



SENTRUM VIR BIBLIOTEEK EN
INLIGTINGSDIENSTE
CENTRE FOR LIBRARY AND
INFORMATION SERVICES

VERVALDATUM/DATE DUE

--	--	--	--

2547

0 0 0 2 4 3 2 4 5 8



19 20000003

001.3072068 HSRC LEXI 9



* 2 4 3 2 4 5 *

Rekenaarfasiliteite vir die hantering van taal

(PROLANG-projek in die LEXINET-program)

BIBLIOTEEK
LIBRARY

RGN

RAAD VIR
GEESTESWETENSKAPLIKE
NAVORSING

HSRC

HUMAN
SCIENCES RESEARCH
COUNCIL



RGN · HSRC

Verlag LEXI-9

Rekenaarfasiliteite vir die hantering van taal

(PROLANG-projek in die LEXINET-program)

W. T. Claassen

Pretoria
Raad vir Geesteswetenskaplike Navorsing
1989

W. T. Claassen, D. Litt., US

Afdeling Leksikologie

Instituut vir Taal- en Kunstenavorsing

Uitvoerende Direkteur: Dr. K. P. Prinsloo

ISBN 0 7969 0674 2 (reeks)

© Raad vir Geesteswetenskaplike Navorsing, 1989

Gedruk en uitgegee deur die RGN

Pretoriusstraat 134

Pretoria

RGN BIBLIOTEEK 1990 2. 27. HSRC LIBRARY	
STANDKODE 001-3672068 HSRC LEXI 9	AANWINSNOMMER 082871

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 belangrikheids-waarde vir Suid-Afrika van die gebied te bepaal en om daarvolgens aanbevelings te doen oor moontlike ontwikkelingstappe.

Hierdie verslag is een van 'n reeks van sewe wat verskillende aspekte van gerekenariseerde taalverwerking dek, soos bespreek in hoofstuk een. Dan is daar ook 'n agste verslag as 'n samevatting van die hele reeks. Al die verslae het die gemeenskaplike kenmerk dat hulle verslae van 'n terreinverkenningssaad is. As sodanig is hulle belangrike inligtings-dokumente en vertrekstukke vir die verdere uitbouing van 'n betreklik nuwe vakgebied in Suid-Afrika, nl. die rekenaarhantering van taal.

Die ondersoek waaroor dit in die onderhawige verslag gaan (die PROLANG-verslag) was 'n verkenning van rekenaarfasiliteite - apparatuur en programmatuur - vir die rekenaarverwerking van taal. Die ondersoek is uitgevoer deur 'n navorsingspan van die Universiteit van Stellenbosch onder leiding van prof. W.T. Claassen. Aan die navorsingspan, en dan veral aan prof. Claassen as projekteier en skrywer van die verslag, spreek ek dank en waardering uit vir die bekwame uitvoering van 'n uitdagende projek.

Ten slotte spreek ek ook waardering uit vir die wyse waarop dr. R. Morris as leier van die reeks projekte in die LEXINET-program haar taak deurgevoer het.

