def ordinal-indef	bepaald hoofdtelwoord onbepaald hoofdtelwoord bepaald rangtelwoord onbepaald rangtelwoord	zes veel zesde hoeveelste
<i>base-form</i>	<onverbogen vorm>		<i>dertien</i> voor <i>dertiende</i>
<i>nominalisation</i>	<verbogen vorm>		Met suffix –en: <i>met z'n zessen</i> .
<i>diminutive</i>	<diminutiefvorm>		Met suffix –jes: <i>met zijn viertjes</i>
<i>determiner</i>	No Yes		<i>elfendertigst</i> alle andere telwoorden
<i>function</i>	all_roles predicative all	subject en alle objecten predicatieve functie subject, alle objecten en predicatief	<i>veel, weinig</i> <i>allebei</i> <i>vijf, dertiende</i>
<i>usage</i>	unmarked prenominal prepost	ongemarkeerd alleen prenominaal zowel pre- als postnominaal	<i>elfendertigst</i> alle rangtelwoorden alle hoofdtelwoorden

OMBI

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	VELDNAAM IN OMBI
<i>wordform</i>	<het telwoord>		
<i>sub-type</i> ^{ANS}	cardinal-def cardinal-indef ordinal-def ordinal-indef	card card ord ord	subtype scherm: morphology
<i>base-form</i>	<onverbogen vorm>	<i>niet opgenomen</i>	
<i>nominalisation</i>	<verbogen vorm>	<i>niet opgenomen</i>	
<i>diminutive</i>	<diminutiefvorm>	<i>niet opgenomen</i>	
<i>determiner</i>		<i>niet opgenomen</i>	
<i>function</i>		<i>niet opgenomen</i>	
<i>usage</i>		<i>niet opgenomen</i>	

6.6 Voorzetsels

De voorzetsels hebben geen morfologische informatie en hun syntactische informatie betreft alleen hun plaats ten opzichte van het nomen. Wel hebben zij semantische informatie; deze wordt in het model weergegeven door het veld *sem-type* (en *résumé*, hetgeen echter voor elke woordsoort wordt ingevuld).

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het voorzetsel>		
<i>sub-type</i>	prepositional prepost	voor het nomen zowel voor als achter het nomen	<i>namens</i> <i>over</i>
<i>sem-type</i>	cause circumstance comparison condition contrast direction duration effect location means other purpose restriction state time	oorzaak omstandigheid vergelijking voorwaarde tegenstelling richting tijdsduur effect of resultaat plaats middel overig doel beperking hoedanigheid tijd	<i>wegens</i> <i>bij</i> <i>tegenover</i> <i>behoudens</i> <i>ondanks</i> <i>naar</i> <i>tijdens het vissen</i> <i>ingevolge</i> <i>onder</i> <i>per</i> <i>namens</i> <i>omwille van</i> <i>behoudens</i> <i>door</i> (iets heen) <i>sinds</i>

OMBI

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	VELDNAAM IN OMBI
<i>wordform</i>	<het voorzetsel>		
<i>sub-type</i>	prepositional prepost	<i>niet opgenomen</i> pre-post	subtype scherm: morphology
<i>sem-type</i>	cause circumstance comparison condition contrast direction duration effect location means other purpose	<i>onveranderd</i>	semantic type scherm: semantics (lexical unit)

	restriction state time		
--	------------------------------	--	--

6.7 Voegwoorden

Net zoals de voorzetsels hebben de voegwoorden geen morfologische informatie en krijgen zij alleen de velden *sub-type* (voor hun grammaticale functie) en *sem-type* erbij.

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
<i>wordform</i>	<het voegwoord>		
<i>sub-type</i> ^{POS}	coord coord-second subord	nevenschikkend nevenschikkend, alleen tweede conjunct onderschikkend	<i>en, zowel als, evenmin als</i> <i>want, dus</i> <i>omdat</i>
<i>sem-type</i> ^{ANS}	cause circumstance comparison concession condition conjunctive contrast degree disjunctive effect exception nonrestrictive other purpose restriction time	oorzaak omstandigheid vergelijking toegeving voorwaarde reeksvormend aaneenschakelende reeksvormer graadaanduidend gevolg tegenstellende reeksvormer effect of resultaat uitzondering uitbreiding overig doel beperking tijd	<i>omdat</i> <i>zonder</i> <i>als</i> <i>hoewel</i> <i>tenzij</i> <i>en</i> <i>maar</i> <i>zo ... dat</i> <i>of</i> <i>zodat</i> <i>behalve</i> <i>laat staan</i> <i>dat (hij nog lang moge leven)</i> <i>opdat</i> <i>behalve</i> <i>voordat, zodra, toen</i>

OMBI

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	VOORBEELDEN/ UITLEG
<i>wordform</i>	<het voegwoord>		
<i>sub-type</i> ^{POS}	coord coord-second subord	coord coord subord	subtype scherm: morphology
<i>sem-type</i> ^{ANS}	cause	<i>onveranderd</i>	semantic type

	circumstance comparison concession condition conjunctive contrast degree disjunctive effect exception nonrestrictive other purpose restriction time		scherm: semantics (lexical unit)
--	---	--	----------------------------------

6.8 Databasetabel functiewoorden

Hieronder volgt de volledige tabel van de functiewoorden (zonder de administratieve velden).

