



182969001F

077716



RGN HSRC

SENTRUM VIR BIBLIOTEEK- EN  
INLIGTINGSDIENSTE

CENTRE FOR LIBRARY AND  
INFORMATION SERVICES

VERVALDATUM/DATE DUE

IBL  
7 APR 1995  
TERUG

00002067874



001.3072068 HSRC LEXI 4



\*206787\*

Die rekenarisering van  
leksikografiese prosesse

WORDNET-projek van die  
LEXINET-program

Verslag LEXI-4

# Die rekenarisering van leksikografiese prosesse

## WORDNET-projek van die LEXINET-program

P H Swanepoel  
R Morris

P H Swanepoel  
R Morris

Afdeling Leksikologie

Instituut vir Taal- en Kunstenavorsing  
Uitvoerende Direkteur: Dr K P Prinsloo

ISBN 0-7969 0674 2 (reeks)  
0 7969 0678 5

© Raad vir Geesteswetenskaplike Navorsing, 1988  
Alle regte voorbehou

Gedruk en uitgegee deur die RGN  
Pretoriusstraat 134, Pretoria

## VOORWOORD

Die Afdeling Leksikologie in hierdie Instituut het teen die einde van 1986 'n ondersoek geloods na die gebruik van die rekenaar in die verwerking van taaldata. In die buiteland vind daar snelle ontwikkelings op die gebied plaas terwyl relatief min aandag in Suid-Afrika daaraan bestee word. Die doel van die ondersoek was om die relevantheid vir Suid-Afrika van die gebied te bepaal en om daarvolgens aanbevelings te doen oor moontlike ontwikkelingstappe.

Hierdie verslag is een van 'n reeks wat die verskillende aspekte van gerekenariseerde taalverwerking dek. Hier gaan dit oor die gebruik van die rekenaar in die leksikografie.

Die ondersoek sou nie uitgevoer kon word sonder die samewerking van 'n aantal individue en instansies wat by leksikografiese aktiwiteite betrokke is nie. Opregte dank word betuig teenoor almal wat aan die navorsingspan die nodige gegewens vir die ondersoek verskaf het en/of kommentaar oor die konsepverslag gelewer het:

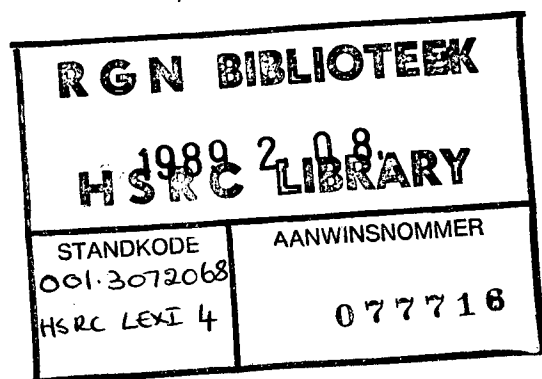
- Die Buro van die Woordeboek van die Afrikaanse Taal
- Rhodes University English Dictionary Unit
- Fort Hare University Dictionary Unit
- Die uitgewers Nasou, HAUM, Tafelberg en Van Schaik
- Die Staatsdrukker en Cape and Transvaal Printers.

Die navorsingspan wat die gegewens verwerk het en die aanbevelings geformuleer het, is gelei deur prof P.H. Swanepoel van die Departement Afrikaans, UNISA. Aan hom, en aan die twee ander buitemedewerkers in die span, prof P.S. Groenewald en dr D.J. Prinsloo, spreek die Instituut groot waardering uit. Dr. Rose Morris, die programleier van die hele Lexinet-ondersoek, is die mede-outeur van hierdie verslag. My dank aan haar vir die aandeel wat sy, saam met prof. Swanepoel, in hierdie verslag gehad het deur die beplanning, deurvoering en afronding daarvan.

K P PRINSLOO

UITVOERENDE DIREKTEUR: INSTITUUT VIR TAAL- EN KUNSTENAVORSING

JULIE 1988



## INHOUDSOPGAWE

Bladsy

1.	INLEIDING	1
1.1	RAAMWERK VAN DIE ONDERSOEK	1
1.2	TAAKOMSKRYWING	2
1.2.1	Navorsingsdoelstellings	2
1.2.2	Agtergrond	3
1.2.3	Werkwyse	5
2.	STANDBESKRYWING	7
2.1	WOORDEBOEKVOORSIENING	7
2.1.1	Afrikaans	7
2.1.2	Engels	10
2.1.3	Afrikatale	14
2.1.4	Indiese tale	21
2.1.5	Europese tale	22
2.1.6	Opsomming	24
2.2	REKENAARGESTEUNDE LEKSIKOGRAFIEPROJEKTE IN SUID-AFRIKA	26
2.2.1	Hoofwoordeboekprojekte	26
2.2.2	Projekte aan universiteite	27
2.2.2.1	Projek: Woordeboek van die Noord-Sotho-taal (voorlopig genoem WNT-Projek)	27
2.2.2.2	Die Franse woordeboekprojek aan die Universiteit van Pretoria	28
2.2.2.3	'n Morfeemwoordeboek - US	28
2.2.3	Uitgewers	29
2.2.4	Drukkers	30
2.2.5	Taaldatakorpuse	31
2.3	VOORUITSIGTE VIR REKENARISERING	32
2.3.1	Die hoofwoordeboekkantore	32
2.3.1.1	Die Woordeboek van die Afrikaanse taal	32
2.3.1.2	The Dictionary of South African English on Historical Principles	33
2.3.1.3	The Xhosa-woordeboek	33
2.3.2	Ander woordeboeke	34
2.3.3	Die uitgewers	34
3.	REKENAARGESTEUNDE LEKSIKOGRAFIE IN DIE BUITELAND	35
3.1	INLEIDING	35
3.2	DIE LEKSIKOGRAFIESE DATABANK(LDB)	35

3.2.1	Tegniese en logiese struktuur van die LDB	36
3.2.1.1	Die teksargief en die leksikon(s)	38
3.2.1.1.1	Logiese en rekenaarstruktuur van die leksikons	42
3.2.2	Die rekenariserings van die gedrukte leksikon	53
3.2.3	Die voordele van rekenariserings	64
3.3	PROFIELE VAN 'N AANTAL LEKSIKOGRAFIESE ONDERNEMINGS	66
3.3.1	Longman-Uitgewers	66
3.3.2	Die New Oxford English Dictionary-Projek (NOED)	75
3.3.3	Die Instituut voor Nederlandsche Lexicologie (INL)	83
3.3.3.1	Die afdeling Woordenboek der Nederlandsche Taal (WNT)	83
3.3.3.2	Die afdeling Thesaurus	84
3.3.4	Die Van Dale Projectontwikkeling (VDP)	89
3.3.4.1	Groot woordenboek hedendaags Nederlands (VDHN)	90
3.3.4.2	Die tweetalige woordeboeke	94
3.4	LEKSIKONS IN (SEMI-) OUTOMATIESE TEKSONTLEDINGSTELSELS	96
3.5	SLOTOPMERKINGS	106
4.	SAMEVATTING EN AANBEVELINGS	107
4.1	SAMEVATTING	107
4.1.1	Die hoofwoordeboekprojekte	107
4.1.2	Kleiner woordeboekprojekte	109
4.1.3	Die huidige infrastruktuur: die rol van akademici	109
4.2	AANBEVELINGS	110
4.2.1	Die hoofwoordeboeke	111
4.2.2	Nie-hoofwoordeboeke	112
4.2.3	Opleiding	113
5.	BIBLIOGRAFIE	115

## HOOFSTUK 1: INLEIDING

### 1.1 RAAMWERK VAN DIE ONDERSOEK

Die ondersoek waaroor daar hier verslag gedoen word (die WORDNET-onderzoek) was een van die projekte in die navorsingsprogram LEXINET en die rekenarisering van taal.

Die LEXINET-program was navorsing wat met steun van die Wetenskaplike Adviesraad onderneem is ten einde rekenaartoepassings op natuurlike-taal-verwerking te ondersoek. Die oorsprong van die besluit om die navorsing te onderneem, was enersyds bevindinge uit vorige RGN-navorsing (wat daarop gedui het dat rekenarisering op taalgebiede in Suid-Afrika min benut word), en andersyds die waarneming dat rekenaartoepassings op die verwerking van taal in die buiteland 'n besonder hoë prioriteit geniet. In 'n aantal lande word daar grootskaals bestee aan die gebied, terwyl Engeland en Japan programme vir gerekenariseerde taalverwerking, en die kunsmatige intelligensie-toepassings wat daaruit voortvloei, tot nasionale ondernemings verklaar het.

Die terrein van gerekenariseerde taalverwerking is wyd. Vir die doeleindes van die LEXINET-onderzoek is dit onderverdeel in sewe deelterreine, soos weerspieël in die titels van die sewe verslae wat nou verskyn (die verslae se noemname staan in hakies):

- Die rekenarisering van leksikografiese prosesse (WORDNET)
- Die rekenarisering van terminografiese prosesse (TERMNET)
- Rekenaargesteuende vertaling (TRANSNET)
- Rekenaartoepassings in die taalwetenskap (TEXTNET)
- Kunsmatige intelligensie en die prosessering van natuurlike taal (AILANG)
- Rekenaarapparatuur en -programmatuur (PROLANG)
- Gerekenariseerde taaldokumentasie-databasisse (DOCNET)

Die sewe deelterreine hou met mekaar op verskillende wyses nou verband. Die samehang tussen hulle kan kortweg geïllustreer word aan die hand van toepassings op die gebied van mens-masjienraakvlakstelsels. 'n Stelsel wat aan gebruikers die moontlikheid bied om in gewone taal navrae aan 'n databasis te stel en om antwoorde in gewone taal te ontvang, moet onder andere komponente bevat wat taal kan ontleed, data kan sorteer en taal kan genereer. Hierdie funksionering veronderstel



- gepaste apparatuur en programmatuur (die gebied wat deur die PROLANG-verslag gedek word);
- een of meer taaldatabasisse (die onderwerp van die DOCNET-verslag);
- een of meer woordeboeke (die TERMNET- en WORDNET-verslae bespreek onder meer ingeboude woordeboeke);
- 'n taalontleider ('parser') en -genereerder (sodanige ontleders is kernkomponente in vertaalprogramme, soos onder andere bespreek in die TRANSNET-verslag);
- kunsmatige intelligensie tegnieke ten opsigte van taalverwerking (die gebied gedek deur die AILANG-verslag).

Hierdie illustrasie aan die hand van 'n bepaalde tipe stelsel is maar een voorbeeld van die samehang tussen die deelterreine. In die verslae word die saak verder bespreek, maar elke terrein word ook in eie reg behandel. So byvoorbeeld is daar in die verslag oor rekenaargesteuende taalondersoek (die TEXTNET-verslag) enersyds sprake van die rekenaar in diens van taalnavorsing en andersyds bespreking van taalnavorsing in diens van gerekenariseerde taalverwerking.

Die aard van die LEXINET-ondersoek, naamlik 'n verkennings- en doenbaarheidstudie, bring mee dat die verslae eerder as inligtingsdokumente, nie as navorsingsverslae nie, beskou kan word. Ter ondersteuning van hierdie inligtingsfunksie word sommige van die verslae gekoppel aan 'n reeks meer tegniese bronstukke wat van die RGN se Afdeling Leksikologie bestel kan word.

Die inhoud van al die verslae word saamgevat in 'n hoofverslag getiteld LEXINET en die rekenarisering van taal (Engelse weergawe: LEXINET and the computer processing of language).

## 1.2 TAAKOMSKRYWING

### 1.2.1 Navorsingsdoelstellings

Die breë doelstellings met die LEXINET-program was om ondersoek in te stel na

- (i) die stand van die gebruik van die rekenaartegnologie in die verwerking van taaldata;
- (ii) huidige leemtes en toekomstige behoeftes in bogenoemde verband;

- (iii) die implikasies van (i) en (ii) vir die bestaande mannekrag, infrastrukturele en finansiële situasie in Suid-Afrika;
- (iv) prioriteitsgebiede vir die verdere ontwikkeling van rekenaar-ondersteunde taalverwerking.

Hierdie breë doelstellings is vir die WORDNET-projek verbesonder na 'n ondersoek na:

- (i) die stand van woordeboekproduksie vir al die hooftaalgroepe in Suid-Afrika;
- (ii) die huidige gebruik van die rekenaartegnologie in Suid-Afrika vir die saamstel, byhou en druk van woordeboeke;
- (iii) die gebruik van die rekenaartegnologie vir woordeboekproduksie oorsee en navorsing oor die gerekenariseerde leksikografie aldaar;
- (iv) die huidige en toekomstige probleme en behoeftes in Suid-Afrika in die breë veld van die algemene leksikografie waarby rekenarisering 'n bydrae tot oplossings kan lewer.

#### 1.2.2 Agtergrond

Bogenoemde navorsingsdoelstellings spruit direk uit:

- (i) 'n aantal probleme wat tans met die voorsiening van gepaste woordeboeke vir al die inligtingsfunksies in ons meertalige taalsituasie in Suid-Afrika ervaar word;
- (ii) die wyse waarop daar in die buiteland met die hulp van die bestaande rekenaartegnologie van hierdie probleme die hoof gebied word.

In die ondersoek na leksikografie-aangeleenthede in Suid-Afrika wat gedurende 1984-1985 uitgevoer is (vgl. Eksteen 1985; voortaan LEXI-1-verslag of LEXI-1-ondersoek), het dit duidelik geblyk dat die geïnstitutionaliseerde leksikografie-ondernemings en die privaatuitgewers nie in die behoefte aan gepaste woordeboeke en woordeboekdienste vir Suid-Afrika kan voorsien nie. Van die faktore wat as direkte oorsaak hiervan geïdentifiseer is, is die volgende:

- (i) die hoë koste verbonde aan die saamstel, byhou en uitgee van al die soorte woordeboeke en die gebrek aan voldoende finansies vir die oprigting en instandhouding van woordeboekondernemings;

- (ii) die gebrek aan opgeleide mannekrag vir woordeboekproduksie en navorsing oor die leksikografie;
- (iii) die bestaande metodes wat vir die saamstel en uitgee van woordeboeke gebruik word, in die besonder die feit dat die rekenaartegnologie nie op uitgebreide skaal vir woordeboekproduksie gebruik word nie.

Uit die LEXI-1-verslag blyk dit ook duidelik dat daar in die algemeen 'n geringskatting bestaan van die noodsaak van adekwate woordeboekvoorsiening in al die hooftaalgroepe in Suid-Afrika vir die instandhouding van effektiewe taaldienste en kommunikasie tussen die verskillende taalgroepe.

Die noodsaak vir die uitvoering van die WORDNET-projek spruit uit die besondere navorsingstemas wat in die LEXI-1-onderzoek ter sake gestel word. In die LEXI-1-onderzoek is daar veral gelet op probleme wat in die staatsondersteunde woordeboekprojekte ervaar word en daar is in die besonder ingegaan op die probleme en behoeftes van woordeboekvoorsiening vir die twee amptelike landstale.

Die leksikografiese aktiwiteite in Suid-Afrika vind oor 'n breë spektrum plaas en word uitgevoer deur verskeie individue, akademiese en nie-akademiese instansies, privaatuitgewers en staatsondersteunde woordeboekkantore en taaldienkantore. Die beperkinge van die LEXI-1-onderzoek het dus die behoefte geskep om 'n omvattender beeld te kry van alle persone en instansies wat aktief aan die produksie van woordeboeke werk. Die leemtes in die woordeboekvoorsiening vir die Afrikatale is nie in diepte ondersoek nie, alhoewel daar in die LEXI-1-verslag klem gelê word op die groeiende behoefte aan sodanige woordeboeke en taaldienste met die uitbreiding van die gebruik van hierdie tale op ekonomiese, politieke, maatskaplike en sosiale vlak, en vanweë die beginsel van moedertaalonderwys wat in Suid-Afrika gehandhaaf word.

Die ongelyke stand van woordeboekvoorsiening vir die Afrikatale moet in die eerste plek in die historiese konteks van die gebrek aan hoëvlakfunksies van hierdie tale gesien word. Daarbenewens bestaan daar allerlei probleme met die standaardisering van die skryfstelsels van verskeie van die Afrikatale. Die uitbreiding van die hoëvlakfunksies van hierdie tale, die toenemende onderrig van die Afrikatale aan nie-moedertaalsprekers, die standaardiseringspogings van die Taalrade en die bestaande kommunikasieprobleme tussen die verskillende taalgroepe het die behoefte aan doeltreffende woordeboekvoorsiening en die identifisering van die behoeftes en probleme in hierdie verband noodsaaklik gemaak.

Gedurende die ondersoek in 1984-1985 is daar ook gevind dat daar by woordeboekinstansies 'n gebrekkige kennis bestaan van die moontlikhede wat die rekenaartegnologie vir die saamstel, byhou en druk van woordeboeke en vir die hergebruik van woordeboekmateriaal bied. Hierdie situasie is vinnig aan die verander met die verspreiding van navorsing oor die rekenaargesteuende leksikografie, die toename in die gebruik van rekenaarset-en-drukapparaat en die gebruik van mikro- en makro-

rekenaars vir die verwerking van teksmateriaal (waaronder die leksikografies-bewerkte materiaal vir woordeboeke). Dit het egter geblyk dat daar in Suid-Afrika nog 'n groot behoefte aan inligting oor die rekenaargesteunde leksikografie bestaan. Vir hierdie doel is daar dus besluit dat daar in die WORDNET-navorsingsverslag aandag aan die voordele van die rekenaargesteunde leksikografie gegee sal word, ten spyte van die vinnige ontwikkelinge op hierdie gebied wat enige so 'n poging noodwendig as onvolledig en verouderd kan laat voorkom.

Een van die eerste take van die WORDNET-projek was egter om te bepaal wat die stand van die rekenaargesteunde leksikografie in Suid-Afrika is, en om dan teen die agtergrond van die gebruik van die tegnologie oorsee te bepaal hoe die rekenaar op 'n effektiewe wyse in die Suid-Afrikaanse leksikografie benut sou kon word om tot 'n verbeterde leksikografiese metodiek en 'n verhoging in die produksie van noodsaaklike woordeboeke te lei.

Daar word tans oorsee verskeie navorsingsprojekte geloods vir die ontwikkeling van outomatiese taalontledingstelsels, waarvan woordeboeke van uiteenlopende aard van die hoofkomponente vorm. Die ontwikkeling van hierdie stelsels en die gebruik daarvan in Suid-Afrika sal in 'n groot mate bepaal word deur die beskikbaarheid van masjienleesbare woordeboeke vir al die hooftale. Die potensiele gebruiksmoontlikhede van hierdie tegnologie is ook ondersoek en as leidraad gebruik vir die bepaling van die toekomstige behoeftes en probleme wat daar in die leksikografiepraktyk in Suid-Afrika ervaar kan word. Vergelyk in hierdie verband ook die TERMNET- en TEXTNET-verslae.

### 1.2.3 Werkwyse

Vir die uitvoering van die basiese navorsing vir die WORDNET-projek is 'n span saamgestel wat uit die volgende lede bestaan het:

Prof P H Swanepoel, Projekleier (Dept. Afrikaans en Nederlands, UNISA)  
Dr R Morris, Programleier (RGN)  
Prof P S Groenewald (Dept. Afrikatale, UP)  
Dr D J Prinsloo (Dept. Afrikatale, UP)  
Mev M Alberts (RGN)  
Mev P Mtintsilana (RGN)  
Mej H Retief (RGN)

Hierbenewens het die volgende persone in 'n adviserende hoedanigheid opgetree:

Dr J Branford (Dictionary of South African English, Rhodes)  
Prof W R G Branford (Dept of Linguistics and English Language, Rhodes)  
Mnr D C Hauptfleisch (Buro van die WAT, US); en  
Mnr P T Mtuze (Xhosa-woordeboek, Univ. van Fort Hare).

Prof P H Swanepoel en dr R Morris het die eindverslag opgestel.

Die materiaal vir die verslag is versamel uit die volgende bronne: bestaande literatuur oor die Suid-Afrikaanse leksikografie, literatuur en verslae oor leksikografie-ondernemings oorsee en uit die LEXO-

dokumentasiedatabasis van die RGN (oor woordeboek-aangeleenthede). Die gegewens in die dokumentasiedatabasis is verder bygewerk met behulp van 'n vraelys wat gestuur is aan persone en instansies wat in die leksikografie betrokke is. Verder het spanlede besoek gebring aan woordeboekkantore, uitgewers en drukkers om met hulle die stand van rekenaargebruik te bespreek.

Vanweë onderlinge konkurrensie tussen leksikograwe is sommiges nie bereid om inligting te verskaf nie. Daar kan dus geen volledige profiel van die rekenaargesteuende leksikografie in Suid-Afrika opgestel word nie, maar daar is gepoog om gegewens oor die aktiwiteite van die meeste groot uitgewers, woordeboekkantore en -projekte en individue vir die ondersoek te bekom. Hierdie gegewens is gebruik om 'n aantal algemene tendense en die probleme en behoeftes in die Suid-Afrikaanse leksikografie te identifiseer.

## HOOFSTUK 2: STANDBESKRYWING

### 2.1 WOORDEBOEKVOORSIENING

Vanweë die ooglopende verskille in die voorsiening van een- en meertalige woordeboeke vir die verskillende taalgroepe in Suid-Afrika, word die gegewens in verband met woordeboekvoorsiening en die leemtes in hierdie verband hier onder per taalgroep verstrekkend en bespreek. Hierdie uiteensetting berus op die bestaande inligting in die LEXO-dokumentasiedatabasis en die inligting wat in hierdie verband met behulp van 'n vraelys bekom is. Die gegewens wat hier verstrekkend word, sluit ook woordeboeke in wat uitdruk is, en in dié opsig sou die gegewens misleidend kon wees ten opsigte van die beskikbare woordeboeke. In die samestelling van die lys is daar nie gepoog om die woordeboeke na hulle leksikografiese gehalte of hulle bruikbaarheid te gradeer of te differensieer nie. Die span het wel in hulle bepaling van die bestaande leemtes in die woordeboekvoorsiening die gehalte en beskikbaarheid van die woordeboekvoorraad in aanmerking geneem. 'n Volledige tipologie van die betrokke woordeboeksoort word gegee in Eksteen (1985:30-35).

#### 2.1.1 Afrikaans

##### 2.1.1.1 Eentalige woordeboeke

sprekwoorde/idiome/uitdrukkings/gesegdes:	4
etimologieë:	4
sinonieme, antonieme:	1
uitspraak:	2
rjm:	
retrograde:	1
verklarend:	19
frekwensietellings:	2
spelling:	2
sakwoordeboek:	1
blokkiesraaiselhulp:	2
skoolwoordeboeke:	5

Uit druk: 7

Lopende projekte: 7

Rekenaarbetrokkenheid: 6

##### 2.1.1.2. Tweetalige woordeboeke met Afrikaans as bron- of doeltaal

ARABIES/AFRIKAANS = 1

BOESMAN (!XU)/AFRIKAANS = 1

DUIITS/AFRIKAANS = 11

ENGELS/AFRIKAANS

sakwoordeboeke:	5
skoolwoordeboeke:	4
spreekwoordeboeke:	3
gebaretaal:	1
afkortings:	1
voorsetsels:	2
verklarend:	1

Uit druk: 19

Lopende projekte: 1

Hersienings: 4

Rekenaarbetrokkenheid: Geen

AFRIKAANS/FANAKALO = 1

AFRIKAANS/FRANS = 2 (1 lopende projek; rekenaarbetrokkenheid)

AFRIKAANS/GRIEKS = 1 (woordelys; rekenaarbetrokkenheid)

AFRIKAANS/HEBREEUS = 1 (lopend)

AFRIKAANS/HOTTENTOTS = 1 (uit druk)

AFRIKAANS/LATYN = 1

AFRIKAANS/NDONGA = 1

AFRIKAANS/NEDERLANDS = 1

AFRIKAANS/NOORD-SOTHO = 5 (1 spelling; 1 word hersien; 1 lopend, rekenaarbetrokkenheid)

AFRIKAANS/PORTUGEES = 5 (3 uit druk)

AFRIKAANS/SUID-SOTHO = 4 (1 skoolwoordeboek; 1 lopend)

AFRIKAANS/TSONGA = 1 (uit druk)

AFRIKAANS/TSWANA = 1 (lopend)

AFRIKAANS/VENDA = 4

AFRIKAANS/XHOSA = 4 (1 uit druk)

AFRIKAANS/ZULU = 3 (1 lopend; rekenaarbetrokkenheid)

2.1.1.3. Meertalige Woordeboeke met Afrikaans as een van die tale

AFRIKAANS/AFRIKATALE/ENGELS = 2

AFRIKAANS/DUIITS/HERERO/NDONGA = 1

AFRIKAANS/DUIITS/NAMA = 1

AFRIKAANS/ENGELS/DUIITS = 1  
AFRIKAANS/ENGELS/FANAKALO = 1  
AFRIKAANS/ENGELS/FRANS = 1  
AFRIKAANS/ENGELS/HERERO = 2  
AFRIKAANS/ENGELS/KWANYAMA = 2  
AFRIKAANS/ENGELS/LATYN = 1  
AFRIKAANS/ENGELS/NAMA/DAMARA = 1  
AFRIKAANS/ENGELS/NDONGA = 1  
AFRIKAANS/ENGELS/NOORD-SOTHO = 11  
AFRIKAANS/ENGELS/POOLS = 1  
AFRIKAANS/ENGELS/SUID-NDEBELE = 1  
AFRIKAANS/ENGELS/SUID-SOTHO = 7  
AFRIKAANS/ENGELS/SWAZI = 1  
AFRIKAANS/ENGELS/TSONGA = 4  
AFRIKAANS/ENGELS/TSWANA = 6  
AFRIKAANS/ENGELS/VENDA = 8  
AFRIKAANS/ENGELS/XHOSA = 9  
AFRIKAANS/ENGELS/XHOSA/ZULU = 1  
AFRIKAANS/ENGELS/ZULU = 5

Uit druk: 4

Lopende projekte: 9

Rekenaarbetrokkenheid: 2

Uit die bostaande gegewens is dit duidelik dat daar in 'n groot mate voorsien word in die behoefte aan een- en meertalige (Afrikaans-Engelse) woordeboeke vir Afrikaans. Daar bestaan egter ook 'n aantal leemtes waarin daar in die toekoms voorsiening gemaak sal moet word, naamlik:

- (i) die versnelde voltooiing en bywerk van die Woordeboek van die Afrikaanse Taal;
- (ii) gegradeerde, eentalige pedagogiese woordeboeke van Afrikaans vir die onderrig van Afrikaans aan nie-moedertaalsprekers;



- (iii) die daarstel van 'n leksikografies-bewerkte korpus van Afrikaans wat as basis kan dien vir die saamstel van 'n reeks brugwoordeboeke met die Afrikatale as doeltaal respektiewelik teiken-taal;
- (iv) gespesialiseerde woordeboeke oor aspekte van die woordeskat van Afrikaans, in die besonder 'n omvattende historiese woordeboek en 'n woordeboek van die streeksvariante van Afrikaans.

Daarbenewens bestaan daar 'n behoefte aan die daarstelling van 'n rekenaartoeganklike, omvattende leksikografiese databasis van Afrikaans wat vir die doeleindes van leksikografiese en leksikologiese navorsing gebruik kan word en wat as basis kan dien vir die saamstel van gespesialiseerde woordeboeke oor Afrikaans. Met die oog op die implementering van outomatiese teksontledingsisteme en die saamstel van rekenaarwoordeboeke vir hierdie doel, sal die behoefte aan so 'n gerekenariseerde leksikografiese databasis in die toekoms toeneem.

### 2.1.2 Engels (slegs plaaslik opgestelde woordeboeke)

#### 2.1.2.1 Eentalige woordeboeke

spreekwoorde/idiome/uitdrukkings/gesegdes:	3
sleng:	1
gebruiks (usage):	1
SA English:	7
sinonieme, antonieme:	1
afkortings, punktuasie:	1
semantiese velde:	1
frekwensietelling:	1
sakwoordeboek:	1
skoolwoordeboeke:	3

Uit druk: Geen  
 Lopende projekte: 5 (1 word hersien)  
 Rekenaarbetrokkenheid: 1

#### 2.1.2.2 Tweetalige woordeboeke met Engels as bron- of doeltaal

##### ENGELS/AFRIKAANS

sakwoordeboeke	= 5
skoolwoordeboeke	= 4
spreekwoordeboeke	= 3
gebaretaal	= 1
afkortings	= 1
voorsetsels (1 word hersien)	= 2
verklarend (uit druk)	= 1

Uit druk: 19  
 Lopende projekte: 1  
 Hersienings: 4

Rekenaarbetrokkenheid: Geen

BOESMAN/ENGELS = 1 (uit druk)

COPI/ENGELS = 1 (lopend)

ENGELS/(LILIMA) KALENGA = 1 (lopend)

ENGELS/AUNI = 1 (uit druk)

ENGELS/FANAKALO = 3 (1 uit druk)

ENGELS/GITONGA = 1 (lopend)

ENGELS/GRIEKS = 2 (albei lopend; 1 rekenaarbetrokkenheid)

ENGELS/HEBREUS = 1 (lopend)

ENGELS/HERERO = 5 (3 uit druk)

ENGELS/KORANA = 1 (uit druk)

ENGELS/KWANYAMA = 5 (1 lopend)

ENGELS/LAMBA = 3 (2 uit druk)

ENGELS/NAMA = 2 (1 lopend; 1 uit druk)

ENGELS/NDEBELE = 1

ENGELS/NDONGA = 1

ENGELS/NEDERLANDS = 1 (uit druk)

ENGELS/NOORD-SOTHO = 12 (1 uit druk; 1 sakwoordeboek)

ENGELS/PORTUGEES = 1 (lopend)

ENGELS/PROTO-BANTOE = 2 (lopend)

ENGELS/SWAZI = 6 (1 lopend; 1 uit druk)

ENGELS/SOTHO = 4 (1 uit druk)

ENGELS/SUID-SOTHO = 21 (8 uit druk; 1 sakwoordeboek; 1 idiome; 1 lopend)

ENGELS/TONGA = 4 (3 uit druk)

ENGELS/TSONGA = 10 (3 sakwoordeboeke; 1 uit druk)

ENGELS/TSWA = 1 (uit druk)

ENGELS/TSWANA = 9 (3 uit druk; 1 lopend)

ENGELS/VENDA = 10 (1 lopend; 2 uit druk)

ENGELS/XHOSA = 19 (3 lopend; 1 skoolwoordeboek; 6 uit druk)

ENGELS/ZULU = 40 (16 uit druk; 1 hersien)

### 2.1.2.3 Meertalige woordeboeke met Engels as een van die tale

AFRIKAANS/ENGELS/DUIITS = 1

AFRIKAANS/ENGELS/FANAKALO = 1

AFRIKAANS/ENGELS/FRANS = 1

AFRIKAANS/ENGELS/HERERO = 2

AFRIKAANS/ENGELS/KWANYAMA = 2

AFRIKAANS/ENGELS/LATYN = 1

AFRIKAANS/ENGELS/NAMA/DAMARA = 1

AFRIKAANS/ENGELS/NDONGA = 1

AFRIKAANS/ENGELS/NOORD-SOTHO = 11

AFRIKAANS/ENGELS/POOLS = 1

AFRIKAANS/ENGELS/SUID-NDEBELE = 1

AFRIKAANS/ENGELS/SUID-SOTHO = 7

AFRIKAANS/ENGELS/SWAZI = 1

AFRIKAANS/ENGELS/AFRIKATALE = 2

AFRIKAANS/ENGELS/TSONGA = 4

AFRIKAANS/ENGELS/TSWANA = 6

AFRIKAANS/ENGELS/VENDA = 8

AFRIKAANS/ENGELS/XHOSA = 9

AFRIKAANS/ENGELS/XHOSA/ZULU = 1

AFRIKAANS/ENGELS/ZULU = 5

ARAMEES/ENGELS/HEBREEUS = 1

BIISA/ENGELS/LALA/LAMBA = 1

BOESMAN/ENGELS = 1

BOESMAN/ENGELS/NAMA = 1

COPI/ENGELS = 1

COPI/ENGELS/TSWA = 1

ENGELS/KALANGA/NDEBELE = 1

ENGELS/LALA/LAMBA = 1

ENGELS/NDEBELE/SHONA = 2

ENGELS/TONGA/NAMBZYA = 1

ENGELS/TONGA/TUMBUKA = 1

ENGELS/SOTHO/XHOSA/ZULU = 1

ENGELS/XHOSA/ZULU = 3

ENGELS/G//ANA/ G/WI = 1

Uit druk: 8

Lopende projekte: 10

Rekenaarbetrokkenheid: 2

In die breë word in gebruikersbehoefte aan eentalige Engelse woordeboeke in Suid-Afrika voorsien deur ingevoerde woordeboeke soos die Oxford-, Webster-, Longmans- en Collins-reeks. Boekhandelaars bepaal volgens verkope watter woordeboeke ingevoer word, wat nie noodwendig beteken dat daar in alle behoeftes voorsien word nie. Daarbenewens bestaan daar 'n behoefte aan woordeboeke wat die immeruitbreidende Suid-Afrikaans-Engels woordeskat dek. 'n Aanduider van hierdie behoefte is dat daar van die Dictionary of South African English, wat in 1978 verskyn het, nou reeds 'n derde uitgawe in die pers is. Die tweede uitgawe is geruime tyd al uitverkoop. Terselfdertyd word die groot woordeboek, die Dictionary of South African English on Historical Principles, waarvan bogenoemde woordeboek die voorloper is, met groot belangstelling ingewag. Die afhandeling van die groot woordeboek word steeds vertraag deur die mannekrag- en verwerkingsprobleme wat in Eksteen (1985) beskryf is. Opsommend kom die probleme hoofsaaklik op twee neer. Eerstens het die Woordeboekenheid by Rhodes, waar die woordeboek opgestel word, tot heel onlangs in 'n onsekere situasie verkeer wat die deurlopende finansiering van die kantoor betref. Die gevolg is dat personeelwerwing met die vooruitsig van permanente aanstellings nie aangegaan kon word nie.

Die tweede hoofprobleem wat by die Rhodes-teenheid ervaar word, hou regstreeks verband met die huidige ondersoek na rekenarisering. Die omvang van die materiaal vir die groot woordeboek bring mee dat die verwerking en die selektering en voorbereiding daarvan vir opname in woordeboekformaat 'n reusetaak is. Daarbenewens is dit 'n besonder

tydrowende proses met die huidige handverwerkingsmetode om byvoegings tot en hersienings van itemopgawes te doen.

'n Leemte wat daar tans ten opsigte van die Engelse woordeboekmark ervaar word, is die gebrek aan 'n reeks gegradeerde, eentalige, pedagogiese woordeboeke vir die onderrig van Engels aan nie-moedertaalsprekers, in die besonder dan vir die moedertaalsprekers van die Afrikatale. Soos uit die bespreking hieronder sal blyk, bestaan daar ook 'n behoefte aan 'n reeks vertalende woordeboeke met Engels en elkeen van die hoof Afrikatale as doeltaal respektiewelik teikentaal.

Die kulturele verskille tussen die sprekers van Engels en die moedertaalsprekers van die Afrikatale word nie op voldoende wyse in die bestaande eentalige pedagogiese woordeboeke ondervang nie.

Uit die bogenoemde sou dit ook as taak van die Rhodes-woordeboekenheid gesien kan word om 'n basiese, leksikografies-verwerkte korpus daar te stel wat as basis kan dien vir die saamstel van bogenoemde twee woordeboeksoorte.