K. P. PRINSLOO

UITVOERENDE DIREKTEUR: INSTITUUT VIR TAAL- EN KUNSTENAVORSING

INHOUDSOPGAWE

1.	INLEIDING	1
1.1	RAAMWERK VAN DIE ONDERSOEK	1
1.2	TAAKOMSKRYWING	2
1.2.1	Navorsingsdoelstellings en werkwyse	2
1.2.2	Navorsingspan	4
1.3	DIE INTERRELASIE VAN REKENAARTOEPASSINGS WAARBY TAAAL TER SPRAKE IS	4
1.4	DIE REIKWYDTE VAN TAAAL	5
1.5	UITLEG VAN DIE VERSLAG	6
1.6	TERMINOLOGIE	7
2.	ALGEMENE TENDENSE IN DIE ONTWIKKELING VAN PROGRAMMATUUR EN APPARATUUR	9
2.1	INLEIDING	9
2.2	ALGEMENE TENDENSE	11
2.3	TENDENSE MET BETREKKING TOT APPARATUUR	14
2.4	TENDENSE MET BETREKKING TOT PROGRAMMATUUR EN KUNDIGHEID	15
3.	VLAKKE VAN DIE HANTERING VAN TAAAL MET BEHULP VAN DIE REKENAAR	18
3.1	INLEIDING	18
3.2	DIE GENERERING VAN TEKS	19
3.3	DIE MANIPULASIE VAN DIE UITERLIKE TAAALTEKENS	20
3.4	DIE HANTERING VAN BETEKENIS EN KONTEKS	21
3.5	DIE REKENAAR-GEBRUIKER-RAAKVLAK	21

4.	AKSIES TEN OPSIGTE VAN DIE VERWERKING VAN NATUURLIKE TAAL	24
4.1	INLEIDING	24
4.2	DIE GENERERING VAN TEKS	24
4.2.1	Direkte (skeppende) generering van teks	24
4.2.1.1	Woordverwerkingsprogramme en meegaande fasiliteite	24
4.2.1.2	Ondersteuning ten opsigte van teksgenerering	31
4.2.2	Meganiese generering van teks	39
4.3	DIE UTTERLIKE VERSORGING VAN TEKS	40
4.3.1	Vorm en voorkoms	40
4.3.2	Afkapping	40
4.4	DIE ORDENING VAN TEKS	41
4.4.1	Berging - inleidende opmerkings	41
4.4.2	Die ordening van taalgegewens in databasisse	42
4.4.2.1	Inleiding	42
4.4.2.2	Veeldoelige databasisse	43
4.4.2.3	Taaldatabasisse	44
4.4.2.4	Toepassingsprogramme	45
4.4.3	Konkordansies, woordelyste en woordeskatstatistiek	45
4.4.3.1	Inleiding	45
4.4.3.2	Konkordansies: basiese uitgangspunte en opsies	46
4.4.3.3	Verskillende konkordansieprogramme	49
4.4.3.4	Konkordansies vir die leek	54
4.5	DIE KOMMENTARIERING OF ANNOTASIE VAN TEKS	55
4.6	TOEGANG TOT DATABASISSE	58
4.6.1	Inleiding	58
4.6.2	Dataherwinning	60
4.6.3	Dokumentherwinning	61
4.6.4	Tesouri (inligtingkundig)	64
4.7	TEKS-AANBIEDING	66
4.7.1	Inleiding	66
4.7.2	Rekenaargesteuende onderrig (RGO)	68
4.7.3	Hiperteks	69
4.7.3.1	Inleiding	69
4.7.3.2	'n Kort omskrywing van hiperteks	70

4.7.3.3	'n Kort historiese oorsig oor hiperteks	72
4.7.3.4	Terminologie in verband met hiperteks	75
4.7.3.5	Belangrike aspekte in verband met hiperteks	77
4.7.3.6	Die struktuur van hiperteksdokumente	78
4.7.3.7	Die lees van hiperteksdokumente	78
4.7.3.8	Die ontwikkeling van hiperteksdokumente	79
4.7.3.9	Hiperteksprogramme	80
4.7.3.10	Toepassings van hiperteks	83
4.7.3.11	Probleme in verband met hiperteks	88
4.8	'N GEÏNTEGREERDE OMGEWING VIR DIE VERWERKING VAN TAAL	91
5.	DIE LOGISTIEK VAN DIE VERWERKING VAN NATUURLIKE TAAL	93
5.1	INLEIDING	93
5.2	STELSELANALISE	93
5.3	NETWERKVORMING	97
5.4	TEKSKODERINGSTANDAARDE	98
5.4.1	Inleiding en agtergrond	98
5.4.2	Verskillende tekstkoderingstandaarde	101
5.4.2.1	Eerste pogings	102
5.4.2.2	Moderne tekstkoderingstandaarde	102
5.4.3	Opsommende opmerkings	106
6.	DIE KEUSE VAN APPARATUUR	107
6.1	ALGEMEEN	107
6.2	DIE KEUSE VAN 'N APPARATUUROMGEWING	108
6.2.1	Betrokkenheid	108
6.2.2	Rekenaaromgewings	108
6.2.3	Oorwegings by die keuse van 'n rekenaaromgewing	110
6.3	MEEGAANDE TOERUSTING	112
6.3.1	Kleurgrafika	112
6.3.2	Optiese skyf	113