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
idnr-functiewrd	teller		Teller.
id-form	<woordvorm>		Woordvorm.
id-number	<nummer van de lezing>		Nummer van de betekenis.
cat	adv adv comb art conj conj comb num prep prep comb pron	bijwoord bijwoord multiwordgroup lidwoord voegwoord voegwoord multiwordgroup telwoord voorzetsel voorzetsel multiwordgroup voornaamwoord	Woordsoort.
sub-type	cardinal-def cardinal-indef coord coord-second definite demonstrative exclamative indefinite interrogative ordinal-def ordinal-indef other	bepaald hoofdtelwoord: num onbepaald hoofdtelwoord: num nevenschikkend: conj nevenschikkend, tweede: conj bepaald: adv/art aanwijzend: pron uitroepend: pron onbepaald: adv/art/pron vragend: adv/pron	Sub-type van de woordsoort. De verklaring wordt gevolgd door de woordsoorten waarop de waarde van toepassing is. Dit veld wordt door alle woordsoorten gebruikt.

	personal possessive prepositional prepost reciprocal reflexive relative subord	bepaald rangtelwoord: num onbepaald rangtelwoord: num other: adv persoonlijk: pron bezittelijk: pron prenominaal: prep pre- èn postnominaal: prep wederkerig: pron wederkerend: pron betrekkelijk: adv/pron onderschikkend: conj	
graph-status	full clitic weak emph	volle of neutrale vorm clitische vorm zwakke vorm benadrukte vorm	Status van de woordvorm; als het veld <i>gr-length</i> dat wordt gebruikt in de rest van het RBN. Hoewel dit een universeel bruikbaar veld is, is het alleen ingevuld voor de aanvullingen op de tabel Leuven.
graph-declined	<verbogen vorm>		Verbogen vorm voor determinatoren.
morpho-type	compderiv compound derivation simplex word group xcompound	samenstellende afleiding samenstelling afleiding ongeleed woord woordgroep idiomatische samenstelling	Morfologisch type.
morpho-form	unmarked prepositional pronominal other	niet van toepassing voorzetselbijwoord pronominaal bijwoord geen van beide vormen	Indeling naar vorm van de bijwoorden.
morpho-base	<onverbogen vorm>		Onverbogen vorm van de (verbogen) ingang.
morpho-nomen	<genominaliseerde vorm>		Genominaliseerde vorm op <i>-en</i> . Gebruikt voor telwoorden en voornaamwoorden.
morpho-dimin	<verkleinvorm>		Verkleinvorm op <i>-jes</i> . Gebruikt voor telwoorden.
syn-disc	Yes/No		Syntactische discontinuïteit: scheidbaarheid.
syn-person	unmarked first second third all	niet van toepassing eerste tweede derde zowel eerste, tweede als derde	Waarde voor de persoon (van de referent) van voornaamwoorden.
syn-number	unmarked sing plur all	niet van toepassing enkelvoud meervoud zowel enkel- als	Waarde voor het getal (van de referent) van voornaamwoorden

		meervoud	
syn-gender	unmarked male female neutre all	niet van toepassing mannelijk vrouwelijk onzijdig zowel mannelijk als vrouwelijk	Waarde voor het geslacht (van de referent) van voornaamwoorden.
syn-concord	unmarked sing-all sing-mf-plur sing-neut quantity plur all	niet van toepassing alle enkelvoud niet-onzijdig enkelvoud + meervoud onzijdig enkelvoud ontelbaar enkelv. + telbaar meerv. alle meervoud alle vormen	Concordantie determiners met het eropvolgende nomen (in tegenstelling tot het getal en geslacht van de referent).
syn-determiner	Yes/No		Determinergebruik.
syn-usage	unmarked prenominal prepost	niet van toepassing voor het nomen zowel voor als achter het nomen	Positie van het niet-zelfstandig voornaamwoord, telwoord en lidwoord ten opzichte van het nomen.
syn-function	unmarked subject object indirect object subj-obj all_roles predicative all	niet van toepassing alleen onderwerp object van werkwoord of voorzetsel alleen indirect object alles behalve indirect object zowel subject als elk object predicatieve functie subject, alle objecten en predicatief	Functie in de zin van de zelfstandig gebruikte voornaamwoorden en telwoorden.
sem-type	<zie de lijsten hierboven>		Semantisch type voor adverbia, conjuncties en preposities.
sem-résumé	<résumé>		Résumé.
sem-crossref	<verwijzing<, verwijzing>>		Kruisverwijzingen naar synoniemen.
prag-origin	<standaard RBN- waarden>		Origin.
prag-style	<standaard RBN- waarden>		Stijl.
prag-connot	<standaard RBN- waarden>		Connotatie.
prag-soc-group	<standaard RBN- waarden>		Sociale groep.
prag-chronol	<standaard RBN- waarden>		Chronologie.
prag-geogr	<standaard RBN- waarden>		Geografische beperking.

prag-subjf	<standaard waarden>	RBN-		Vakgebied.
prag-freq	<standaard waarden>	RBN-		Frequentie.