### 2.1.3 Afrikatale

#### 2.1.3.1 Eentalige woordeboeke

HOTTENTOTS = 1 (2 dele)

NOORD-SOTHO = 2

SUID-SOTHO = 2

TSWANA = 4

VENDA = 6

XHOSA = 3

ZULU = 15

Uit druk: 11

Lopende projekte: 3

Rekenaarbetrokkenheid: Geen

#### 2.1.3.2 Tweetalige woordeboeke met 'n Afrikataal as bron- of doeltaal

BOESMAN (!XU)/AFRIKAANS = 1

AFRIKAANS/FANAKALO = 1

AFRIKAANS/HOTTENTOTS = 1 (uit druk)

AFRIKAANS/NDONGA = 1

AFRIKAANS/NOORD-SOTHO = 5 (1 spelling; 1 word hersien; 1 lopend; 1 rekenaarbetrokkenheid)

AFRIKAANS/SUID-SOTHO = 4 (1 skoolwoordeboek; 1 lopend; 1 uit druk)

AFRIKAANS/TSONGA = 1 (uit druk)

AFRIKAANS/TSWANA = 1 (lopend)

AFRIKAANS/VENDA = 4 (3 uit druk)

AFRIKAANS/XHOSA = 4 (1 uit druk)

AFRIKAANS/ZULU = 3 (1 lopend; rekenaarbetrokkenheid)

BOESMAN/DUIITS = 1

BOESMAN/ENGELS = 1

COPI/ENGELS = 1 (lopend)

COPI/PORTUGÊES = 1

DUIITS/HERERO = 4

DUIITS/KORANA = 1 (uit druk)

DUIITS/KWANYAMA = 1

DUIITS/NAMA = 8 (1 skoolwoordeboek)

DUIITS/SOTHO = 1 (uit druk)

DUIITS/VENDA = 1 (uit druk)

DUIITS/XHOSA = 1 (uit druk)

ENGELS/(LILIMA) KALENGA = 1 (lopend)

ENGELS/AUNI = 1 (uit druk)

ENGELS/FANAKALO = 3 (1 uit druk)

ENGELS/GITONGA = 1 (lopend)

ENGELS/HERERO = 5 (3 uit druk)

ENGELS/KORANA = 1 (uit druk)

ENGELS/KWANYAMA = 5 (1 lopend)

ENGELS/LAMBA = 3 (2 uit druk)

ENGELS/NAMA = 2 (1 uit druk; 1 lopend)

ENGELS/NDEBELE = 1  
ENGELS/NDONGA = 1  
ENGELS/NOORD-SOTHO = 12 (1 sakwoordeboek; 1 uit druk)  
ENGELS/PROTO-BANTOE = 2 (lopend)  
ENGELS/SWAZI = 6 (1 lopend; 1 uit druk)  
ENGELS/SUID-SOTHO = 21 (8 uit druk; 1 sakwoordeboek; 1 idiome; 1 lopend)  
ENGELS/TONGA = 4 (3 uit druk)  
ENGELS/TSONGA = 10 (3 sakwoordeboeke; 1 uit druk)  
ENGELS/TSWA = 1 (uit druk)  
ENGELS/TSWANA = 9 (3 uit druk; 1 lopend)  
ENGELS/VENDA = 10 (1 lopend; 2 uit druk)  
ENGELS/XHOSA = 19 (3 lopend; 1 skoolwoordeboek; 6 uit druk)  
ENGELS/ZULU = 40 (6 uit druk, 1 hersien)  
FRANS/ZULU = 1 (uit druk)  
KWANYAMA/PORTUGEES = 1  
PORTUGEES/TSONGA = 1  
PORTUGEES/TSWA = 2  
SOTHO/TSWANA = 1 (uit druk)  
XHOSA/ZULU = 2 (1 uit druk)  
Uit druk: 55  
Lopende projekte: 18  
Rekenaarbetrokkenheid: 2

#### 2.1.3.3 Meertalige woordeboeke

AFRIKATALE = 1 (verklarend)  
AFRIKAANS/DUIITS/HERERO/NDONGA = 1  
AFRIKAANS/DUIITS/NAMA = 1  
AFRIKAANS/ENGELS/FANAKALO = 1  
AFRIKAANS/ENGELS/HERERO = 2

AFRIKAANS/ENGELS/KWANYAMA = 2  
AFRIKAANS/ENGELS/NAMA/DAMARA = 1  
AFRIKAANS/ENGELS/NDONGA = 1  
AFRIKAANS/ENGELS/NOORD-SOTHO = 11  
AFRIKAANS/ENGELS/SUID-NDEBELE = 1  
AFRIKAANS/ENGELS/SUID-SOTHO = 7  
AFRIKAANS/ENGELS/SWAZI = 1  
AFRIKAANS/ENGELS/AFRIKATALE = 2  
AFRIKAANS/ENGELS/TSONGA = 4  
AFRIKAANS/ENGELS/TSWANA = 6  
AFRIKAANS/ENGELS/VENDA = 8  
AFRIKAANS/ENGELS/XHOSA = 9  
AFRIKAANS/ENGELS/XHOSA/ZULU = 1  
AFRIKAANS/ENGELS/ZULU = 5  
BIISA/ENGELS/LALA/LAMBA = 1  
BOESMAN/ENGELS/NAMA = 1  
COPI/ENGELS/TSWA = 1  
COPI/PORTUGEES/TSONGA = 2  
DUIITS/HERERO/KWANYAMA/NDONGA = 1  
DUIITS/HERERO/NAMA/NDONGA = 1  
ENGELS/KALANGA/NDEBELE = 1  
ENGELS/LALA/LAMBA = 1  
ENGELS/NDEBELE/SHONA = 2  
ENGELS/SOTHO/XHOSA/ZULU = 1  
ENGELS/TONGA/NAMBZYA = 1  
ENGELS/TONGA/TUMBUKA = 1  
ENGELS/XHOSA/ZULU = 3



ENGELS/G//ANA/ G/WI = 1

PORTUGEES/TSONGA/TSWA = 1

Uit druk : 24

Lopende projekte : 8

Rekenaarbetrokkenheid: 2

Alhoewel bogenoemde lys 'n groot aantal woordeboeke vermeld, is vele daarvan uit druk.

Die gegewens toon dat daar vir elke Afrikataal 'n behoefte aan spesifieke woordeboeksoorte bestaan.

Kategorie 1(a): 'n Groot omvattende verklarende woordeboek.

So 'n woordeboek sal ongeveer die ekwivalent wees van die Woordeboek van die Afrikaanse Taal (WAT), d.w.s. een wat verskeie dele beslaan en slegs in die Afrikataal geskryf is, maar uiters omvattend is. Die betekenis van woorde sal uitvoerig in die betrokke Afrikataal omskryf word. Woorde sal waarskynlik gewoonweg alfabeties ingeskryf word en nie volgens woordstamme nie. Hierdie woordeboek moet beslis rekenaarmatig aangepak word sodat kleiner woordeboeke (sien sekere ander kategorieë hieronder) daaruit gereduseer kan word.

Kategorie 1(b): Verklarende handwoordeboek.

'n Groot verklarende handwoordeboek, wat direk uit kategorie 1(a) gereduseer word, sal alle inskrywings van kategorie 1(a) bevat minus sekere woorde. 'n Tipiese voorbeeld van so 'n woordeboek is die HAT (Verklarende Handwoordeboek van die Afrikaanse Taal).

Kategorie 2: Vertalende woordeboek(e).

'n Groot vertalingswoordeboek Afrikaans-Afrikataal-Afrikaans (een deel) en 'n ekwivalente Engels-Afrikataal-Engels (een deel). Hierdie woordeboek sal direk uit kategorie 1 gereduseer word en sal alle inskrywings van kategorie 1 bevat minus sekere woorde soos byvoorbeeld tegniese terme. Ten einde hierdie relatief groot aantal inskrywings in een deel (elk vir Afrikaans en Engels) te kan akkommodeer, sal in die inskrywing slegs trefwoorde met hulle vertaalekwivalente opgeneem word.

Kategorie 3: 'n Naslaanwoordeboek.

'n Groot woordeboek Afrikataal-Afrikaans/Engels in een deel waarin woorde volgens stamme gerangskik is en geskik sal wees vir Afrikataalstudente. So 'n woordeboek moet verskeie vormvariante, afleidings, kombinasies van afleidings, ensovoorts gee. 'n Tipiese

voorbeeld van só 'n woordeboek is die Groot Noord-Sotho woordeboek van D Ziervogel en P C Mokgokong.

Kategorie 4: 'n Pedagogiese woordeboek.

Hierdie woordeboek moet van hanteerbare formaat (plus-minus 600 bladsye) wees en as 'n tweerigtingwoordeboek (Afrikaans-Afrikataal-Afrikaans, een deel, en ook Engels-Afrikataal-Engels, een deel) opgestel word. Inskrywings sal alfabeties georden wees. Die Afrikaanse of Engelse weergawe sal saam met kategorie 3 deur studente gebruik word.

Kategorie 5:

Benewens 'n versameling terminologiewoordeboeke van die vernaamste vakgebiede, bestaan daar ook 'n behoefte aan kleiner gespesialiseerde woordeboeke van aspekte van die woordeskat van die volgende tale:

Noord-Sotho  
Suid-Sotho  
Tswana  
Zulu  
Xhosa  
Swazi  
Venda  
Tsonga  
Ndebele.

'n Evaluering van die huidige stand van woordeboeke in die Afrikatale, dui op die volgende:

Noord-Sotho

Kategorieë 2 en 4 redelik voldoende, behoefte aan woordeboeke van kategorieë 1(a), 2 en 5.

Suid-Sotho

Kategorie 4 redelik voldoende, behoefte aan woordeboeke van kategorieë 1(a), 1(b), 2, 3 en 5.

Tswana

Behoefte in alle kategorieë.

Zulu

Kategorieë 1(b), 3 en 4 redelik voldoende, behoefte aan woordeboeke van kategorieë 1(a), 2 en 5.

Xhosa

Kategorieë 1(a) en 4 redelik voldoende, behoefte aan woordeboeke van kategorieë 2, 3 en 5.

Swazi

Behoefte in alle kategorieë.

Venda

Kategorie 4 redelik voldoende, behoefte aan woordeboeke in kategorieë 1(a), 1(b), 2, 3 en 5.

Tsonga

Behoeft in alle kategorieë.

Suid-Ndebele

Behoeft in alle kategorieë.

Ten opsigte van Kategorie 4 is die voorsiening vir Suid-Sotho, Xhosa en Venda wat woordeboeke van die tipe Engels/Afrikataal betref, redelik voldoende; daar is egter behoefte aan woordeboeke van die tipe Afrikaans/Afrikataal.

Uit bostaande blyk dit duidelik dat daar 'n behoefte bestaan aan 'n omvattende, eentalig verklarende woordeboek vir elkeen van die Afrikatale. Die vestiging van woordeboekkantore en die daarstelling van omvattende eentalig verklarende woordeboeke vir die Afrikatale sou meer as net 'n simboliese manifestasie van die kulturele en taalidentiteit van die afsonderlike taalgroepe wees omdat dit in 'n aantal diepgevoelde kommunikasiebehoefte sou voorsien. 'n Omvattende leksikografies-bewerkte databasis van elkeen van hierdie tale sou verder as basis kon dien vir die genereer van die kleiner een- en meertalige woordeboeke van kategorieë waaraan daar ook 'n dringende behoefte bestaan.

Tans word daar aan verskeie universiteite in Suid-Afrika gewerk aan die saamstel van woordeboeke om in die bestaande leemtes te voorsien:

#### Universiteit van Pretoria

Woordeboek van die Noord-Sotho taal. Noord-Sotho/Afrikaans.  
(Verklarend, beginstadium, gerekenariseerd.)

#### Potchefstroomse Universiteit vir CHO

Algemene woordelys. Afrikaans/Tswana - Tswana/Afrikaans. (Daar is reeds ongeveer 8 000 Tswana items.)

#### Universiteit van Kaapstad

English/Southern-Sotho Vocabulary. (In voorbereiding.)

#### Universiteit van Suid-Afrika

Northern-Sotho General Vocabulary for beginners.  
Noord-Sotho/English/Afrikaans.

Praktiese lys van Woorde en Uitdrukkings in Noord-Sotho. Practical list of Words and Expressions in Northern-Sotho. Noord-Sotho/English/Afrikaans.

A vocabulary of Xhosa with English equivalents. Xhosa/Engels-woordeboek.

Tswana-woordeboek. Afrikaans/Engels/Tswana-woordeboek.

Universiteit van Fort Hare.

Xhosa/English/Afrikaans dictionary.

Universiteit van die Oranje-Vrystaat

Bukantswe e maleme-pedi. SeSotho/SeAfrikaans- SeAfrikaans/SeSotho.

Universiteit van Zululand

Afrikaans/Zulu/Engels-woordeboek. (Algemene woordeskat, tuisrekenaar.)

Zulu verklarende woordeboek.

Universiteit van die Noorde

Engels/Venda-woordeboek. (In manuskripvorm - Juta.)

Randse Afrikaanse Universiteit

Noord-Sotho/Afrikaans-Afrikaans/Noord-Sotho-woordeboek. (Hersiening.)

#### 2.1.4 Indiese tale

'n Algemene woordeboek oor die Indiese dialek van Engels wat in Suid-Afrika gebruik word, is deur R Mesthrie saamgestel. Dié woordeboek is tans in die pers en die titel sal waarskynlik South African Indian English Dialect Dictionary wees. Die doel met die woordeboek is om bepaalde aspekte van 'n hoofdialek van Suid-Afrikaanse Engels so sistematies moontlik te beskryf. Die woordeboek bevat hoofsaaklik dialektiese woordeskatinskrywings asook idiomatiese gebruik, uitspraakaanduiding en sekere grammatikale inligting. Die woordeboek is op mikrorekenaarskyf vasgelê.

Dié woordeboek word as glossariumbylaag by die Dictionary of South African English van Jean Branford, bygevoeg.

Daar is nog geen woordeboeke oor die variëteite van die Indiese tale in Suid-Afrika saamgestel nie. Daar word wel aan 'n woordeboek gewerk oor Tamil, soos wat dit in Indië gebruik word. Die samesteller is mnr Gounden van die Departement Tamil, Telugu, Hindi, Gujarati, van die Universiteit Durban-Westville.

Dr Rajend Mesthrie, Departement Linguistiek, Universiteit van Kaapstad, beplan 'n moontlike woordeboek oor Natalse Hindi, soos wat dit oor die afgelope 125 jaar ontwikkel het. Hy beplan ook 'n moontlike woordeboek oor een van die interessante maar minder bekende Indiese tale van die Kaap, naamlik Kokani. Daar sal waarskynlik geringe mikrorekenaarbetrokkenheid by die saamstel van die woordeboeke wees, aangesien die Departement Linguistiek slegs van die woordverwerkingsaspekte van rekenaars gebruik maak.

### 2.1.5 Europese tale

Daar word in Suid-Afrika ook gewerk aan die saamstel van woordeboeke van die verskillende tale van Europa. Die meeste van hierdie woordeboeke is vertalend van aard met Afrikaans of Engels as doeltaal respektiewelik teikentaal. Die meeste van hierdie woordeboeke is gerig op die onderrig en verwerwing van die betrokke tale op primêre, sekondêre en tersiêre onderrigvlak. Die volgende gegewens dui die stand van die woordeboekvoorsiening vir hierdie tale aan.

#### 2.1.5.1 Eentalige woordeboeke

GRIEKS = 1

LATYN = 2 (1 lopend)

NEDERLANDS = 1 (uit druk)

#### 2.1.5.2 Tweetalige woordeboeke

DUITS/AFRIKAANS = 11

AFRIKAANS/FRANS = 2

AFRIKAANS/GRIEKS = 1

AFRIKAANS/HEBREEUS = 1

AFRIKAANS/LATYN = 1

AFRIKAANS/NEDERLANDS = 1

AFRIKAANS/PORTUGEES = 4

BOESMAN/DUIITS = 1  
COPI/PORTUGEES = 1  
DUIITS/HERERO = 4  
DUIITS/KORANA = 1  
DUIITS/KWANYAMA = 1  
DUIITS/NAMA = 8  
DUIITS/SOTHO = 1  
DUIITS/VENDA = 1  
DUIITS/XHOSA = 1  
ENGELS/GRIEKS = 2  
ENGELS/HEBREEUS = 1  
ENGELS/NEDERLANDS = 1  
ENGELS/PORTUGEES = 1  
FRANS/ZULU = 1  
KWANYAMA/PORTUGEES = 1  
PORTUGEES/TSONGA = 1  
PORTUGEES/TSWA = 2

Uit druk: 18  
Lopend: 6  
Rekenaarbetrokkenheid: 2

#### 2.1.5.3 Woordeboeke met 'n Europese taal as een van die komponente

AFRIKAANS/DUIITS/HERERO/NDONGA = 1  
AFRIKAANS/DUIITS/NAMA = 1  
AFRIKAANS/ENGELS/DUIITS = 1 (idiome; uit druk)  
AFRIKAANS/ENGELS/FRANS = 1 (lopend; rekenaarbetrokkenheid)  
AFRIKAANS/ENGELS/LATYN = 1 (kernwoordelys)  
AFRIKAANS/ENGELS/POOLS = 1 (rekenaardrukstuk)  
ARAMEES/ENGELS/HEBREEUS = 1 (lopend)

COPI/PORTUGEEES/TSONGA = 2 (uit druk)

DUIITS/HERERO/KWANYAMA/NDONGA = 1 (uit druk)

DUIITS/HERERO/NAMA/NDONGA = 1 (uit druk)

PORTUGEEES/TSONGA/TSWA = 1 (uit druk)

Uit druk: 6

Lopend: 2

Rekenaarbetrokkenheid: 2

Aangesien 'n woordeboek 'n bekendstellingsmiddel is vir 'n taal sou dit sinvol wees om meer aandag te skenk aan die saamstel van woordeboeke met een van die twee amptelike landstale of Afrikataal as bron- of doeltaal en 'n westerse taal as die ander komponent. Die Portugees-Afrikaanse woordeboek bedien byvoorbeeld die groot Portugese bevolking se kommunikasiebehoefte. Dié woordeboek voldoen ook aan die vereistes wat aan 'n naslaanwerk vir studiedoeleindes gestel word. Daar is soortgelyke behoeftes aan woordeboeke met Italiaans en met Spaans as komponente en die bestaande woordeboeke met Duits en Afrikaans behoort hersien te word.

#### 2.1.6 Opsomming

In Suid-Afrika is die private verskaffers van leksikografiedienste en -produkte 'n besonder belangrike bondgenoot van die eentalige nie-private leksikografie-instansies (of hoofwoordeboeke) wat finansiële gedeeltelik of in die geheel deur die staat ondersteun word.

Die ryk verskeidenheid woordeboeksoorte, die aantal spraakgemeenskappe en die gebruiksdoeleindes waarvoor daar reeds deur hierdie bondgenootskap voorsiening gemaak word, asook die leemtes in hierdie verband, blyk duidelik uit die voorgaande standbeskrywing.

Die maatskaplike, wetenskaplike en kultuurpolitieke belang van die hoofwoordeboeke vereis dat daar in 'n breër (nasionale) leksikografiebeleid nie alleen voorsiening gemaak moet word vir die voortgesette instandhouding van die bestaande hoofwoordeboeke nie, maar ook vir die ontwikkeling van ander soortgelyke projekte, in die besonder vir die Afrikatale.

Die maatskaplike belang van die hoofwoordeboeke lê daarin dat hulle leksikografiehulpmiddele skep wat kommunikasie tussen die lede van 'n taalgemeenskap vergemaklik en help met die verspreiding van taalkennis en taalnorme.

In teenstelling met die projekte van handelsondernemings (die private projekte) bou die hoofwoordeboeke 'n uitgebreide materiaalversameling op wat 'n kwalitatiewe basis kan vorm vir alle woordeboekarbeid oor die betrokke tale. Staatsfinansiering is noodsaaklik sodat die materiaal los van ekonomiese oorwegings met wetenskaplike grondigheid bestudeer en bewerk kan word.

Die wetenskaplike voordeel lê daarin dat van die hoofwoordeboekondernemings woordeboeke saamstel wat - in vergelyking met soortgelyke private produkte - uitgebreider, vollediger en grondiger gestruktureerde beskrywings van die trefwoorde gee. So 'n hoofwoordeboek vorm 'n onmisbare deel van die taalkundige beskrywing van 'n taal omdat werk aan al die fasette van die leksikologie van 'n taal gewoonlik die intensiefste deur hierdie instansies verrig word.

Die kultureel-politieke belang lê daarin dat 'n hoofwoordeboek uitdrukking gee aan en 'n bevestiging is van die kulturele identiteit van 'n taalgemeenskap (vgl. Geeraerts 1984:9-11).

Die groot nadeel van die hoofwoordeboeke is dat dit besonder lank duur voordat dit in die geheel afgehandel word of gedeeltelik in afsonderlike bande beskikbaar kom en gebruik kan word.

Die tydsfaktor bring mee dat:

- (1) hierdie projekte al hoe meer kos om in stand te hou en te voltooi;
- (2) die bruikbaarheid van dele van woordeboekreekse ingeperk word as gevolg van kruisverwysings na dele wat nog nie beskikbaar is nie;
- (3) die beplanning en uitvoering van ander belangrike projekte al hoe langer uitgestel word.

Die tydsfaktor is nie in alle gevalle die gevolg van die wetenskaplike bewerking van die versamelde data nie, maar van die omstandigheid dat die grootste deel van die leksikografiedokumentasie nog met die hand verrig word.

Soos in afdeling 3 van hierdie verslag aangetoon word, kan die wetenskaplike kwaliteit van leksikografiebeskrywing en die produksievermoë van die leksikograaf met behulp van relatief goedkoop rekenaarfasiliteite verhoog word.

Die grootste nadele verbonde aan die situasie dat die databasisse van die hoofwoordeboeke nog nie in masjienleesbare vorm beskikbaar is nie, is die volgende:

- (1) die bestaande, versamelde data (wat meestal 'n duur belegging verteenwoordig), is in 'n groot mate onbruikbaar vir navorsingsdoelendes omdat dit ontoeganklik is;
- (2) die bestaande inligtingseenhede oor leksikale items kan nie maklik en goedkoop vir ander nodige woordeboekprojekte (bv. sinoniem- en etimologiese woordeboeke) onttrek word nie;
- (3) die byhou van die woordeboeke plaas 'n verdere las op die personeel en vertraag die voltooiing van bestaande projekte verder;



- (4) die databasis of geselekteerde onderdele daarvan kan nie as 'n geheel vir verdere bewerking in ander leksikografieprojekte (bv. twee- of meertalige woordeboeke) beskikbaar gestel word nie.

Die organisatoriese, mannekrag- en finansiële probleme wat in hierdie verband ervaar word, word in meer besonderhede bespreek in afdeling 4 van hierdie verslag.

## 2.2 REKENAARGESTEUNDE LEKSIKOGRAFIEPROJEKTE IN SUID-AFRIKA

In hierdie afdeling word daar 'n uiteensetting gegee van die gebruik van die rekenaartegnologie in Suid-Afrika vir die produksie van woordeboeke.

### 2.2.1 Hoofwoordeboekprojekte

- (a) Die Woordeboek van die Afrikaanse Taal (Stellenbosch)

By die Buro van die Woordeboek van die Afrikaanse Taal is daar teen einde 1987 begin met 'n vier-jaar rekenariseringsplan. Rekenaars is aangekoop en die personeel ontvang opleiding. Stappe word ook gedoen om 'n kantoorpublikasie-stelsel aan te koop sodat materiaal drukgereed aan die uitgewer gelewer kan word. Die omskepping van die WAT se kaartjie-stelsel na 'n woordbank en databasis moet nog beplan word. Die vooruitsigte vir rekenarisering word in 2.3.1 hieronder bespreek.

- (b) Die Dictionary of South African English on Historical Principles (Rhodes)

Die Rhodes-eenheid het in die afgelope jaar die eerste stappe tot rekenarisering gedoen. Mikro-rekenaars, met koppeling aan die Universiteit se hoofraam, word reeds gebruik vir die rekenarisering van die woordelys vir die groot woordeboek en van die bibliografie. Die verwerking van inskrywings direk op rekenaar is die volgende stap. Intussen het die eenheid die rekenaarbande van die General Readers' Dictionary of South African English van Oxford University Press ontvang. Dit sal die verwerking van die volgende uitgawe van die woordeboek aansienlik vergemaklik. Die hoofstap, wat rekenarisering betref, is egter die insleuteling van al die materiaal wat woordeboeke ten grondslag lê. Hierdie groot onderneming word nog nie in die vooruitsig gestel nie.

- (c) Die Xhosa-Engels-Afrikaans Woordeboek (Fort Hare)

Die woordeboekspan by Fort Hare is tans besig met die publikasiefase van die woordeboek. Rekenarisering van die kaartstelsel word nog nie beplan nie, maar 'n eerste woordverwerker is onlangs aangekoop. Toekomsplanne word in 2.3.1 hieronder genoem.

## 2.2.2 Projekte aan universiteite

### 2.2.2.1 Projek:      Woordeboek van die Noord-Sotho-taal (voorlopig genoem WNT-Projek)

Hierdie woordeboekprojek word onderneem in opdrag van die Noord-Sotho Taalraad en is vir uitvoering toegewys aan die Departement Afrikatale van die Universiteit van Pretoria. Die projek is as 'n navorsingsprojek geregistreer onder leiding van prof P S Groenewald en dr D J Prinsloo.

Die WNT moet uiteindelik 'n omvattende verklarende woordeboek vir die Noord-Sotho taal wees - met ander woorde 'n hoofwoordeboek. Dié woordeboek moet dieselfde formaat aanneem en soortgelyk wees aan die WAT en dus vir die Noord-Sotho-taal dieselfde waarde hê as wat die WAT vir die Afrikaanse taal het, naamlik 'n volledige dokumentasie van die taaleie.

Proeflope en beplanning is gedoen in 1987. Die samestelling van die WNT het amptelik op 1 Januarie 1988 begin. Die WNT sal uiteindelik net soos die WAT etlike volumes beslaan en die verwagte voltooiingsdatum sal ongeveer 2003 wees.

Die woordeboekprojek word volledig rekenaarmatig aangepak. Datavaslegging geskied op 'n mikrorekenaar met die leksikonprogram wat deur die Nasionale Vakterminologiesdiens (NVD) verskaf word. (Vergelyk die uiteensetting van hierdie stelsel in die TERMNET-verslag.) 'n Mikrorekenaar met 'n harde skyf word gebruik. Inskrywings word per disket aan die NVD gelewer wat dit deur middel van die hoofraamrekenaar in die nasionale termbank stoor. Groot verwerkings, sistematiesing en die uitdruk van data word ook deur die NVD gedoen. Benewens opname in die termbank word alle data ook deur die UP hoofraamrekenaar gerugsteun. Vanweë die besondere veelsydigheid van die programmatuur is dit moontlik om van tyd tot tyd sakwoordeboeke/woordelyste uit te druk en selfs as woordeboeke te publiseer na gelang van waar die dringendste behoefte op 'n bepaalde tydstip bestaan, byvoorbeeld 'n tweetalige sakwoordeboek of 'n omvattende tweetalige verklarende woordeboek. Benewens die aspek van omvattende rekenariserings (soos hierbo uiteengesit), verskil die WNT van die WAT onder meer ook in dié opsig dat die projek nie alfabeties (d.w.s. vanaf a tot uiteindelik by z) aangepak word nie, maar eerder *temagerig*. Die WNT sal uiteraard alle bestaande woordeboeke vir Noord-Sotho omvat maar veel verder gaan ten opsigte van die optekening van dialektiese terme, plant-, boom-, voël- en diername. Dit sal ook alle woorde insluit wat tot op datum in Noord-Sotho literêre werke verskyn en wat nog nêrens in 'n woordeboek opgeneem is nie. As eerste groot baken op die pad van die WNT word gestreef na 'n groot Afrikaans-Noord-Sotho-Afrikaans of 'n Engels-Noord-Sotho-Engels woordeboek. Daar is aanvanklik gehoop om deur middel van befondsing deur 'n instansie soos die RGN die WNT op groot skaal te kon aanpak. Uiteindelik is die samestellers aangewese op 'n beskeie begin met 'n toekenning van R1 000 en 'n tydelik deeltydse assistent vir 12 ure per week.

Die vaslegging van die data sal op só 'n wyse geskied dat die inskrywings uiteindelik op verskeie maniere gesistematiseer kan word:

- (i) alfabeties deur naamwoorde volgens die prefiks in te sleutel;
- (ii) verskillende tipes kodering te gebruik ten einde byvoorbeeld 'n naslaansubwoordeboek te genereer wat op grond van woordstamme gesistematiseer is en wat die volledige afleidingsvorme ten opsigte van elke stam sal bied.

Subwoordeboeke kan waar toepaslik en na gelang van behoefte op 'n kontrakbasis vir instansies of die privaatsektor saamgestel word ten einde aan 'n bepaalde behoefte te voldoen.

#### 2.2.2.2 Die Franse woordeboekprojek aan die Universiteit van Pretoria

'n Volledige programpakket is deur die Dept. Rekenaardienste van UP ontwerp vir die saamstel van 'n Franse woordeboek deur die Dept. Frans. Besonderhede oor die rekenaarapparatuur en -programmatuur wat vir die uitvoering van die projek gebruik word, is egter nie vrylik beskikbaar nie.

#### 2.2.2.3 'n Morfeemwoordeboek - US

Dr L G de Stadler en Prof M W Coetzer is aan die Universiteit van Stellenbosch besig om 'n skrif-tot-spraakomsetter vir Afrikaans te ontwikkel. In hierdie ondersoek is die vaslegging van die verhouding tussen skrif en spraak uiteraard belangrik. Die twee navorsingsterreine wat afgebaken moet word is die volgende:

- (a) die omsetting van skrif na spraak;
- (b) die omsetting van spraak na skrif.

Aan albei hierdie terreine word tans aandag bestee. Die ontwikkeling van 'n spraak-na-skrifomsetter is veel moeiliker as die omgekeerde, veral as gevolg van die geweldige interpersoonlike verskille in fonetiese representasie wat binne een taal van spreker tot spreker bestaan.

In die omsetter word verskillende komponente, waarvan een 'n morfeemwoordeboek is, ingebou. Hierdie morfeemwoordeboek speel 'n belangrike rol in die ontwikkeling van die skrif-tot-spraakomsetter.

Alhoewel die gebruik van 'n volledige woordeboek nie moontlik is nie, word daar tans gewerk aan die gebruik van 'n morfeemwoordeboek saam met 'n program wat morfologiese analyses doen (die MF-analiseerder). Aanvanklik is alle morfeme met hulle fonetiese transkripsies en sintaktiese kategoriemerkers in die morfeemwoordeboek opgeneem. Die

morfeemwoordeboek wat opgestel is, bevat tans ongeveer 10 000 morfeme. Dit sluit heelwat afleidings met gebonde stamme soos vibreer en konstipeer in, aangesien die morfonologiese komponent van die omsetter nog nie hierdie gevalle kan hanteer nie. Soos die morfologiese komponent van die omsetter verbeter word, sal die morfeemwoordeboek sover moontlik gesuiwer word van hierdie afleidings en sal slegs die gebonde stamme in die woordeboek opgeneem word.

### 2.2.3 Uitgewers

Die lede van die navorsingspan het besoeke gebring aan die volgende uitgewers: Nasou, HAUM, Tafelberg en Van Schaik. Die hoofdoel van die besoeke was om die LEXINET-projek en die leksikografiese dokumentasiesstelsel van die RGN aan hulle bekend te stel en om te bepaal in hoe 'n mate die rekenaartegnologie tans in die saamstel, byhou, uitgee en hergebruik van woordeboeke gebruik word.

Vanweë die onderlinge konkurrensie tussen die uitgewers, is die meeste van hulle nie bereid om gegewens oor spesifieke woordeboeke of oor die presiese gebruik van die bestaande rekenaartegnologie te verskaf nie. Uit die gesprekke is dit egter moontlik om 'n paar algemene tendense vas te stel:

- (i) gerekenariseerde set- en drukstelsels (soos die ATEX-STELSEL) word deur sommige uitgewers vir woordeboekproduksie gebruik. Dit bring mee dat manuskripmateriaal vir die set- en drukproses op rekenaar geplaas word en deels gekodeer is, in die besonder met set- en fontkodes. Die feit dat manuskripte op rekenaar is, bring verder mee dat hersieninge en bywerkings aan die masjienleesbare manuskripte aangebring sal word en dat verdere hergebruike van die woordeboekmateriaal oorweeg kan word;
- (ii) individuele outeurs wat aan die saamstel van woordeboeke werk, begin in toenemende mate hulle manuskripverwerking op mikrorekenaars uitvoer en die bewerkte manuskrip op diskette aan die uitgewers verskaf;
- (iii) uitgewers begin toenemend besoeke bring aan woordeboekinstansies oorsee om op die hoogte te kom van die gebruik van die rekenaartegnologie en om die toepassingsmoontlikhede daarvan vir hulle eie leksikografiese aktiwiteite te bepaal.

Die uitgewers ondervind in hierdie verband ook 'n aantal probleme.

Die beperkte omvang van die huidige woordeboekmark bring mee dat groot finansiële insette aan rekenaarapparatuur en -programmatuur, rekenaar-

kundigheid en die rekenaaropleiding van leksikograwe en skrywers nog nie as lonend beskou word nie.

Woordeboekmanuskripte word op die oomblik nog in hulle finale vorm op rekenaar geplaas en slegs gekodeer met die nodige set- en fontkodes. Dit bring mee dat die geselekteerde onttrekking van dele van die materiaal vir latere hergebruik nog nie moontlik is nie. Die ontwikkeling van die nodige programmatuur vir die outomatiese ontleding en kodering van die woordeboekmateriaal word op die oomblik ook nog nie oorweeg nie.

Uitgewers word dikwels gekonfronteer met die onversoenbaarheid van die rekenaartale en -stelsels wat die direkte koppeling tussen mikro-rekenaars, hoofraamfasiliteite en die set-en-drukapparaat bemoeilik.

Die daarstel van 'n omvattende gerekenariseerde leksikografiese databasis (vgl. 3.3.2., die Oxford University Press-model) is, vanweë die beperkte woordeboekmark, vir Suid-Afrikaanse uitgewers nie moontlik nie.

#### 2.2.4 Drukkers

##### (a) Die Staatsdrukker

Die rekenaar word reeds by die Staatsdrukker gebruik, maar slegs vir administratiewe doeleindes, byvoorbeeld vir die beplanning van drukwerksskedules. Die drukker is nie aan die hoofraam gekoppel nie.

Dit het reeds gebeur dat manuskripte in disketvorm aan die Staatsdrukker verskaf is. Ten einde die probleem van onversoenbare rekenaarprogramme en -tale te oorkom, is daar van 'n omsetter ('converter') gebruik gemaak. Dit het die probleem egter geensins opgelos nie. Daar is selfs ook van optiese lesers gebruik gemaak - sonder veel sukses. (Daar is reeds van die omsetter en optiese leser ontslae geraak.)

Indien 'n disket met die rekenaarstelsel van die Staatsdrukker versoenbaar is, kan dit deur die rekenaar gedruk word. Die rekenaar is egter nie met die setter verbind nie - die produk is dus steeds 'n rekenaardrukstuk.

Indien 'n woordeboek (bv. die WAT) by die Staatsdrukker gedruk word, word dit dus steeds geset (met galei-, bladsy- en finale bladsyproewe). Daar word van die Hastech-setstelsel (Hendrix) gebruik gemaak. Wanneer 'n woordeboek geset word, is daar 'n stadium waar die woordeboek in die setrekenaar se geheue gestoor word. Die data word egter net daar bewaar totdat die finale setproses afgehandel is. Daarna word die data op film vasgelê en as film bewaar vir latere gebruik by herdrukke.

## (b) Cape and Transvaal Printers

CTP volg in 'n groot mate dieselfde prosedures en metodes as die Staatsdrukker omdat hulle ook van die Hendrix-stelsel (setter) gebruik maak.

Die CTP ontvang opdragte hoofsaaklik op een van die volgende wyses:

- die getikte manuskrip word deur CTP op die setter ingesleutel en galeiproewe word gedruk wat geproeflees moet word;
- gerekenariseerde projekte word via 'n telefoonmodem oorgespoel. Dit word met 'n spesiale stelsel verwerk tot bladsyproewe wat ook geproeflees moet word; of
- die manuskrip word in disketvorm verskaf. Die disket moet verwerk word om aanpasbaar gemaak te word. Sommige diskette lewer groter probleme as ander. Dit is byvoorbeeld verkieslik dat die woorde/terme reeds in alfabetiese volgorde moet wees. Deur vooraf met die programmeerders van CTP te skakel, kan die voornemende leksikograaf van meet af aan die korrekte kodes gebruik. Dit blyk dat CTP dit verkies om gerekenariseerde manuskripte te ontvang - mits daar vooraf oor die toekenning van set- en fontkodes reëlins getref kon word.

### 2.2.5 Taaldatakorpusse

Uit navorsing van die RGN het dit geblyk dat daar aan verskillende universiteite alreeds omvattende nie-gerekenariseerde en gerekenariseerde datakorpusse van gesproke en geskrewe Afrikaans versamel is. Vergelyk die onderstaande lys.

#### a) Nie-gerekenariseerd:

Universiteit van Pretoria: Omgangsafrikaans van Danville, Villieria, Waterkloof en Eersterus (op kasset).

Potchefstroomse Universiteit vir CHO: Getranskribeerde tekste van opnames van taalvariasiedata.

Universiteit van Stellenbosch: Gespreksafrikaanskopus.

Universiteit van die Oranje Vrystaat: Versameling van historiese tekste en bandopnames.

#### b) Gerekenariseerd:

RGN: Kindertaaldatabasis

Randse Afrikaanse Universiteit: Gesproke Afrikaans.

## 2.3 VOORUITSIGTE VIR REKENARISERING

### 2.3.1 Die hoofwoordeboekkantore

#### 2.3.1.1 Die Woordeboek van die Afrikaanse Taal

Soos genoem in 2.2.1 word daar vanaf 1988 eerste stappe gedoen om die Buro van die WAT se verwerking van leksikografiese materiaal te rekenariseer. Rekenaars is reeds aangekoop en die meeste van die personeel het opleiding in woordverwerking ontvang. Daar word ondersoek ingestel na kantoorpublikasiestelsels en na die daarstelling van koppeling in lokale-areanetwerke (LAN's) en met die hoofraamstelsel.

Die Buro is daarvan terdeë bewus dat 'n aantal kernsake betreffende die volledige oorskakeling na rekenarisering verdere beplanning vereis. Dit is noodsaaklik om presies vas te stel wat die uiteindelijke aard, omvang en aanwending van die data sal wees. Hierdie faktore sal riglyne verskaf vir programmatuurontwerp en vir die argitektuur van die woordbank wat in die rekenariseringsproses opgebou kan word. Alvorens die projek fyn uiteengesit word, kan die finansiële implikasies nie bereken word nie. Na aanleiding van rekenariseringsprojekte in die buiteland is dit egter duidelik dat aansienlike fondse nodig sal wees, dus sal die projek fasegewys onderneem moet word.

Die belangrikheid van versoenbaarheid met die stelsels van ander instansies in die land word in die Buro se beplanningsdokumente bespreek maar in hierdie stadium word die saak moontlik te eng omskryf. Versoenbaarheid met die oog op dienslewering, byvoorbeeld aan taalkantore, en met die oog op publikasieprosesse, is klaarblyklik van groot belang. Die klempasing is hier egter op die afvoerprosesse terwyl dit ewe belangrik sou wees om op die mikrovlak van databasisstrukture versoenbaarheid met ander databasisse in die land na te streef. Indien daar (oor die lang termyn) verskeie woordbanke in die land opgebou word, en indien die woordbanke op die mikrovlak betreklik eenders is, sou twee- en meertalige woordeboeke van allerlei aard snel saamgestel kon word.