7.	DIE BESKIKBAARHEID VAN PROGRAMMATUUR	114
7.1	DIE OPDRAG	114
7.2	DIE PROBLEEM	114
7.3	DIE PLAASLIKE ONTWIKKELING VAN PROGRAMMATUUR	115
7.3.1	Belemmerende faktore	116
7.3.2	Die strategiese belang van plaaslike ontwikkeling	117
7.3.3	Alternatiewe vir ontwikkeling	118
7.4	KUNDIGHEID EN ADVIES	118
7.5	SAMEVATTING	119
8.	SAMEVATTENDE OPMERKINGS EN AANBEVELINGS	120
9.	BIBLIOGRAFIE	124
10.	BYKOMENDE LITERATUURLYS	134
11.	SAKEREGISTER	136

HOOFSTUK 1: INLEIDING

1.1 RAAMWERK VAN DIE ONDERSOEK

Die ondersoek waaroor hierdie verslag handel (die PROLANG-ondersoek) 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 natuurliketaal-verwerking te ondersoek. Die oorsprong van die besluit om die navorsing te onderneem, was enersyds bevindings 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-ondersoek is dit onderverdeel in sewe deelterreine, soos weerspieel in die titels van die sewe verslae wat nou verskyn (die verslae se noemname staan in hakies):

- Die rekenarisering van terminografiese prosesse (TERMNET)
- Die rekenarisering van leksigrafiese prosesse (WORDNET)
- Rekenaargesteunde vertaling (TRANSNET)
- Rekenaartoepassings in die taalwetenskap (TEXTNET)
- Kunsmatige intelligensie en die prosessering van die natuurlike taal (AILANG)
- Rekenaarfasiliteite vir die hantering van taal (PROLANG)
- Gerekenariseerde taaldokumentasie-databasisse (DOCNET)

Die sewe deelterreine hou met mekaar op verskillende wyses noue 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 taalontleder (*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 rekenaartocpassings in die taalwetenskap (die TEXTNET-verslag) enersyds sprake van die rekenaar in diens van taalnavoring en andersyds bespreking van taalnavoring in diens van gerekenariseerde taalverwerking.

Die aard van die LEXINET-onderzoek, 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 en werkwyse

Die doel van die PROLANG-onderzoek was om die omgewing van programmatuur en apparatuur wat geskik is vir die verwerking van natuurlike taal te ondersoek.

Hiermee saam sou ook die probleme aan die orde kom wat verbonde is aan die beskikbaarstelling van programmatuur vir die soort navorsing en implementering wat in die ander verslae behandel word.

Die tweeledigheid van die eerste aspek van die opdrag moet dadelik uitgewys word: by "natuurliketaal-verwerking" kan dit enersyds gaan oor die navorsing oor taal en andersyds oor die implementering op 'n groot verskeidenheid gebiede (ook alledaagse toepassings) van prosesse waarin die rekenaar die taalgegewens manipuleer.