6.9 De tabel van de combinatorics

De beschrijving van de combinatoriek is voor alle functiewoorden gelijk. Deze is op enkele punten na gelijk aan de standaard RBN-combinatoricetabel²⁴.

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
idnr-functiewrd	<integer>		Link met het functiewoord.
id-canform	<canonieke vorm>		Standaard vorm van vaste verbindingen.
id-can-cat	AdvP AP NP PP S VP	adverbial phrase adjectival phrase noun phrase prepositional phrase sentence verb phrase	Soort zinsdeel dat de canonieke vorm is.
id-textform	<contextuele vorm>		De vaste verbinding in een zin of een vrij voorbeeld
id-text-cat	AdvP AP NP PP S VP	als bij [id-can-cat]	Soort zinsdeel dat de textuele vorm is.
syn-type	free fixed	vrij voorbeeld vaste verbinding	Het type geeft de vastheid van de voorbeeldzin aan.
syn-subtype	gracol idiom lexcol pragma proverb	grammaticale collocatie idiomatische verbinding lexicale collocatie pragmatisch gebruik zegswijze	Indien het om een vaste verbinding gaat moet hier het soort vaste verbinding worden aangegeven.
crossref	<kruisverwijzing>		Dit veld mag een opsomming van kruisverwijzingen bevatten.
entrywordref	<verwijzing>		Verwijzing naar een ander woord dat dit voorbeeld in zijn volledigheid bevat.
Entrywordref cat	adj adv noun num	bijvoeglijk naamwoord bijwoord zelfstandig naamwoord telwoord	Woordsoort van het woord waarnaar verwezen wordt.

²⁴ De lexicale en grammaticale collocaties worden niet verder gespecificeerd met collocators.

	prep verb	voorzetsel werkwoord	
sem-meandes	<betekenisomschrijving>		Betekenisomschrijving of uitleg bij het voorbeeld.
pragmatiek	<alle pragmatische labels>		Dezelfde pragmatiek als overal in het RBN.

==OMBI==

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	VELDNAAM IN OMBI
id-canform	<canonieke vorm>	<i>onveranderd</i>	canonical form scherm: example, syntax
id-can-cat	AdvP AP NP PP S VP	<i>onveranderd</i>	canonical form scherm: example, syntax
id-textform	<contextuele vorm>	<i>onveranderd</i>	category scherm: example, syntax
id-text-cat		<i>niet opgenomen</i>	
syn-type	free fixed	<i>onveranderd</i> <i>niet opgenomen</i>	example type scherm: example, syntax
syn-subtype	gracol idiom lexcol pragma proverb	<i>onveranderd</i>	example type scherm: example, syntax
crossref	<kruisverwijzing>	<i>niet opgenomen</i>	.
entrywordref	<verwijzing>	<i>niet opgenomen</i>	
Entrywordref cat		<i>niet opgenomen</i>	
sem-meandes	<betekenisomschrijving>	<i>onveranderd</i>	definition scherm: example, semantics
pragmatiek	<alle pragmatische labels>	<i>onveranderd</i>	pragmatics scherm: example, semantics

Referenties voor functiewoorden

- [Van Eynde, 1999] *Part-of-Speech Tagging*, Frank Van Eynde, 1999. Intern rapport van het project Corpus Gesproken Nederlands.
- [ANS, 1997] *Algemene Nederlandse Spraakkunst*, 1997. Martinus Nijhoff Uitgevers Groningen, Wolters Plantijn Deurne.

7. Interjecties²⁵

In het gehele Referentiebestand Nederlands nemen de interjecties een zeer kleine plaats in. Niet alleen met betrekking tot de omvang van de macroselectie, maar ook wat betreft de invulling van de microstructuur zijn de interjecties minder uitgebreid. Het totaal aantal interjecties is circa 150.

Dit document bevat de beschrijving van de interjecties door middel van een descriptiemodel dat voornamelijk is gebaseerd op de *Algemene Nederlandse Spraakkunst* (tweede druk, 1997).

Voor de interjecties of tussenwerpsels onderscheiden we twee aspecten bij de beschrijving van deze woorden: het eerste aspect heeft betrekking op de betekenis van de interjectie; het tweede aspect betreft de plaats van de interjectie in de zin.