Afdeling 3.2.1.1 verskaf 'n uiteensetting van databasisstrukture vir leksikografiese versamelings. In 3.3.2, 3.3.3.2 en 3.4 word die saak verder bespreek.

Die koste verbonde aan die ontwikkeling van meet af van 'n eie stelsel is hoog (vgl. 2.3.1.2). Waar moontlik moet daar dus gebou word op bestaande modelle, waarvan enkeles in hierdie verslag beskryf word en ander in die literatuur (vgl. die Bronnelys). In die bespreking hieronder van rekenariseringsplanne by die Rhodes-kantoor word daar voorgestel dat die Oxford English Dictionary projek (vgl. 3.3.2) waarskynlik 'n baie geskikte model is.

### 2.3.1.2 The Dictionary of South African English on Historical Principles

Beplanning vir die omskakeling na 'n volledig gerekenariseerde stelsel is by die Rhodes-eenheid die verste gevorder. Soos genoem (vgl. 2.2.1.), word die rekenaar reeds gebruik vir take rondom die opstel van die woordeboek, nl. die woordelysindeks en die bibliografie, en sal die rekenaarverwerking van inskrywings waarskynlik binnekort aangepak word. Rekenarisering van die materiaal waaruit 'n woordeboekteks saamgestel word, is egter die sleutel tot werklik effektiewe gebruik van die rekenaar in die leksikografie. Terselfdertyd is dit ook 'n veel omvattender, veel meer ingewikkelde en duurder onderneming as enige ander aspek van rekenarisering op die gebied (vgl. 3.2.1.1 vir 'n uiteensetting van databasisstrukture vir leksikografiese materiaal en 3.3.2, 3.3.3.2 en 3.4 vir verdere bespreking).

In die huidige situasie by die Rhodes-eenheid kan rekenarisering van die materiaal nie onderneem word nie. Die gebrek aan personeel in permanente aanstellings en die druk om die groot woordeboek af te handel, is hoofknelpunte. Oor die effens langer termyn is die vooruitsigte vir volledige rekenarisering van die Eenheid se stelsel egter waarskynlik beter as by die ander woordeboekkantore. Die Eenheid se verbintenis met Oxford University Press, die uitgewers van die Rhodes-woordeboeke, skep die moontlikheid dat OUP met Rhodes Universiteit 'n ooreenkoms sal aangaan om OUP se stelsel te bekom.

OUP is tans besig met die rekenarisering van die groot Oxford English Dictionary, en items uit die Rhodes-materiaal oor Suid-Afrikaanse Engels word in die OUP-woordeboeke en -databasisse opgeneem. 'n Ooreenkoms met Rhodes Universiteit om die OUP-stelsel te mag gebruik, sou vir OUP die voordeel inhou dat die Suid-Afrikaanse materiaal in die gewenste vorm opgeneem kan word in die OUP-databasisse.

Nêrens elders ter wêreld bestaan daar in die leksikografie 'n rekenariseringsprogram wat met die OUP-stelsel ten opsigte van omvang en deskundigheid vergelyk nie (vgl. 3.3.2). Die ontwikkelingskoste beloop ongeveer 7 miljoen pond. Dit spreek vanself dat lande met kleiner woordeboekondernemings dit moeilik kan oorweeg om van vooraf eie stelselontwikkeling aan te pak. Daarom sou 'n Rhodes-OUP-ooreenkoms van besondere belang wees, en sou dit 'n logiese verdere stap wees om so 'n ooreenkoms te probeer uitbrei na ander woordeboekkantore in Suid-Afrika.

### 2.3.1.3 Die Xhosa-woordeboek

Die woordeboekkantoor by Fort Hare is tans besig met die publikasie-proses van die Xhosa-woordeboek. Dit sal 'n geruime tyd duur voordat al drie die dele afgehandel is. Die finansiering van die voortsetting van die kantoor se werk ná die publikasie van die woordeboek gaan 'n knelpunt wees, en so ook die finansiering van rekenarisering.



### 2.3.2 Ander woordeboeke

In die geval van hoofwoordeboeke is 'n redelik ingewikkelde rekenaarstelselontwerp nodig omdat die materiaal in databasisvorm verwerk moet word. Waar dit by kleiner woordeboeke slegs om die opstel van 'n woordeboek gaan, word woordeboeke toenemend met behulp van die rekenaar opgestel. Die universiteite en ander groot instansies met uitgebreide rekenaarfasiliteite kan programmeringshulp verleen.

Die ontwikkeling van leksikografiese programme wat op tuisrekenaars gebruik kan word en na hoofraamstelsels omgewerk kan word, is onlangs deur die RGN se Afdeling Leksikologie geloods.

### 2.3.3 Die uitgewers

Die toenemende belangstelling van uitgewers in gerekenariseerde publikasieprosesse is 'n aanduiding dat daar oplossings gesoek sal word vir die huidige probleme (vgl. 2.2.3. hierbo). Uitgewers kan egter nie spesifieke inligting oor toekomsplanne verskaf nie.

## HOOFSTUK 3: REKENAARGESTEUNDE LEKSIKOGRAFIE IN DIE BUITELAND

### 3.1 INLEIDING

Die rekenaar word alreeds die afgelope aantal jare op 'n koste-effektiewe wyse oorsee deur leksikografie-ondernemings gebruik vir die saamstel, bewerking en byhou van meerdoelig-bruikbare, masjienleesbare leksikografiese databasisse op grond waarvan 'n verskeidenheid leksikografiese produkte op verskillende datadraers aan uiteenlopende gebruikersgroepe beskikbaar gestel word.

'n Groot deel van die huidige taal-en-rekenaarnavorsing word gewy aan die ontwerp van stelsels vir die gesofistikeerde outomatiese en semi-outomatiese ontleding van masjienleesbare tekstmateriaal. Leksikons van allerlei soorte en omvang vorm saam met 'n reeks taalspesifieke grammatikas die hoof-taalkundige komponente van hierdie stelsels wat vir die grammatiese ontleding en vertaling van tekste en vir inligtingstelsels ingespan word. (Vergelyk die bespreking in 3.3) Batori en Weber (1986:106) wys in hierdie verband daarop dat die daarstelling van die leksikons vir hierdie stelsels die duurste komponent daarvan uitmaak. Hieruit volg die noodsaak daarvan dat daar die nouste samewerking tussen leksikografie-instansies en die ontwikkelaars van hierdie stelstels moet wees, in die besonder dat eersgenoemde in die beplanning en strukturering van leksikografiese databasisse met die leksikografiese vereistes van hierdie stelsels rekening moet hou.

Teen die agtergrond van die voorafgaande word die volgende sake nou behandel:

- (i) die algemene struktuur en funksies van die gerekenariseerde leksikografiese databank (LDB) en van die voordele wat dit vir leksikografiewerk inhou (vgl. 3.2);
- (ii) die struktuur en funksies van 'n aantal spesifieke leksikografie-instansies (vgl. 3.3);
- (iii) die struktuur van die leksikons in outomatiese en semi-outomatiese teksontledingstelsels.

In die beplanning van 'n leksikografiese databasis en teksontledingstelsel is die ontwerp van die formate van die verskillende leksikons die eerste belangrike stap omdat besluite hieroor in 'n groot mate die omvang, kwaliteit en uiteindelijke meerdoelige bruikbaarheid van hierdie stelsels bepaal.

### 3.2 DIE LEKSIKOGRAFIESE DATABANK (LDB)

Twee hoofsoorte LDB's word onderskei op grond van die soort taaldata wat hulle berg, prosesseer en beskikbaar stel, naamlik

- (i) die algemene leksikografiese databank met grammatiese inligting oor die lekseme van die algemene woordeskat van 'n taal (soos byvoorbeeld dié van Longman, Van Dale, die New Oxford English Dictionary (NOED) en die Instituut voor Nederlandse Lexicologie (INL));
- (ii) die terminologiedatabank (soos byvoorbeeld dié van TEAM, NORMATERM, LEXIS, DANTERM en die databank van die Nasionale Vakterminologiediens) wat een- of meertalige inligting bevat oor die terminologieë van verskillende vakgebiede (vgl. in hierdie verband Swanepoel (1986) en die TERMNET-verslag).

Wat hulle tegniese struktuur betref, stem hierdie twee soorte databanke in 'n groot mate ooreen, maar in die bespreking hierna word in die besonder gelet op die algemene leksikografiese databank.

### 3.2.1 Tegniese en logiese struktuur van die LDB

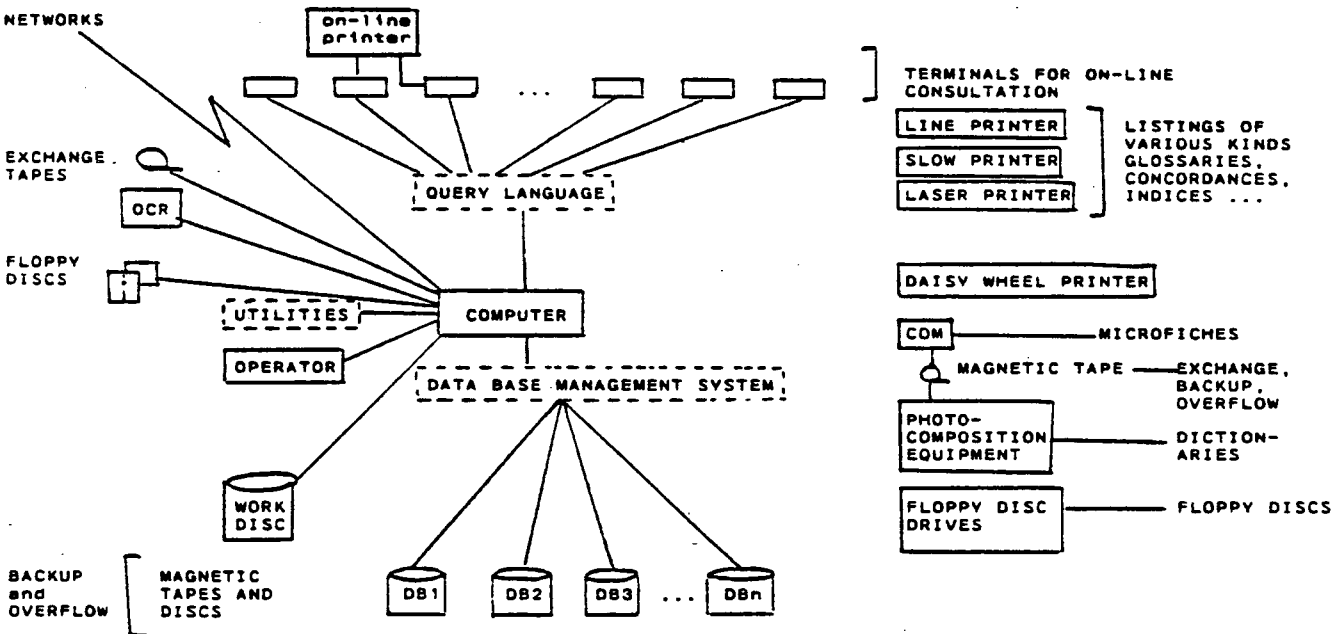
Enige gerekenariseerde leksikografiese databank bestaan uit 'n stelsel rekenaarapparatuur en -programmatuur vir die invoer en herwinning van linguistiese data. Die tegniese struktuur van die individuele linguistiese databanke toon 'n aantal verskille vanweë die besondere apparatuur en programmatuur wat gebruik word, maar in die breë word daar by die oprigting van 'n LDB vir die onderstaande komponente voorsiening gemaak:

- (i) hoofraamfasiliteite met 'n groot genoeg geheuekapasiteit vir die saamstel, berging en bewerking van 'n omvattende leksikografiese databank;
- (ii) terminale wat gekoppelde toegang tot die hoofraam bied vir navrae en die enkel en gebundelde herwinning en bewerking van onderdele of die hele inhoud van die databasis;
- (iii) 'n versameling drukkers (lyndrukkers, laserdrukkers, ens.) vir die uitdruk van bewerkte onderdele van die inhoud van die databasis (bv. alfabetiese lyste, konkordanse, indekse, sleutelwoord en -kontekslyste, retrogradelyste en spesifieke leksikale inskrywings);
- (iv) fotosetapparatuur vir die opmaak, set en druk van woordeboeke of lyste in hardekopievorm.

'n Illustrasie van een so 'n tegniese stelsel is dié wat in Sager en McNaught (1980c:17-19) vir die voorgestelde Britse linguistiese databank uiteengesit word (vgl. Fig. 3.1). In hierdie struktuur word daar ook voorsiening gemaak vir die inskakeling van die LDB by 'n omvattender netwerkstruktuur en vir drie hoofstelle programmatuur, naamlik

- (a) vir gekoppelde navrae via terminale (via die sg. 'query language');
- (b) 'n databasisbeheerstelsel vir die struktuurering van die databasis;
- (c) 'n versameling nutsprogramme ('utilities') vir roetinetake of vir bewerkinge van die databasis.

Fig. 3.1: British LDB Configuration



Die databasis, dit wil sê die versameling leksikografiese data, word in Fig. 3.1 weergegee as 'n versameling kleiner databasisse  $(Db_1 - Db_n)$  wat op magnetiese bande en skywe geberg word.

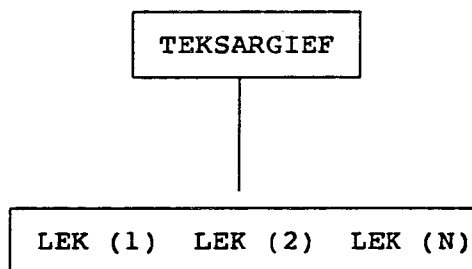
### 3.2.1.1 Die teksargief en die leksikon(s)

In die Leksikografiese databank word die databasis gewoonlik logies gestruktureer in twee hoofkomponente, naamlik;

(a) 'n teksargief;

(b) 'n leksikon (of 'n versameling leksikons) waar elkeen 'n kleiner databasis in die omvattender databasisse uitmaak. (Vgl. Fig. 3.2).

Fig. 3.2: Die struktuur van die LDB



Die omvang van die leksikografiese data vir die saamstel van die meeste woordeboeke en die programmatuur wat vir die bewerking daarvan nodig is, bring mee dat 'n rekenaar met 'n groot geheuekapasiteit vir dié doel nodig is. Dit geld in die besonder in gevalle waar die teksargief en die onderskeie leksikons met mekaar geskakel is en die hele databasis in bewerkinge betrek word.

Die onderhoud van die LDB val uiteen in twee take:

(i) die onderhoud van die databasis, dit is die bywerk, hersiening, leksikografiese verwerking en verspreiding van die data in die basis (d.i. die leksikon(s) en die teksdatabasis);

(ii) die onderhoud van die rekenaarstelsel, dit is instandhouding van die rekenaarstelsel en die ontwerp en realisering van die nodige programmatuur en apparatuur vir die doeleindes van (i).

Waar eersgenoemde die taak van die leksikograwe is, is laasgenoemde die taak van die rekenaardeskundiges en ingenieurs.

(a) Die teksargief

Die teksargief, dit wil sê 'n teks- of dokumentdatabasis, kan in masjienleesbare of nie-masjienleesbare vorm bestaan, dit wil sê deel uitmaak van die rekenaardatabasis en met die res van die onderdele van die databank geskakel wees of apart bestaan, byvoorbeeld in die vorm van 'n versameling dokumente in 'n biblioteek, waar hierdie dokumente dan basiese inligting bevat waarna daar in die leksikale inskrywings van die leksikons verwys word of waarop die grammatiese beskrywing in die leksikale inskrywings berus.

In die geval van sommige LDB's bestaan daar twee sodanige gerekenariseerde teksdatabasisse:

- (i) 'n databasis met die sitatemateriaal wat gebruik is/word in die saamstel van die leksikale inskrywings van die onderskeie rekords, dit is sitate wat direk in sodanige leksikale inskrywings aangehaal word (vgl. die bespreking van die struktuur van die databasis van die Instituut vir Nederlandse Leksikologie en dié van die New Oxford English Dictionary-projek in 3.3);
- (ii) 'n databasis bestaande uit teksmateriaal van uiteenlopende aard wat onder andere gebruik word vir die genereer van basiese leksikografiese materiaal, maar wat ook onafhanklik gebruik word vir verskeie ondersoeke na taalgebruiksaspekte van 'n besondere taal (vergelyk in hierdie verband die bespreking van die databasis van die afdeling Tesaurus van die Instituut vir Nederlandse Leksikologie (INL) en die Van Dale-projek in 3.3, en die TEXTNET-verslag).

Die teksmateriaal word in die geval van (ii) saamgestel uit verteenwoordigende gesproke en geskrewe tekste van 'n taal/tale (bv. getranskribeerde gesprekke, romans, tydskrifte en koerantkopie) op verskillende datadraers en in verskillende media (bv. gedrukte en geskrewe tekste op papier, tekste op ponsbande en -kaarte, mikrofiche of op magnetiese skywe of bande).

Afhangende van die datadraer waarop en die medium waarin die data vir toevoer toeganklik is, bestaan daar verskillende outomatiese en semi-outomatiese moontlikhede vir die masjienleesbaarmaak van die data, die verwerking daarvan en die opname daarvan in die teksargief/-dokumentasieargief. Vergelyk die uiteensetting in 3.3 van die toevoermoontlikhede van die Afdeling Tesaurus van die Instituut vir Nederlandse Leksikologie.

In die keuse van tekste word daar gepoog om 'n omvattende, verteenwoordigende en sistematiese teksversameling daar te stel wat as basis kan dien vir die genereer van basiese leksikografiese materiaal vir opname in die leksikon(s) (d.i. trefwoorde, sitate en voorbeelde), vir die

distilleer van 'n wetenskaplike grammatiese beskrywing van die gekose woordeskatitems en vir die kontrole van die volledigheid van die leksikons.

Vir hierdie doel word verskillende rekenaarbewerkinge op die teksdata uitgevoer, waarvan die vernaamste hieronder kortliks bespreek word (vergelyk verder die bespreking van die TEXTNET-verslag).

Met die invoer van 'n teks word dit gekodeer met behulp van etikette ten opsigte van 'n aantal algemene tekseienskappe soos die stilistiese kenmerke (bv. gesproke, geskrewe, formeel, informeel, literêr, nie-literêr of koeranttaal), die tematiese bereik (bv. politiek, sport, geskiedenis of chemie) en bibliografiese besonderhede (outeur, titel, datum). Met die latere herwinning van teksdele (woorde, stukke of sinne) kan hierdie gegewens ook daarmee saam herwin word, en vir beskrywing van spesifieke leksikale items gebruik word.

Met behulp van (i) algemene teksbewerkingsprogramme, vir die bewerking van tekste onafhanklik van die tekstipe en die taal van die teks, en (ii) spesiale teksbewerkingsprogramme, vir spesifieke tekssoorte of tekste in 'n besondere taal (vgl. Brustkern, Lenders en Willée (1981) vir die beskrywing van sodanige programme), kan daar verskillende soorte lysste van die ortografiese woorde in die teksbank gegenereer word:

- (i) 'n omvattende alfabeties geordende lys van die tekswoorde in die tekskorpus. Hierdie lysste dien as basis vir die saamstel van 'n leksikon en as kontrolemiddel om die volledigheid van bestaande leksikons te bepaal. Verder kan sulke lysste per tekstipe gegenereer word vir die vasstelling van leksikaal-stilistiese merkers;
- (ii) sleutelwoord-in-kontekslyste. 'n Lys kan gegenereer word van alle tekswoorde waarin elke tekswoord as sleutelwoord met 'n gespesifiseerde hoeveelheid konteks gegee word. Hierdie lysste word gebruik vir 'n omvattende grammatiese beskrywing van elke sleutelwoord (veral die sintakties-sintagmatiese eienskappe en betekenisonderskeidings daarvan, vir die identifisering van homograwe en vir die keuse van verteenwoordigende voorbeeldmateriaal en sitate. Waar die tekskenmerke van die sleutelwoord gegee word (vgl. die bespreking hier bo), is die stylkenmerke van 'n woord en die bibliografiese besonderhede van sitate dus ook voorhande;
- (iii) frekwensielyste van tekswoorde per teks, tekstipe of die tekskorpus as 'n geheel. Hierdie lysste word veral gebruik vir die keuse van woordeskatitems vir opname in spesifieke leksikonsoorte en vir die empiriese begronding van die keuse van beskrywingsetikette wat die spreiding en frekwensie van leksikale items aandui (bv. ongewoon, seldsaam, en nie algemeen);

- (iv) retrogradelyste. Saam met alfabeties geordende lyste word die retrogradelyste gebruik vir die beskrywing van die morfologiese eienskappe van leksikale items. Dit is egter ook moontlik om met 'n stringsoekfasiliteit (bv. deur 'n spesifieke morfeem aan te gee) volledige morfologiese paradigmas vir ontleding te genereer. Hierdie lyste word ook gebruik as bron vir voorbeeldmateriaal om die morfologiese eienskappe van trefwoorde te illustreer en om die spellingvariasie van morfeme in afleidings en komposita te bepaal;
- (v) lemmatiseringslyste. Daar bestaan reeds 'n aantal programme met behulp waarvan die fleksie-morfologiese verwante woordvorme van 'n teks na hulle grondvorm (lemmavorm of lekseemvorm) herlei kan word. Met behulp van 'n lemmatiseringsprogram is dit dus moontlik om te bepaal hoeveel en watter lemmas in die tekskorpus voorkom. Hierdie programme is egter vir hulle toepassing afhanklik van 'n bestaande leksikon van lemmavorme en 'n reeks algoritmes vir die morfologiese ontleding van tekswoordvorme (vgl. die bespreking in 3.4).

Uit die voorafgaande behoort dit duidelik te wees dat 'n wetenskaplik geselekteerde en geanaliseerde bank tekstmateriaal 'n belangrike rol in die 'empiriese' begroning van 'n leksikon speel. Die vinnige, betroubare en omvattende soektogte van tekstmateriaal vir leksikografiese verwerking en die manipulasiemoontlikhede van die teksdata maak 'n tekstebank 'n onmisbare deel van die versameling van leksikografiese materiaal vir die saamstel en byhou van 'n LDB. Vergelyk verder die bespreking in 3.3.

#### (b) Die leksikons

'n Leksikon in die LDB is 'n leksikale databasis waarin die lekseme van een of meer tale en hulle geassosieerde grammatiese inligting as 'n versameling leksikale inskrywings (LI's) gestruktureer word.

Met die daarstelling van 'n LDB word daar gewoonlik 'n meerdoelig-bruikbare hoofleksikon saamgestel en op grond hiervan dan 'n reeks (kleiner) leksikons gegenereer wat in die databasis as aparte eenhede opgeneem word. Die versameling leksikons kan ook uitgebrei word deur die toevoeging van masjienleesbare leksikons wat van ander leksikografie-instansies ontvang word.

Alhoewel die hoofleksikons in die meeste gevalle aanvanklik vir 'n spesifieke projek ontwerp en saamgestel word (byvoorbeeld vir die uitgee van 'n bepaalde woordeboek), het die hoë koste van so 'n onderneming (byvoorbeeld die insleutel van gedrukte tekste, die koste verbode aan apparatuur en die ontwikkeling van programmatuur) in die verlede daartoe gelei dat die hoofleksikons uitgebrei en vir verskeie nuwe projekte hergebruik word. Van hierdie hergebruikte sluit onder meer die volgende in: die uitgee van verskillende uitgawes en hersiene uitgawes van dieselfde leksikon in verskillende formate en op verskillende data-



draers, die ontwerp van kleiner of gespesialiseerde woordeboeke op grond van onderdele van die omvattende leksikografiese databasis, die ontwerp van 'n tesourus-uitgawe, en die saamstel van die eentalige deel van vertalende woordeboeke. Elke leksikon wat op grond van die bestaande databasis gegenereer word, word gewoonlik as 'n aparte entiteit in die LDB opgeneem met die oog op die latere hergebruik daarvan.

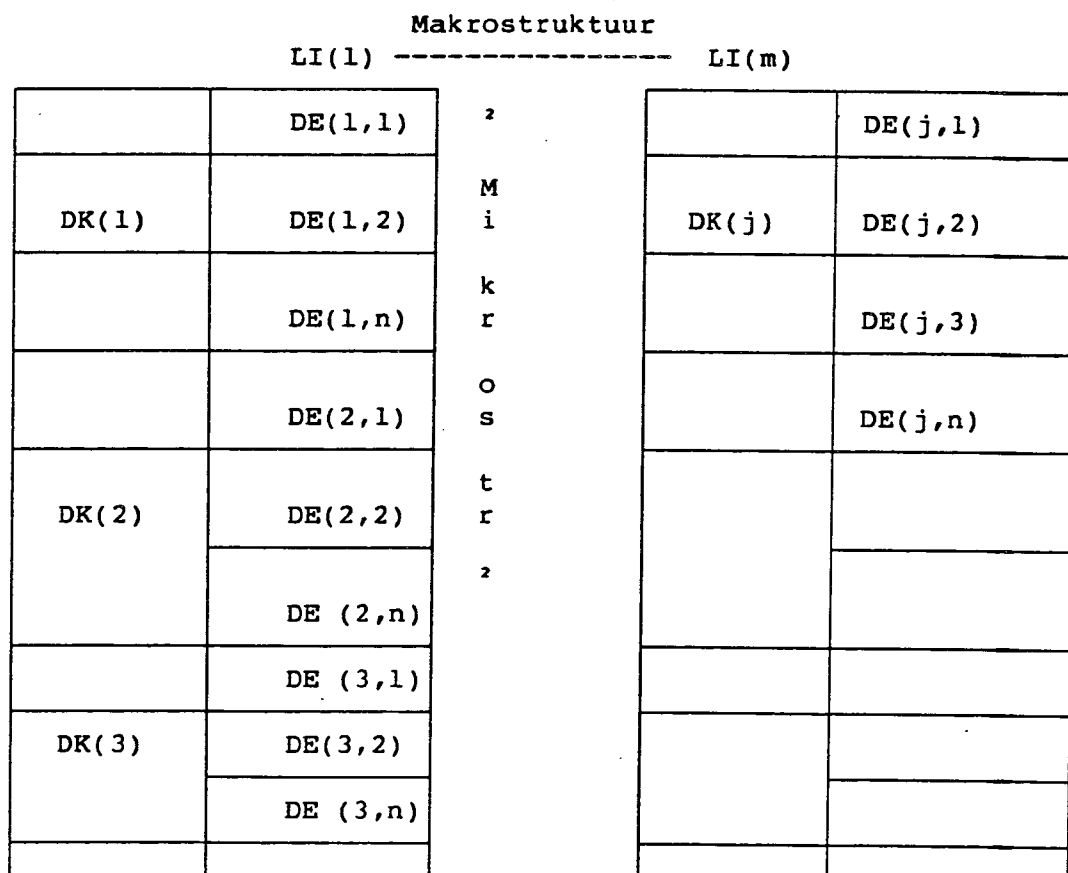
### 3.2.1.1.1 Logiese en rekenaarstruktuur van die leksikons

Die hoofdeel van enige leksikon (gerekenariseerde leksikon of die tradisionele gedrukte woordeboek) is op twee vlakke gestruktureer, nl.

- (i) op die makrostruktuurvlak as 'n versameling leksikale inskrywings (LI's);
- (ii) op die mikrostruktuurvlak, met ander woorde op die vlak van die enkel leksikale inskrywing, as 'n versameling grammatiese data-elemente uit verskillende grammatiese datakategorieë wat 'n enkele lekseem beskryf.

Skematies voorgestel:

Fig. 3.3: Die logiese struktuur van die leksikon.



(LI = leksikale inskrywing, DK = datakategorie en DE = data-element)

In ooreenstemming hiermee word die masjienleesbare leksikon op die makrostrukturele vlak fisies georganiseer as 'n lêer wat saamgestel word as 'n versameling rekords. Elke rekord word op die mikrostrukturele vlak verdeel in 'n aantal datavelde, ooreenkomstig die aantal leksikografiese datakategorieë wat onderskei word. Hierdie datavelde kan 'n vaste lengte hê (uitgedruk in terme van 'n vaste aantal karakters), waardeur dit moontlik word om die data-elemente van elke veld op grond van hulle posisie in die rekord te identifiseer, of hulle kan van 'n veranderlike lengte wees (in welke geval die data-elemente van elke datakategorie gewoonlik met 'n kategorie-etiket geïdentifiseer moet word). Skematies voorgestel:

Fig. 3.4: Skematiese voorstelling van die struktuur van die masjienleesbare leksikon

REKENAARLÊER

RR 1	DV .....
	1
	DV .....
	2
RR 2	DV .....
	1
	DV .....
	2
RR m	DV .....
	3
	DV .....
	j
RR m	

(RR=rekenaarrekord en DV=dataveld)

Ten opsigte van die masjienleesbare leksikon word dit as 'n vereiste gestel dat (i) elke data-element wat in 'n leksikale inskrywing opgeneem word, uniek onderskei en/of geïdentifiseer moet word sodat dit direk aanspreekbaar en afsonderlik herwinbaar is; en (ii) dat elke leksikale inskrywing 'n uniforme en vaste struktuur moet hê. Ten opsigte van die enkodering van die data-elemente word die vereiste gestel dat dit sover moontlik (eenduidig) geformaliseer (en dus ekonomies) weergegee moet word. Enkele voorbeelde ter illustrasie.

In die leksikale inskrywingsmatrys van BONNLEX word die datakategorieë geëtiketteer (vgl. WORTLAUT, LEMMA, ens.), en die subverdeling van elke dataveld of subkategorie aangedui (vgl. die fleksie-morfologiese kategorie wat verder onderverdeel word in kasus, numerus, genus, graadtrappe, ens.).



Die datakategorieë wat deur elke syfersimbool verteenwoordig word, is die volgende:

1=LEMMA

Die inskrywingsvorm van die lekseem (woord).

2=WOORDKLAS

Die volgende hoofwoordsoorte word onderskei: naamwoord, werkwoord, adjektief, setsel, voegwoord, voornaamwoord, kwantifiseerder, telwoord, lidwoord en tussenvoegsel.

3=TIPE

Die besondere subkategorie van die woordsoort van die trefwoord word aangedui, byvoorbeeld dat 'n werkwoord 'n hoofwerkwoord of 'n hulpwerkwoord is.

4=FUNKSIE

Die sintaktiese funksie van die woord word aangedui, byvoorbeeld dat 'n werkwoord oorganklik is of dat 'n kwantifiseerder as 'n determineerder optree.

5=KOMPLEMENTSTRUKTUUR

Hier word aangedui met watter komplemente die besondere woord verbind, byvoorbeeld: met 'n dat-klous.

6=GEBRUIK MET

Hier word aangedui met watter ander woorde die naamwoord, werkwoord, adjektief en bywoord vas of opsioneel verbind. In die geval van voornaamwoorde, telwoorde en kwantifiseerders word hulle telbaarheidseenskappe aangedui.

7=OORBLYWENDE INFO

Inligting wat in 'n woordeboek gegee word, maar wat nie direk van belang vir die ontledingstaak is nie, word hier aangedui.

## 8=FLEKSIE-MORFOLOGIE

Die inligting van hierdie posisie word deur die morfologiese komponent aangegee sodat dit vir die meeste trefwoorde leeg sal wees of later ingevul word.

Die data-elemente van elke datakategorie word met behulp van 'n stel letter- en syfersimbole weergegee. Die geformaliseerde aard van die inligting word geïllustreer in die volgende LI's:

harass \* V \* M \* T \* l \* -- \* WV5 \* --  
1      2      3      4      5      6      7      8

(waar '--' aandui dat 'n datakategorie leeg is, en

V = werkwoord

M = hoofwerkwoord'

T = oorganklik

l : kan met 'n naamwoord(stuk) gekomplementeer word

WV5 : kan met -ing verbind om 'n adjektief te vorm).

Verskeie faktore bepaal die keuse van die besondere leksikale items en hulle geassosieerde datakategorieë wat vir opname in die LI's van 'n leksikon gekies word. Van die belangrikste hiervan is die volgende:

- (i) die besondere funksies wat die leksikon moet verrig as komponent van 'n leksikografiese databasis (of van 'n ontledingsstelsel);
- (ii) die stratum/strata van die woordeskat waaruit lemmas vir opname gekies word en die grammatiese inligting wat met hierdie lemmas geassosieer word;
- (iii) die datakategorieë wat in die LI's onderskei word in die leksikografiese bronne wat vir die saamstel van die rekenaarleksikon gebruik word (hier in die besonder bestaande, gedrukte leksikons);
- (iv) die grammatiese inligtingsbehoefte van al die potensiële gebruikersgroepe van die leksikon(s);
- (v) internasionale en nasionale riglyne en standaarde vir die keuse en strukturering van leksikale inskrywings op magnetiese band (vgl. in hierdie verband ISO 1982).

Naas die grammatiese datakategorieë, bevat LI's ook data-elemente wat vir die administrasie van elke rekord ingevoer word (bv. die datum van samestelling of laaste bywerking, naam van die samesteller, aantekeninge of bronverwysings).

Die noodsaak vir die fynere subkategorisering van spesifieke grammatiese datakategorieë is dikwels 'n taalspesifieke aangeleentheid. Ter illustrasie: in Duits vereis die beskrywing van die naamwoord, die adjektief en die werkwoord 'n verfynde subkategorisering van die fleksie-morfologiese kategorie, maar nie in Afrikaans nie.

Een van die moderne neigings in die Engelstalige leksikografie is dat 'n LDB as 'n leksikon-cum-grammatika gekonspieer word. In die praktyk word dit veral gemanifesteer in die uitbreiding van die grammatiese beskrywing van die leksikale items wat as trefwoorde in die leksikon opgeneem word, en dan in die besonder in die sogenaamde pedagogiese woordeboeke, dit is eentalige woordeboeke vir nie-moedertaalsprekers van 'n betrokke taal.

'n Volledige beskrywing van die grammatiese inligting in vyf sodanige leksikons word gegee in Lemmens en Wekker (1986). Op grond van hulle ondersoek kom hulle tot 'n omvattende aanbeveling vir die sintaktiese beskrywing van die hoofwoordsoorte (werkwoorde, naamwoorde en adjektiewe). Hier onder word daar kortliks aangedui hoe die sintaktiese inligting in hulle aanbeveling in die leksikale inskrywings vir werkwoorde en in die leksikon as 'n geheel uiteengesit word.

In die LI word die sintaktiese eienskappe direk na die trefwoord gegee in gekodeerde vorm (bv. do/NP; io/NP) met daarnaas dan voorbeelde ter illustrasie van die besondere sintaktiese struktuur. Vergelyk die twee voorbeelde in Fig. 3.6. Die syferkodes wat die sintaktiese inligting voorafgaan, verwys na 'n aparte beskrywing van die werkwoordsoorte, soos weergegee in Tabel I (vgl. Lemmens en Wekker 1986:102-104); as sodanig illustreer hierdie materiaal ook die wyse waarop daar op die makrostrukturele vlak relasies tussen die individuele LI's en ander dele van die leksikon gelê word.

Fig. 3.6:

**bring** /.../ v **brought** /.../ 1 [ d0/NP; i0/NP + d0/NP objm- to or for, d0/NP + Ac/AdvP,PP ] to come with or lead: bring me the book/bring the book to me - The soldier's brave deeds brought him honour and glory - John brought Mary some flowers/brought some flowers for Mary; Can you bring your sister back home? - Bring your friend to the party - The prisoner was brought before the judge - The beauty of the music brought tears to her eyes. --see USAGE

2 [ d0/NP,NP to-inf ] to cause or lead to: Spring rains bring summer flowers - What brought you to do it? - He could never bring himself to kill an animal or bird.

3 [ d0/NP; i0/NP + d0/NP ] to sell or be sold for: This old car will bring about 10 pounds - What/How much would a new car bring?; The pictures he sells bring him 12,000 pounds a year.

4 [ d0/NP often + PP ] (against) law to make officially: The policeman brought a charge against the fast driver - The neighbours brought a complaint against the noisy family.

5 [ d0/NP,NP- ing ] to cause to come: One sad letter from his wife brought many offers of help/brought him crying home - A phone call brought him hurrying to Leeds.

6 [ d0/NP + Ac/AdvP,PP ] to cause to reach a certain state: to bring someone low (in defeat) - bring them in/out/back/together - to bring someone into being/into action/to an end/to his knees (in defeat).

...