In die ander verslae oor die verskillende deelterreine is telkens aandag gegee aan bepaalde gebiede, byvoorbeeld kunsmatige intelligensie (Roux, 1988). In dié verslae het dit uiteraard in die eerste plek gegaan oor sekere teoretiese aspekte en oor 'n verankering van daardie aspekte binne bepaalde teorieë van die algemene linguistiek, asook oor toepassings waarin dit uiting vind. Noodwendig moes daar in dié verslae ook aandag gegee word aan rekenaarprogramme wat ter sprake kom in verband met daardie betrokke toepassingsgebied (of van die databanke wat daardeur opgebou is of geoperasionaliseer

word). Die PROLANG-verslag wil nie oorkoepelend aandag gee aan al die programme wat in die deelgebiede behandel word nie. Dit wil veeleer

- die ander ondersoeke aanvul met agtergrondinligting ten opsigte van rekenaartoerusting (apparatuur en programmatuur) wat daar as vanselfsprekend aanvaar is of nie direk ter sake was nie en
- bykomende inligting verskaf oor toepassings en toepassingsgebiede wat buite die gebied van die ander verslae val (alhoewel dit ten nouste daarmee mag saamhang) en wat 'n plek moet kry in 'n verslag waarin die hele wye gebied van die verwerking van natuurlike taal ter sprake is.

Dit kan in hierdie verslag ook nie gaan oor 'n beskrywing van die presiese, huidige stand van programme waarin taal geoperasionaliseer word of wat vir navorsing oor taal gebruik kan word nie. Die mark ontwikkel te vinnig om vir hierdie moontlikheid voorsiening te maak: nuwe produkte kom tot stand en bestaendes word uitgebrei om vir nuwe moontlikhede voorsiening te maak. Dit is veeleer sinvol om die leser in te lig en te oriënteer oor

- die skopus van bestaende programme vir die verwerking van natuurlike taal,
- die soort gebiede waarop hierdie verwerking kan plaasvind,
- die moontlikhede met betrekking tot die verwerking van natuurlike taal wat nog nie in bestaende programme ontgin word nie en ten opsigte waarvan daar in die toekoms sinvolle ontwikkelings verwag kan word,
- tendense van ontwikkeling en
- standaarde wat op dié gebied blyk rigtinggewend te wees.

Baie programme wat nog in 'n ontwikkelings stadium is, kan moontlik binnekort in produksie kom; baie huidige programme kan van die toneel verdwyn, aangesien hulle nie aan die eise van die verwerking van natuurlike taal voldoen nie. Dit is juis belangrik om die keuse van programmatuur vir die verwerking van taal te laat plaasvind op die basis van die tendense van ontwikkeling. In 'n aparte hoofstuk word aandag gegee aan dié tendense; dit sal blyk dat verskeie van dié tendense herken kan word in konkrete toepassings, maar ook andersom, dat die sinvolheid van verskeie ontwikkelings bevraagteken moet word in die lig van dié tendense.

Hierdie verslag wil dus inligting verskaf vir heeltemal uiteenlopende groepe mense, byvoorbeeld

- gewone rekenaargebruikers wat belangstel in die vraag of die programme wat hulle gebruik, reg laat geskied aan die verwerking van natuurlike taal of wat daarin belangstel om in 'n groter mate die moontlikhede van ander/nuwe programme te ontgin,

- navorsers op 'n groot verskeidenheid gebiede wat met tekskorpora te make het en die rekenaar vir dié doël wil aanwend of
- navorsers op die gebied van taal en die letterkunde.

Kortom: dié verslag wil aan almal wat belangstel in die wyse waarop natuurlike taal op of deur middel van die rekenaar hanteer word, die geleentheid gee om hulle eie konsepte in dié verband te verruim en krities in gesprek te tree met onlangse ontwikkelings. Dit geld selfs met betrekking tot 'n oënskynlik voor die hand liggende en neutrale toepassing soos woordverwerking.