7.1 Indeling met betrekking tot betekenis

Met betrekking tot de lexicale betekenis is de klasse van interjecties verdeeld in twee aparte groepen: woorden waarbij een duidelijk onderscheid gemaakt kan worden tussen vorm en betekenis zoals bij de meeste woorden, en woorden waarbij dit niet het geval is maar waarbij vorm en betekenis samengaan. Tot deze laatste verzameling worden voornamelijk de klanknabootsingen gerekend. Klanknabootsingen, zoals *tsjoeketsjoeke*, verwijzen naar noch noemen iets abstracts of concreets; evenmin vervullen zij een functie in de zin zoals de functiewoorden dat doen. Het zijn slechts talige representaties van geluiden die op zichzelf geen betekenis hebben. De interjectie *tsjoeketsjoeke* kan derhalve gekarakteriseerd worden door ofwel “geluid dat een trein maakt” ofwel simpelweg “klanknabootsing”. Voor het RBN is gekozen voor de tweede, simpelere omschrijving.

De overige interjecties hebben wel (communicatieve) betekenis; zij vormen de grootste groep. Binnen deze groep kan weer een onderverdeling worden gemaakt, namelijk tussen interjecties met een voornamelijk communicatieve functie zoals een verzoek of aansporing, en interjecties die zonder specifieke context gebruikt kunnen worden en grotendeels emotionele uitingen zijn.

7.1.1 Interjecties met voornamelijk communicatieve functie

Deze groep interjecties wordt gekarakteriseerd door de interactie tussen spreker en toehoorder die ermee geassocieerd is, hier benoemd met “communicatieve functie”. De spreker wil de toehoorder iets laten weten of tracht hem/haar iets te laten doen. Er zijn dus minimaal twee participanten in het gesprek.

In grote lijnen zijn er weer vier groepen binnen deze groep: mededelingen, uitingen met als doel de hoorder tot handelen te laten overgaan (bijvoorbeeld bevelen), vragen en pragmatische formules voor interactie met anderen. Hieronder volgen een paar functies met voorbeelden (de lijst is niet uitputtend).

FUNCTIE	DRUKT UIT	VOORBEELD
mededeling	afkeuring	<i>foei, 't</i> ²⁶

²⁵ Tabelnaam RBN versie 1.0 (1998) interjecties en voorbeelden
RBN versie 2.0 (2005) bjwoorden+functiewoorden en fw-voorbeelden

	ontkenning	<i>nee, welnee</i>
	bevestiging	<i>ja, zeker, nou (en of), o.k.</i>
	twijfel	<i>nou, hm, tsja, sorry?</i>
verzoek	bevel	<i>halt, ho, mars</i>
	aansporing	<i>kom, toe</i>
Vraag	verzoek om herhaling	<i>sorry, pardon, hè, excuseer</i>
	verzoek om instemming	<i>hè, niet(waar)</i>
formule	groet (afscheidsgroet; begroeting)	<i>goedemorgen, welterusten</i>
	gelukwensen	<i>gefeliciteerd, proficiat</i>
	overhandiging	<i>alstublieft</i>
	verontschuldiging	<i>sorry, pardon</i>
	goedkeuring	<i>bis, bravo</i>

7.1.2 Interjecties met expressieve functie

De interjecties in deze laatste groep hebben geen communicatieve maar een expressieve functie; met andere woorden: de spreker doet geen beroep op een toehoorder maar geeft een (ongerichte) uiting van zijn/haar emoties. Een toehoorder is niet essentieel in deze context; dit geldt met name voor krachttermen.

Deze interjecties worden vaak gebruikt om emotionele reacties op een gebeurtenis of uitspraken van een andere spreker uit te drukken. Hier volgen enkele voorbeelden.

EMOTIE	VOORBEELD
Pijn	<i>au</i>
berusting	<i>ach</i>
opluchting	<i>goddank, hèhè</i>
herkenning	<i>aha</i>
verbazing, verwondering	<i>oh, gunst, jemig</i>
(negatieve emoties als) frustratie, teleurstelling, ontgoocheling	<i>verrek, god, godverdomme, fuck</i>

7.2 Gebruik en plaats in de zin

Hoewel interjecties met name los voorkomen, is het ook mogelijk om ze in combinatie met een zin te gebruiken. Hun plaats ten opzichte van die zin is voor de meeste wel beperkt tot vóór of achter de eigenlijke zin en een enkele interjectie mag tussen twee zinsdelen in de zin voorkomen.

7.2.1 Interjecties voor of achter de zin

In deze gevallen wordt de interjectie direct voor het eerste zinsdeel geplaatst:

- 1) bah, wat is die soep al koud.
- 2) goh, wat is het hier gezellig.
- 3) afijn, om kort te gaan: men wil naar huis.

of direct na het laatste zinsdeel:

- 4) het is nog niet gelukt hoor.