**decelerate** /.../ v [ Ø; d0/NP ] to (cause to) go slower: Before you decelerate you should check in your rear-view mirror; We decelerated the engine long before we came to a stop -- compare ACCELERATE (1)



Fig. 3.7:

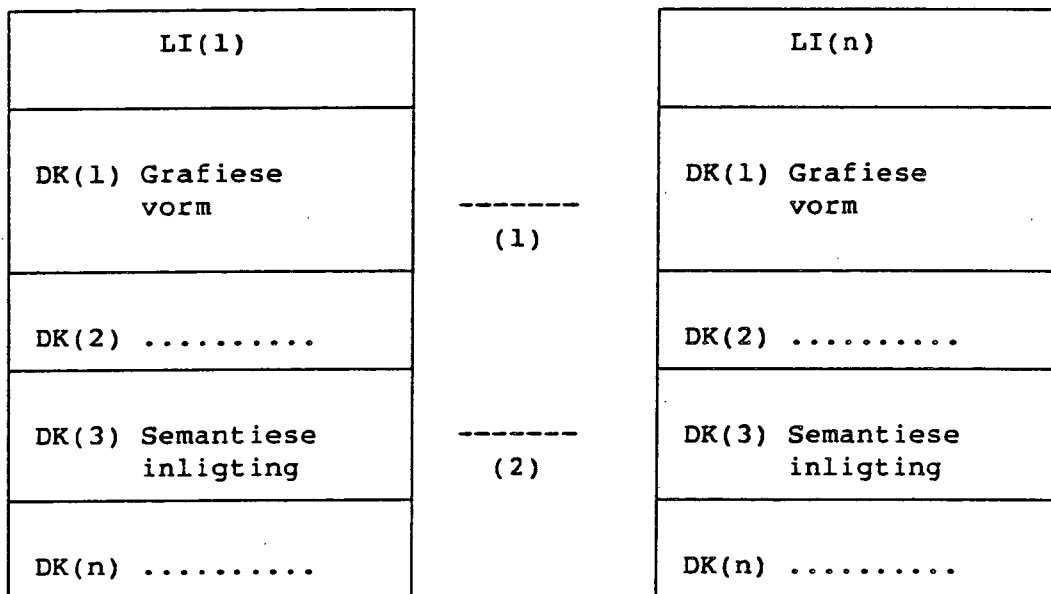
Basic VP-structures	Verb types	Examples
1.V $\emptyset$ (A)	Intransitive	John snores John drinks
2.V $\left\{ \begin{array}{l} / \text{NP} \\ \text{Pc} / \text{AdjP} \\ / \text{PP} \\ / \text{Cl} \end{array} \right\}$ (A)	Copula	John is $\left\{ \begin{array}{l} \text{a fool} \\ \text{foolish} \\ \text{in a good mood} \end{array} \right\}$  John is to be envied (to-inf.pass.)  John grew old The milk turned sour  They got rich John became an actor
3.V $\left\{ \begin{array}{l} / \text{Loc} / \text{AdvP} \\ / \text{PP} \end{array} \right\}$ (A)	Loc Intransitive	John is $\left\{ \begin{array}{l} \text{outside} \\ \text{in the garden} \end{array} \right\}$ John lives in Paris
4.V $\left\{ \begin{array}{l} / \text{PP} \\ \text{Ac} / \text{Cl} \\ / \text{AdvP} \\ / \text{NP} \end{array} \right\}$ (A)	Ac Intransitive	John belongs to a left-wing party. John condescended $\left\{ \begin{array}{l} \text{to bribery} \\ \text{to help us} \end{array} \right\}$  He lived to be 90 This car drives smoothly  It weighs 10 pounds It lasted (for) three hours
5.V $\left\{ \begin{array}{l} \text{Ac} - \text{Pc} / \text{NP} \\ / \text{AdjP} \end{array} \right\}$ (A)	Complex Intransitive	John died a Catholic Susan married young
6.V $\left\{ \begin{array}{l} \text{dO} / \text{NP} \\ / \text{Cl} \end{array} \right\}$ (A)	Monotransitive	John admires Susan  John likes $\left\{ \begin{array}{l} \text{to sing} \\ \text{singing} \end{array} \right\}$ (to-inf) (-ing)  John hopes that you will help him (fin)  John saw Susan $\left\{ \begin{array}{l} \text{leave} \\ \text{leaving} \end{array} \right\}$ (NPinf) (NP-ing) John kept me waiting (NP-ing) John believes the dog to be hungry (NPto-inf)
7.V IO/NP + dO/NP (A)	Ditransitive	John gave Susan the book

Fig. 3.7(vervolg)

<p>8.V dO/NP + {Ac-Pc / NP / PP} (A)</p>	<p>Complex Monotransitive</p>	<p>She painted the door {green all colours}</p> <p>John saw Susan drunk He flung the door open</p> <p>They found John {a fool foolish in a good mood}</p> <p>They appointed him chairman They christened the baby John They beat the prisoner unconscious</p>
<p>9.V dO/NP + {Ac / PP / AdvP} (A)</p>	<p>Ac Monotransitive</p>	<p>John put the money {in a box there}</p> <p>John's firm posted him to Singapore John reminds me of his father John worded the letter carefully</p>

In die gerekenariseerde leksikons is die afsonderlike datakategorieë van gelykwaardige status vir die manipulasie van die leksikale inskrywings of van die onderdele daarvan. Dit is byvoorbeeld moontlik om die LI's alfabeties te orden op grond van die ortografiese vorm van die trefwoorde, of semanties op grond van die data-elemente wat in die betrokke datakategorie opgeneem is (bv. volgens sinonieme of antonieme), of deur 'n kombinasie van hierdie ordeningsmoontlikhede (vgl. Fig. 3.8). Daarby is dit ook moontlik om LI's met gemeenskaplike data-elemente van dieselfde datakategorieë as 'n groep te onttrek, byvoorbeeld al die leksikale inskrywings van oorganklike werkwoorde, of om slegs die data-elemente van 'n spesifieke datakategorie te onttrek, byvoorbeeld al die definisies van 'n gespesifiseerde versameling leksikale items. Hierdie manipulasievermoë van die leksikografiese data is een van die grootste voordele van die rekenaarleksikon.

Fig. 3.8: Ordeningsrelasies



((1) dui 'n alfabetiese relasie volgens 'n alfabetstelsel aan en (2) 'n semantiese relasie volgens 'n semantiese/konseptuele stelsel)

Op die mikrostrukturele vlak kan daar verder onderskei word tussen die leksikografiese struktuur van 'n leksikale inskrywing, dit is die keuse, enkodering en plasing van die grammatiese data-elemente, en die afhanklikheidsstruktuur van die leksikale inskrywing, dit is die onderlinge grammatiese relasies tussen die data-elemente van 'n enkele leksikale inskrywing. Op hierdie vlak is daar veral van twee relasies sprake, naamlik die redundansie en voorspelbaarheid van data-elemente.

Ter illustrasie: as 'n leksikale item na sy woordsoort as 'n onoorganklike of 'n koppelwerkwoord geklassifiseer word, dan kan dit (in

Afrikaans) nie ook 'n lydende vorm hê nie; of as 'n leksikale item in Duits met 'n hoofletter gespel word, dan is die kans besonder hoog dat dit na die woordsoort daarvan 'n naamwoord sal wees. Redundansies van hierdie aard word in sommige leksikons met behulp van 'n stel oortolligheids- /redundansiële of 'n stel algoritmes ondervang om sodoende ekonomie ten opsigte van die beskrywing van die data te bewerkstellig.

### 3.2.2 Die rekenarisering van die gedrukte leksikon

In die saamstel van die leksikon(s) van 'n LDB word daar in hoofsaak op bestaande gedrukte (nie-masjienleesbare) leksikons en leksikons wat reeds op magnetiese band is, staat gemaak.

Die masjienleesbaar maak van 'n gedrukte leksikon en die strukturering daarvan as meerdoelig-buikbare databasis is 'n omvattende taak (vgl. die bespreking in 3.3). Die vernaamste stappe in hierdie proses word hier onder kortliks bespreek.

#### (1) Die beplanningsfase

In die beplanning van 'n LDB moet daar vooraf 'n duidelike profiel saamgestel word van

- (a) die funksies van die (afsonderlike) leksikon(s) (vgl. die bespreking in 3.4);
- (b) die omvang en struktuur van die beoogde gerekenariseerde leksikon(s) (vgl. die bespreking in 3.4);
- (c) die omvang en struktuur van die leksikons wat as basis vir die saamstel van die gerekenariseerde leksikon(s) sal dien;
- (d) die apparatuur en programmatuur wat benodig sal word vir die herstrukturering van die gedrukte leksikons in die struktuur van die gerekenariseerde leksikon.

#### (2) Die opvolgfase

In die opvolgstap moet daar 'n diepte-ontleding van die makro- en die mikrodatastruktuur van die beoogde gerekenariseerde leksikon en van die bestaande gedrukte leksikons gemaak word.

Op die makrostruktuurvlak behels dit ten opsigte van altwee soorte leksikons dat daar 'n uiteensetting gegee word van (a) die hoofdele van die leksikons, hulle makrostrukturele ordening ten opsigte van mekaar en die kruisverwysings tussen hulle, en (b) die interne struktuur van elkeen van die hoofdele.

Op die mikrostrukturele vlak moet daar 'n duidelike uiteensetting gegee word van die struktuur van elke komponent van die hoofdele. Wat die

versameling leksikale inskrywings betref, moet die datastruktuur van die leksikale inskrywing(s) gespesifiseer word:

- (i) uit watter grammatiese en nie-grammatiese datakategorieë 'n LI opgebou is;
- (ii) op watter wyse die datakategorieë onderling vormlik van mekaar onderskei word;
- (iii) hoe die data-element(e) van elke datakategorie ten opsigte van mekaar in die LI(s) georden is/moet wees;
- (iv) in watter vorm die data-elemente van elke kategorie geënkodeer en/of van mekaar onderskei is/moet word.

Hierdie inligting is van wesenlike belang vir die ontwerp van die programmatuur wat nodig is vir die outomatiese en semi-outomatiese ontleding van die LI's van die masjienleesbare vorm van 'n gedrukte leksikon en vir die omskakeling daarvan na die struktuur van die LI's van die gerekenariseerde leksikon.

Die struktuur van die LI van die rekenaarleksikon en die oorwegings en vereistes wat in hierdie verband 'n rol speel, word in besonderhede bespreek in die voorafgaande dele van hierdie verslag (vgl. 3.2). Hierna word daar slegs in meer besonderhede ingegaan op aspekte van die ontleding van die datastruktuur van die LI's in die gedrukte leksikon, geïllustreer aan die hand van die LI's in die Handwoordeboek van die Afrikaanse Taal (HAT).

Vooraf egter eers 'n aantal algemene opmerkinge oor die datastrukturering in die gedrukte woordeboek.

Vergeleke met die masjienleesbare leksikon, is die gedrukte leksikon as 'n statiese gegewe georganiseer. In die gedrukte leksikon word die datakategorieë waartoe die afsonderlike data-elemente behoort en die grense tussen hulle met behulp van 'n verskeidenheid beskrywingskonvensies uitgedruk, byvoorbeeld die plasing van data-elemente, verskillende tipografiese style (swaardruk, skuinsdruk, gewone druk, bokas en onderkas), 'n verskeidenheid lees- en skryftekens, soos kommas, punte, dubbelpunte en aksenttekens, letter- en syfersimbole, of 'n kombinasie van enige van hierdie middele.

Ter wille van 'n ekonomiese beskrywing word die grammatiese inligting in die gedrukte woordeboek op allerlei ander maniere gekompakteer, byvoorbeeld:

- (i) daar word gesteun op 'n verskeidenheid afkortings en verkortings van gewone woorde in die beskrywingstaal (vgl. die formulering van definisies en aantekeninge oor die gebruikaspekte van 'n leksikale item);

- (ii) soms word homonieme binne 'n enkele leksikale inskrywing beskryf;
- (iii) sommige inligting word gewoon aan die afleidingsvermoë van die gebruiker op grond van die aangebode data-elemente oorgelaat, soos byvoorbeeld afleidings oor die vaste leksikale elemente waarmee 'n trefwoord kan verbind (hierdie gegewens kan afgelei word uit die gebruiksvoorbeelde wat ter toeligting van 'n definisie gegee word);
- (iv) bepaalde inligting word oorgedra deur 'n afwesigheid van enige vorm van beskrywing, soos die feit dat die weglating van 'n woordsoortetiket daarop dui dat die trefwoord tot die kategorie naamwoorde behoort of dat die gebrek aan 'n etiket voor 'n betekenisbeskrywing impliseer dat dit die letterlike betekenis van die trefwoord weergee.

Die interpretasie van die grammatiese inligting in die LI van die tradisionele gedrukte woordeboek kan vir die gebruiker dikwels probleme oplewer, veral as die interpretasie van die inligting reeds kennis van die besondere taal vereis (vgl. byvoorbeeld die spelling van afleidings en samestellings van die trefwoord).

Woordeboekmakers kan egter besonder slordig en inkonsekwent in die saamstel van die LI's van 'n woordeboek wees, of heelwat aan die onderskeidingsvermoë van die gebruiker oorlaat - 'n feit wat die outomatiese interpretasie van die inhoud van 'n woordeboek bemoeilik. Enkele algemene probleme is die volgende:

- (i) die afwesigheid van 'n uniforme leksikale inskrywingstruktuur;
- (ii) inkonsekwente plasing van die data-elemente van verskillende datakategorieë;
- (iii) inkonsekwente gebruik van die beskrywingskonvensies;
- (iv) die saambundeling van inligting in 'n enkele leksikale inskrywing oor twee of meer lekseme.

Die makrostrukturele ordening van die LI's in HAT kan aan die hand van die volgende twee stelling voorbeelde geïllustreer word:

Fig. 3.9: Makrostrukturele ordening van LI's in HAT

A.

LI 1	kor rek'sie, (-s). 1. Verbetering van 'n fout, on-noukeurigheid: 'n Korreksie aanbring. 2. Handleiding van te verbeter, te korrigeer: Die korreksie van skoolopstelle, van drukproewe is tydrowend. korreksie;...teken, ...werk.
LI 2	kor rek si o neel', b.nw. en bw. Verbeterend, wat tot verbetering dien: Korreksionele straf.
LI 3	kor rek tief'. I s.nw. (korrektiewe). Middel tot verbetering van 'n fout: Dit moet beskou word as 'n korrektief op die gebrekkige optrede van die staat. II b.nw. Wat verbeter; verbeterend: Korrektiewe behandeling.
LI 4	kor rek'tor. (-e, -s). Persoon wat drukproewe verbeter; proefleser.
LI 5	kor'rel. I s.nw. 1. Klein, ronderige liggaampie waaruit sommige stowwe bestaan of waarin 'n stof verdeel is: 'n korrel sout, suiker. 2. Saad van graan e.a. vrugte: Koringare met vet korrels. 'n Druiwetros met gesonde korrels. 3. Klein hoeveelheid, krummel: Ek het g'n korreltjie suiker meer nie. 4. Inmekaargedraaide bossie hare soos by Bantoes en party Kleurlinge. 5. Knoppie van die visier op 'n vuurwapen. II ww. (gekorrel). 1. Mik na, rig op. rigting soek om te skiet: Hy korrel na die bok. Ook fig.: Jy korrel verniet na my, ek kan die werk nie doen nie. 2. Korrels afpluk van 'n tros druiwe: Pa het ons belet om te korrel. 3. Korrels vorm. as korrels verskyn. 4. Tot korrels maak. korrel: ...agtig, ...sneeu, ...suiker, ...struktuur.

B.

**ka ros' 1, (-se; -sie).** (*Hott.*) 1. Mantel van gebreide (skaap)velle met die hare aan, of so 'n vel as mantel by Suid-Afrikaanse inboorlinge: 'n *Karos dra, omhang*. 2. Abbavel. 3. Velkombers: 'n *Karos van tiervel oor die bed hê*.

**ka ros' 2, (-se).** (*veroud.*) 1. Rytuig, reiskoets op vier wiele. 2. Staatskoets vir plegtige geleent-hede. [ F. *carosse*; It. *carozza* L. *carrus* kar &.

In Fig. 3.9 is die afsonderlike LI's alfabeties georden volgens die ortografie van die trefwoord/lemma wat in swaardruk elke LI inlei. Vergelyk:

korreksie  
korreksioneel  
korrektief  
korrektor  
korrek

Wat die rangskikking van gevalle van funksiewisseling en homograwe betref, word die volgende beleid gevolg:

(a) Funksiewisseling: die data-elemente van elke funksievariant word in 'n enkele LI groepeer, maar van mekaar onderskei met behulp van Romeinse syfers, byvoorbeeld:

**korrektief.** I. s.nw .....  
II. b.nw .....

(b) Homograwe: elke homograaf word in 'n aparte LI beskryf, dit wil sê homograwe word as aparte trefwoorde opgeneem en van mekaar onderskei in die alfabeties geordende reeks met behulp van Arabiese syfers bo die trefwoord; byvoorbeeld:

**karos** ,(-se; -sie).  
**karos** ,(-se).

Fleksie-morfologies verwante woordvorme word tot 'n enkele lekseem gereduseer wat dan as trefwoord optree; die woordvormlike variante van 'n fleksie-morfologiese paradigma word aangedui met behulp van die aangee van die betrokke fleksiemorfeme, byvoorbeeld: die fleksie-morfologiese klasse in A word gekompakteer tot die beskrywings in B:

Al. korreksie	2. korrektor	3. karos
korreksies	korrektore	karosse
	korrektors	karossie

B1. *korreksie*, (-s) 2. *korrektor*, (-e, -s) 3. *karos*, (-se; -sie)



Lede van die morfologiese paradigma van die trefwoord wat nie fleksie-morfologies verwant is nie (d.i. afleidings en komposita) word in swaardruk aan die einde van 'n LI geplaas as hulle betekenis deursigtig of selfverklarend is, byvoorbeeld:

a. korreksie, .....

*korreksie: ..teken, ..werk*

b. korrel, .....

*korrel: ..... ..agtig, .. sneeu, .. suiker, .. struktuur*

Afleidings en komposita wat nie-deursigtig is nie of betekenis-spesialisasie vertoon, word as aparte trefwoorde in afsonderlike LI's opgeneem en beskryf, byvoorbeeld:

LI kort. ....

1

*kort: .. erig, .. heid, .. heidshalwe.*

LI kortaf, .....

2

*kortafheid.*

LI kortasem(ig), .....

3

*kortasemigheid.*

LI kortbeen,

4

LI kortbegrip.

5

Die inligting oor fleksie-morfologies verwante woordvorme en die woordvorme van 'n nie-fleksiemorfologiese verwante morfologiese paradigma word dus op 'n duidelik onderskeibare manier aangebied ten opsigte van die plasing (begin resp. einde) en die enkoderingsvorm daarvan, vergelyk:

a. fleksiemorfologies verwante reeks:

*trefwoord, (-affiks, -affiks)*

of:

*trefwoord, (-affiks; -affiks)*

b. nie-fleksiemorfologies verwante reekse:

**TREFWOORD** .....

*..... morfeem, .. morfeem.*

(Vergelyk verder die bespreking hieronder.)

Uit die voorafgaande is dit moontlik om te bepaal watter beleid die leksikonsamestellers gevolg het ten opsigte van die volgende sake rakende die woordvorme van 'n taal:

- (i) die keuse van lemmas/lemmavorme wat as trefwoorde van LI's dien;
- (ii) die opname en enkodering van inligting oor fleksiemorfologies-verwante woordreekse en nie-fleksiemorfologies verwante woordreekse van dieselfde morfologiese paradigma;
- (iii) die groepering van inligting oor funksievariante en homonieme in dieselfde respektiewelik verskillende LI's;
- (iv) die ordening van die afsonderlike LI's.

Die leksikografiese beleid onderliggend aan al bogenoemde sake steun egter ook op 'n aantal algemeen-teoretiese uitgangspunte ten opsigte van kernkonsepte, soos 'woord', 'lekseem', 'woordvorm' en '(nie-)verwante woordvorme', waarvan daar ook 'n duidelike uiteensetting gegee moet word (vgl. verder Swanepoel (1986), Hoofstuk I, 1.3.1-1.3.2).

Die LI's van HAT het nie 'n eenvormige datastruktuur nie. In die ontleding kan daar dus slegs 'n uiteensetting gegee word van die struktuurmoontlikhede. Vir hierdie doel kan die LI in drie hoofdele verdeel word, naamlik.

- (i) die hoofwoordgroep, wat strek vanaf die trefwoord wat die LI inlei tot die eerste definisiesin;
- (ii) die semantiese definisierreeks wat op die hoofwoordgroep volg;
- (iii) morfologiese inligting.

Vergelyk die onderstaande voorbeeld:

a. oudertous, b.nw. (nie alg.)

Outyds, uit die mode, ouderwets

Hoofwoordgroep  
(HWG)

Definisierreeks  
(DR)

b. plegstatig, b.nw. en bw.

HWG

Baie plegtig en statig:  
Plegstatig binnetree.

DR

Plegstatigheid

MI

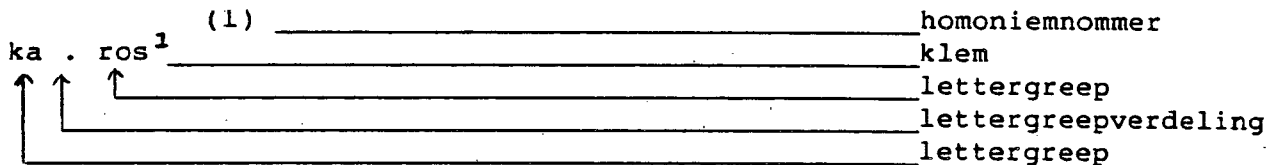
(i) - (iii) verteenwoordig die hoofgroepering. Elkeen van hierdie omvattender datakategorieë kan egter in kleiner datakategorieë verdeel word. Allerlei onkonsekwenthede ten opsigte van die rangskikking van die data-elemente van die verskillende datakategorieë in die hoofgroepe kom voor. Die onderstaande gee slegs 'n (onsistematiese en onvolledige) aanduiding van die moontlike datakategorieë van die hoofwoordgroep van die hoofwoordsoorte (nw., ww. en adj.).

(a) Hoofwoordgroep

Data-elemente van die volgende datakategorie kom voor:

- (1) woordvorm (swaardruk);
- (2) lettergreepverdeling (skryfwyse m.b.v. 'n punt);
- (3) klemtoon (aksent: bo en na die beklemtoonde lettergreep);
- (4) homoniemnommer (Arabiese syfer)

Voorbeeld



(5) vormvariante

Voorbeeld

karoobos(sie) (swaardruk; hakie)  
 koringstaf, koringstok (swaardruk; komma)  
 kartelig, .... Ook kartelrig. (swaardruk; tweede variant  
 voorafgegaan deur Ook )

kasjet, .... Ook cachet.

kasmier. Sien KASJMIER (swaardruk; bokas, voorafgegaan deur  
 Sien )

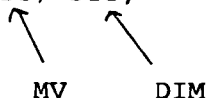
(6) fleksie-morfologiese inligting

Voorbeeld

korrektor, (-e, -s)  
 ↑                    ↑  
 MV                    MV

(tussen hakies; koppelteken  
 plus affiks in gewone druk;  
 komma om data-elemente van  
 dieselfde datakategorie te  
 onderskei)

karos, (-se;-sie)



(dubbelpunt om date-elemente van verskillende datakategorieë te onderskei)

karteer, ww. (gekarteer)

(verledetydsvorm van 'n ww. met ge-)

berg, ww. (-)

(koppelteken: verledetydsvorm sonder ge-)

(7) woordsoort

Voorbeeld

korreksioneel, b.nw. en bw.  
korrektief. I s.nw.  
korrel, ... II ww.

(afkorting in gewone druk)

(8) skryfwyse (aantekening)

Voorbeeld

kaalwa. (*ook los geskrywe*) (tussen hakies; kursief)

(9) etikette t.o.v. domein, gebruiksveld, styl, frekwensie, chronologie ens.

Voorbeeld

goëlb<sup>1</sup>. (*krieket*)  
generlei, .... (*veroud.*)  
geneugte, .... (*verh.*)  
lui<sup>1</sup>, .... (*w.g.*)  
brekfis. (*angl.*)  
con animo (*L., mus.*).

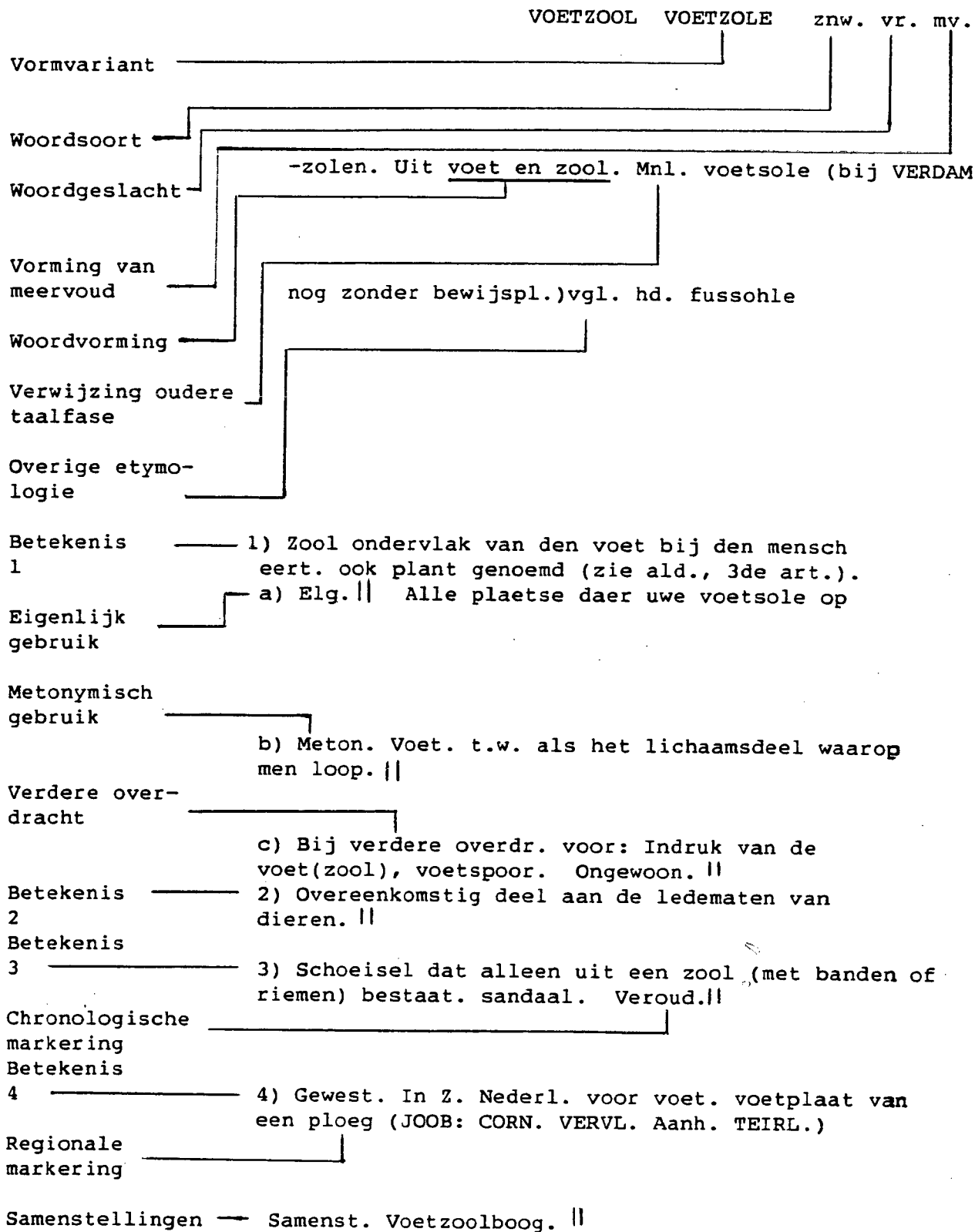
(tussen hakies; kursief)

Die komplekse aard van die strukturering en enkodering van die data-elemente in die gedrukte leksikon, blyk duidelik uit die onderstaande ontleding van die struktuur van die LI **korreksie** in HAT en van **voetsool** in die **Woordenboek der Nederlandsche Taal** (vgl. Van Sterkenburg 1984:79).

Fig. 3.10: Leksikografiese struktuur van 'n LI in HAT

<u>DATA-ELEMENT</u>	<u>DATAKATEGORIE</u>	<u>OMVATTENDER DATAKATEGORIE</u>
	woordvorm	
	lettergreep- vedeling	Skryfwyse
kor- <u>rek sie</u>	klemtoon	Fonetiek
(-s)	meervoudsmorfeem	Fleksiemor- fologie
1. Verbetering van 'n fout, onnoukeurigheid:	___ definisie ___ sinoniem	Semantiek
'n Korreksie aanbring	___ voorbeeld	
2. Handeling van te verbeter, te korrigeer:	___ definisie ___ sinoniem	Semantiek
Die korreksie van skool- opstelle, van drukproewe is tydwend.	___ voorbeeld	
korreksie: ..teken, ..werk	stam stam stam	Morfologie

Fig. 3.11: Structuur van 'n LI in die WNT



Die tweede stap in die rekenarisering van die gedrukte leksikon behels die plasing van die inhoud van die leksikon op magnetiese skywe of bande sodat dit masjienleesbaar is en bewerkinge daarop met programmatuurpakkette uitgevoer kan word.

Waar die tipografiese gehalte van die woordeboek duidelik genoeg is, is dit moontlik om die inhoud van die woordeboek met behulp van 'n optiese karakterleser (OCR) op magneetband te plaas en om dan die nodige foutekontrolle met behulp van gewone redigeringspakkette op terminale uit te voer. Indien die tipografiese gehalte van die leksikon dit nie toelaat nie, word die hele inhoud van die woordeboek per hand ingesleutel - 'n proses wat soms verkies word omdat dit dan moontlik is om die inhoud van die woordeboek alreeds deels te struktureer en te kodeer volgens die voorafbepaalde datastruktuur van die LI's van die gerekenariseerde leksikon.

Vergelyk verder die bespreking van die rekenariseringsprojek van die New Oxford English Dictionary in 3.3.

### 3.2.3 Die voordele van rekenarisering

Uit die voorafgaande het die voordele wat die gebruik van die rekenaar vir leksikografiese aktiwiteite inhou, telkens ter sprake gekom. Die vernaamste voordele word hieronder puntsgewys herhaal:

- (i) Een van die voordele van moderne rekenaarsisteme is dat hulle groot hoeveelhede data kan ontvang en in klein maat kan berg, volgens programinstruksies komplekse bewerkinge outomaties, gekontroleerd, vinnig en op 'n akkurate wyse kan uitvoer, en dat die resultate van hierdie bewerkinge weer geberg en in verskillende vorme beskikbaar gestel kan word.
- (ii) Die gebruik van masjienleesbare tekste hou dié voordeel in dat die leksikale inligting op 'n hele tekskorpus eerder as op geselekteerde sinate van die leksikograaf gebaseer kan word, en dat kontrole oor die gebruike van leksikale items (bv. frekwensie en spreiding), sistematies en vergelykenderwys uitgevoer kan word.

Gemeet aan die tyd wat dit neem om inligting met die hand te versamel, op te teken en te struktureer, bied die gebruik van die bestaande rekenaartegnologie vir die leksikograaf dié voordeel dat hy in hierdie fase van die dokumentasie van leksikale items sy take met groter spoed en akkuraatheid kan verrig.

Tyusbesparing bring 'n verhoging van produktiwiteit mee en skep die geleentheid vir die leksikograaf om meer inligting te versamel. Die moontlikheid om hele tekskorpusse as bronnemateriaal te gebruik, hou ook dié voordeel in dat die leksikograaf die betroubaarheid en kwaliteit van die versamelde inligting kan verhoog.

Die voordele van die gebruik van rekenaarhulp wat spoed, akkuraatheid en verhoogde produktiwiteit betref, hoef nie opgeweeg te word teen leksikografiese vereistes soos omvangrykheid van dekking, betroubaarheid en die kwaliteit van bewerking nie. Trouens, die gebruik van die rekenaar bied juis die moontlikheid om in 'n groter mate aan hierdie vereistes te voldoen.

- (iii) Met die hulp van 'n woordverwerkingspakket kan leksikale inskrywings direk op die rekenaarskerm saamgestel, foutekorleksies aangebring en die nodige kontrole en proefleeswerk onmiddellik gedoen word (vgl. Alberts 1983:122). Die feit dat hierdie gegewens daarna outomaties in die rekenaargeheue opgeneem kan word, skakel latere kontrole uit.
- (iv) Dit is veral ten opsigte van woordeboekproduksie dat 'n gerekenariseerde woordbank verskeie voordele inhou en allerlei nuwe moontlikhede bied.

In 'n omvattende woordbank is elke leksikon 'n meestersversameling van inligting oor die betrokke leksikale item waaruit verskillende woordeboeke feitlik outomaties saamgestel kan word.

Die aantal ontsluitingsmoontlikhede waarvoor die bestaande programmatuur voorsiening maak, bring mee dat allerlei gespesialiseerde woordeboeke (na gelang van spesifieke gebruikersbehoefte) op grond van dieselfde databasis saamgestel en relatief goedkoop uitgegee kan word (bv. in die vorm van rekenaardrukstukke).

Die databasis kan in sy geheel of in onderdele vir ander leksikografie-projekte (bv. vir die opstel van vertalende woordeboeke) beskikbaar gestel word. Hierdie moontlikhede hou ook die voordeel in dat grootskaalse duplisering in die samestelling van databasisse vir leksikografieprojekte uitgeskakel word (vgl. byvoorbeeld Van Sterkenburg, Martin en Al (1982); Moore (1983); Michiels, Mullenders en Noël (1982); Norling-Christensen (1982)).

- (v) 'n Groot voordeel wat die rekenaar vir die verdere bewerking en toevoeging van die inligting oor 'n spesifieke leksikale item bied, is dat afsonderlike lêers maklik en vinnig vir die bewerking en kontrole van ander lêers opgeroep kan word, en dat besondere inligtingskategorieë dadelik in 'n verskeidenheid lêers toeganklik is. Hierdeur word omslagtige kontroleprosedures direk uitgeskakel, en die leksikograaf het onmiddellik toegang tot die inligting op grond waarvan hy sy toevoegings of veranderings in 'n individuele lêer kan motiveer.



- (vi) Die veranderinge wat die woordeskat van enige taal voortdurend ondergaan en die opspoor van ontbrekende inligting oor spesifieke leksikale items bring mee dat die LI'S van 'n woordbank voortdurend bygehou moet word. Inligting van hierdie aard en die veranderinge kan vinnig en teen 'n baie lae koste op 'n gerekenariseerde lêer aangebring word.

As die basiese materiaal vir 'n woordeboek eers in masjienleesbare vorm beskikbaar is, kan resente inligting nog tot voor publikasie ingesleutel word en materiaal vir hersiene en uitgebreide herdrukke kan maklik toegevoeg word.

### 3.3 PROFIELE VAN 'N AANTAL LEKSIKOGRAFIESE ONDERNEMINGS

#### 3.3.1 Longman-Uitgewers

Longman is 'n uitgewersmaatskappy wat reeds sedert 1975 die rekenaar inspan vir die produksie van woordeboeke, dit is vir die prosessering van leksikale materiaal, die saamstel van woordeboeke en die fotoset en publikasie van woordeboeke.

Naas die rekenaarspan wat verantwoordelik is vir die apparatuur en programmatuur wat in die produksie van hulle woordeboeke gebruik word, bestaan die woordeboekspan van Longman verder uit die volgende: 'n groep leksikograwe vir die (intellektuele) verwerking van die leksikografiese materiaal, 'n groep navorsers en ontwerpers, wat die nodige navorsing oor gebruikersbehoefte doen en verantwoordelik is vir die beplanning en ontwerp van leksikografiese produkte, en 'n versameling buite-lesers (sg. 'readers') wat tekste (koerante, tydskrifte en ander publikasies) lees met die doel om nuutskeppinge te versamel en om die betekenisverandering van bestaande woorde te monitor.

Alle leksikografiese materiaal, in die besonder al die masjienleesbare leksikons wat ook in hardekopie-vorm al uitgegee is, word in 'n sentrale databasis geberg en op allerlei maniere bewerk en herbewerk vir nuwe leksikografiese produkte. Die inhoud van die sentrale databasis word aangevul met die materiaal wat deur die groep lesers versamel word. Hierdie materiaal word in bewerkte vorm in 'n sitatedatabasis geberg.

#### (a) Databasisorganisasie

Die leksikografiese databasis van Longman bestaan uit:

- (i) 'n sentrale gerekenariseerde databasis met al die leksikons wat deur Longman in hardekopievorm gepubliseer is;
- (ii) 'n sitatedatabasis.

(i) Die sentrale databasis

Elke leksikon word as 'n afsonderlike lêer in die databasis geberg. Elke leksikon is as 'n meesterleksikon gestruktureer, dit wil sê die inhoud daarvan is so volledig moontlik gekodifiseer en gestruktureer sodat elkeen as databasis hergebruik kan word vir die produksie van nuwe woordeboeke. Die aard en omvang van sommige van die gebergde leksikons blyk uit Fig. 3.12.

Tabel 3.12: Longman-databasis

Longman Dictionary of Contemporary English	Longman Dictionary of English Idioms	Longman Lexicon of Contemporary English	Longman Dictionary of Phrasal Verbs
1978 1303 pp. 55,000 LIS	1979 387 pp. 4500 Idiome en hulle afleidings	1981 910 pp. 15,000 LIS	1983 744 pp. 12,000 LIS

Longman Active Study Dictionary of English	Longman Dictionary of Business English	Longman Dictionary of American English	Longman Dictionary of Scientific Usage
1982 492 pp. 12,500 LI's	1983 766 pp. 38,000 LI's	1983 792 pp. 38,000 LI's	1979 684 pp. 10,000 LI's

(LI's = leksikale inskrywings)

## (ii) Die sitatebasis

Die sitatebasis word saamgestel uit die leksikografiese materiaal wat deur 'n groep buite-lesers op kontrakbasis verskaf word. Vir hierdie doel is 'n verskeidenheid tekstipes geselekteer wat 'n baie breë spektrum van die woordeskat van Engels dek en oor 'n verskeidenheid onderwerpe handel. Voorkeur word gegee aan tekste in die algemene spreektaal wat vir neologismes en betekenisveranderinge by bestaande woorde gekontroleer word (bv. koerant- en tydskrifartikels oor 'n verskeidenheid populêre onderwerpe, gebruikershandleidings vir allerlei produkte en populêre geskrewe vakartikels).