1.2.2 Navorsingspan

Die span wat die PROLANG-projek uitgevoer het, het bestaan uit personeellede en vennote van die Eenheid vir Rekenaartoepassings op die Taal en Teks van die Ou Testament aan die Universiteit van Stellenbosch. Hierdie Eenheid het in die afgelope paar jaar, in samewerking met oorsese vennote, heelwat werk in die toepassing van teksverwerkingsprogramme in Suid-Afrika gelewer. Die navorsers wat by die PROLANG-projek betrokke was, is:

Prof. W.T. Claassen (projekleier) (US)

Mnr. F. van Wyk (US)

Dr. R. Morris (programleier) (RGN)

1.3 DIE INTERRELASIE VAN REKENAARTOEPASSINGS WAARBY TAAL TER SPRAKE IS

Daar is soveel aspekte verbonde aan die verwerking van natuurlike taal in of deur rekenaarprogramme dat dit baie moeilik word om 'n "ordenende beginsel" te vind vir 'n verslag van hierdie aard waarin 'n wyë spektrum van rekenaartoepassings ter sprake kom.

In die afbakening van die opdrag is genoem dat die aspekte van implementering en navorsing ten nouste met mekaar saamhang wanneer dit gaan om natuurliketaal-verwerking. Trouens, daar is 'n interessante wisselwerking tussen hierdie fasette: dieselfde prosesse wat in die een geval met die oog op navorsing plaasvind, vind in die ander geval (soms onsigbaar op die agtergrond) plaas in 'n program wat vir 'n heel ander doel geskryf is. Programme wat oënskynlik 'n alledaagse funksie (of 'n spesifieke funksie in die besigheidswêreld) het, kan soms berus op 'n sekere vlak van taalverwerking of kan self op hulle beurt vir die taalnavorsers belangrike instrumente wees met die oog op sy eie navorsing.

Om die verslag in te deel op grond van die vraag of dit oor navorsing of implementering gaan, is dus op grond van hierdie noue wisselwerking nie moontlik nie.

Dit sou ook oorweeg kon word om die verslag te orden uit die oogpunt van die vlak van rekenaarverwerking wat telkens plaasvind, byvoorbeeld die intensiteit of die gesofisti-

keerdheid daarvan, die vraag of dit gaan om die uitwendige manipulerings van gegewens of om betekenis, ensovoorts. Hierdie onderskeidingsbeginsel is baie sinvol, aangesien dit ook die beperkinge van bestaande programme aan die lig laat tree. Dit is egter ook weer onvoldoende in die mate dat daar in bepaalde programme telkens verskeie vlakke van verwerking plaasvind.

Dit is dus duidelik dat die "ordenende beginsel" van 'n verslag soos hierdie een van vele kan wees – afhangende van die uitgangspunt ten opsigte van die materiaal waarmee gewerk word. Dit het geblyk die beste te wees om die veelheid van toepassings te bespreek uit tweeklei gesigspunt, naamlik

- a) die vlak van die verwerking van natuurlike taal en
- b) die soort aksie wat telkens met die taalgegewens uitgevoer word (waarby die bedoeling nie is om weer al die programme te bespreek wat in die ander verslae genoem is nie).

1.4 DIE REIKWYDTE VAN TAAI

In baie besprekings oor die rekenaarmatige verwerking van natuurlike taal word hoofsaaklik aandag gegee aan die produk van die proses van taalgebruik vir die doeleindes van kommunikasie. Hierdie produk het dan gewoonlik alreeds vorm gekry in gesproke of geskrewe teks. Die aspek van betekenis is vanselfsprekend hierby inbegrepe en in 'n groot verskeidenheid toepassings word dus ook aandag gegee aan dié aspek.

Een gebied wat egter in terme van die rekenaarmatige verwerking van natuurlike taal nog te min aandag geniet het, is dié van die proses van die generering of totstandkoming (nie histories bedoel nie) van taal as kommunikasiemedium, wat hier breedweg beskryf kan word as die gebied van die kognitiewe en die epistemologiese, asook die psigodinamika.