²⁶ De zogenaamde click.

- 5) hij is heelhuids thuisgekomen, goddank.
6) mooi hè?

In veel gevallen is het niet mogelijk om een interjectie die voor de zin staat ook achter de zin te gebruiken:

- 7) * hoor, het is nog niet gelukt.
8) * wat is het hier gezellig, goh.

7.2.2 Interjecties binnen de zin

Een zeer klein aantal van de interjecties kan ook tussen twee zinsdelen voorkomen, echter zonder deel uit te maken van de syntactische structuur van de zin. De interjectie vormt dan min of meer een zelfstandige zin die wordt afgebakend door pauzes. Klanknabootsingen binnen de zin zijn in gebruik vergelijkbaar met directe rede.

- 9) het leek haar o zo mooi.
10) de hond doet "woef-woef".

Deze groep interjecties lijkt groter vanwege zinnen als “hij viel pats-boem in het water” en “een kleine sprong en ze zijn hup bovenaan”. In deze gevallen is de interjectie echter homoniem met een bijwoord: het geeft vaak de snelheid of abruptheid van de handeling aan.

7.3 Syntaxis

De enige syntactische informatie heeft betrekking op de plaats van de interjectie ten opzichte van de zin. Hiervoor worden drie velden met Ja/Nee-waarden gebruikt:

VELD	BETEKENIS	<i>ho</i>	<i>hoor</i>	<i>alstublieft</i>
syn-position-pre	voor de zin	Ja	Nee	Ja
syn-position-post	achter de zin	Nee	Nee	Ja
syn-position-inter	in de zin	Nee	Ja	Ja

Voor deze verzameling velden geldt het volgende:

- mininaal één veld moet als waarde “Ja” hebben;
- het is mogelijk om alle velden op “Ja” te zetten;
- interjecties die alleen los voorkomen, zoals krachttermen en klanknabootsingen, hebben alleen “Ja” voor [syn-position-pre];
- als de interjectie in een interjectiegroep gebruikt kan worden (“jewel hoor”, “ho zeg”) gaat het om de positie binnen deze groep. Uitzonderingen zijn combinaties met *o* of *oh* (als in “o jee”) en met komma’s geschakelde interjecties (“één, twee, drie, mars!”).

7.4 Het resumé

Interjecties krijgen geen betekenisomschrijving zoals de lexicale woorden, maar veelal een resumé met een functionele omschrijving zoals “groet”, “ontkenning” en “drukt o.a.

[berusting/verwondering] uit”. In een enkel geval is een synoniem of een omschrijving wel mogelijk.

7.5 Pragmatiek

Bij de pragmatische labels vormt het vaktaallabel een uitzondering met betrekking tot de interjecties. Dit label wordt niet gebruikt omdat interjecties juist voornamelijk in de algemene taal voorkomen.

De overige pragmatische labels met betrekking tot herkomst, stijl, connotatie, tijdsbeperking, geografische beperking en sociale groep, worden op dezelfde wijze en met dezelfde waarden gebruikt als in de rest van het RBN.

7.6 Example Units

Voor de meeste interjecties is het niet eenvoudig om EU's op te geven die de betekenis of functie illustreren, omdat de contexten zeer uiteenlopend kunnen zijn en/of de interjecties voornamelijk los voorkomen. Voor groeten, krachttermen en klanknabootsingen zijn dan ook vaak geen EU's opgenomen. In andere gevallen, zoals bij de interjecties zonder communicatieve functie die meer dan één emotie kunnen uitdrukken, is meer dan één EU wenselijk en ook mogelijk.

Interjecties lenen zich niet bijzonder goed voor collocaties en/of idioom en als gevolg daarvan zijn vrijwel alle voorbeeldzinnen gecodeerd als “free”.

7.7 Databasetabel interjecties

Hieronder volgt de databasetabel met de veldnamen en hun uitleg. De onderstepte waarden zijn de waarden die als standaard worden ingevuld.

VELDNAAM	WAARDEN	BETEKENIS VAN DE WAARDEN	UITLEG VAN HET VELD
Idnr-interjectie	<teller>		Teller.
id-form	<interjectie>		De interjectie.
id-cat	Int		De woordklasse.
id-number	<volgnummer>		Het betekenisnummer.
syn-position	<u>pre</u> post inter	voor de zin achter de zin tussen de zinsdelen	De mogelijke positie van de interjectie ten opzichte van de zin.
sem-resume	<resumé>		Het resumé.
prag-origin	<u>norig</u> <taal>	inheems woord taal waaruit het woord afkomstig is	Herkomst.
prag-style	<u>nstyle</u> informal formal slang vulgar	neutrale stijl informeel formeel slang grof	Stijl.
Prag-connotation	<u>nconnot</u> offens pejor euphem	geen connotatie beledigend pejoratief, negatieve connotatie	Connotatie.

	jocular	eufemistisch schertsend	
Prag- socGroup	<u>gensoc</u> child sold stud romcath	alle taalgebruikers kindertaal soldatentaal studententaal rooms-katholiekentaal	Sociale groep.
prag- chronology	<u>synchron</u> oldfash neolog	hedendaags ouderwets neologisme	Tijdsmarkering.
prag- geography	<u>gengeo</u> belg dialect	alle regio's Vlaams dialect	Regio.