Die buitelesers lewer die leksikale materiaal op voorgeskrewe vorms aan Longman waarop die volgende gegewens verskaf word:

- die betrokke leksikale item;
- 'n voorgeskrewe lengte konteks waarin die betrokke leksikale item voorkom;
- volledige bibliografiese besonderhede oor die bron.

Hierdie materiaal word dan deur die leksikografiespan bewerk (bv. betekenisveranderinge word omskryf en nuutskeppings word vir hulle frekwensie en stylaard gekontroleer). Die verwerkte materiaal word dan in die sitatebasis geberg en gebruik om die leksikale inskrywings van die gebergde leksikons by te hou of om nuwe publikasies te genereer (bv. die woordeboek van neologismes wat voorberei is vir The Guardian).

## (b) Databasisbewerkinge

Gegee dat die struktuur van 'n woordeboek met die insleutel van die leksikale inskrywings vasgevang word in 'n kodesisteesem, is die materiaal geskik vir teks- en woordprosesseringsprogramme. Hiermee is dit moontlik om te bepaal wat die omvang van die beskikbare materiaal in 'n lêer is, om 'n aantal kontroleprosedures ten opsigte van die materiaal uit te voer, om wysiginge aan die bestaande materiaal aan te bring en om die inhoud van 'n lêer vir hergebruik te ontgin (vgl. Moore (1983)). Met die ontwerpte programmatuur van Longman is dit moontlik om ten opsigte van enige leksikon te bepaal wat die omvang daarvan is (bv. die aantal inskrywings en die aantal definisies) en watter leemtes onderdele van die materiaal vertoon. Statistiek hieroor word gebruik vir die opstel van werkprogramme, die bepaling van die hulpbronne en om leemtes in die materiaal aan te vul.

'n Aantal programme word ingespan vir die kontroleer van spesifieke aspekte van die materiaal in 'n lêer, soos die kontroleer van die woordeskat wat in definisies gebruik word. Hierdie fasiliteit maak dit moontlik om foute in die definiëringswoordeskat te monitor en te hersien. Kontrole van die kruisverwysings in leksikale inskrywings word uitgevoer met behulp van programmatuur wat bepaal of elke kruisverwysing binne die materiaal wel 'n verwysing het. Dit geskied deur lyste te

genereer waarin elke kruisverwysing gegee word met die materiaal waarna daar verwys word. Bogenoemde programmatuur maak dit dus moontlik om 'n woordeboek met groter akkuraatheid te produseer en om sirkulariteit in die definisies uit te skakel. Hierdie kontrolemechanismes is onontbeerlik in die gevalle waar afgeleide publikasies vervaardig word en die materiaal daarvoor saamgestel word deur die materiaal van 'n bestaande woordeboek te verkort, te wysig of uit te brei.

Stringsoekprogramme word ingespan om foute in leksikale inskrywings op te spoor. As 'n fout in hierdie verband gevind word, word die parameters daarvan eenduidig beskryf en die korreksies geskied dan met behulp van gegenereerde lysse waarin die betrokke woord (in konteks) of teksdeel aangegee word. Tekswysigings word dan uitgevoer met stringsoekprosedures en vervangingsroetines.

Die grootste voordeel van die gerekenariseerde databasis is die hergebruiksmoontlikhede van die gebergde materiaal. As die besonderhede van 'n nuwe woordeboek eers vasgelê is en genoeg bruikbare materiaal in die databasis beskikbaar is, word 'n beskrywing van die verlangde materiaal as parameters van 'n algemene ekstraksieprogram/herwinningsprogram ingevoer wat 'n gedrukte kopie van hierdie materiaal lewer. Hierdie proewe word dan gekontroleer, die parameters van die herwinningsprogram gewysig (indien nodig) en 'n proef vir finale redigeerwerk (kontroleprosedures en wysigings) gelewer. Die afgehandelde materiaal word dan gestuur deur 'n koppelvlakprogram vir die finale setwerk. Hierdie program wys alle struktuurkoderinge in die teks uit en laat net dié deur wat tipografiese inligting bevat sodat die blaaie van die publikasie outomaties opgemaak kan word.

### (c) Die navorsingskomponent

Die taak van die navorsingskomponent is om op grond van opnames oor gebruikersbehoefte leksikografiese produkte te ontwerp wat aan hierdie behoeftes voldoen. Die resultate van hierdie ondersoek het 'n direkte invloed op die struktuur van woordeboeke in die geheel en op die inhoud en struktuur van die leksikale inskrywings in 'n woordeboek. Van die vernaamste manifestasies hiervan is die volgende:

- (i) die uitbreiding van die leiding wat daar vir die gebruiker verskaf word oor die maksimale benutting van die aangebode leksikale materiaal;
- (ii) die opneem van uitgebreide grammatikale aantekeninge in die vorm van subgrammatikas;
- (iii) die toename in die ensiklopediese karakter van die woordeboek, byvoorbeeld deur 'n groter hoeveelheid illustrasie materiaal en bylaes op te neem waarin inligting verskaf word oor taalkultuurspesifieke aspekte (byvoorbeeld 'n tabel wat familieverwantskappe, militêre range, mate en gewigte, ens. toon);

- (iv) die uitbreiding van die aantal en inhoud van die inligtingsvelde in leksikale inskrywings (bv. uitgebreide gebruiksaantekeninge en voorbeelde wat die gebruik van 'n woord in al sy betekenisonderskeidings illustreer);
- (v) die gebruik van 'n gekontroleerde definiëringswoordeskat in definisies.

(ii) bring mee dat die onderskeid tussen leksikon en grammatika al hoe meer vervaag en (iii) dat die verskil tussen linguistiese woordeboek en ensiklopedie al hoe geringer word. Die uitbreiding van die inligting in 'n woordeboek bring ook mee dat daar om redes van ekonomie in die beskrywing toenemend op die kodifisering van die aangebode inligting gesteun moet word (bv. deur die gebruik van letterkodes wat verwys na spesifieke sintaktiese patrone wat in 'n afsonderlike grammatika opgeneem is) en dat tipografiese konvensies maksimaal ontgin moet word.

Die uitbreiding van die inhoud van en aantal inligtingskategorieë in 'n woordeboek hou die voordeel in dat (i) hierdie leksikografiese produkte meer geskik is vir taalonderrig- en taalverwerwingsprogramme; en (ii) dat hierdie leksikons optimaal benut kan word vir die saamstel van leksikons vir ander outomatiese teksverwerkingsisteme (bv. vir outomatiese vertaalsisteme). Vergelyk in hierdie verband die bespreking in Michiels, Mullenders en Noël (1982).

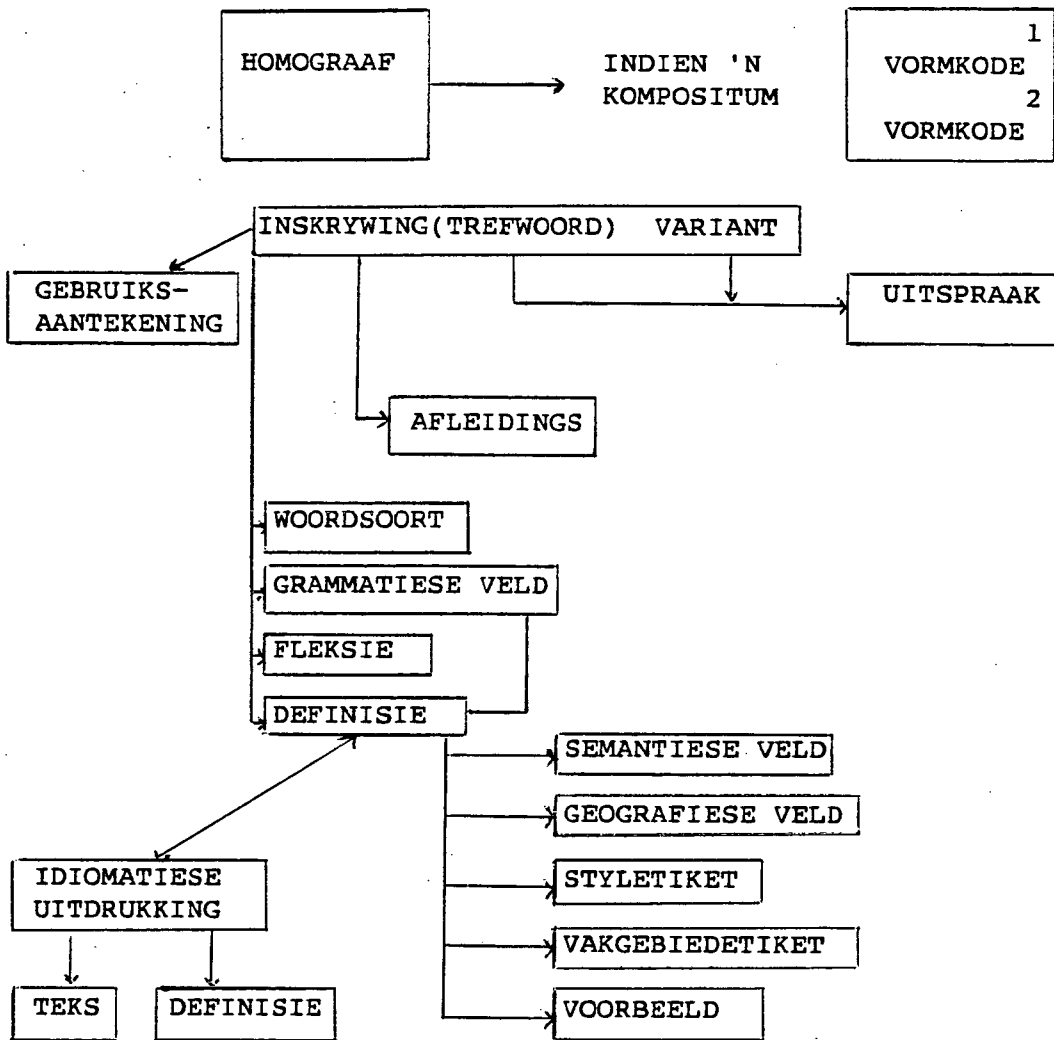
#### (d) Voorbeeldmateriaal

Die inhoud en struktuur van die leksikale inskrywings in die Longman Dictionary of Contemporary English (LDOCE) word geïllustreer in Figure 3.13 en 3.14.

Fig. 3.15 illustreer 'n gegeneerde lys leksikale inskrywings wat volgens bepaalde parameters van die sentrale databasis van Longman herwin is en vir leksikografiese bewerking gereed is.

Fig. 3.16 illustreer die logiese struktuur van die leksikale inskrywings in die Longman Dictionary of English Idioms (LDOEI).

Fig. 3.13: Logiese struktuur van 'n leksikale inskrywing in die LDOCE



- 1) Aanduiding van bepaalde eienskappe soos die aantal woordvorme in die kompositum en die aan- of afwesigheid van koppeltekens.
- 2) Kodering van die woordsoort waartoe die eerste en die laaste woordvorm van die kompositum behoort.

Vergelyk Michiels, Mullenders en Noël (1982:204).

Fig. 3.14: Voorbeeldinskrivings uit die LDOCE

<p>hos-pi-tal-ize, -ise /'hɒspɪtəl-aɪz/ 'hɑ-/ v [T1 often pass.] to put (a person) into hospital: <i>He broke a leg and was hospitalized for a month</i> —ization /,hɒspɪtəl-aɪ'zeɪʃən/ ,hɒspɪtəl-ə'zeɪ-/ n [U; (C)]</p> <p>host<sup>1</sup> /həʊst/ n [GC (of)] a large number</p> <p>host<sup>2</sup> n 1 a man who receives guests (note the phr. host to): <i>He acted as host to his father's friends.</i> (fig.) <i>the host country for the Olympic Games</i> 2 <i>old use or humor</i> an innkeeper (note the phr. mine host) 3 an animal or plant on which some lower form of life is living as a PARASITE</p> <p>host<sup>3</sup> v [T1] to act as host at (a party, friendly meeting, etc.)</p> <p>host<sup>4</sup> n [the + R] (often cap.) the holy bread eaten in the Christian service of Holy COMMUNION (2)</p> <p>hos-tage /'hɒstɪdʒ/ 'hɑ-/ n 1 a person kept by an enemy so that the other side will do what the enemy wants: <i>The man with the gun took the child with him as a hostage and no one could do anything for fear the child would be killed</i> 2 give hostages to fortune to accept responsibilities that may make it hard to act freely in the future: <i>Francis Bacon believed that a married man had given hostages to fortune because he had to consider his wife and children before taking any risks</i> 3 take (someone) hostage to catch (someone) from the other side to use as a hostage</p> <p>hos-tel /'hɒstl/ 'hɑ-/ n 1 a building in which certain types of person can live and eat, as for students, young people working away from home, etc. 2 YOUTH HOSTEL</p> <p>hos-tel-er, AmE usu. hos-tel-er /'hɒstəls/ 'hɑ-/ n a person travelling from one HOSTEL (2) to another</p> <p>hos-tel-ry /'hɒstəlri/ 'hɑ-/ n <i>old use</i> an inn</p> <p>host-ess /'həʊstəs/ n 1 [C] a female host 2 [C] <i>old use or humor</i> a female innkeeper (note the phr. mine hostess) 3 a young woman who acts as companion, dancing partner, etc., in a social club 4 [C; (N)] AIRHOSTESS</p>	<p>hot cross bun /,hɒt'krɒsbʌn/ n a sort of small cake made of bread with a cross-shaped mark on top, which is eaten on Good Friday, just before EASTER</p> <p>hot dog /,hɒt'dɒg/ n a special sort of long red SAUSAGE in a bread ROLL (2)</p> <p>USAGE In <i>informal AmE</i>, the expression Hot dog! (/,hɒt'dɒg/) can be used for showing approval or pleased surprise: "You mean we're really going? Hot dog!"</p> <p>ho-tel /həʊ'tel/ n a building where people can stay if they pay a certain amount of money each night</p> <p>USAGE One BOOKS (1) or RESERVES (2) a double or single room in a hotel, with bath/without bath.</p> <p>ho-tel-i-er /həʊ'teliə, -iə/ n a man who keeps a hotel</p> <p>hot flush /,hɒt'flʌʃ/ AmE usu. hot flash /,hɒt'flʌʃ/ n [usu. pl.] a sudden feeling of heat in the skin, esp. as experienced by women at about the time (MENOPAUSE) when they cease to be able to bear children</p> <p>hot-foot<sup>1</sup> /,hɒt'fʊt/ 'hɒt'fʊt/ adv <i>informal</i> (of movements) fast and eagerly: <i>We ran hotfoot to find out the news</i></p> <p>hotfoot<sup>2</sup> v hotfoot it <i>informal</i> to move fast: <i>We hotfooted it down the street</i></p> <p>hot-head /'hɒthead/ 'hɒt-/ n a person who does things in haste, without thinking —ed /,hɒt'hedʒd/ adj [Wa2] —edly /,hɒt'hedʒdli/ adv</p> <p>hot-house /'hɒthaus/ 'hɒt-/ n -houses /,hɒt'haʊzɪz/ 1 a warm building where flowers and delicate plants can grow; GREENHOUSE 2 hothouse plant a delicate or sensitive person who must be treated more carefully than other people</p> <p>hot line /,hɒt'laɪn/ n a direct, usu. telephone, line between heads of government, to be used at times of great difficulty, esp. when war is threatened</p> <p>hot-ly /'hɒtli/ 'hɒtli/ adv 1 in anger and with force 2 closely and eagerly (often in the phr. hotly</p>
--	---

Fig. 3.15: Herwinning van databasismateriaal volgens aangeduide parameters (trefwoord, woordsoort, definisie, voorbeelde met aanduiding van die bron (Longman/Merriam))

---

FULL ALPHA PRINT NEW LADYBIRD

HDW = yesterday

+DEF (on) the day before this one

+DFC 2 / /DA...../../..../...../../60/LONGMAN...../

+DEF only a short time ago: the fashions of yesterday

+RUN Yesterday

+RPS adj

---

+HDW 030189 yet 1

+POS adv

+DFC 1 / /DA...../../..../...../../50/LONGMAN...../

+DEF at this moment: so far; still; at a future time; even; in

+DEF addition; again

+DFC 2 / /DA...../../..../...../../50/LONGMAN...../

+DEF as yet up to this moment

---

+HDW 030205 yolk

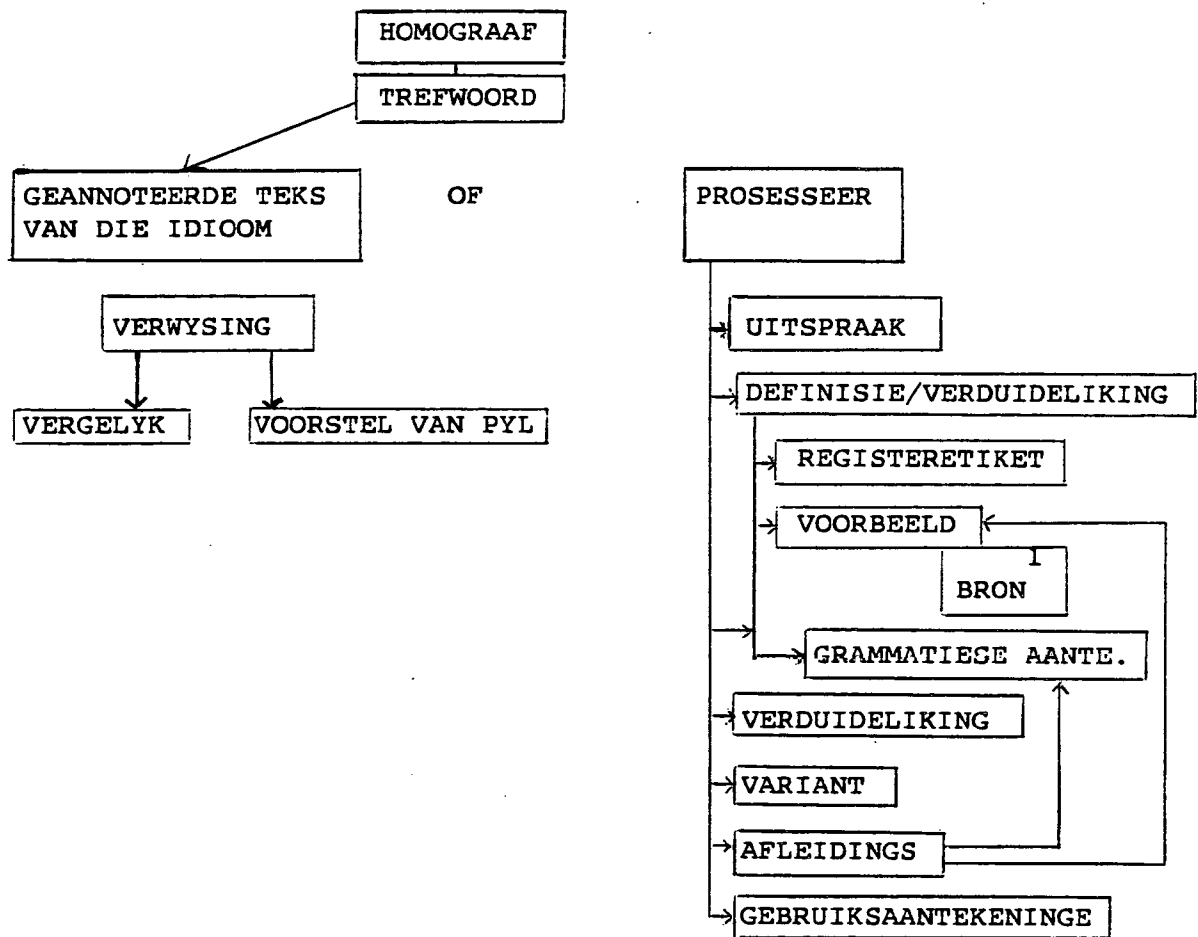
+POS n

+DFC 0 / /MY.MA...../../..../...../../50/MERRIAM

=DEF the yellow central part of an egg



Fig. 3.16: Logiese struktuur van 'n leksikale inskrywing in die LDOEI



1) Net vir voorbeelde wat nie nuut geskep is nie en wat lekseme kan bevat wat nie in die definiëringswoordeskat voorkom nie.

Vergelyk Michiels, Mullenders en Noël (1982:208).

### 3.3.2 Die New Oxford English Dictionary-Projek (NOED)

In gedrukte vorm dek die Oxford English Dictionary (OED) die woordeskat van Engels vanaf 1150 tot vandag. Die twaalf hoofdele en vier bylaes ('Supplements') waaruit die OED saamgestel is, beslaan 21,000 gedrukte bladsye en bevat 400,000 definisies in 60,000,000 woorde en illustrasies. Na die voltooiing van die hoofdele in 1928 is hulle bygewerk met behulp van die vier bylaes. Dit het meegebring dat die leksikale inskrywings van ongeveer 10-20 persent van die trefwoorde van die geheel oor die hoofdele en die bylaes versprei is. Die laaste van die bylaes het in 1986 verskyn.

Met die voltooiing van die laaste bylae (en met die geleentheid van die honderdste verjaardagviering van die OED) het die uitgewers Oxford University Press (OUP) besluit om 'n omvattende rekenariseringsprogram aan te pak

- (i) om die twaalf hoofdele en die vier bylaes van die OED tot 'n enkele databasis te integreer en om die geïntegreerde databasis in hardekopie-vorm vry te stel;
- (ii) om die bestaande sitatebasis wat die betekenis en gebruik van die opgenome woorde in die OED illustreer, tot 'n gerekenariseerde sitatebasis van die Engelse letterkunde uit te bou;
- (iii) om 'n ontwikkelingsprogram aan te pak waarin die geïntegreerde woordeboekdatabasis bygewerk en met 'n aantal satelietdatabasisse aangevul sal word, die inhoud van die databasisse op verskillende datadraers vrygestel sal word en waarin ander toegangsmoontlikhede tot en vrystellingsmoontlikhede van die databasisse ontwikkel sal word.

Naas die leksikografie- en bemarkingspan van OUP is die vernaamste vennote in die onderneming die volgende: IBM (vir die ontwikkeling en verskaffing van die benodigde apparatuur en programmatuur), die Universiteit van Waterloo (vir verdere navorsing in verband met die rekenariseringsprogram) en International Computaprint Corporation (ICC) (vir die insleutel van die hele woordeboekteks).

Die hoofdoel van die eerste fase van die projek is om die geïntegreerde databasis daar te stel en die geheel in hardekopie-vorm te publiseer. In die onderstaande dele word daar in nadere besonderhede ingegaan op die teksontledingsfase van die eerste deel van die projek.

(a) Teksontleiding

As aanloop tot die rekenarisering en integrasie van die hoofdele en die bylaes van die OED is daar in 'n agt maande projek

- (i) 'n ontleding gemaak van die inhoud en struktuur van die woordeboekteks as 'n geheel, van die inligtingsvelde waaruit die leksikale inskrywings van die hoofdele en die bylaes saamgestel is en van die wyse waarop die inligtingsvelde leksikografies en tipografies gemerk word;
- (ii) 'n volledige koderingstelsel (die sg. 'Generalised Mark-up System') ontwerp op grond van die struktuurelemente waaruit die leksikale inskrywings bestaan en die tipografiese konvensies wat in die saamstel van die woordeboek in die hoofdele en in die bylaes gevolg is;
- (iii) 'n voorlopige metode ontwerp vir die kodering van die teks vir die insleutelfase;
- (iv) 'n ontleding- en benoemingsprogram (d.i. 'n 'parser') ontwerp wat die leksikale inskrywings outomaties in hulle samestellende inligtingsvelde en -subvelde verdeel en benoem. Vir hierdie doel is die voorlopige koderingsstelsel wat op tipografiese en struktuurkenmerke steun, vertaal in 'n suiwer strukturele stelsel.

Vir die integrasie van die leksikale inskrywings van die hoofdele en die bylaes, die latere bewering van hierdie inskrywings en die herwinning van onderdele van die databasis was dit nodig

- (i) om die inligtingsvelde van die leksikale inskrywings te identifiseer en apart te kodeer sodat hulle deur die rekenaarsstelsel onderskei kan word;
- (ii) om vas te stel hoe die inligtingsvelde gemarkeer word sodat bepaal kon word watter inligtingsvelde reeds in die insleutelfase gekodeer moes word en watter outomaties met behulp van 'n ontledings- en benoemingsprogram gekodeer sou kon word.

Meer as veertig datakategorieë is in die leksikale inskrywings van die hoofdele en die bylaes van die woordeboek onderskei en aan elkeen is 'n etiket uit die koderingstaal (GML) toegeken.

Enige woordeboek bevat 'n mengsel tekstuele en strukturele inligting wat gewoonlik deur tipografiese konvensies (bv. die swaardruk van lemmas en die skuinsdruk van definisies) en die plasingsvolgorde van die inligtingsvelde gemarkeer word. Waar die tipografiese konvensies konsekwent toegepas word en die struktuurering van 'n leksikale inskrywing vas is, is dit moontlik om sekere van die inligtingsvelde (of die data-elemente waaruit hulle opgebou is) outomaties te identifiseer of om die aan- of afwesigheid van bepaalde data-elemente te voorspel op grond van die aan- of afwesigheid van ander data-elemente. In die eerste fase van die ontleding van die teks kon hierdie middele dus eksplisiet geformaliseer en gebruik word vir die eerste ontleding en benoeming van die onderdele waaruit die leksikale inskrywings saamgestel is. Nie alle inligtingsvelde en hulle inhoud word egter op die manier gemarkeer nie en gevolglik was dit nodig om sekere inligtingsvelde van etikette te voorsien (bv. op grond van die inhoud van bepaalde teksgedeeltes in die leksikale inskrywings).

Die kompleksiteit van die datastruktuur van die leksikale inskrywings en die probleme met die herkenning van die onderdele van inskrywings het meegebring dat die ontleding en benoeming van die leksikale inskrywings in twee fases moes verloop: met die insleutel van die teks en met die hulp van die outomatiese ontleding- en benoemingsprogramme. Die verskille in die leksikografiese struktuur van die inskrywings in die hoofdele van die woordeboekteks en die bylaes het verder meegebring dat aparte ontledings- en benoemingsprogramme vir die twee soorte leksikale inskrywings ontwerp moes word en dat die ontleding van die woordeboekteks dus in twee fases moes geskied.

Die ontledingsprogramme is ook so ontwerp dat hulle die indekse en kruisverwysings tussen onderdele van die teks vir die integrasieproses outomaties genereer.

#### (b) Databasisbewerking

Die insleutel en prekodering van die woordeboekteks word uitgevoer deur ICC wat die resultaat aan OUP lewer op magnetiese skywe (ongeveer veertig in totaal). Die swak tipografiese gehalte van die eerste hoofdele van die woordeboek het meegebring dat optiese lesers nie vir die invoer van die teks gebruik kon word nie. Die personeel van OUP het direkte toegang tot die woordeboekteks met behulp van terminale. Galeiproewe van die teks word gedruk, geproeflees en aan ICC vir korreksies teruggestuur. Tans word 'n foutgrens van 7 uit 'n 100,000 karakters gehandhaaf. Die gekorrigeerde teks word stuk vir stuk deur die ontledings- en benoemingsprogramme gestuur wat die afvoer daarvan na die databasis oorplaas vir die latere integrasieprogram. Hierdie ontleding van die teks verloop vir 80 % outomaties; die res word intellektueel verwerk en ontleed.

Die integrasie van die leksikale inskrywings van die hoofdele van die woordeboek en die bylaes bestaan daaruit dat

- (i) nuwe leksikale inskrywings in die bylaes by dié van die hoofdele ingesorteer word (dit geld vir ongeveer die helfte van die inhoud van die bylaes);
- (ii) dat die (opsigself onvoltooide) leksikale inskrywings van die bylaes aan die leksikale inskrywings van die hoofdele toegevoeg word of as basis gebruik word om die leksikale inskrywings van die hoofdele te wysig.

Proewe van die resultate van die integrasieprosedure word uitgedruk en vir foute gekontroleer. Spesifiek ontwerpte redigeringsprogramme word vir hierdie doel ingespan, asook vir die intellektuele integrasie van leksikale inskrywings wat nie deur die integrasieprogramme hanteer kan word nie.

Die goedgekeurde materiaal word na finale redigering op magnetiese band geplaas en deur 'n gesofistikeerde setprogram gestuur wat die bladsye outomaties opmaak en vir finale publikasie gereedmaak.

#### (c) Bewerking en ontwikkeling

Die feit dat die inligtingsvelde van die leksikale inskrywings in die elektroniese databasis gestruktureer en volledig benoem is en direk en sistematies toeganklik is, bring mee dat dit moontlik is om in alle nuwe uitgawes van die woordeboek foute te korrigeer, inligting aan te vul of by te werk en om nuwe leksikale inskrywings aan die geheel toe te voeg.

In die eerste fase van die rekenariseringsprojek is daar ook 'n opname van gebruikersbehoefte en die potensiële gebruike van die databasis gemaak. Die vraelyste wat vir hierdie doel uitgestuur en terugontvang is, word tans ontleed om te bepaal watter entiteite, attribute en relasies die databasis sal moet besit om die grootste toepassingsmoontlikhede daarvan op die mees ekonomiese wyse daar te stel.

Die gebruiksmoontlikhede van die databasis sal in die toekoms uitgebrei word deur die ontwikkeling van 'n aantal satellietdatabasisse, byvoorbeeld:

- (i) 'n gerekenariseerde argief van die sitate wat nie in die hoofdatabasis opgeneem is nie. Tans is daar nog ongeveer drie miljoen sulke sitate. Hierdie databasis sal vir taal- en letterkundenavorsing ingespan kan word, vir die opstel van frekwensielyste en vir die ontwikkeling van 'n verbeterde sisteem van styltipering;
- (ii) woordeboeke van die nasionale en streeksvariante van Engels;
- (iii) 'n tesaurus wat gebaseer sal word op die Glasgow Historical Thesaurus of English en wat aangevul sal word met die definisies uit die NOED-databasis;

(iv) 'n databasis van eiename.

Daar word tans ook voorsien dat die databasis aangevul sal kan word met lyntekeninge (vir illustrasie materiaal by definisies), die toevoeging van kaarte, planne en diagramme, die gebruik van bewegende beelde om prosesse te illustreer en oudio-materiaal om uitspraak weer te gee.

Wat die ontwikkeling van programmatuur betref, sal daar in die toekoms toegangs- en herwinningsmoontlikhede tot die hoofdatabasis en die satellietbasisse vir gebruikers ontwerp moet word. Hieronder val programmatuur om buite-gebruikers toegang tot die databasisse te gee, redigeringsprogramme vir leksikograwe en programmatuur wat die semantiese skakels in die bestaande databasisse sal implementeer vir die saamstel en toegang tot die tesourus. Verder is dit nodig om vrystellingsmoontlikhede van die hele woordeboek of onderdele daarvan te ontwikkel. Van die datadraers wat vir hierdie doel ingespan sal word, is gedrukte kopieë, laser-diskette, magnetiese bande en mikrofilm. Daar word ook daaraan gewerk om gebruikers gekoppelde toegang tot die databasis te gee deur die netwerk NEXIS of DIALOG.

(d) Voorbeeldmateriaal

Fig. 3.17 illustreer op vereenvoudigde wyse die relasie tussen 'n leksikale inskrywing in die hoofdeel van die OED en 'n ooreenstemmende, aanvullende leksikale inskrywing in die bylae. Die integrasieproses behels onder meer dat die onderdele van die leksikale inskrywing van die hoofdeel aangevul of gewysig moet word soos aangedui in die leksikale inskrywing in die bylae. Vergelyk die OED se hoofdeel en bylae vir die volledige inskrywing onder die trefwoord Map.

Die rekenarisering van die hoofdele en bylaes van die OED vereis dat die teks van 'n leksikale inskrywing (deels outomaties, deels met handwerk) volgens 'n koderingstaal ontleed moet wees. OUP het in samewerking met IBM 'n algemene koderingsstelsel vir die ontleding van leksikografiese tekstmateriaal ontwerp. Hierdie stelsel is in die ontleding van die OED toegepas en dit dien tans voor die Britse standaard-instituut vir standaardisering. Hierdie koderingsstelsel word as Fig. 3.18 hier opgeneem.

Fig. 3.17: Integrasie van leksikale inskrywings

**Map** (mæp), *sb.*<sup>1</sup> Also 6-7 *mappe*, 6-8 *mapp*. [ad. L. *mappa*, in class. L. 'table-cloth, napkin', but in med. L. used *transf.* in the combination *mappa mundi* (see MAPPEMONDE). Cf. the synonymous OF. *mappe* (rare; also in Rousseau c1770, Sp. *mapa*, Pg. *mapa*, G. *mappe* (obs.: the mod. sense 'portfolio' is not directly connected).]

L. A representation of the earth's surface or a part of it, its physical and political features, etc., or of the heavens, delineated on a flat surface of paper or other material, each point in the drawing corresponding to a geographical or celestial position according to a definite scale or projection.

A hydrographical map is now more usually called a *chart* (formerly *chart*).  
 1527 R. THORNTON in Hakluyt *Voy.* (1589) 257 To make a bigger and a better *mappe*. 1589 G. HARVEY *Pierres's Supper. Wks.* (Grosart) II. 130 The great *Mappe* of Mercator. 1601 SHAKS. *Twel. N.* III. II. 54 He does smile his face into more lynes, then is in the new *Mappe*, with the augmentation of the Indies. 1605 N. CARPENTER *Geog. Del.* I. vii. (1633) 166 A Geographical *Mappe* is a plaine Table, wherein the Lineaments of the Terrestrial Spheare are expressed. 1760 JOHNSON *Idler* No. 97 ¶ 5 A rivulet not marked in the *mappe*. 1867 W. W. SMITH *Coal & Coal-mining* 44 On examination of a geological map it will be seen that [etc.].

b. *transf.* † A table, chart (obs.). Also (*nonce-use*), applied to a mental conception of the arrangement of something.  
 1606 [FEATLEY] in *Parall. To Rdr. A. II*, Errors... which collected into a small *map*, they exhibit. 1855 BAIN *Senses & Int.* III. I. § 13 (1664) 198 By a hurt on the ribs we come to connect feelings in the chest with the place on our *map* of the body.

c. A tract of country spread out like a map.  
 1784 COWPER *Task* I. 321 A spacious *map* Of hill and valley interpos'd between.

d. A figure resembling a map in form or outline.  
 1822 GOOD *Sturdy Med.* IV. 571 Mottled and drif. Scalliness in diffuse *maps* of irregular outline, and diverse colours.

2. *fig.* A detailed representation in epitome; a circumstantial account of a state of things. Very common in the 17th c.; now rare or obs.  
 1596 CRESS *Pembrace Fe. cxlii. i.* My voice. [O Lord], before thy face: my *mappe* is laich. 1607 ROWLAND *Cure it ever.* ¶ Who in her face a *Map* of sorrow wear. A countenance compos'd all mournful sad. 1647 SALTHERN *Paradise Lost* 11. 157 So as *map* is all created excellency in the *map* of *the* *world*. 1721 BRYAN *Letts. & A. Serms* 1244: III. 27, I can't know the *map* of their situation.

† b. The embodiment or incarnation (of a virtue, vice, character, etc.); the very picture or image of. (No Sp. *mapa*.) (Obs. (Common in the 17th c.)

1591 H. SMITH *Sinf. Man's Search Six Serms.* (1614) E. What were man if he were once left to himself? A *map* of misery. 1606 CHAPMAN *Mons. D'Olivo Plays* 1873 L. 200 Farewell the true *mappe* of a gull. 1698 FAYER *Acc. E. India & P.* 13 They are the absolute *map* of sordidness, fawning hardy, and professing fairly.

† c. An aggregation, multitude. Obs.  
 1598 GREGG *Declarat. Wks.* (Grosart) XIV. 109 In whose high thoughts A *map* of many values is unshrind. 1597 MIDDLETON *Wid. Soliman* xv. 22 My soul, saith he, is but a *map* of shew. No substance, but a shadow for to please.

† 3. [? After L. *mappa*; cf. G. *mappe* portionio.]  
 ? A wrapper. Obs.  
 1608 TOPSELL *Serpents* 220 But some then will demand, where had Pope Alexander... that *map* or net at Rome wherein (it is said) the napkin of our Saviour Christ is preserved.

4. *attrio. and Comb.*, as *map-graver*, *-maker*, *-making*, *-monger*, *-mounter*, *-seller*; *map-like* adj.; *map butterfly*, a butterfly with map-like markings; *map-flapping Nil.*, the process of transmitting by flag-signals the outline of a *map* (or other drawing); *map lichen*, a lichen, *Lecidea geographica*, the thallus of which has markings resembling a *map*; *map-measurer*, an instrument for measuring distances on *maps* (Knight *Dict. Mech.* 1875); *map-meter* = prec.; *map-turtle*, an American turtle, *Malaclemmys geographica*, called from the markings of the shell (*Cent. Dict.*). 1894 E. H. ATTKEN *Naturalist on Prout* 50 The delicately devised \**Map Butterfly*, *Cyrestis thycodanus*. 1886 Longm. *Mag. Feb.* (art.) \**Map-flapping*. 1662 EVELYN *Chalcogr. Contents, Chart* and \**Map-graver*. 1796 NEMICH *Polygl. Lex. Nat. Hist.*, \**Map-lichen*, *Lichen geographica*. 1775 ROMANS *Florida* App. 77 Our wise \**map-makers*... have corrupted it into *Ponio bay*. 1807 PARKER *Catal. Instruments* 30 Oplisometer or \**Map Meter*. 1639 FULLER *Holy War* xiv. (1840) 267 A great \**mapmonger*... undertook to travel over England by help of his *maps*. 1858 SIMMONDS *Dict. Trade*, \**Map-mounter*, a workman who backs *maps* with canvas, varnishes and fixes them on rollers [etc.]. 1710 Lond. *Gas*. No. 4695/4 Sold by C. Browne, Print and \**Map-seller*.