Die gebiede van die kognitiewe en die epistemologiese geniet vandag wydverspreid die aandag, hoewel nog nie genoegsaam binne die raamwerk van die rekenaarmatige verwerking van inligting nie. (Vergelyk 3.5 en 4.7.3.5c.) Die gebied van die psigodinamika (veral met betrekking tot die transformasieteorie op die gebied van die psigodinamika) binne die gebied van die rekenaarmatige verwerking van inligting het egter nog nie genoegsaam die aandag geniet nie. Heim (1987) gee baie sistematies aandag aan dié aspek van die psigodinamika vir soverre dit betrekking het op dié gebied van woordverwerking. Die opsteller van hierdie verslag het al dikwels die term "logistiek" gebruik as verwysing na die prosesse wat Heim onder die term "psigodinamika" behandel. Die term "logistiek" word dan gebruik in 'n ander betekenis as dié van die daarstelling en ordening van die uiterlike middele wat vir die uitvoering van 'n bepaalde aksie nodig is. Dit verwys in dié verband eerder daarna dat daar deur die gebruik van rekenaarmatige middele op die kognitiewe en intellektuele vlakke ander aspekte van die haalbaarheid en uitvoerbaarheid van die omgang met (onder andere) tekstmateriaal op die spel kom en dat hierdie aspekte ook 'n invloed het op die resultate van die betrokke prosesse.

Dit is verder duidelik dat hierdie kognitiewe en epistemologiese fasette nie net op die vlak van die generering van teks 'n rol speel nie, maar ook met betrekking tot die hele proses van die menslike omgang met talige materiaal. Waar hierdie prosesse plaasvind by wyse van, of ondersteun word deur, rekenaarmatige prosesse, verdien dit uiteraard ook die aandag binne die raamwerk van 'n ondersoek van hierdie aard.

Trouens, dit kan gesê word dat dit die dringende aandag verdien, aangesien die tussenkoms van die rekenaar die prosesse in so 'n mate versnel of in intensiteit laat toeneem dat dit baie belangrik is dat die volle impak daarvan vroegtydig besef en daarvoor voorsiening gemaak sal word. In verskeie resente studies word hierdie sake aangespreek, byvoorbeeld Capurro (1986) in verband met vakinligting, Dolan (1988) in verband met die epistemologiese en hermeneutiese aspekte verbonde aan toepassings soos hiperteks en Heim (1987) in verband met woordverwerking.

1.5 UITLEG VAN DIE VERSLAG

Op grond van die bogenoemde uiteensetting word die verslag in die volgende vorm aangebied:

- a) In Hoofstuk 2 word aandag gegee aan algemene tendense met betrekking tot die ontwikkeling van programmatuur en apparatuur (en die wyer raamwerk waarbinne dié tendense geplaas moet word).
- b) In Hoofstuk 3 gaan dit oor die verskillende vlakke waarop natuurliketaal-gewens verwerk kan word.
- c) In Hoofstuk 4 word aandag gegee aan die verskillende soorte aksies wat met betrekking tot taalgewens uitgevoer kan word (besonderhede hieroor volg hieronder).
- d) In Hoofstuk 5 word aandag gegee aan enkele aspekte wat betrekking het op die logistiek van die verwerking van natuurlike taal, o.a. aan tekstkodering-standaarde.
- e) In Hoofstuk 6 word enkele opmerkings gemaak oor die beskikbaarheid van programmatuur.
- f) In Hoofstuk 7 word die implikasies vir die keuse van apparatuur kortliks aangedui wat volg uit die soort verwerking wat in Hoofstuk 4 beskryf is.
- g) In Hoofstuk 8 word 'n samevatting en aanbevelings gebied.

Hoofstuk 4 word ingedeel volgens die aksies wat uitgevoer kan word. Die volgende word onderskei, sonder om daarmee voor te gee dat dit omvattend of uitputtend is:

- i) Die generering van teks.
- ii) Die uiterlike versorging van teks.