De tabel voor de Example Units is identiek aan die van de functiewoorden. Zie 6.9

8. GEOGRAFISCHE NAMEN

8.1 Bestanden

Er zijn twee bestanden van geografische namen: het "grote" bestand en het basisbestand. Het basisbestand is een subset van het grote bestand met een selectie van de belangrijkere en frequentere geografische namen (voor meer details zie Macrostructuur).²⁷

	Nomina	Adjectiva	Totaal
Groot Bestand	2680	640	3320
Basisbestand	1428	473	1901

8.2 Macrostructuur

8.2.1 **Bronnen**

Voor de buitenlandse plaatsnamen, inwoneraanduidingen en adjectieven: Nederlandse Taalunie, *Buitenlandse aardrijkskundige namen in het Nederlands*, SDU Uitgevers, Den Haag 1996.

Voor de Nederlandse en Belgische plaatsnamen: atlas, wegenkaarten en introspectie.
Voor de bijbehorende inwoneraanduidingen en adjectieven: woordenboeken Nederlands.

8.2.2 **Groot bestand**

Buitenlandse plaatsnamen: de continenten, de wereldzeeën, alle staten van de wereld en hun hoofdsteden. Op basis van het genoemde werk van de Nederlandse Taalunie zijn bij de andere geografische aanduidingen alle exoniemen opgenomen. Exoniemen zijn plaatsnamen die een Nederlandse, d.w.z. een andere dan in het betreffende land gebruikelijke vorm hebben. Met uitzondering van de staten en hun hoofdsteden zijn er geen endoniemen (plaatsaanduidingen die in de vorm worden gebruikt die ook in het desbetreffende land geldt) opgenomen. Plaatsen zoals Lyon, München of New York zijn dus NIET in het bestand aan te treffen. Historische plaatsnamen zoals b.v. Babel zullen evenmin worden aangetroffen.

Plaatsnamen in België en Nederland: Alle provincies, hun hoofdsteden en alle steden met meer dan 100.000 inwoners. Eveneens opgenomen zijn de belangrijkere en bekendere regio's, eilanden, rivieren en kanalen.

In binnen- en buitenland zijn bij alle landen, provincies, regio's, eilanden en steden ook het bijbehorende bijvoeglijke naamwoord en de aanduidingen voor inwoner en inwonerster opgenomen. Bij andere entiteiten (b.v. gebergtes, rivieren) is het respectievelijke bijvoeglijke

²⁷ Ten opzichte van de rest van het RBN zijn twee velden toegevoegd: *geografische naam* (waarde: *ja/nee*), *basis* (waarde: *ja/nee*). In OMBI_RBN zijn deze velden niet overgenomen.

naamwoord opgenomen voor zover het in het Nederlands bestaat (b.v. Pyrenees, Rijns).

Verder bevat het grote bestand nog

- de belangrijkste talen van de wereld
- de belangrijkste dialecten in het Nederlandse taalgebied
- de planeten van ons zonnestelsel

8.2.3 Basisbestand

Het basisbestand bevat de meest frequente en de belangrijkste aardrijkskundige aanduidingen uit het grote bestand:

- continenten en wereldzeeën
- staten van de wereld en hun hoofdsteden met uitzondering van kleine, relatief onbekende staten b.v. in Oceanië en het Caraïbisch gebied.
- buitenlandse gebergtes, rivieren, regio's, steden etc. als ze door hun omvang of hun cultureel, historisch of toeristisch belang als algemeen bekend kunnen worden verondersteld (b.v. Siberië, Florence, Eufraat, Gardameer).
- alle Nederlandse en Belgische namen uit het grote bestand
- namen van inwoners van Belgische en Nederlandse plaatsen, van alle staten van de wereld en algemeen gebruikelijke namen zoals Scandinaviër, Parijzenaar, enz.
- namen van inwonsters die niet systematisch van het adjectief kunnen worden afgeleid. (bv. Française). N.B. De vrouwelijk inwonersnamen zijn verder niet in de basisselectie opgenomen. Zij kunnen van de verbogen vorm van het bijbehorende bijvoeglijke naamwoord worden afgeleid.
- In het basisbestand staan ook de meeste talen, dialecten en planeten uit het grote bestand.