Hence *Map-press a.*, without *maps*.  
 1659 FULLER *App. Inj. Junoc.* I. 5 Mr. Camden's *Britania*. His first Edition was a *Babe* in a little.. Octavo... the third, a *Youth* in a Quarto (but *Map-press*). 1889 T. HODGKIN *Dynasty of Theodosius* 15. Their deficiency of light cavalry prevented them from obtaining, in those *mapless* days, the much-needed information.

**Map, sb.**<sup>1</sup> Add: L. e. *fig.* In recent phrases: *Off the map*: out of existence; into (or in) oblivion or an insignificant position; of no account; obsolete. *On the map*: in an important or prominent position; of some account or importance; in existence (see also quotes.).

(a) 1904 W. H. SMITH *Promoters* II. 54 When she [Caribage] wouldn't let up, the only thing left was to wipe her off the *map*. 1911 R. D. SAUNDERS *Col. Tadmuster* vii. 99 A good set-to is the best way... to put a stop to quarrelling. It just wipes the whole thing off the *map*. 1914 *Grand Mag.* Jan. 429/2 [He] had been so harried by the Federal officers that he had faded off the *map*. 1922 *Tatler* 6 Sept. 386/1 A man who owns a lot of coaches... said that the big staff... was off the *map* as far as he was concerned. 1924 W. M. RAINE *Troubled Waters* xiii. 243 'Anything new, Mation?'... 'Don't forget we've been off the *map* most three weeks' 1928 *Weekly Dispatch* 13 May 9/6 Cochineal insects, except for making tinctures to colour jellies, are practically off the *map* to-day.

(b) 1920 P. G. WOODROUSE *Damsel in Distress* vii. 93 What I mean to say is, you are on the *map*. You have a sporting chance. 1924 W. M. RAINE *Troubled Waters* xii. 205 Didn't know you knew I was on the *map*. You're sure honouring me. 1926 BENNETT *Lord Raingo* L. xii. 60 Some say if there's two members of the War Cabinet, it isn't Andrew Clyth and Tom Hogarth—it's Andrew Clyth and Andrew Clyth... But that isn't so. Tom's on the *map* all right.

2. (Add quotes.)  
 1781 COWPER *Hope* 607 He draws upon life's *map* a zig-zag line, That shows how far 'tis safe to follow sin. 1899 W. E. H. LECKY (*title*) *The Map of Life*.

4. *map-roller*; *map-dro*, artillery-dro in which *maps* are used for laying the guns.  
 1928 *Encycl. Brit.* XXX. 252/2 The precision with which \**map fire* could be carried out. 1851 C. CIST *Cincinnati* 245 Shade and \**map-roller*, turning in ivory, done in a superior style.

Fig. 3.18a: Koderingstelsel van die NOED

*NEW OED MARKUP SYSTEM - TAG LIST AND DESCRIPTION IBM			
tag	attributes	tag description	entity content
ENTRY	status xref	start of an entry	HWGP, LABS.VFL.ETYM.SIGNIF
HWGP		Headword Group	HWLEM. POS. PRON
HWLEM	status	Headword Lemma	#Lemma, HOM
HOM		Homonym Number	#Homno
PRON		Pronunciation	(.#Phon.#Str,#Txt,)
POS		Part Of Speech	#Post. HOM
LABS		Label list	LAB.#Txt
LAB		Label	#Lab
VFL		Variant Forms List	#Txt. VF,XRA,VDAT
VDAT		Variant Forms Date	#Date
VF		Variant Form	#Lemma
XRA	sup	Cross-reference Address	XLEM.POS.SNUM
XLEM	status	Cross-reference Lemma	#LEMMA.HOM
SNUM		Sense Number	#Sno
ETYM		Etymology	#Txt.XRA.NOTE.CF.DAT
CF		Cited Form	#Lemma
DAT		Date in Etymology or Sense	#Date
NOTE		Note in Etymology or Sense	#Txt.XRA.CF.DAT.BLEM (Sense only)
SIGNIF		Signification	QPARA.#TXT.SEN.QUOT. DERIV
SEN	num	Sense Section	LAB.#Txt.POS.XRA.BLEM. ILEM.NOTE
	para		
	status		
DERIV		Derivative Sense Section	LAB.#Txt.POS.XRA.BLEM. ILEM.NOTE
BLEM	status	Bold Lemma	#Lemma
ILEM	status	Italic Lemma	#Lemma
QPARA		Quotation Paragraph	LABS.#Txt
QUOT		Quotation	QFDAT.AUTH.WK.QTXT
QDAT		Quotation Date	#Date
AUTH		Quotation Autho	#Txt
WK		Quotation Work	#Txt
QTXT		Quotation Text	#Txt



Fig. 3.18b: NEW OED MARKUP SYSTEM - SPECIAL CHARACTER MNEMONICS

mnemonic	description	ICC code
&mac.	floating macron	+1
&uml.	floating umlaut	+2
&acu.	floating acute	+3
&grave.	floating grave	+4
&circ.	floating circumflex	

### 3.3.3 Die Instituut voor Nederlandsche Lexicologie (INL)

Die INL is in 1969 in die lewe geroep met as hooftake

- (i) die versnelde voltooiing van die Woordenboek der Nederlandsche Taal (WNT), waarmee daar in 1864 begin is;
- (ii) die byeenbring van 'n so volledig moontlike versameling Nederlandstalige teksmateriaal wat as basis kan dien vir die saamstel van 'n reeks nuwe sinchrone woordeboeke.

Vir hierdie doel is die aktiwiteite van die INL in twee afdelings georganiseer:

- (i) die afdeling WNT, wat verantwoordelik is vir die voltooiing van die WNT op grond van die bestaande sitatebasis;
- (ii) die afdeling Thesaurus, wat verantwoordelik is vir die aanlê en ontwikkel van 'n gerekenariseerde taaldatabank van Nederlands vir die periodes wat nie deur die WNT gedek word nie.

#### 3.3.3.1 Die afdeling Woordenboek der Nederlandsche Taal (WNT)

Die verdragings met die voltooiing van die WNT en die hoë koste van die hele onderneming het daartoe gelei dat die vernaamste leksikografiese taak van hierdie afdeling na 1976 gesien is as die versnelde voltooiing van die WNT op grond van die bestaande handgeskrewe sitatebasis. Die voorstelle wat hiervoor opgevolg is, word volledig bespreek in Van Sterkenburg (1976 en 1984:75 e.v.).

Sedert 1983 word mikrorekenaars ook in die leksikografiese prosedures en die uitgee van nuwe aflewings ingespan ten einde die voltooiing van die WNT te bespoedig en om alreeds die nuwe dele van die WNT in masjienleesbare vorm daar te stel vir latere invoer aan die woorddatabank van die afdeling Thesaurus. Na verwagting sal die laaste aflewering van die WNT teen die jare negentig voltooi wees. Die leksikologiese taak van die afdeling WNT is die opbou van 'n volledige sitatebasis van die bronne wat vir die reeds verskene dele van die WNT gebruik is en die seleksie en ordening van sitate vir die oorblywende dele van die WNT. Hierdie taak spruit daaruit dat alle bronnemateriaal vir die uitgawes van voor 1958 (d.i. dele A-Q) vernietig is en die feit dat die bestaande sitatebasis allerlei leemtes vertoon.

Tot en met 1976 het die sitatebasis bestaan uit elf verskillende versamelings wat as basis vir die elf verskene dele gedien het. Soos Van der Voort van der Kleij (1976:58-68) aantoon, vertoon die bestaande databasis geen uniforme, permanente versamelingsbeginsel nie: die versameling is willekeurig opgebou uit die beskikbare bronnemateriaal en daar bestaan allerlei leemtes ten opsigte van die verskillende sitate van die woordeskat van die gekose taalperiode (1500-1921).

Tans bestaan die sitatebasis uit 'n versameling alfabeties geordende verwerkte sitate van die dele wat reeds verskyn het en sitate vir die dele tot en met die letter Z wat nog nie verwerk is nie. Vanweë die hoë koste daaraan verbonde is daar besluit om die sitatebasis van die WNT voorlopig nie op rekenaar te plaas nie maar om dit wel in geskrewe vorm te behou en toeganklik te maak.

'n Studie is reeds uitgevoer na die moontlikheid om die tradisioneel gesette dele van die WNT (d.i. dele wat voor 1983 verskyn het) te digitaliseer en te rekenariseer. Vir hierdie doel is daar 'n ontleding gemaak van die leksikografiese struktuur van die inskrywings in deel A om te bepaal of

- die invoer van die dele van die WNT tegnies moontlik is;
- ongedateerde sitate gedeeltelik outomaties van hulle datering voorsien sien kan word;
- die inligtingsvelde van die leksikale inskrywings outomaties geselekteer en gekodeer kan word;
- die WNT as 'n gerekenariseerde inligtingsbron kan funksioneer.

Die aard en omvang van die soort leksikale inskrywing in die WNT word in Fig. 3.11 (p.63) geïllustreer; die leksikografiese struktuur van die WNT word in besonderhede bespreek in Van Sterkenburg (1984:66-74).

Die ondersoek na die rekenariseringsmoontlikhede van die WNT is deur 'n onafhanklike maatskappy uitgevoer en hulle aanbevelings sal binnekort aan die betrokke ministerie voorgelê word.

### 3.3.3.2 Die afdeling Thesaurus

Die hoofsaak van die afdeling Thesaurus is die uitbou en onderhoud van die gerekenariseerde woordargief. Die taaldatabank is van meet af beplan om 'n verteenwoordigende versameling Nederlandstalige tekstmateriaal te bevat uit die taalperiodes wat nie deur die WNT gedek word nie. Hierdie databasis moet dien vir die wetenskaplike bestudering van Nederlands en as leksikografiese basismateriaal vir die saamstel van 'n aantal sinchrone woordeboeke wat saam 'n historiese woordeboek van Nederlands moet vorm. Die leksikologiese taak van hierdie Afdeling is dus die versameling, verwerking, berging, bewerking en aanvulling van masjienleesbare Nederlandse tekste (vgl. Van Sterkenburg 1984:88). Die leksikografiese taak van die afdeling Thesaurus bestaan daarin dat dit hierdie teks- materiaal tot basiese leksikografiese materiaal moet verwerk. In die praktyk kom dit daarop neer dat die woordvorme van tekste gelys en voorsien word van inligting oor die woordsoort, datering, lokalisering, stylaanduiding en konteks daarvan.

Die verdeling van sitate na betekenis, die formuleer van definisies vir elke betekenisonderskeiding, die seleksie van sitate wat die formele,

semantiese en grammatiese geskiedenis van elke vorm en betekenis in die tyd, ruimte en genre die beste illustreer, bly nog steeds hoog gekwalifiseerde mensewerk (vgl. Van Sterkenburg 1984:89).

Die afdeling Thesaurus tree ook in raadgewende hoedanigheid op vir buite-instansies oor allerlei sake rakende die versameling en bewerking van tekste en die apparatuur en programmatuur wat vir hierdie doel benodig word. Die woordbank word bewerk en geberg met behulp van 'n aantal mikrorekenaars Tans bevat dit 60 miljoen woordvorme (waaronder 'n outomatiese leksikon met 720,000 inskrywings) wat na 200 miljoen uitgebrei sal word sodra hoofraamfasiliteite beskikbaar word. Die uitbouing van die woordbank het tot op hede geskied op grond van 'n aantal projekgerigte onderneminge, waarvan die belangrikste hier onder bespreek word.

(a) TACON (Taalkundige Analyse en Codering van het Oudere Nederlands)

Die doel van hierdie projek is die aanlê van 'n taaldatabank van Nederlands vir die periode 850-1350. Vir hierdie doel is die versamelde tekste van dr M Gysseling, die Corpus van Middelnederlandse Ambtelijke Bescheiden (of: die Corpus Gysseling), en 'n aantal tekste uit die periode 1301-1315 met die rekenaar verwerk tot 'n databasis van ongeveer een en 'n half miljoen woordvorme. Die Corpus Gysseling is met 'n rekenaargestuurde fotografiese setsisteem gepubliseer. Vir die saamstel van die woorddatabank is alfabetiese, frekwensie- en retrograde-indekse van al die woordvorme gegenereer en 'n versameling kontekskaarte opgebou waarin 'n hoeveelheid natuurlike teks rondom elke woordvorm gegee word. Die databasis sal uitgebrei word met die verwerking en publikasie van die reeks Middelnederlandse Literêre Handskrifte en 'n aantal glosse.

Die versamelde materiaal in die woorddatabank moet as basis dien vir die volgende publikasies:

- (i) 'n Vroegmiddelnederlandse Woordeboek
- (ii) 'n Atlas van Vroegmiddelnederlandse Taalvariante
- (iii) 'n Etimologiese Woordeboek van Oud- en Middelnederlands
- (iv) Studies oor die Middelnederlandse Morfologie.

(b) ALEX (Automatisch Lexicon)

Werksaamhede aan ALEX val uiteen in twee afdelings:

- (i) die moderne taalbank;
- (ii) deskriptiewe projekte op grond daarvan, met name 'n Morfologiese Handboek. (Vgl. INL 1985:39 vir besonderhede oor die Morfologiese Handboek.)

Die taaldatabank van Moderne Nederlands is in 1982 aangelê en het in 1986 reeds 60 miljoen woordvorme bevat. Die versameling en bewerking van die tekste hiervoor word hier onder beskryf na aanleiding van die uiteensetting in Portier en Pijnenburg (1984).

(a) Materiaalseleksie

Die tekstmateriaal vir die taaldatabank word verwerf

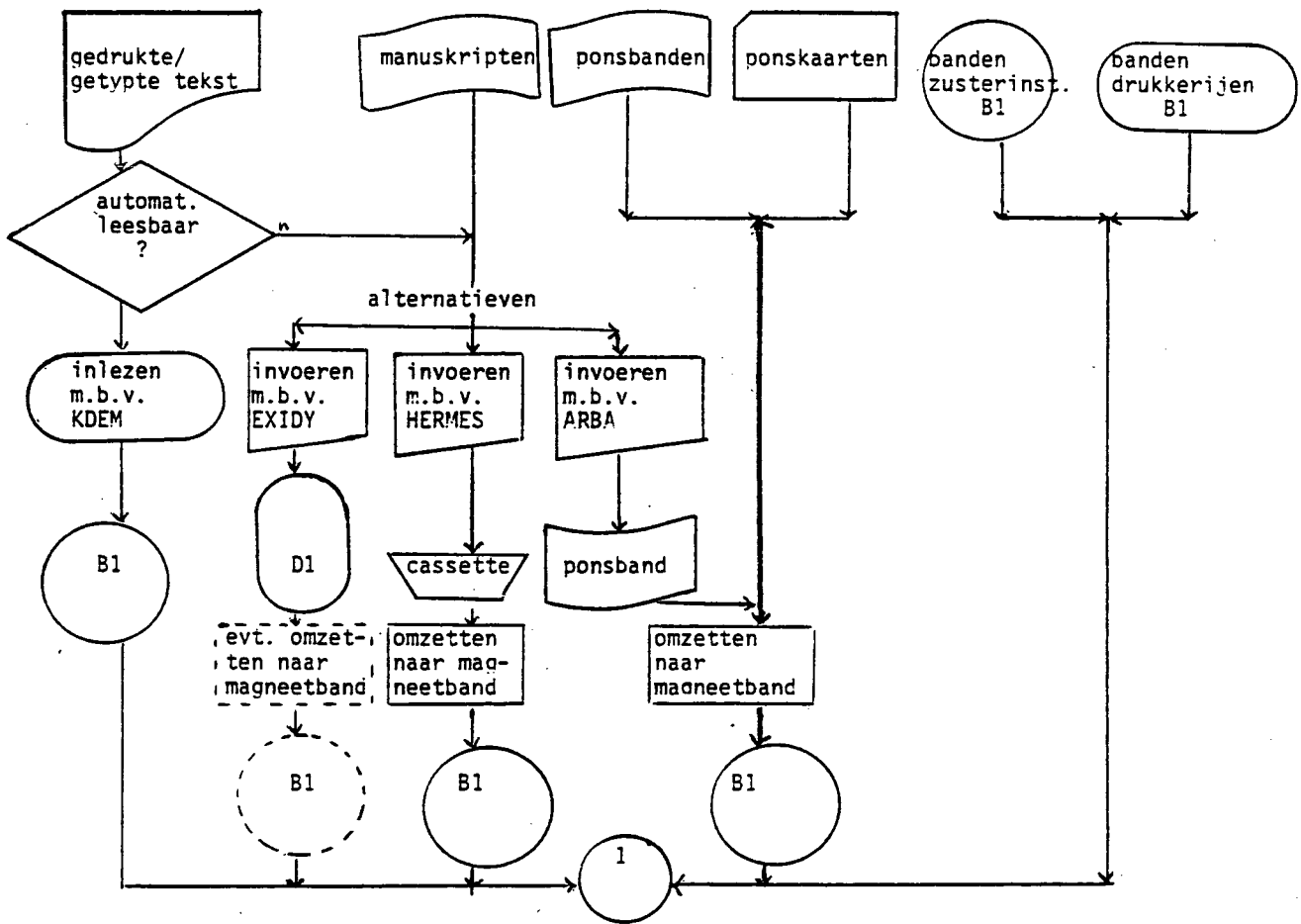
- (i) deur die uitwisseling van masjienleesbare materiaal met ander institute;
- (ii) deur die opname van materiaal wat deur die grafiese industrie beskikbaar gestel word;
- (iii) deur tradisioneel gesette materiaal met behulp van 'n optiese leser (KDEM) in te lees.

In die geval van materiaal wat reeds in masjienleesbare vorm is, vind die seleksie van die materiaal vir opname in die woordbank eers plaas op grond van die gelewerde produkte. Ten opsigte van die materiaal wat die Afdeling self inlees, word daar gepoog om materiaal te versamel uit alle deelgebiede van die maatskappy waarin kommunikasie met behulp van die gedrukte media plaasvind. Vir die keuse van hierdie materiaal word daar gesteun op die SISO-rubrisering (wat weer steun op die UDK-stelsel, wat deur die meeste biblioteke gebruik word) sodat ten minste een verteenwoordigende werk per vakgebied vir opname in die databasis gekies kan word.

Tot op hede is daar nog hoofsaaklik net gedrukte boeke met behulp van die KDEM ingelees (d.w.s. voorbeelde van die geskrewe taal). Alle strata van die woordeskat van Nederlands is nog nie gedek nie vanweë die seleksiekriteria wat die Afdeling handhaaf, of vanweë die feit dat daar nog nie oor presiese riglyne vir die opname van tekste vir besondere onderdele van die woordeskat (bv. die gesproke taal, streeksvariante en reklametaal) besluit is nie. Enkele van die probleme in hierdie verband is die volgende: 'n groot deel van die gepubliseerde Nederlandse romans is vertalings (terwyl voorkeur gegee word aan oorspronklike Nederlandse werke); beskikbare skoolhandboeke bevat baie herhalings van dieselfde inhoud; masjienleesbare koerantkopie bevat baie illustrasies, reklametekste, afkortings van woordvorme in die advertensies, druk- en spelfoute, tabelle met syfers; daarbenewens steun hierdie instansies op foutiewe programmatuur vir die afbreek van woorde.

Figuur 3.19 toon die wyse aan waarop data in verskillende media en op verskillende datadraers in die databasis opgeneem word.

Fig. 3.19: Opname van data in die taaldatabank



Vergelyk Van Sterkenburg (1985 : 39).

## (b) Materiaalbewerking

Voordat 'n teks ingelees word, moet daar vooraf bepaal word watter probleme die teks vir die outomatiese bewerking daarvan kan lewer. Daar moet bepaal word hoe die volgende hanteer moet word: bronverwysings, spesiale lettertekens, onversoenbaarhede tussen die afbreek van woorde in die teks en die wyse waarop die woorde deur die programme van die databasis afgebreek word, hoe variante en voetnotas in die teks verwerk moet word, hoe leestekens en eiename gekodeer moet word, en verder moet daar bepaal word of die ortografie in standaardvorm is of nie. Daarbenewens moet die optiese leser voor die inlees van die teks ook eers vir die lees van die besondere teks geoefen word. Gegee dat bogenoemde probleme vooraf uitgeskakel kan word en dat die nodige voorbereiding van die teks uitgevoer is, kan die KDEM ('n omnifont optiese leser) ongeveer 40,000 tekens (in verskillende lettertypes) per uur inlees. Die ingelese materiaal word op 'n beeldskerm geprojekteer sodat dit moontlik is om alle metatekstuele inligting wat nie opgeneem moet word nie (bv. opskrifte) uit te wis en om die bladsynommers van die oorspronklike teks in te voeg vir die latere verwerking van die bibliografiese besonderhede van sitate. Op die skerm word daar ook 'n lys geprojekteer van alle gevalle waarvan die interpretasie van lettertekens moontlik betwyfel kan word. Na 'n teks ingelees is, word dit via 'n 'RS23'-verbinding na 'n mikrorekenaar gestuur waar korreksies aan die ingelese teks (bv. foutiewe interpretasies van lettertekens of letterkombinasies en die verkeerde los- en vasskryf van woorde) met behulp van gewone woordverwerkingsprogramme aangebring word.

Van elke ingelese teks word daar 'n voorlopige woordvormlys ('typelist') gemaak wat met 'n werkleksikon vergelyk word. Woorde wat nie in die werkleksikon voorkom nie, word gelys en gekontroleer voordat hulle aan die werkleksikon toegevoeg word. Foutiewe vorme (wat nie in die werkleksikon voorkom nie) word op die ingelese teks gekontroleer en gekorrigeer. Hierna word daar 'n frekwensielys van die woordvorme van 'n teks gemaak en dié word saam met die frekwensielyste van die ander tekste gebruik om 'n kumulatiewe frekwensielys vir die hele taaldatabank saam te stel. Hierdie lys lewer inligting oor die woordvoorraad en frekwensieverdeling van die taaldatabank.

Naas die frekwensie van die woordvorme word daar vir elke teks ook lysie van die volgende besonderhede gegenereer: die aantal karakters, gemiddelde woordlengte (tans 6.6 karakters), die aantal woordvorme, en die gemiddelde aantal woordvorme per woordtipe (of lekseem). Hierdie gegewens word gebruik om te bepaal hoeveel lekseme aanwesig sal wees in 'n boek met 'n bepaalde omvang. Tans is dit so dat die toevoeging van nuwe tekste meestal daartoe lei dat die sitate vir elke woordvorm uitgebrei word maar nie dat die aantal nuwe woorde uitgebrei word nie. 'n Vergelyking tussen die woordvorme van 28 tekste het byvoorbeeld getoon dat die oorvleueling tussen tekste ten opsigte van hulle woordvorme so hoog soos 62,40 % is (vgl. Portier en Pijnenburg 1984:352).

### (c) Berging en herwinning

Om die gebruiksmoontlikhede van die data so plooibaar as moontlik te hou, word die ingesele materiaal so ver as moontlik intakt gehou en alle nuwe toegevoegde inligting in aparte bergingseenhede opgeneem. Laasgenoemde bring mee dat nuwe inligting aan die bank toegevoeg kan word sonder om die struktuur van die inhoud van die databank telkens te verander en dat die hoeveelheid ongevraagde inligting wat met elke navraag saamgegee word, tot die minimum beperk kan word. Die verbinding tussen aparte bergingseenhede (en dus ook tussen 'n bergingseenheid en die taaldatabank as 'n geheel) word gelê met behulp van 'n aantal wysers (d.i. verwysings na fisieke adresse op 'n skyf) wat aandui by watter element 'n bepaalde gegewe behoort. Relasies word byvoorbeeld gelê tussen 'n woordvorm en die teks waarin dit voorkom en tussen tekste en die SISI-rubrisering. Hierdeur word dit moontlik om te bepaal of 'n spesifieke woordvorm in 'n teks voorkom en watter woordvorme uniek aan sekere tekstipes is. Hierdie bergingsmetode het die voordeel dat nuwe inligting (bv. uitgebreide linguistiese inligting) voortdurend aan die bank toegevoeg kan word en in die taaldatabank as 'n geheel geïntegreer kan word.

Naas die konkordansie wat die INL vir die Lexicon Latinitatis Neerlandicae Medii Aevi saamgestel en gepubliseer het (vgl. INL 1985:41), het die Instituut ook die basiese materiaal vir 'n spelkontroleprogram vir Nederlands aan IBM gelewer (vgl. INL 1985:40).

#### 3.3.4 Die Van Dale Projectontwikkeling (VDP)

Die VDP is in 1976 in die lewe geroep met die doel om 21 woordeboeke in drie reekse van 7 elk saam te stel. Die eerste reeks bestaan uit 7 hardebandwoordeboeke, naamlik 'n woordeboek van hedendaagse Nederlands, drie tweetalige woordeboeke met Nederlands as brontaal en Frans, Duits en Engels as teikental en drie tweetalige woordeboeke met Frans, Duits en Engels as die brontale onderskeidelik en Nederlands as teikental (voltooi in 1986). Die tweede reeks bestaan uit skooluitgawes van bogenoemde sewe woordeboeke en die derde reeks uit sagtebanduitgawes daarvan.

Die saamstel en uitgee van die eerste reeks is voorafgegaan deur die saamstel van 'n gerekenariseerde databasis van hedendaagse Nederlands en die leksikografiese bewerking van hierdie materiaal tot die reeks leksikale inskrywings vir die Van Dale Hedendaagse Nederlands. Die leksikale inskrywings van hierdie woordeboek het uiteindelik as basis gedien vir die mikrostruktuur van die tweetalige woordeboeke met Nederlands as teikental en as die beginpunt vir die makrostruktuur van die woordeboeke met Nederlands as brontaal. Die ontwerp van die woordeboeke is voorafgegaan deur 'n uitgebreide ondersoek na die gebruikersbehoefte van die teikenmark.

In die onderstaande word daar in meer besonderhede ingegaan op die samestelling van die gerekenariseerde databasis en op die leksikografiese struktuur van die leksikale inskrywings in die verskillende



woordeboeke. Vergelyk in hierdie verband ook die bespreking in Van Sterkenburg, Martin en Al (1982).

### 3.3.4.1 Groot woordenboek hedendaags Nederlands (VDHN)

#### (a) Materiaalseleksie

Vir die saamstel van die VDHN is 'n gerekenariseerde korpus hedendaagse Nederlands (vanaf 1950) aangelê wat soos volg saamgestel is:

- (i) 'n redusering van die leksikale inskrywings van die 'Dikke van Dale' (wat op die WNT steun, 230,000 trefwoorde bevat en die woordeskat sedert 1830 dek);
- (ii) 'n mini-korpus van alledaagse leesstof;
- (iii) opgetekende neologismes en betekenisveranderinge van bestaande woorde wat deur 'n groep buite-lesers verskaf is;
- (iv) 'n korpus sinonieme wat uit die definisies van die 'Dikke van Dale' gegenereer is.

Vir die bewerking van die 'Dikke van Dale' is die hele woordeboek (tiende druk) op rekenaar geplaas. Die hoofbewerking het bestaan uit die redusering van die 230,000 trefwoorde tot ongeveer tagtig na negentig duisend toe. Die seleksie moes plaasvind vanweë die beoogde omvang van die woordeboek en vanweë die buite-leksikografiese oorweging van die hoë koste verbode aan die produseer van 'n omvattende woordeboek. Drie Nederlandssprekende leksikograwe het onafhanklik van mekaar aan die trefwoorde en hulle betekenisonderskeidings 'n gradering tussen 1 en 5 toegeken op grond van hulle sinchrone aard om so te bepaal watter lemmas in 'n woordeboek van hedendaagse Nederlands opgeneem sou kon word. Trefwoorde met die etiket 'gepas' is met behulp van die rekenaar geskei van die lemmas en betekenisonderskeidings wat as onvanpas of verouderd bestempel is (bv. eiename, terme oor seilbootvaart, streeksvariante, terminologie uit die koloniale verlede en deursigtige samestellings en afleidings). Laasgenoemde groep is nie in die nuwe woordeboek opgeneem nie.

Vanweë die gebrek aan 'n betroubare frekwensietelling van 'n korpus gesproke en geskrewe Nederlands, is die frekwensie van elke trefwoord en die betekenisonderskeidings daarvan op grond van die valensie daarvan en die ervaring van die leksikograwe (subjektiewe valensie) bepaal.

Bogenoemde geëkstraheerde korpus is aangevul met 'n korpus tekste van alledaagse leesstof. Die breë spektrum taalgebruik wat hierdeur gedek is, blyk uit die volgende lysie tekste wat in die databasis opgeneem is. (Vgl. Van Sterkenburg, Martin en Al 1982:224): Wat iedereen over drugs zou moeten weten, Drukwerk helpt de wereld draaien, De Stadskrant, De geheelonhouder, Muziek deze Maand, de Microgids ('n weeklikse

televisieprogramtydskrif), Fundamenteel beleggen, Uitleg ('n tweeweklikse tydskrif van die Nederlandse Departement van Opvoeding), Auto-Kampioen ('n maandblad vir motoriste), en Mensen van Nu ('n populêre maandblad).

Die tekste wat deur die buite-lesers gemonitor is vir nuutskeppinge en die betekenisverandering van bestaande woorde kom onder andere uit die volgende gebiede: die ruimtevaart, die elektronika, die feminisme, die ekonomie, mediese sorg, omgewingstudies, die opvoeding en die gebied van werkersunies.

(c) Invoer- en bergingsformaat

'n Beskrywingsraamwerk is ontwikkel om al die leksikale inskrywings van die VDHN op 'n eenvormige manier te hanteer sodat die onderdele sistematies toeganklik sou wees vir die rekenaar vir die genereer van afgeleide produkte (soos die sagteband- en skooluitgawes). 'n Leksikale inskrywing in die VDHN het die volgende struktuur:

- (i) tegniese inligting:
  - woord
  - vormvariante
  - verwysings en uitspraak
  - inligting oor frekwensie
  - grammatiese inligting
  - etikette wat geld vir alle betekenis
  - 'n verwysing na 'n register spreekwoorde;
- (ii) betekenisbeskrywing:
  - alle moontlike gebruikstoepassinge
  - sinonieme;
- (iii) voorbeelde van die grammaties-sintaktiese kombinasie-moontlikhede van die trefwoord in die verskillende betekenisonderskeidings wat dit vertoon.

Vergelyk in hierdie verband Figure 3.20 en 3.21.

Die finale gedrukte kopie van die Van Dale Hedendaags Nederlands bevat 95,000 leksikale inskrywings bestaande uit twaalf en 'n half miljoen tekens op 1,600 gedrukte bladsye.

Fig. 3.20: Voorbeeldinschrijving uit Van Dale Hedendaagse Nederlands

<u>Markt</u>	<f>	<14; -en>	<→ spr. 153, 376 >
<u>0.1</u>	openbare verkoop en koop op een daarvoor bestemde plaats in de open lucht, waar goederen aanwezig zijn		
<u>0.2</u>	plaats waar markt wordt gehouden		
<u>0.3</u>	de uitwisseling van allerlei producten die ergens plaats heeft		
<u>0.4</u>	prijs waartegen iets ergens wordt verhandeld		
<u>0.5</u>	< hand. > geneigdheid om te kopen; intensiteit van de handel		
◆			
<u>2.2</u>	de Grote -		
<u>2.3</u>	de Europese -		
<u>2.5</u>	de - was stil		
<u>3.4</u>	de - bederven; een stijgende -, een dalende -		
<u>6.1</u>	de - op het Waterlooplein; naar de - gaan; op de - kopen, ter - gaan bij iemand a) bij hem kopen b) bij hem voorlichting, hulp enz. zoeken; in de - zijn voor, gezadigd zijn voor; aan de - komen, aanbieden; van alle -en thuis zijn, overal raad op weten		
<u>6.4</u>	de - van eieren; goed in de - liggen a) gevraagd zijn b) <fig.> een goede naam hebben;		
<u>9.1</u>	bij het scheiden van de -, leert men de kooplui kennen, iemand's ware karakter blijkt pas als het erop aankomt		

(14=nw, manlik en vroulik).

Vergelyk van Sterkenburg, Martin en Al (1982 : 225-226).

Fig. 3.21: Voorbeeldkolom uit Van Dale Hedendaags Nederlands

- Info** (de - (v.)) (inf.) 0.1 (verkorting van) informatie.
- Infobox** (de -; -eh) (com.) 0.1 soort (openbare) telefooncel met beeldscherm en toetsenbord waarmee informatie (b.v. toeristische informatie en hotelreserveringen) uit een (viditel)computer kan worden gevraagd en zichtbaar gemaakt.
- In Forma** (Lat.) 0.1 in (goede) vorm.
- Info'raast** (de - (m.); -maten) 0.1 informatie verwerkende automaat  
⇒ *computer*.
- Informal'ering** (-isering) (de - (v.); -en) 0.1 ontvormelijking.
- Informal'teit** (de - (v.); -en) 0.1 het niet in overeenstemming zijn met de voorgeschreven vorm ⇒ *onvormelijkheid* 0.2 afwijking van de voorgeschreven vorm 0.3 (euf.) onregelmatige handeling van een ambenaar, ambtsvergriep.
- Info'raast** (de - (m.); -en) 0.1 iem. die inlichtingen verstrekt ⇒ *zegsman*.
- Informa'teur** (de - (m.); -s) 0.1 beambte belast met het inwinnen van inlichtingen op sociaal gebied 0.2 politicus die onderzoekt of een kabinetsformateur met een bepaalde opdracht kans van slagen zou hebben.
- Info'rmatica** (de -) 0.1 leer van het verzamelen en de verwerking van informatie van gegevens met behulp van computers ⇒ *informatie-mathematica*.
- Info'rmaticus** (de - (m.); -tici) 0.1 deskundige op het gebied van de informatica.
- Info'rmatie** (f) (de - (v.); in bet. 2 -s) 0.1 alles wat als bericht, als overdracht van kennis of gegeven een persoon of een informatieverwerkend toestel bereikt 0.2 inlichting ⇒ *mededeling, bericht* 0.3 het vragen, zich verschaffen van kennis of inzicht ♦ 2.1 (biol.) *geneesheer*; een overstelpende hoeveelheid -; *nuttige, waardevolle* - 2.2 *nadere* - (s) is (zijn) te verkrijgen bij; om *nadere* - *verzoeken* 3.1 - *doorgeven* 3.2 - *s inwinnen*; - *verschaffen* 6.1 *een schat aan* - 6.3 *te uwer* -; *ter* -.
- Info'rmatiebalie** (de - (v.); -s) 0.1 balie, desk, toonbank waar inlichtingen verstrekt worden.
- Info'rmatiebank** (de - (v.)) (comp.) 0.1 gestructureerde hoeveelheid informatie op schijf of magneetband ⇒ *data-base*.
- Info'rmatiedrager** (de - (m.); -s) 0.1 middel tot het overdragen van informatie, met name aan een toestel ♦ 0.1 *ponskaarten als* -.
- Info'rmatief** (f) (bn.) 0.1 tot in- of voorlichting dienend 0.2 informatie, vooral veel informatie bevattend ♦ 1.1 *een - gesprek* 1.2 *een - artikel*.
- Info'rmatie-industrie** (de - (v.)) 0.1 industrie die alle activiteiten omvat op het terrein van de informatieverwerking.
- Info'rmatiemarkt** (de -; -en) 0.1 verzameling marktkramen waar b.v. bedrijven of instellingen informatie verschaffen door middel van folders, films, mondelinge uitleg e.d..
- Info'rmatiemechanica** (de - (v.)) 0.1 informatica.
- Info'rmatiestroom** (de - (m.); -stromen) 0.1 gestadige toevloed van informatie.
- Info'rmatiesysteem** (het -; -systemen) (comp.) 0.1 geheel van mensen, apparatuur en procedures binnen een organisatie dat gegevens verzamelt, verwerkt, opslaat, terugwekt en kan oproepen ten dienste van gebruikers.
- Info'rmatietechnologie** (de - (v.)) 0.1 kennis omtrent methoden, technieken en technische hulpmiddelen voor het opslaan en automatisch verwerken van informatie.
- Info'rmatietheorie** (-theorie) (de - (v.)) 0.1 theorie betreffende de in de dingen zelf aanwezige informatie, voor zover van betekenis voor communicatiesystemen, deel van de cybernetica.
- Info'rmatietijdperk** (het -) 0.1 periode in de geschiedenis van de mensheid waarin de samenleving sterk beheerst wordt door de automatische verwerking en overdracht van informatie met behulp van micro-elektronica.
- Info'rmatietoon** (de - (m.)) (com.) 0.1 afwisselend hoge en lage toon die men hoort bij het kiezen van een nog niet verstrekt nummer.
- Info'rmatieverwerking** (de - (v.)) (tech.) 0.1 de verwerking van gegevens met behulp van computers om hiermee nieuwe informatie te verkrijgen.
- Info'rmatie'ring** (-ering) (de - (v.); g.m.v.) 0.1 toenemend gebruik van automatische informatieverwerking ♦ 1.1 *de - van de samenleving*.
- Info'rmatrice** (de - (v.); -s) 0.1 vrouw die (beroepshalve) inlichtingen verstrekt.
- Info'rmeel** (bn., bw.) 0.1 niet in overeenstemming met de voorgeschreven vorm ⇒ *onvormelijk, onofficieel* 0.2 voorlopig, vrijblijvend ♦ 1.1 *een informele ontvangst* 1.2 *informele besprekingen*.
- Info'rmeer** (f) (informeerde, h. geïnformeerd)
- I (onov. ww.) 0.1 (iemand) inlichtingen inwinnen ⇒ *navragen, vragen* ♦ 5.1 *ik heb ernaar geïnformeerd* 5.1 *informeer eens waar je wezen moet*;
- II (ov. ww.) 0.1 inlichtingen verschaffen, kennis geven van ⇒ *inlichten* ♦ 4.1 *zij heeft zich daarover terdege geïnformeerd*.

### 3.3.4.2 Die tweetalige woordeboeke

Die databasis van die Van Dale Hedendaags Nederlands het gedien

- (i) as die gemeenskaplike inligtingsbron vir die mikrostruktuur van die tweetalige woordeboeke met Nederlands as teikentaal;
- (ii) as die gemeenskaplike beginpunt vir die afgrensing van die makrostruktuur van die tweetalige woordeboeke met Nederlands as brontaal.

Vir die keuse van trefwoorde is daar gesteun op die frekwensieontledings van die Nederlandse taalvorme. Die hergebruik van die Nederlandse databasis hou dié voordeel in dat daar vir die tweetalige woordeboeke 'n woordeskat gekies is wat die Nederlandse wêreldbeeld weerspieël (soos dié leksikaal vergestalt word) en nie die van die ander teikentale nie.

Die dele van die reeks tweetalige woordeboek het 'n uniforme formaat, 'n uniforme stelsel tipografiese konvensies word gebruik en die leksikale inskrywings het 'n vaste leksikografiese struktuur. Elke uitgawe bevat die volgende:

- (i) 'n voorwoord met 'n gebruikersgids;
- (ii) 'n grammatika in die vorm van 'n alfabetiese leksikon wat 'n verduideliking gee van die taalkundeterminologie wat in die gebruikersgids en die leksikale inskrywings gebruik word.

Figuur 3.22 illustreer die leksikografiese struktuur van hierdie tweetalige woordeboeke.

Die strukturering en kodering van die inligting in die leksikale inskrywings bring mee dat allerlei byprodukte van die gemeenskaplike databasis gegenereer kan word. Die volgende word byvoorbeeld beplan:

- (i) 'n woordeboek met sinonieme en pseudo-sinonieme met die eentalige Nederlandse woordeboek as basis;
- (ii) 'n woordeboek met idiome in Nederlands en Engels;
- (iii) ander publikasies van spesifieke inligtingsvelde uit die leksikale inskrywings in die een- en tweetalige woordeboeke.

Fig. 3.22: Van Dale Tweetalige Woordeboeke

Frans-Nederlands

<u>poche</u>	[pɔʃ]	↔	↔
I	<1>	inf.	<u>0.1</u> <u>pocket</u> ;
II	<2>		
<u>0.1</u>	<u>zak</u>	<n kleding>	
<u>0.2</u>	<u>uitgezakke plooi</u>	⇒ bobbel;	
<u>0.3</u>	<u>zak</u>	<verpakking >;	
<u>0.4</u>	<u>vak</u>	<n tas, portefeuille >;	
<u>0.5</u>	<u>(vang)net</u>	⇒ sleepnet;	
<u>0.5</u>	<geol.>	<u>holte</u> ⇒ put, laag;	
<u>0.7</u>	<biol.>	<u>buidel</u> ⇒ krop;	
<u>0.8</u>	<fig.>	<u>haard</u> ⇒ punt;	
◆			
<u>1.1</u>	argent de --	<u>zakgeld</u> ;	
	lampe de --	<u>zaklantaarn</u> ;	
	livre de --	<u>pocket</u> ;	
	les mains dans les poches	<u>moeiteloos, zonder een</u> <u>poot uit te steken</u> ;	
	théâtre de --	<u>vestzaktheater</u> ;	
<u>1.6</u>	-- de gaz	<u>gasbel</u> ;	
	-- de pétrole	<u>olielaag</u> ;	
<u>1.8</u>	-- de résistance	<u>verzetshaard</u> ;	
<u>3.1</u>	acheter, vendre chat en --	<u>een kat in de zak kopen,</u> <u>verkopen</u> ;	
	<inf.>	n'avoir pas sa langue dans sa -- <u>niet</u> <u>op z'n mondje gevallen zijn</u> ;	
		n'avoir pas les yeux dans sa -- <u>z'n</u> <u>ogen niet in z'n zak hebben</u> ;	
		faire les poches à qn. <u>iebands zakken</u> <u>doorzoeken, leeghalen, rollen</u> ;	
		mettre qn. dans sa -- <u> iemand in z'n zak steken</u> ;	
<u>3.2</u>	elle avait des poches sous les yeux	<u>zij had wallen</u> <u>onder haar ogen</u> ;	
	ton pantalon fait des poches aux genoux	<u>er zitten</u> <u>knieën in je pantalon</u> ;	
<u>6.1</u>	<inf.>	l'affaire est <u>dans la -- de zaak is</u> <u>rond, bekeken</u> ;	
	payer de sa --	<u>uit eigen zak betalen</u> ;	
	<inf.>	en être de sa -- <u>er (geld) bij in-</u> <u>schieten</u> ;	
	il a 100 F en --	<u>hij heeft 100 F op zak</u> ;	
<u>8.1</u>	je le connais comme ma --	<u>ik ken hem door en door</u> .	

Vergelyk Van Sterkenburg, Martin en Al (1982 : 231-232).

### 3.4 LEKSIKONS IN (SEMI-)AUTOMATIESE TEKSONTLEDINGSTELSLS

In die beplanning van enige LDB moet daar ook voorsiening gemaak word vir die potensiële hergebruik van die een- en meertalige leksikons wat 'n kardinale onderdeel vorm van die semi-outomatiese en outomatiese stelsels vir teksontledingstake.

Verskillende teksontledingstelsels kan op grond van hulle ontledingsfunksies en -vermoëns onderskei word, byvoorbeeld stelsels vir

- (i) die morfologiese ontleding en lemmatisering van die ortografiese woorde van 'n teks;
- (ii) die sintaktiese woordsoortbenoeming van tekswoorde;
- (iii) die sintaktiese ontleding van teksstukke en -sinne;
- (iv) die funksioneel-semantiese interpretasie van tekseenhede;
- (v) die vertaal van 'n teks.

Sommige stelsels word ontwerp om net enkele van hierdie funksies uit te voer, ander die hele reeks, soos wat die uitvoer van sekere funksies direk afhanklik is van die resultate van 'n voorafgaande ontleding.

Die invoer van hierdie stelsels is masjienleesbare teksmateriaal. Die hoofaalkomponente van die stelsels is een of meer (sub)grammatikas en een of meer leksikons, waar laasgenoemde dikwels ingebed is in 'n omvattender kennisdatabasis.

Die grammatikas word hoofsaaklik saamgestel as versamelings konteks-vrye of konteks-sensitiewe herskryfreëls van die tipe

A ---- B (C)

wat die strukturele reëlmatighede van 'n besondere taal op elke vlak van grammatiese ontleding weergee (die vlak van die ortografie, fonetiek, fonologie, morfologie, sintaksis en die semantiek). Die stel herskryfreëls word met programme omgeskakel na 'n herkenningsstelsel wat die teks-eenhede op elke vlak kan identifiseer en benoem, en dit word genererend gebruik om die nodige struktuurontledings (bv. in die vorm van boomdiagramme) te lewer.

Die leksikons in hierdie stelsels word, soos in die geval van die LDB, gestruktureer as 'n versameling leksikale inskrywings met die nodige grammatiese inligting vir die uitvoer van besondere ontledingstake. Dit spreek vanself dat die omvang van die ontledingstake bepalend is vir die hoeveelheid grammatiese inligting wat in die leksikonstelsels opgeneem sal word. In die eenvoudigste gevalle kan so 'n leksikon net inligting oor die ortografiese variante van lekseme bevat (bv. vir die kontrole van die spelling in tekste of vir 'n lemmatiseringsprogram); in

omvattender stelsels word die leksikografiese databasis egter as 'n woordeboek-cum-grammatika opgestel waarin die hele spektrum grammatiese inligting opgeneem word wat met die lekseme van 'n taal/tale geassosieer word.

Die omvang en struktuur van die leksikons van hierdie stelsels word hieronder kortliks geïllustreer.

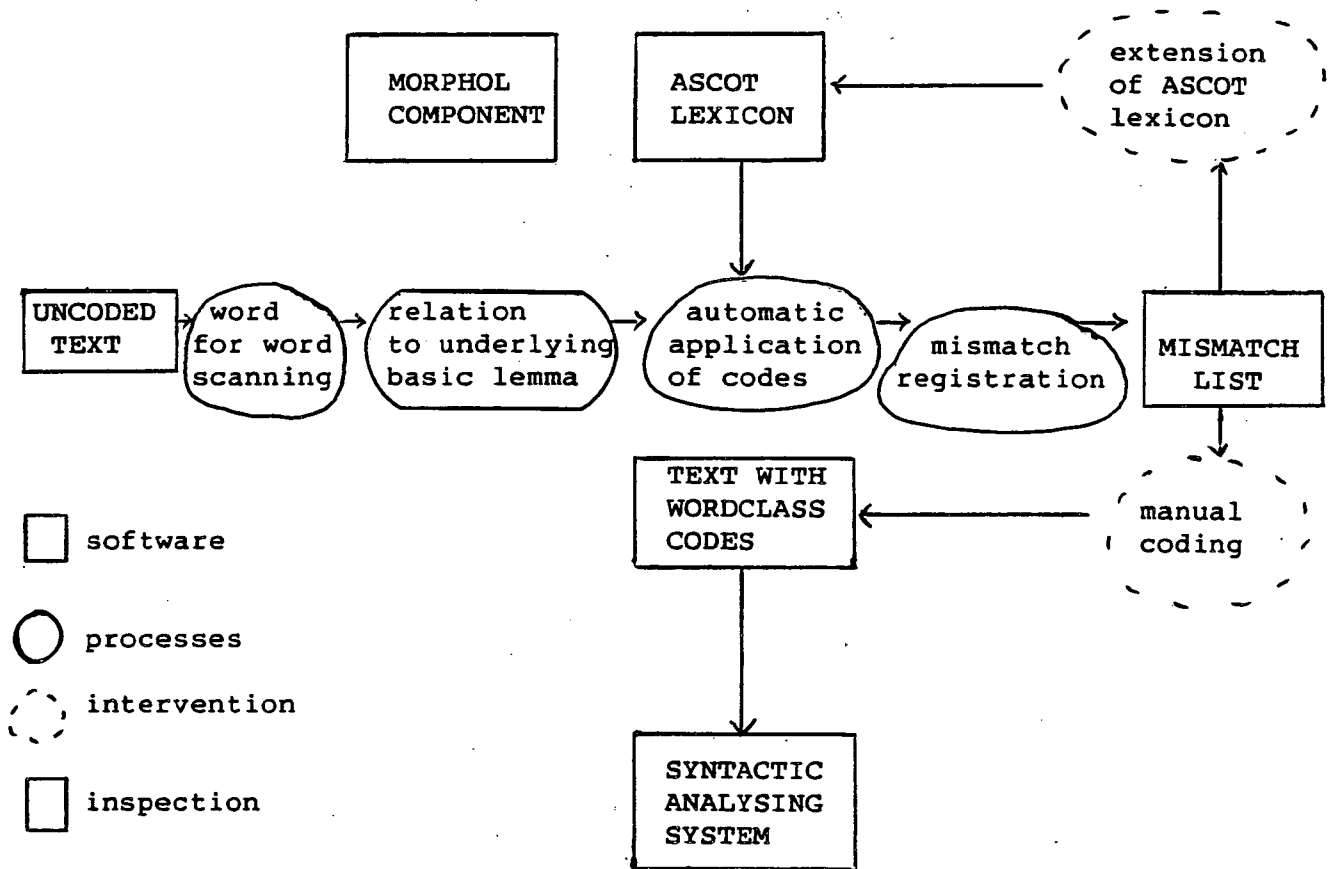
Die ASCOT-stelsel is 'n teksontledingstelsel vir die outomatiese sintaktiese (woordsoortlike) kodering van die ortografiese woordvorme van 'n teks. Die struktuur van die eentalige leksikon van hierdie stelsel word bespreek in 3.2.

Die invoer van die stelsel is 'n ongekodeerde teks waarvan die ontleding en kodering woord-vir-woord geskied. Elke woord word via die morfologiese komponent geanaliseer en na die lemmavorm daarvan gereduseer alvorens die woordsoort daarvan volgens die leksikale inskrywings in die ASCOT-leksikon nageslaan en toegeken word.

Die woordvorme wat nie in die leksikon gevind word nie, word as 'n aparte lys gegenereer en intellektueel gekodeer. Hierdie lys word dan, na die beskrywing van elke lemma weer met die ASCOT-leksikon geïntegreer. Vergelyk die onderstaande uiteensetting.



Fig. 3.23: Die ASCOT-stelsel.

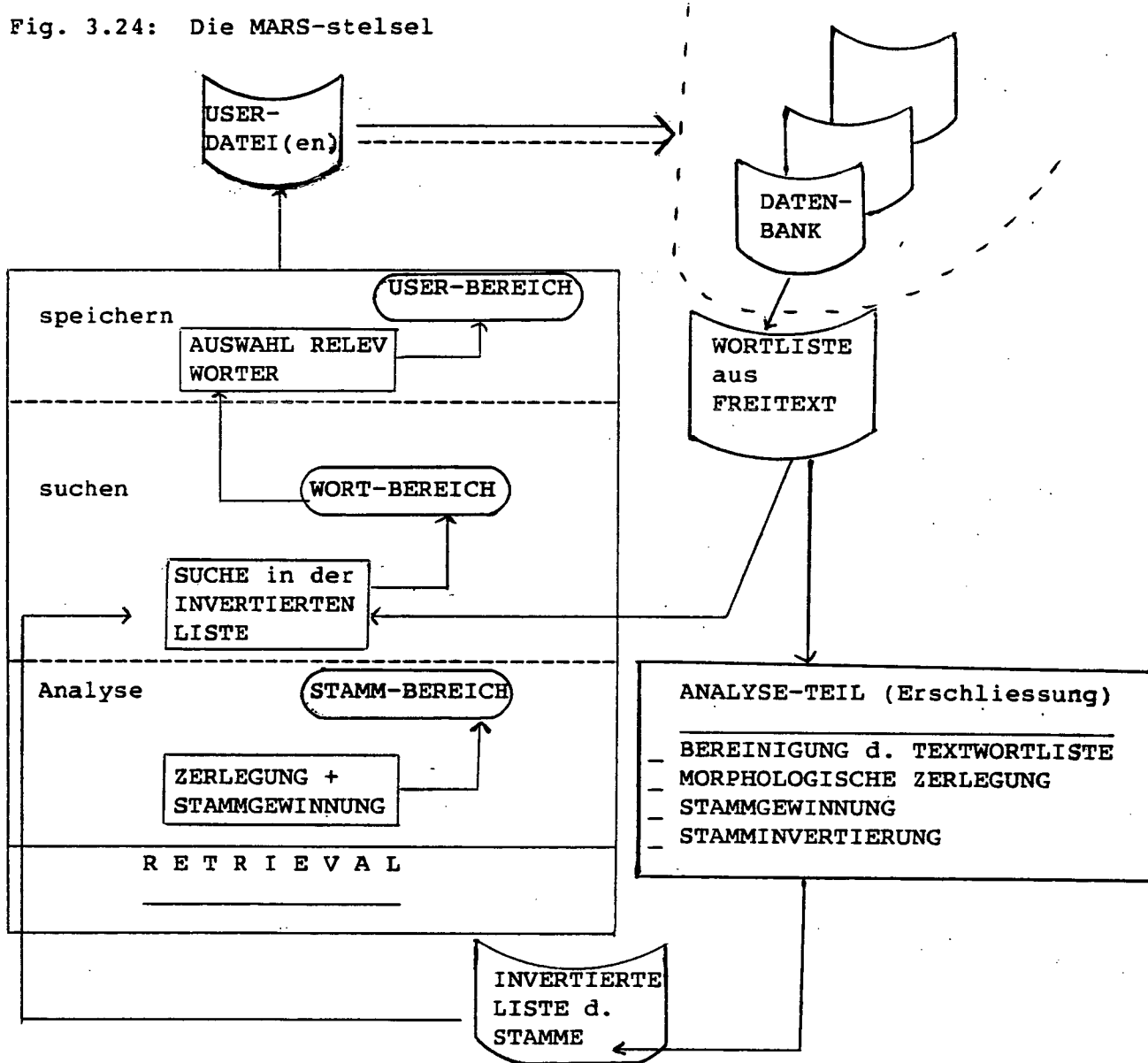


Die lemmatisering van ortografiese tekswoordvorme is een van die basiese funksies van die meeste teksontledingstelsels. Hierdie funksie onderlê byvoorbeeld ook die MARS-inligtingstelsel, wat ontwerp is om 'n gebruiker van 'n inligtingstelsel met die formulering van soekvrae en die herwinning van inligting op hierdie soeke te help.

Die basiese funksie van die stelsel is om die soekwoorde wat die gebruiker insleutel met die soekwoorde (deskriptore) in die databank te vergelyk en om die geassosieerde inligting op 'n beeldskerm aan die gebruiker te lewer. Vir die herwinning van die inligting moet die soekwoord(e) van die gebruiker eers gelemmatiseer word (vgl. die ANALYSE-TEIL van die stelsels hier onder) in ooreenstemming met die lemmavorme van die databank.

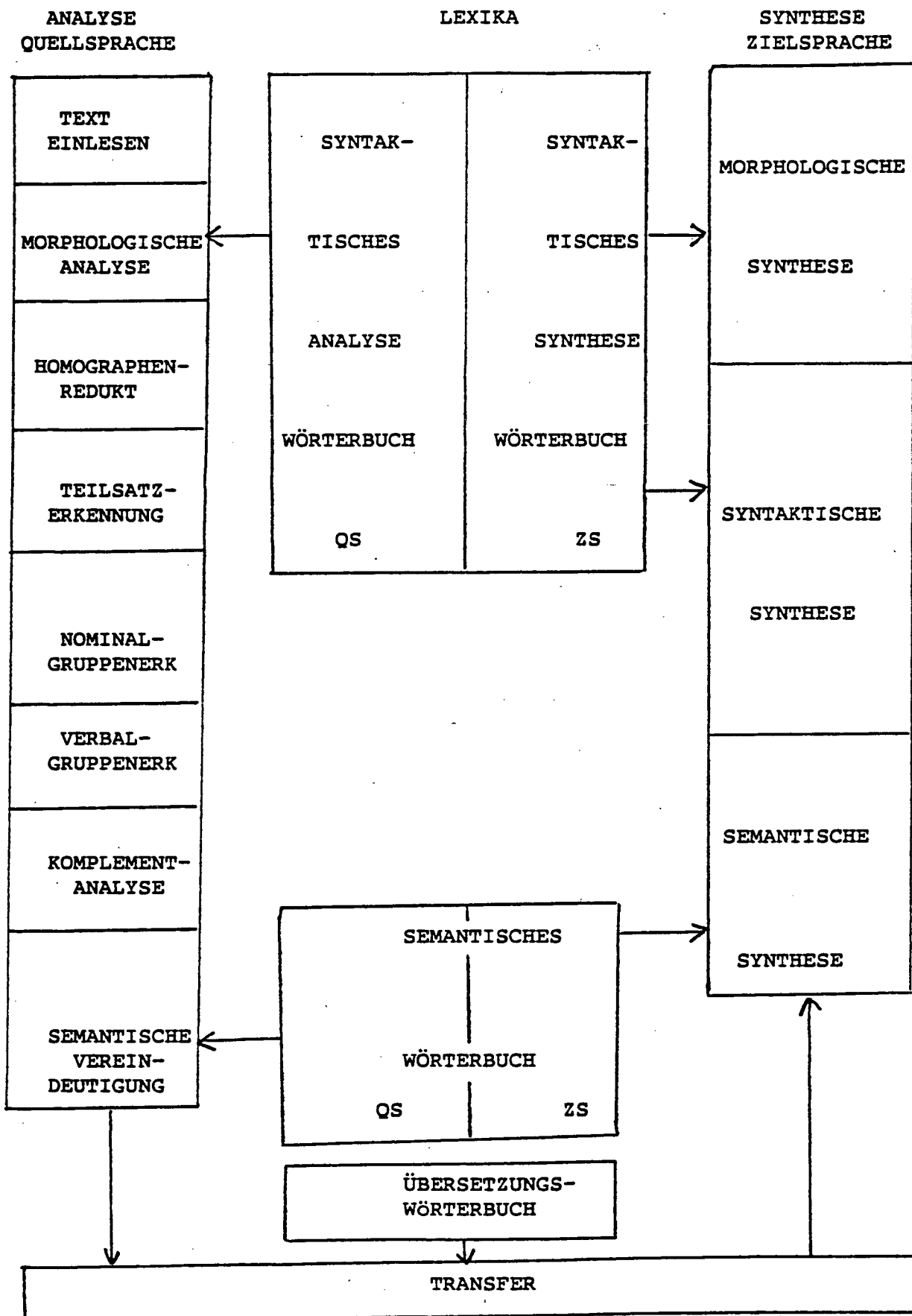
Vergelyk die onderstaande uiteensetting en die bespreking in Schwarz en Thurmair (1986:157-217).

Fig. 3.24: Die MARS-stelsel



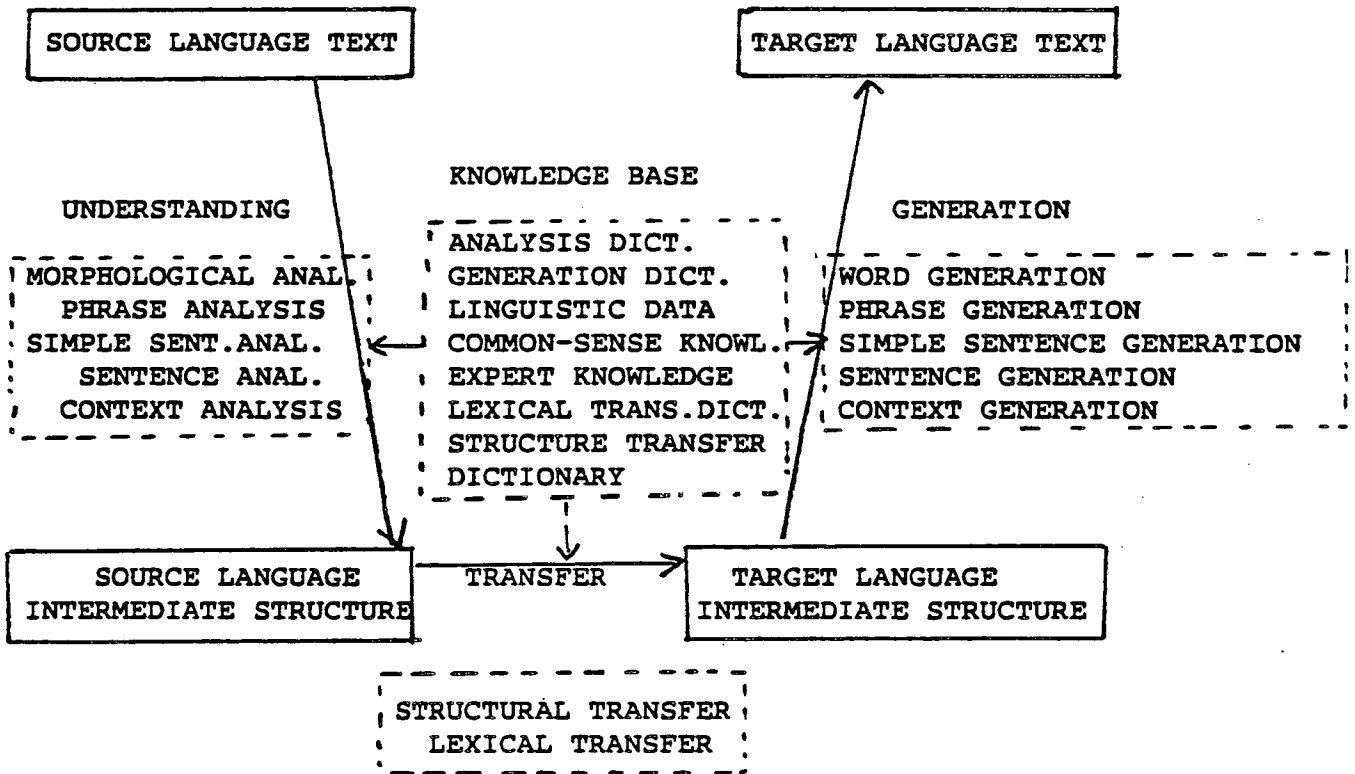
'n Verskeidenheid een- en meertalige leksikons word in outomatiese vertaalstelsels gebruik in die ontleding, vertaling en generering van die vertaalde teks. Vergelyk in hierdie verband byvoorbeeld die versameling een- en meertalige leksikons van die eksperimentele vertaalstelsel SUSY 1 wat aan die Universiteit van Saarbrücken ontwikkel is. Hierdie tweede-generasie-vertaalstelsel bevat naas 'n sintaktiese en 'n semantiese leksikon vir sowel die brontaal as die doeltaal, ook nog 'n vertalende leksikon wat die resultate van die ontleding van die brontaal na die doeltaal vertaal.

Fig. 3.25: Die vertaaltstelsel SUSY 1



In omvattender vertaalstelsels vorm die versameling leksikons 'n onderdeel van die kennisbasis van die stelsel. Naas die genoemde een- en meertalige algemene leksikons, word die benodigde vakkennis van die kennisbasis gestalte gegee in die vorm van 'n versameling meertalige vakleksikons. Vergelyk in hierdie verband die skematiese uiteensetting van die LUTE-vertaalstelsel en die bespreking in Bátori en Weber (1986).

Fig. 3.26: Die LUTE-vertaalstelsel



Wat die saamstel van die leksikons vir hierdie stelsels betref, is dit nodig om daarop te wys dat uiteenlopende oorwegings van krag kan wees in die saamstel en enkodering van die algemene gerekenariseerde leksikografiese databasis wat vir die gewone gebruiker bedoel is en die wat vir gebruik in gerekenariseerde teks-ontledingsisteme ontwerp word. In die algemene leksikografiese databasis word daar byvoorbeeld ook data-elemente betrek wat bedoel is as hulpmiddel vir die gebruiker om sy besondere taalprobleme te help oplos; naas inligting oor die grammatiese eienskappe van 'n leksikale item, kan daar byvoorbeeld ook allerlei gebruiksaantekeninge (pragmatiese inligting) opgeneem word en in 'n vorm wat vir die gewone gebruiker die toeganklikste is, naamlik meestal in die vorm van ongeformaliseerde beskrywings in natuurlike taal.

Vir die outomatiese ontleding van tekste is dit egter ook noodsaaklik dat elke ortografiese woord ontleed moet kan word. Dit bring mee dat alle relevante grammatiese inligting vir hierdie doel in die leksikons verstrekk en in 'n vorm geënkodeer moet word waarin dit deur die rekenaarstelsel geïdentifiseer, geïnterpreteer en vir die ontledingstake gebruik kan word.

Naas die herstrukturering van die inligting in bestaande gerekenariseerde leksikons, moet die inligting daarin dan ook dikwels aangevul en geformaliseer word. Wat eersgenoemde betref, is dit veral die morfologiese en sintaktiese datakategorieë wat in meer besonderhede beskryf en veel omvattender gesubkategoriseer moet word as wat normaalweg die geval is in die saamstel van 'n LDB vir die genereer van leksikons vir die menslike gebruiker.

Vir die ontwikkeling van die vertaalsistiem SYSTRAN word daar in Deville, Gebruurs en Jaspaert (1985) en Jaspaert (1984) dus voorgestel dat alle affikse ook as trefwoorde in die leksikon opgeneem moet word en dat 'n leksikale inskrywing met 'n affiks as 'n trefwoord die volgende data-elemente moet bevat:

- (i) kategoriernaam en -soort;
- (ii) kategoriale eienskappe;
- (iii) morfo-sintaktiese inligting;
- (iv) diakritiese eienskappe;
- (v) subkategorisasiematrjks;
- (vi) morfo-semantiese eienskappe;
- (vii) morfonologiese weergawe.

Die inligting in (i)-(v) bepaal die spreiding van 'n affiks in morfologiese strukture en die sintaktiese en semantiese eienskappe van woorde wat deur afleiding met die affiks gevorm is; die datakategorie (vi) gee 'n beskrywing van die wyse waarop die eienskappe van 'n stam kan verander as gevolg van die toevoeging van die besondere affiks, soos byvoorbeeld die verandering van die kasusmatrjks van 'n stam; (vii) bevat alle inligting oor die morfonologiese variasie van 'n affiks.

Wat die uitbreiding van die sintaktiese beskrywing van lemmas betref, word sintaktiese inligting soos in eentalige pedagogiese grammatika vereis. Vergelyk in hierdie verband die bespreking in 3.2 en Lemmens en Wekker (1986).

Die formalisering van die semanties-funksionele kenmerke van leksikale items is een van die grootste take in die saamstel van 'n leksikon vir 'n stelsel wat tekste outomaties moet interpreteer. Die ongeformaliseerde definisies in die gewone leksikons kan nie deur die rekenaar geïnterpreteer of direk vir interpretasiefunksies gebruik word nie. Gevolglik word uitgebreide stelsels semanties-funksionele rolle (kasusrolle, byvoorbeeld agens, onderganer, instrument, ens.) gedefinieer en komplekse versamelings semantiese primitiewe opgestel om die betekenis van leksikale items te definieer. Vergelyk in hierdie verband die onderstaande kasusrolle en versameling semantiese primitiewe

wat in Nagao, Tsujii en Nakamura (1986) vir 'n Japannees-Engels vakwoordeboek vir 'n vertaalstelsel voorgestel word.

Fig. 3.27: Kasusrolle vir Engels

- |                   |                   |
|-------------------|-------------------|
| 1. AGent          | 17. RANge         |
| 2. Causal-POtency | 18. COMpaRison    |
| 3. EXPeriencer    | 19. TOOl          |
| 4. OBJect         | 20. PURpose       |
| 5. RECipient      | 21. Space-FROm    |
| 6. ORigin         | 22. Space-AT      |
| 7. SOURce         | 23. Space-TO      |
| 8. GOAL           | 24. Space-THrough |
| 9. CONtent        | 25. Time-FROm     |
| 10. PARTner       | 26. Time-AT       |
| 11. OPPonent      | 27. Time-TO       |
| 12. BENeficiary   | 28. DURation      |
| 13. ACCompaniment | 29. CAUSE         |
| 14. ROLe          | 30. CONdition     |
| 15. DEGree        | 31. RESult        |
| 16. MANner        | 32. CONcessive    |

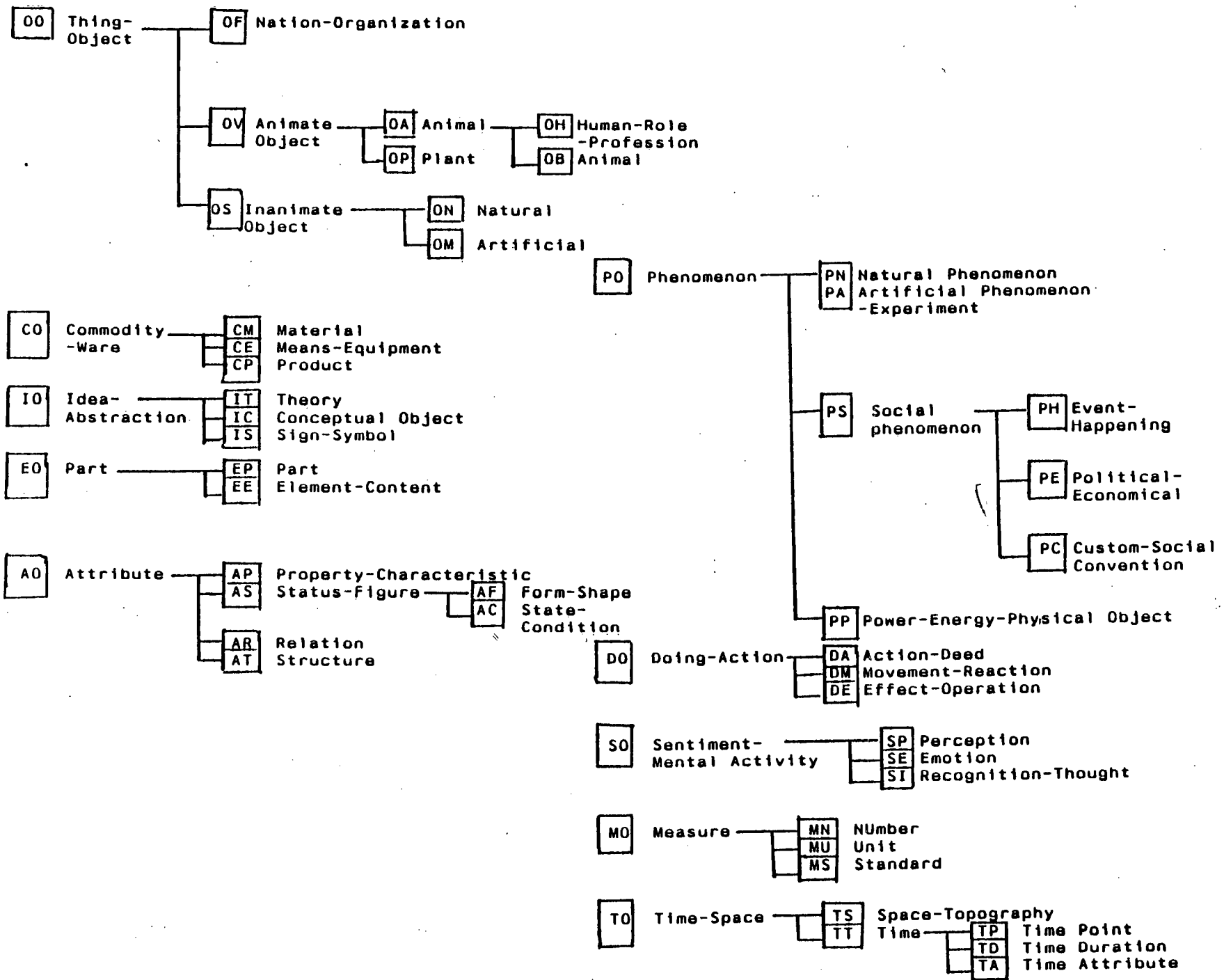


Fig. 3.28: Semantiese primitiewe vir naamwoorde



Die verskille wat dikwels voorkom in die interpretasie van die semantiese funksie van sintaktiese stukke, het dit genoodsaak dat omvattende algoritmes vir leksikografiese opgestel moet word om 'n eenvormige ontleding en beskrywing op die semantiese vlak te kan uitvoer. Vergelyk in hierdie verband byvoorbeeld die algoritme in Ngao, Tsujii en Nakamura (1986:294) en verder die bespreking in Steiner (1985) en Zelinsky-Wibbelt (1986).

### 3.5 SLOTOPMERKINGS

Die voorgaande bespreking van oorsese modelle toon eensyds dat die rekenarisering van leksikografiese ondernemings 'n verskeidenheid fasette behels, en andersyds dat omskakeling na rekenarisering op verskillende wyses aangepak kan word. Die hoofasette is die rekenarisering van bestaande en nuwe woordeboekteks, die ontwikkeling van 'n woordeboekteks, die ontwikkeling van 'n woordbank en die rekenarisering van 'n sitatebasis. In die geval van Oxford word omskakeling na rekenarisering as een intensiewe en omvattende onderneming aangepak. Dit is moontlik gemaak deur Oxford University Press se reuse finansiële sterkte en die aansienlike steun wat deur die Britse Regering, Universiteit van Waterloo (Kanada) en IBM (Verenigde Koninkryk) verleen word. In die geval van Longmans is rekenarisering stapsgewys bewerkstellig, beginnende met die rekenarisering van bestaande tekste. Die INL pak rekenarisering ook stapsgewys aan, op 'n projekbasis en met verskillende insette deur die onderskeie afdelings..

Op grond van die inligting wat in hierdie hoofstuk aangebied is, asook die bronne wat genoem word, sou Suid-Afrikaanse leksikografiese ondernemings na gelang van bepaalde eie behoeftes en vermoëns, rekenariseringstrategieë kon ontwerp.

## HOOFSTUK 4: SAMEVATTING EN AANBEVELINGS

### 4.1 SAMEVATTING

Die hoofogmerke van hierdie ondersoek was om vas te stel of die voordele van die rekenarisering van leksikografiese prosesse in Suid-Afrika doeltreffend benut word, om inligting te verskaf oor die gebruik van die rekenaartegnologie in leksikografie-projekte in die buiteland, en om voorstelle te doen oor toekomstige ontwikkelings in die Suid-Afrikaanse leksikografie-bedryf.

Ten spyte van die uitgebreide gebruik van die rekenaartegnologie in die leksikografie in die buiteland, het dit uit die ondersoek geblyk dat die voordele van die rekenaartegnologie op 'n uiters beperkte wyse in Suid-Afrika benut word. Die huidige stand van die leksikografiebedryf in Suid-Afrika het verder getoon dat daar in die aanbevelings oor rekenarisering enersyds onderskei moet word tussen die bestaande hoofwoordeboekprojekte en ander projekte, en andersyds tussen leksikografieprojekte in die Afrikatale en dié in die twee amptelike tale.

#### 4.1.1 Die hoofwoordeboekprojekte

Daar is drie hoofwoordeboekprojekte: die Woordeboek van die Afrikaanse Taal (die Universiteit van Stellenbosch), die Dictionary of South African English on Historical Principles (Rhodes-Universiteit) en die Xhosa-woordeboek (die Universiteit van Fort Hare).