Aantal entries *

* aantallen zonder verwijzentries (i.t.t. par. 1)

	Groot bestand	Basisbestand
Landen	225	192
Hoofdsteden	224	162
Inwoners	483	348
Inwonsters	460	3
Geoniemen NL	73	73
Geoniemen BE	45	45
Talen/dialecten	115	95
Andere geoniemen	788	354
Totaal Nomina	2418	1277

Adjectiva	517	371
------------------	------------	------------

8.3 Nomina

De microstructuur van de geografische nomina is in principe identiek aan de structuur van de "gewone" zelfstandige naamwoorden. Ze worden met behulp van de zelfde frames of velden beschreven. Hieronder worden daarom alleen die velden besproken waarbij de beschrijving van geografische namen van de beschrijving van gewone nomina afwijkt. Voor een verdergaande beschrijving van de gegevensstructuur wordt naar de algemene documentatie van de nomina in het RBN verwezen (hoofdstuk 1). Geografische namen die uit woordgroepen bestaan, zoals *Golf van Bengalen* en *Benedenwindse Eilanden*, worden als geheel in de macro opgenomen.

8.3.1 Identification

Voor de Belgische en Nederlandse plaatsnamen wordt de algemeen gebruikelijke vorm aangegeven (atlas, wegenkaarten e.d.). De spelling van de bijbehorende inwoneraanduidingen volgt de spelling in de verschillende woordenboeken. In een enkel geval geven de verschillende bronnen verschillende spellingen op. Dan zijn beide spellingen opgenomen (hoofdentree en Spelvar).

Bij de buitenlandse aardrijkskundige namen wordt de door de Nederlandse Taalunie aanbevolen spelling aangehouden.

8.3.2 Graphemics

8.3.2.1 Spelvar en Formvar

Waar spelling- of vormvarianten voorkomen, zijn deze respectievelijk als Spelvar of Formvar opgenomen.²⁸

De voornaamste redenen voor spellingvarianten zijn verschillen in de spellingnorm tussen Vlaanderen en Nederland en verschillende spellingen in verschillende bronnen.

Vormvarianten komen voor bij enkele binnen- en buitenlandse plaatsnamen en vaak bij afgeleide inwoneraanduidingen (b.v. Albaniër en Albanees).

8.3.3 Syntax

8.3.3.1 Gender

In principe wordt het in de woordenboeken opgegeven geslacht aangehouden. Bij landen, provincies, steden e.d. wordt alleen "n" opgegeven ondanks dat het steeds meer gebruikelijk is ook met het vrouwelijke voornaamwoord "haar" naar deze onzijdige entiteiten te verwijzen (Vlaanderen en haar regering, Parijs met haar boulevards).

8.3.3.2 Article

De meeste plaatsnamen zijn opgenomen zonder lidwoord. Bij plaatsnamen die obligatorisch een lidwoord hebben is dit opgegeven (de Achterhoek, het Albertkanaal). Voor inwoneraanduidingen is altijd het lidwoord "de" opgegeven. Voor talen staan altijd "het".

²⁸ Spelling- en vormvarianten zijn niet opgenomen in OMBI-RBN.

8.3.4 Semantics

8.3.4.1 Reference

Alle plaatsnamen en talen hebben een unieke referentie en krijgen de waarde PROPER. De inwonersnamen hebben een niet-unique referentie en dus de waarde COMMON.

8.3.4.2 Countability

De plaatsnamen en talen zijn UNCOUNT, de inwoneraanduidingen COUNT.

8.3.4.3 Semantic Type

Plaatsnamen hebben de waarde PLACE, inwoneraanduiding de waarde HUMAN en talen de waarde NONDYN.

8.3.4.4 Shifts

Bij staten, territoria, steden, provincies en regio's met een officieel bestuur is een shift PLACE>INSTITUT gespecificeerd omdat in deze gevallen de plaatsnaam ook voor het bestuur van de plaats kan staan. Alle andere plaatsnamen, de inwonersnamen en de talen hebben de waarde NOSHIFT.

8.3.4.5 Resumé

Het resumé is voor plaatsnamen consequent in de vorm "x in y" gehouden, waarbij x de aard van de plaats en y de grotere geografische eenheid aanduidt (land in Afrika, meer in Zwitserland etc.). Het resumé voor inwoneraanduidingen is altijd in de vorm "inwoner van y" of "inwoner van y".

Deze consequente representatie maakt het mogelijk om op een eenvoudige manier naar entiteiten als landen, hoofdsteden of inwoners te zoeken (zoek in Resumé naar "x" of "inwoner") of om aan een bepaalde geografische eenheid gerelateerde namen te vinden (zoek in Resumé naar "in /van y").

8.3.4.6 Definition

De definitie komt in de meeste gevallen overeen met het resumé. Daarbij is "x", respectievelijk "inwoner/inwoner" GENUS PROX. en de rest van het resumé DIFF. SPECIF. Waar nodig bevat DIFF. SPECIF nog aanvullende informatie.