Die WAT- en die DSAEHP-projekte ontvang deurlopende staatsteun, dus het hulle bestaansekuriteit. Die voortsetting van die Xhosa-projek is in die weegskaal omdat fondse uit die privaatsektor en van die RGN nie meer beskikbaar is nie en die Universiteit van Fort Hare nie die koste kan dra nie. Soos uitgewys in die vorige ondersoek (vgl. Eksteen 1985), kan hoofwoordeboekprojekte nie sonder staatsteun onderneem en voortgesit word nie. Dit is by gebrek aan staatsteun dat daar, met die uitsondering van Xhosa, nie hoofwoordeboeke vir die Afrikatale opgestel is nie; selfs die Xhosa-woordeboek is in hierdie stadium van heel beskeie omvang vergeleke met die WAT. Daar bestaan egter die behoefte aan 'n hoofwoordeboek vir elk van die hooftale om in die behoeftes van die betrokke taalgroep en van die breë samelewing te voorsien. Die SAUK is 'n voorbeeld van 'n gebruiker wat hierdie behoefte terdeë aanvoel.

Soos uitgewys in die voorgaande bespreking, is een van die treffendste eienskappe van die rekenaar die vermoë om groot hoeveelhede data te orden en toeganklik te maak. Die omvang van die databasisse wat vir hoofwoordeboekprojekte, en selfs vir woordeboeke van kleiner omvang, ontwikkel word, is so groot dat die data slegs rekenaarmatig hanteer kan word vir gereelde bywerking en beskikbaarstelling.

Benewens die organisatoriese en produksie-voordele, hou die gerekenariseerde bedryf van hoofwoordeboekprojekte groot voordele ten opsigte van mannekrag- en finansiële besparing in. Dit verminder die behoefte aan ondersteunende werkkragte aansienlik, verhoog die

produktiwiteit van leksikograwe en stel die onderneming in staat om deurlopend verkoopbare produkte en dienste te lewer terwyl die hoofproduk nog ontwikkel word. Hierdie aspekte van die saak verdien deeglike oorweging. Dit moet opgeweeg word teen die aanvanklike koste van omskakeling na rekenarisering.

By die WAT word die woordeboek oor 'n baie lang termyn ontwikkel. Die bestaande versameling woordeboekmateriaal verteenwoordig egter 'n groot belegging wat nie geredelik en gerieflik vir moontlike kopers van materiaal (bv. uitgewers), ander gebruikers en vir gebruik vir ander projekte deur die WAT toeganklik is nie. Alhoewel dit 'n besonder uitdagende taak sal wees om volledig na rekenarisering oor te skakel terwyl daar voortgewerk word aan die woordeboekdele wat nog nie verskyn het nie, sal die alternatief (nl. beperkte gebruik van die rekenaar, slegs vir die verwerking van die teks van die publikasies, totdat al die dele afgehandel is) op die lange duur onekonomies wees. Hoe gouer die Buro se materiaal in 'n woordbank opgeneem kan word, hoe groter sy geleenthede om kommersiële voordeel te trek uit sy besit, ten opsigte beide van die verkoop van materiaal en van koste-effektiewe opdatering en instandhouding.

Die ondersoek het verder uitgewys dat die twee staatsondersteunde hoofwoordeboekkantore voordelig geplaas is vir omskakeling na rekenarisering. Die universiteite waaraan hulle verbonde is, bedryf reeds vir navorsings- en administratiewe doeleindes op uitgebreide skaal rekenaarstelsels en hulle beskik oor aansienlike kundigheid wat die programmering van sodanige projekte betref.

Met die Rhodes-projek is die eerste stappe al gedoen om die rekenaar in te span. Tot dusver is eintlik nog net by perifere sake uitgekom, soos die ordening van die werk aan die woordeboek. Min aandag is nog geskenk aan die verwerking van die leksikografiese materiaal self. Uit die bespreking van oorsese modelle in hierdie verslag, blyk egter dat watter omskakelingstrategie ook al gekies word, die omskakeling binne die raamwerk van 'n oorkoepelende plan moet geskied. Dit is byvoorbeeld teenproduktief en tydverkwistend om volgens die voorkeure van individuele leksikograwe op klein skaal materiaal te begin insleutel en om dan later uit te vind dat die stelsel nie aan woordbankvereistes voldoen nie.

Daar moet voortdurend rekening gehou word met die versoenbaarheid van die stelsel met die stelsels van ander instansies met wie materiaal uitgeruil wil word, soos in die geval van die ooreenkoms tussen die Rhodes-kantoor en Oxford University Press om materiaal aan laasgenoemde te verskaf vir sy internasionale woordeboeke. Verder is daar die moontlikheid dat materiaal van kantore wat eentalige woordeboeke opstel, in die ontwikkeling van meertalige woordeboeke herbenut sou kon word.

#### 4.1.2 Kleiner woordeboekprojekte

Uit die ondersoek het dit geblyk dat privaatuitgewers in samewerking met akademiëci betreklik voldoende voorsien in die behoefte aan woordeboeke van kleiner omvang vir Afrikaans en Engels. Daar is ook duidelik tekens dat die uitgewersbedryf in Suid-Afrika in toenemende mate die rekenaar-tegnologie vir leksikografiese doeleindes begin inspan.

Die rekenariserings van die WAT en die DSAEHP sal in die toekoms die basis kan skep vir 'n nuwe samewerking tussen privaatuitgewers en hierdie woordeboekkantore wat die uitruil, gebruik en herbewerking van leksikografiese materiaal betref.

Die opname van die bestaande kleiner woordeboeke vir die Afrikatale het aangetoon dat daar 'n behoefte aan bykans elke soort woordeboek bestaan vir gebruik in die formele en die nie-formele onderrig van en in die betrokke tale.

Met die Noord-Sotho-woordeboekprojek van die Universiteit van Pretoria en die hersiening van die Noord-Sotho-woordeboek van die Randse Afrikaanse Universiteit is daar 'n begin gemaak met die rekenariserings van leksikografiese prosedures en tegnieke vir die Afrikatale. Die sukses van hierdie twee projekte sal in die toekoms bepalend wees vir die verdere benutting van die rekenaar-tegnologie in soortgelyke projekte.

Die uitgewers se belangstelling in die benutting van die rekenaar-tegnologie vir die saamstel van woordeboeke en die twee projekte in die Afrikatale het ook die behoefte aan teoretiese en tegniese inligting oor hierdie aspek van die leksikografie onderstreep, asook die behoefte aan opgeleide leksikograwe vir al die taalgroepe.

#### 4.1.3 Die huidige infrastruktuur: die rol van akademiëci

Met die uitsondering van die drie hoofwoordeboeke wat hier bo bespreek is, toon die standbeskrywing in hierdie verslag dat die produksie van algemene woordeboeke byna uitsluitlik in die hande van akademiëci is. Uitgewers benut in 'n groot mate op 'n ad-hoc basis die dienste van akademiëci vir die leksikografie-projekte wat hulle inisieer, of hulle onderneem die uitgee van woordeboeke wat deur akademiëci van verskillende taaldepartemente, en in enkele gevalle deur taalpraktisyne, onderneem is.

Hierdie situasie is 'n natuurlike uitvloeisel van die feit dat die saamstel van 'n woordeboek gespesialiseerde kennis van die betrokke tale en van die leksikografië-teorie en -praktyk vereis, en van die feit dat die lede van die taaldepartemente hulle aan 'n universiteit in 'n situasie bevind waar hulle noue kontak met ander taalkenners en toegang tot die biblioteek- en rekenaar-fasiliteite het.

Hierdie organisatoriese struktuur hou die voordeel in dat die bestaande mannekrag vir die voorsiening in leksikografiese produkte en dienste maksimaal benut word. Daarbenewens is studentehulp 'n belangrike

mannekragbron. Finansiëel hou die situasie die voordeel in dat bestaande infrastrukture benut word en dat die universiteite rekenaardienste aan die betrokke akademici beskikbaar stel. Die universiteite is die aangewese tuiste vir leksikografieprojekte omdat sulke groot instansies oor die rekenaarfasieliteit beskik vir die ontwikkeling van woord- en termbanke. Oorsese voorbeelde toon dat die ontwikkeling van sodanige banke noodsaaklik is vir die optimale beskikbaarstelling van leksikografiese materiaal.

Die huidige bedeling is veral bevorderlik vir die produksie van Afrikataalwoordeboeke. Onder nie-akademici bestaan daar 'n nog groter gebrek aan leksikograwe wat hulle met Afrikatale bemoei as in die geval van Engels en Afrikaans. Gevolglik is die Afrikatale Departemente 'n belangrike bron vir die saamstel van woordeboeke. Finansiëel is dit ook vir die Afrikatale nodig om hulp vanuit die universiteite te verkry. Afgesien van steun vir woordeboeke vir gebruik in die swart primêre onderwys, is dit die Staat se amptelike beleid om slegs steun aan projekte vir die twee amptelike tale te verleen. Die selfregerende en nasionale state het tans nie die middele en vermoëns om beduidend te voorsien in woordeboekbehoefte nie. Hulle is op vele ontwikkelingsterreine afhanklik van buitehulp, ook op die gebied van taalontwikkeling.

Die rol wat die universiteite tans in die Suid-Afrikaanse leksikografiebedryf speel, hou egter nie net voordele vir die bedryf in nie. Onder die nadele tel die feit dat leksikografieprojekte meestal op 'n ad-hoc, deeltydse basis geskied. Dit bring mee dat projekte vir 'n besondere taal beperk word tot dié van kleiner omvang. Daar word nie werklik wetenskaplik aandag gegee aan die reële behoeftes en aan die sistematiese uitwissing van leemtes nie. 'n Verdere nadeel van die situasie is dat die voltooiing van lopende leksikografieprojekte vertraag word omdat die meeste taaldepartemente nie oor die nodige finansies beskik om voltydse personeel, selfs op ondersteunende vlak, aan te stel om deurlopend aan hierdie projekte te werk nie.

Ten spyte van die nadele wat genoem is, is die skep van woordeboeke as 'n nuwe-produk in die aktiwiteite van akademici lewensbelangrik vir die leksikografiebedryf. Daarom is dit onrusbarend vir die bedryf dat akademici toenemend onder druk staan om slegs betrokke te wees by aktiwiteite wat subsidie genereer. Woordeboekbou word in sommige kringe glad nie as navorsing beskou nie. Selfs met rekenaarondersteuning, is die saamstel van 'n woordeboek 'n tydrowende proses wat ook heelwat taalnavorsing vra. Hierdie aktiwiteit kompeteer dus moeilik met ander aktiwiteite in 'n situasie waar daar druk is om navorsingsprodukte op 'n gereelde grondslag voort te bring. Die toekoms van die leksikografiebedryf staan dus onder bedreiging.

#### 4.2 AANBEVELINGS

By die opstel van die onderstaande aanbevelings was die sentrale uitgangspunt dat die huidige beleid in die land ten opsigte van die prioriteitsvlak wat aan navorsing en navorsingsverwante aktiwiteite toegeken word, bepaal wat aanbeveel kan word.

Ooreenkomstig die doel van die ondersoek, spreek die aanbevelings hoofsaaklik die rekenariseringsaspek van die leksikografiebedryf aan. Aangesien rekenarisering slegs in bepaalde kontekste haalbaar is, word sekere omringende faktore ook genoem.

#### 4.2.1 Die hoofwoordeboeke

##### **Aanbeveling 1: Rekenarisering by die DSAEHP en die WAT**

**Op grond van voorbeelde uit die buiteland word daar aanbeveel dat die DSAEHP- en WAT-kantore die volgende riglyne vir beplanning oorweeg:**

- dit is koste-oneffektief om rekenarisering uit te stel, die aanvanklike redelik hoë uitgawes ten spyte;
- alhoewel omskakeling vanweë finansiële en mannekragprobleme stapsgewys onderneem moet word, is dit gerade om 'n oorkoepelende raamwerk vir die hele onderneming eers op te stel; dit is veral belangrik om te verseker dat die twee hoofbene - produksie van woordeboekteks en databasisbou - albei binne die stelselontwerp geakkommodeer kan word;
- terwyl elke instansie volgens eie behoeftes en diensleweringfunksies sy stelsel sal ontwerp, is dit wenslik om versoenbaarheid en koppelingsmoontlikhede met ander sentra binnelands en buitelands deeglik te ondersoek;
- in die lig van die finansiële en mannekragimplikasies van stelselontwikkeling is dit wenslik om soveel moontlik op bestaande kennis en kundigheid in die buiteland te steun; onderhandelings met die Oxford English Dictionary Computerization Project se bestuurder - wat reeds deur die Rhodes-kantoor in die vooruitsig gestel word - sal moontlik kan lei tot 'n voordelige ooreenkoms vir die gebruik en aanpassing van die Oxford-stelsel;
- befondsing vir die volledige oorskakeling na gerekenariseerde stelsels sal waarskynlik nie maklik gevind kan word nie maar met die voorlegging van 'n deeglike beplanningstuk, en spesifieke voorstelle vir fase-indeling en tussentydse sperdatums, kan gefaseerde befondsing dalk bewerkstellig word; die gesamentlike bemoontliking van die Oxford-projek deur die Britse regering en die privaatsektor is 'n noemenswaardige voorbeeld.

##### **Aanbeveling 2: Die Xhosa-woordeboekprojek**

**Daar word aanbeveel dat die Universiteit van Fort Hare steun ontvang vir die rekenarisering van die woordeboekmateriaal.**

Aangesien die enigste omvangryke versameling woordeboekmateriaal in die Afrikatale van Suid-Afrika tans op handgeskrewe kaarte by die Universiteit van Fort Hare bestaan, word daar aanbeveel dat die rekenarisering van hierdie waardevolle versameling as 'n saak van wyer belang as net vir die Universiteit beskou word. Die versameling sal vanweë die aard en omvang daarvan, internasionale belangstelling wek indien dit deur rekenarisering toeganklik sou wees.

#### **Aanbeveling 3: Hoofwoordeboeke vir die ander Afrikatale**

Die behoeftestudie toon dat daar vir elkeen van die Afrikatale binne die Republiek se grense 'n hoofwoordeboekkantoor tot stand gebring behoort te word. Vir die huidige ontbreek die middele daartoe. Daar word egter aanbeveel dat daar aan die universiteite in die betrokke geografiese gebiede finansiële steun gegee word om enersyds 'n langtermyn ontwikkelingsplan vir 'n hoofwoordeboek te ontwikkel, en andersyds genoeg steun verleen word om die projek van stapel te stuur.

#### **Aanbeveling 4: Die bestuur en tegniese bedryf van hoofwoordeboeke**

Daar word aanbeveel dat die tegnologie van die inligtingsera meer uitgebreid aangewend word.

Daar word aanbeveel dat hoofwoordeboekprojekte volgens die vereistes en tegnologie van die inligtingsera bedryf moet word. Dit impliseer onder meer dat bewerkte materiaal so vinnig as moontlik bemark moet word en ook in elektroniese vorm vir potensiële gebruikers, (byvoorbeeld vir die Beltel-stelsel, die SAUK, uitgewers, asook vervaardigers van teksverwerkingsprogramme) beskikbaar moet wees.

#### **4.2.2 Nie-hoofwoordeboeke**

In die samevatting hierbo oor die omstandighede waarin die leksikografiebedryf hom bevind, was die hoofpunt die feit dat die bedryf oorwegend steun op die deeltydse leksikografiese aktiwiteite van akademië en op die ondersteuning wat universitêre infrastrukture aan hierdie aktiwiteite verleen. Dit blyk egter ook dat die huidige subsidie-stelsel vir universiteite woordeboekprojekte ernstig gaan benadeel. Die aanbevelings hieronder hou rekening met die genoemde feite.

**Aanbeveling 5: Die leksikografiebedryf moet nie losgemaak word van die universiteit nie.**

Dit is op mannekrag-, finansiële en infrastrukturele vlak buite die land se bereik om 'n onafhanklike leksikografiebedryf te ontwikkel. So 'n stap is verder onwenslik omdat leksikografiese aktiwiteite en die taaldepartemente mekaar onderling aanvul en omdat universiteite reeds oor hoofraamfasiliteite beskik.

**Aanbeveling 6:** Leksikografieprojekte moet nie binne die raamwerk van die bestaande subsidie- en ander fondstoekenningsmeganismes met ander navorsingsprojekte kompeteer nie.

Leksikografieprojekte is ontwikkelings- eerder as navorsingsgerig. Ontwikkeling staan in Suid-Afrika hoog op die prioriteitslys en daar bestaan 'n verskeidenheid finansiële bronne vir die steun van ontwikkelingswerk (waaronder die Ontwikkelingsbank). Taalontwikkeling word in Suid-Afrika egter nog nie na waarde geskat nie en dus nie uit ontwikkelingsfondse gesteun nie.

In die vorige RGN-ondersoek na leksikografiese aangeleenthede in Suid-Afrika (Eksteen 1985) is daar voorgestel dat 'n fonds spesifiek vir woordeboekprojekte naamlik deur die RGN geadministreer kan word op dieselfde basis as die RGN se administrasie van die nasionale begroting vir universiteitsnavorsing. Die behoefte aan so 'n fonds neem toe as gevolg van die groter verkeer tussen die tale van die land, dus die groter aanvraag na woordeboeke vir verskillende doeleindes.

**Aanbeveling 7:** Leksikograwe aan universiteite moet hulle leksikografiese aktiwiteite verder ontgin vir subsidie-doeleindes.

Bogenoemde kan byvoorbeeld bewerkstellig word deur die skryf van vakkundige artikels oor die leksikologiese bevindinge wat voortspruit uit die opstel van 'n woordeboek. Fondse wat sodoende verdien word, sal egter nie projekbedryfskoste ten volle dek nie, veral nie as die reële koste van rekenaardienste hierby ingereken word nie.

**Aanbeveling 8:** Leksikografieprojekte moet uit die staanspoor rekenaar-matig aangepak word en die leksikografies bewerkte taalmateriaal en onbewerkte taalkorpusse moet in 'n gerekenariseerde teks-en-woordbank by die betrokke instansie opgeneem word.

Naas die beskikbaarstelling van verkoopbare gedrukte produkte bied so 'n stap die geleentheid dat daar kumulatief aan die instansie se leksikografiese besit uitgebou kan word, ten behoeve van sowel interne as eksterne gebruikers.

Die ontwerp van stelsels vir woordeboekprogramme en vir woordbanke moet in oorleg met alle potensiële vennote geskied ten einde onversoenbaarheid en ander beperkings op die uitruilbaarheid en verkoopbaarheid van die materiaal so ver moontlik uit te skakel. Daar is reeds vermeld dat die ontwikkelingskoste verbonde aan stelselontwerp vir leksikografiese databasisse (woordbanke) hoog is, en dat die regstelling van foute in stelselontwerp groot kostes kan meebring.

#### 4.2.3 Opleiding

**Aanbeveling 9:** Taalstudente moet aangemoedig word om kursusse in die rekenaarwetenskap te volg. Vir hierdie doel word verder aanbeveel dat universiteite in hulle regulasies vir die keuse van vakke voorsiening moet maak vir 'n kombinasie van taalkursusse en rekenaarwetenskap vir graaddoeleindes.



Die beperkte gebruik van die rekenaartegnologie in die leksikografiebedryf in Suid-Afrika kan hoofsaaklik toegeskryf word aan 'n gebrek aan kundigheid. Daar word dus aanbeveel dat die taaldepartemente aan universiteite die voorbeeld van talle taaldepartemente in die buiteland navolg, naamlik om studente aan te moedig om kursusse in die rekenaarwetenskap te volg en om in die taaldepartemente aandag te gee aan rekenaartoepassings in die taalkunde en die toegepaste taalkunde. Hierdie aanbeveling word in die ander verslae in die LEXINET-reeks ook gedoen en onderstreep die behoefte aan die uiteindelijke instelling van tersiêre kursusse in die rekenaarlinguistiek.

## BIBLIOGRAFIE

- AGRICOLA, E.J., SCHILDT, VIEHWEGER, D. 1982. *Wortschatzforschung heute*. Leipzig: VEB Verlag.
- AKKERMAN, E. et al. 1985. *Designing a computerized lexicon for linguistic purposes. Ascot Report No. 1*. Amsterdam: Costerus.
- AL, B.P.F., VAN STERKENBURG, P.G.J. (reds.). 1978. *Wetenskap en woordenskat*. Muiderberg: Coutinho.
- ALBERTS, M. 1983. *Aspekte van die rekenarisering van vakwoordeboeke*. Ongepubliseerde M.A.-verhandeling, Universiteit van Pretoria.
- AL-KASAMI, A.M. 1977. *Linguistics and bilingual dictionaries*. Leiden: Brill.
- AMSLER, R.A. 1980. *The structure of the Merriam-Webster Pocket Dictionary*. Ongepubliseerde proefskrif, The University of Texas at Austin.
- ANDERSON, M.E. 1978. *Roger's Thesaurus: an explanation of its purpose and a study of some applications of its principles*. Ongepubliseerde proefskrif, Case Western Reserve University.
- ASTRÖM, K. 1983. Aspects of term bank operation. In: Snell (ed.) 1983:179-185.
- BAMMESBERGER, A. (ed.) 1985. *Problems of Old English lexicography*. Regensburg: Verlag Friedrich Pustet.
- BENSON, M. 1985. Collocations and idioms. In: Ilson (ed.) 1985:61-68.
- BLAHA, H. 1980. Aufbau und Nutzungsmöglichkeiten einer Normen-Terminologie-Datenbank. *Fachsprache*, 2.4:146-155.
- BLATT, A., FREIGANG, K.H., SCHNUTZ, K.D., THOME, G. 1985. *Computer und Übersetzen; Eine Einführung*. Hildesheim: Georg Olms Verlag.
- BOTHA, R.P., SINCLAIR, M. (reds.) 1985. Norme vir taalgebruik *Spil Plus*, 10 (1985):269-291.
- BRINKMAN, K.H., TANKE, E. 1976. The TEAM program system and international cooperation in terminology. In: Infoterm 1976: 180-186.
- BRUSTKERN, J., HESS, K.D. The BONNLEX Lexicon System. In: Goetschalckx and Rolling (eds.) 1982:35-40.
- BRUSTKERN, J., REUDERS, W., WILLÉE, G. 1981. *Handbuch der Programmbibliothek zur linguistischen und philologischen textverarbeitung*. Wesdeutsche Verlag.
- BRUSTKERN, J., SCHULZE, W. 1983(a). *The structure of the word data base for the German Language*. Ongepubliseerde ms.

- BRUSTKERN, J., SCHULZE, W. 1983(b). *Towards a cumulated word data base for the German Language*. Ongepubliseerde ms.
- CLUVER, A.D. de V., SCHEFFER, C.J. (reds.) 1984. *RGN-ondersoek na vaktaalaangeleenthede in Suidelike Afrika: Verslag van die ondersoek na organisatoriese aangeleenthede op vaktaalgebied*. Pretoria: Raad vir Geesteswetenskaplike Navorsing. Verslag TERM-I.
- COETZEE, N., DU PLOOY, N.F., CLUVER, A.D. DE V. 1985. *Tweetalige Rekenaarwoordeboek; Bilingual Computer Dictionary*. Johannesburg: McGraw-Hill.
- DEBILLE, L., MASEREEUW, P., MEIJS, W. 1985. *Investigations into Dutch morphology*. Interim report submitted to the Commission of the European Communities. Ongepubliseerde ms.
- DE TOLLENAERE, F. 1963. *Nieuwe wegen in de Lexicologie. Verhandelingen van de Koninklijke Nederlandsche Akademie van Wetenschappen*. Nieuwe Reeks, 71.1. Amsterdam.
- DIETRICH, R. 1973. *Automatische Textwörterbücher*. Tübingen: Max Niemeyer Verlag.
- EKSTEEN, L.C. (red.) 1985. *Leksikografie-aangeleenthede in Suid-Afrika*. Pretoria: RGN. Verslag Lexi-1.
- ENGEL, G., MADSEN, B.N. 1984. From dictionary to data-base. In: Hartmann (ed.) 1984:339-344.
- FOX, M.S., BEBEL, D.J., PARKER, A.C. 1980. *The automated dictionary*. Computer, 13.7:35-48.
- GALINSKI, C., NEDOBITY, W. 1986. *Eine terminologische Datenbank als Managementinstrument*. Infoterm 3-86. Wenen: Infoterm.
- GEERAERTS, D., JANSSENS, G. 1982. *Wegwijs in woordenboeken*. Assen: Van Gorcum.
- GOETSCHALCKX, J., ROLLING, L. (eds.) 1982. *Lexicography in the electronic age. Proceedings of a symposium held in Luxembourg, 7-9 July, 1981*. Amsterdam: North Holland.
- GOFFIN, R. 1982. Linguistic criteria to evaluate terminology banks. In: Goetschalckx and Rolling (eds.) 1982: 159-168.
- HARTMANN, R.R.K. (ed.) 1983. *Lexicography: Principles and Practice*. Academic Press.
- HARTMANN, R.R.K. (ed.) 1984. *International conference on lexicography. Lexeter '83. Proceedings*. Tübingen: Max Niemeyer.
- HAUPTFLEISCH, D.C. 1985. Taalnormering en woordeboeke: 'n Praktykgerigte perspektief. In: Botha en Sinclair (reds.) 1985: 211-237.

- HEINZE, G., en BRUSTKERN, J. n.d. Das Lexikonsystem ALEXSYS. IKP. *Arbeitsberichte, Abteilung LDV*, nr. 6.
- HESS, K., BRUSTKERN, J., LENDERS, W. 1983. *Maschinenlesbare deutsche Wörterbücher*. Tübingen: Max Niemeyer Verlag.
- HILL, C.P. 1985. Alternatives to Dictionaries. In: Ilson (red.) 1985:115-122.
- HJULSTAD, H. 1986. *Norwegian Council for Technical Terminology. Program Documentation*. Ongepubliseerde ms.
- HOFFMAN, E. 1983. Stages in the life cycle of LEXIS. In: Snell 1983: 186-191.
- HOHNHOLD, I. 1984. *Die Terminologie-Datenbank TEAM im sprachendienst Siemens München*. Mitteilungsblatt für Dolmetscher und Übersetzer, 3/30:1.
- ILSON, R. 1984. The communicative significance of some lexicographical conventions. In: Hartmann 1984 (ed.) 80-85.
- ILSON, R. (ed.) 1985. *Dictionaries, lexicography and language learning*. Oxford: Pergamon Press.
- INFOTERM. 1976. *International cooperation in terminology*. Infoterm series 3. Verlag Dokumentation.
- INFOTERM. 1982. *ISO Handbook*. Geneve: ISO
- INFOTERM. 1986. *Second Infoterm Symposium. Networking in terminology. International co-operation in terminology work*.
- INSTITUUT voor Nederlandse Lexicologie 1985. *Jaarboek van de Stichting Instituut voor Nederlandse Lexicologie; overzicht van het jaar 1984*. Leiden: Secretariaat van de Stichting.
- ISO. 1982. *ISO Standards Handbook*. Geneve.: ISO
- ISO 1951-1973. *Lexicographical symbols particularly for use in classified defining vocabularies*. Geneve: ISO.
- ISO 2788-1974. *Documentation - Guidelines for the establishment and development of monolingual thesauri*. Geneve: ISO.
- ISO/DIS 704-. 1985. *Principles and methods of terminology*. Geneve: ISO.
- ISO/DIS 6156. *Magnetic tape exchange format for terminological/lexicographical records (MATER)*. Geneve: ISO.
- ISO/R 639-1967. *Symbols for languages, countries and authorities*. Geneve: ISO.
- ISO/R 704-1968. *Naming principles*. Geneve: ISO.

- ISO/R 860-1968. *International unification of concepts and terms.* Geneve: ISO.
- ISO/R 919-1969. *Guide for the preparation of classified vocabularies. (Example of method).* Geneve: ISO.
- ISO/R 1087-1969. *Vocabulary of terminology.* Geneve: ISO.
- ISO/R 1149-1969. *Layout of multilingual classified vocabularies.* Geneve: ISO.
- ISO/TC 46/GT 5 *Guidelines for the establishment of thesauri; multilingual thesauri.* (Proposals by the French member committee)
- JASPAERT, L. 1984. *Matters of Morphology; Eurotra morphology legislation.* Ongepubliseerde ms.
- KLEINBART, P. *Prolegomenon to 'intelligent' thesaurus software.* Journal of Information Science 11:45-53.
- KNOWLES, F.E. 1984. Dictionaries and computers. In: Hartmann (ed.) 1984:301-313.
- KUHLEN, R. 1986. *Informationslinguistiek.* Tübingen: Max Niemeyer Verlag.
- LAWSON, V. (red.) 1985. *Tools for the trade; Translating and the computer 5.* Aslib. Alden Press: Oxford.
- LEMMENS, M., WEKKER, H. 1986. *Grammar in English Learners' Dictionaries.* Tübingen: Max Niemeyer.
- Semantics (Volumes 1 and 2).* Cambridge: University Press.
- LYNE, A.A. 1984. On rooting out words with inflated frequencies in word-counts of specialized registers. In: Hartmann (ed.) 1984:362-370.
- LYONS, J. 1977. *Semantics (Volumes 1 and 2).* Cambridge: University Press.
- MCNAUGHT, J. 1982(a). Specialised Lexicography in the context of a British Linguistic Data Bank. In: Goetschalckx and Rolling (eds.) 1982:171-184.
- MCNAUGHT, J. 1982(b). The role of terminological relationships in encoding and accessing systematic multilingual specialized computer dictionaries. *Multilingua*, 1-1:53-54.
- MCNAUGHT, J. 1983. *The generation of term definitions from on-line terminological thesaurus.* Ongepubliseerde ms.
- MARTIN, W.J.R., AL, P.J.G., VAN STERKENBURG, P.J.G. 1983. On the processing of a text corpus. In: Hartmann (red.) 1983:77-87.

- MENTRUP, W. (ed.) 1982. *Konzepte zur Lexikographie; Studien zur Bedeutungserklärung in eiensprachigen Wörterbüchern*. Tübingen: Max Niemeyer.
- MICHIELS, A., MULLENDERS, J., NOËL, J. 1982. The Longman-Liege Project. In: Goetschalckx and Rolling (eds.) 1982:201-209.
- MOORE, K. 1983. Text processing in dictionary compilation. In: Snell (ed.) 1983:134-136.
- MORIK, K. 1982. *Überzeugungssysteme der Künstlichen Intelligenz*. Tübingen: Max Niemeyer Verlag.
- MORRIS, R. 1985. Enkele aspecte rakende die finansiering en bestuur van woordeboekprojekte. In: Eksteen (red.) (1985) pp. 46-66.
- NAGAO, M., TSUJII, J., NAKAMURA, J. 1986. Terminology dictionary for machine translation. In: Infoterm Series 8: 286-293.
- NEDOBITY, W. 1984. *Die Bedeutung der systematischen Terminologiearbeit für den Aufbau von Wissenbanken und andere Expertensystemen*. Infoterm 15-84. Wenen: Infoterm.
- NEDOBITY, W. 1985(b). *Terminology and artificial intelligence*. Infoterm 1-85. Wenen: Infoterm.
- NEDOBITY, W. 1985(b). *Classification, Artificial Intelligence and Cognitive Psychology*. Infoterm: 9-85. Wenen: Infoterm.
- NEUBAUER, F. 1980. *Die Struktur der Expikationen in deutschen einsprachigen Wörterbüchern*. Hamburg: Helmut Buscke.
- NEUHAUS, H.J. 1985. Design options for a lexical database of Old English. In: Bammerberger (ed.) 1985:197-209.
- NEW OXFORD ENGLISH DICTIONARY. Project. *Progress, Plans and Ambitions*. Ongepubliceerde ms.
- NEW OXFORD ENGLISH DICTIONARY. 1986. *Newsletter*. 9: Januarie 1986.
- NORLING-CHRISTENSEN, O. 1982. Commercial lexicography on the threshold of the electronic age. In: Goetschalckx and Rolling (eds.) 1982:211-219.
- OXFORD ENGLISH DICTIONARIES. *Persvystelling*. 15 Mei 1984/16 Mei 1984.
- PIJNENBURG, W.J.T., HULST, H.V.D., MOORTGAT, M. 1978. *Gestitutionaliseerde lexicologie*. In: Al and Van Sterkenburg (eds.) 1978:72-85.
- PORTIER, M.J., PYPENBURG, W. 1984. Naar een taaldatabank van het Nederlands. In: Handelingen van het acht en dertigste Nederlands filologencongres; Nijmegen, 16-17 April 1984. Amsterdam: APA-Holland Universiteits Pers: 345-353.

- SAGER, J.C., MCNAUGHT, J. 1980(a). *Feasibility study of the establishment of a terminological data bank in the U.K.* CCL/UMIST nr. 81/8, British Library Report nr. 5642.
- SAGER, J.C., MCNAUGHT, J. 1980(b). *Selective survey of existing linguistic data banks in Europe.* CCL/UMIST, nr. 81/9. British Library Report nr. 5643.
- SAGER, J.C., MCNAUGHT, J. 1980(c). *Specifications of a linguistic databank for the U.K.* CCL/UMIST, nr. 81/10. British Library Report nr. 5644.
- SCHAEDER, B. 1981. *Lexicographie als Praxis und Theorie.* Tübingen: Max Niemeyr.
- SCHOONEES, P.C. 1971. *HAT. Verklarende Handwoordeboek van die Afrikaanse Taal.* Klerksdorp: Voortrekkerpers.
- SCHMITZ, K-D. 1986. *Automatische Segmentierung natüersprachiger Sätze.* Hildesheim: George Olms Verlag.
- SCHULZ, J. 1979. Eine Terminologiedatenbank für Übersetzer TEAM. *Philips Terminology Bulletin*, 8.1/2.
- SCHWARZ, C., THURMAIR, G. (eds.) 1986. *Informationslinguistische Texterschliessung.* Hildesheim: Georg Olms Verlag.
- SCHWARZE, C., WUNERLICH, D. (eds.) 1985. *Handbuch der Lexikologie.* Athenaum Verlag.
- SCHWERDTFEGER, A.M. 1986. *Meerduidigheid in Afrikaans.* Ongepubliseerde proefskrif, UNISA.
- SIPEL, C.J. 1984. *Computer Dictionary: The A-Z of microcomputers.* Longman: Harlow.
- SNELL, B. (ed.) 1983. *Term banks for tomorrow's world: Translating and the computer.* London: Aslib.
- STEINER, E. 1985. *Interface structure - proposals for a tightening of ELS-3 procedures.* Ongepubliseerde ms.
- SWANEPOEL, P.H. 1985. 'n Sentrale gerekenariseerde woordbank. In: Eksteen (red.) 1985:67-83.
- SWANEPOEL, P.H. 1986. *Leksikonstrukture vir gerekenariseerde linguistiese databanke.* Pretoria: UNISA.
- TERMNET 7-82 rev. Draft. *Guidelines for the recording of terminological data (for machine processing).* TermNet.
- TEUBERT, W. 1984. Setting up a lexicographical data-base for German. In: Hartmann (ed.) 1984:425-429.
- VAN DER VOORT VAN DER KLEIJ, J.J. 1977. De Studie van Nederlandse woordenboeken. In: Van Sterkenburg (red.) 1977:287-293.

- VAN STERKENBURG, P.G.J. 1984. *Van woordenlijst tot woordenboek*. Leiden: Brill.
- VAN STERKENBURG, P.G.J. 1985-1986. *Algemene inleiding in die Lexicologie*. Leiden: Brill.
- VAN STERKENBURG, P.G.J. (red.) 1976. *De Nederlandse lexicologie tussen handwerk en machine*. Groningen: H.D. Tjeenk Willink.
- VAN STERKENBURG, P.G.J., MARTIN, P.W., AL, B. 1982. A new Van Dale project: bilingual dictionaries on one and the same monolingual basis. In: Goetschalckx en Rolling (eds.) 1982: 223-237.
- VOLLNHALS, O. 1982(a) Fachwörterbücher aus Datenbank. *Börsenblatt*, 60, 1982:1663-1666.
- VOLLNHALS, O. 1982(b). *Technical dictionaries retrieved from a data-base*. META, 27.2:157-166.
- VOLLNHALS, O. 1984. Utilization of a commercial linguistic data-base system for electronic storage and automated production of dictionaries. In: Hartmann (ed.) 1984:430-434.
- VOLLNHALS, O. 1985. TEAM: ein systemüberblick. *Le Langage et l'Homme*, 59-70-75.
- WAHRIG, G. 1973. *Anleitung zur grammatisch-semantischen Beschreibung lexikalischer Einheiten*. Tübingen: Max Niemeyer.
- WEBER, H.J. 1974. *Mehrdeutige Wortformen im heutigen Deutsch*. Tübingen: Max Niemeyer Verlag.
- WEINER, E. 1984. The New Oxford English Dictionary. *EURALEX BULLETIN*, 1.2:3-5.
- WERSIG, G. 1978. *Thesaurus-Leitfaden*. München: Verlag Dokumentation Saur KG.
- WILLÉE, G. 1979. Das Bonner Lexikonsystem BONNLEX1. *IKP-Arbeitsberichte, Abteilung LDV*, nr. 6 (ungepublizierte ms.)
- WILLÉE, G. 1982. Das Programmsysteme LEMMA 2 - eine Weiterentwicklung von 'LEMMA'. *IKP-Arbeitsberichte, Abteilung LDV*, nr. 2 (ungepublizierte ms.)
- WÜSTER, E. 1985. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Wenen: Infoterm.
- ZELINSKY-WIBBELT, C., MEULLER-V. BROCHOWSKI, A. 1986. *An attribute grammar as a generating device for a system of semantic features*. Unpublizierte ms.
- ZGUSTA, L. 1971. *Manual of Lexicography*. Den Haag: Mouton.
- ZIMMERMANN, H. 1972. *Das Lexicon in der maschinellen Sprachanalyse*. Athenäum Verlag.



Doc nr 206786  
Copy nr 206787