8.3.5 Combinatorics

Bij enkele aardrijkskundige nomina zijn vrije en vaste verbindingen, collocaties etc. opgenomen. De definitie en de verwijstructuur zijn identiek aan die van "gewone" nomina.

8.4 Adjectiva

De microstructuur van de geografische adjectieven is in principe identiek aan de structuur van de "gewone" bijvoeglijke naamwoorden. Ze worden met behulp van dezelfde frames of velden beschreven. Hieronder worden daarom alleen die velden besproken waarbij de beschrijving van geografische adjectieven van de gewone beschrijving afwijkt. Voor een verdergaande beschrijving van de gegevensstructuur wordt naar de algemene documentatie van de adjectieven in het RBN verwezen (hoofdstuk 2).

8.4.1 Identification

De spelling van geografische adjectieven volgt de spelling in de verschillende woordenboeken en - voorzover ze daar zijn opgenomen - van de Nederlandse Taalunie. In enkele gevallen geven de verschillende bronnen verschillende spellingen op. Dan zijn beide spellingen opgenomen (hoofdentree en Spelvar).

8.4.2 Graphemics

8.4.2.1 Spelvar en Formvar

Waar spelling- of vormvarianten voorkomen, zijn deze respectievelijk als Spelvar of Formvar opgenomen. De voornaamste redenen voor spellingvarianten zijn verschillen in de spellingnorm tussen Vlaanderen en Nederland en verschillende spellingen in verschillende bronnen. Vormvarianten komen bij de geografische adjectieven relatief vaak voor (b.v. Aaks en Akens, Guatemalaans en Guatemalteeks).

8.4.3 Semantics

8.4.3.1 Semantic Type

Alle geografische adjectieven hebben de waarde PLACE

8.4.3.2 Resumé en Definition

Resumé en definitie zijn bewust kort gehouden: "van X" waarbij X voor de plaatsnaam of evtl. een bevolkingsgroep staat (b.v. "van de Slaven" bij "Slavisch"). Waar nodig bevat de definitie aanvullende informatie.

8.4.4 Combinatorics

Ook bij de aardrijkskundige bijvoeglijke naamwoorden zijn vrije en vaste verbindingen, collocaties etc. opgenomen. De definitie en de verwijfsstructuur zijn identiek aan die van "gewone" bijvoeglijke naamwoorden.

8.5 OMBI

Het bestand is opgedeeld in het basisbestand en het bestand "extra". Het basisbestand is gelijk aan bovengenoemd basisbestand; het extra-bestand is het "grote" bestand zonder het basisbestand. Op grond van bovengenoemde aantallen en beschrijvingen kan besloten worden het complete bestand dan wel slechts de basisselectie over te nemen.

Uiteraard kan er ook uit het basisbestand dan wel uit extra bestand geschrapt worden. Zo bevat het basisbestand bij de opgegeven staten ook hun hoofdstad. In een bilinguaal bestand kan hiervan afgeweken worden. Toevoegingen (bijv. van extra naar basis) kunnen eveneens uit contrastieve of andere overwegingen gedaan worden. (bijv. *Kosovaars* van extra naar basis).

Voor de beschrijving van de velden in OMBI wordt verwezen naar de de handleiding van de RBN-nomina en de RBN-adjectiva.

Bijlage 1: Samenstelling subcorpus 38Mi Corpus INL

Het corpus dat geraadpleegd is in het kader van het RBN, is een selectie uit het INL 38.000.000 corpus.

De selectie bestaat uit de volgende bronnen:

I.	Varied	
1.	Mixed/Newspapers	
	De Standaard	2.607.579
1.	Mixed/Written-to-be-Spoken	
	8 uur Journaal	2.641.862
2.	Health/health	
	Waal Malefijt e.a. 'Migraine en Hoofdpijn'	17.672
	Luijterink 'Reiki-groeiboekje'	16.020
2.	Health/psychology	
	Ostaden - de Jong 'Droomsymbolen'	21.578
3.	Leisure	
	Playboy	317.321
4.	Humanities/Philosophy	
	Rijntjes 'Groei naar waarheid'	47.572
	Snel 'Op weg naar Tao'	88.994
5.	Science/Environment/Magazines	
	Waddenbulletin	303.599
5.	Science/Environment/Books	
	Nieuwe strategieën	61.559
5.	Science/Astronomy	
	Zenit	197.311
6.	Society/Politics	
	Rooie Vrouwen	144.203
6.	Society/Politics	
	Troonredes	36.188
6.	Society/Social Studies	
	Klaveren 'Sociaal Ontwerpen'	41.509
	Sips 'Vrouwen...'	17.949
6.	Society/Politics/Reported Speech	
	Verslagen Tweede Kamer	1.869.099
II	Uit de Meppeler Courant zijn ca. 1.500.000 tokens geselecteerd, waaronder <u>niet</u> de sportbijlagen.	
	Newspaper	
	Meppeler Courant (tot +/- 1.5Mi)	3.360.000