- iii) Die ordening van teks.
- iv) Die kommentariëring van teks.
- v) Toegang tot teks.
- vi) Teks-aanbieding.
- vii) 'n Geïntegreerde omgewing vir die verwerking van taal.

Sekere fasette wat in die ander verslae behandel is, kan beskou word as 'n dimensie ten opsigte van natuurlike taal en die verwerking daarvan (byvoorbeeld kunsmatige intelligensie) en kan in werklikheid betrekking hê op 'n hele aantal van die onderafdelings van dié hoofstuk.

Die besprekings by die onderskeie hofies is nie ewe lank nie. Dít moet nie as 'n aanduiding van die relatiewe belangrikheid van die betrokke toepassingsgebied beskou word nie. Die toepassings op sommige gebiede is net minder bekend as dié op ander gebiede, behels meer deel-terreine, het meer aandag in publikasies ontvang en verdien in 'n verslag soos dié 'n bespreking in meer besonderhede.

Aan die einde van die bibliografie word 'n verdere lys literatuur verskaf (hoofsaaklik oor hiperteks). Hierdie literatuur is deur die opsteller van die verslag eers verkry nadat die verslag voltooi is en dit kon dus nie in die teks opgeneem word nie. Dit word volledigheidshalwe verskaf vir verdere gebruik deur die leser.

1.6 TERMINOLOGIE

Dit bly 'n probleem om geskikte terminologie in Afrikaans te vind vir die verskillende prosesse betrokke by die verwerking van taal. In 'n poging om wel geskikte terminologie te vind, word dikwels verskillende terme gebruik, wat tot groot verwarring aanleiding kan gee. Die benadering wat hier gevolg word, is om duidelike en veelseggende terminologie te kies. In die meeste gevalle word die bekende Engelse terminologie ook vermeld.

Oor die volgende kan algemene opmerkings gemaak word:

- a) Die term **teks** word dikwels gebruik in plaas van taal. Die bedoeling is dat dit dan om 'n stuk samehangende teks gaan. In gevalle waar dit net gaan om die verwerking van losstaande woorde, word daar uiteraard van woorde gepraat.
- b) Verskeie van die programme wat hier genoem word, hang saam met **woordverwerking**. Die benaming **teksverwerking** word dikwels verkies, op grond daarvan dat dit dan eerder gaan oor die verwerking van groter tekste. Op grond van die feit dat dit in Afrikaans algemeen ingang gevind het, word die benaming **woordverwerking** in hierdie verslag verkies.
- c) Die term **operasionalisering** word soms gebruik ter aanduiding daarvan dat 'n bepaalde proses met teks plaasvind. Dié term word gebruik by gebrek aan 'n

meer geskikte en veelsydige term: in sommige gevalle kan wel gepraat word van die verwerking, die aanbieding of die lees van 'n teks. Sekere programme (byvoorbeeld hiperteksprogramme) maak egter voorsiening vir 'n verskeidenheid aksies ten opsigte van die taalgegewens, afhangende van die aksie van die gebruiker. Die term operasionalisering blyk 'n nuttige en neutrale keuse te wees, aangesien dit dan telkens ruimte laat vir dit wat tipies deur 'n bepaalde program gedoen word.

HOOFSTUK 2: ALGEMENE TENDENSE IN DIE ONTWIKKELING VAN PROGRAMMATUUR EN APPARATUUR

2.1 INLEIDING

Die potensiaal van die rekenaar vir toepassings op die gebied van natuurlike taal is alreeds vroeg ná die ontwikkeling van die rekenaar ingesien. Kort na die Tweede Wêreldoorlog is daar begin met die verwerking van nie-administratiewe data met behulp van die rekenaar: die eerste werklik omvangryke projek met betrekking tot taalgegewens was dié van Roberto Busa, waarin 'n konkordansie op die werke van Thomas Aquinas opgestel is. Vir besonderhede vergelyk Hockey (1980, p. 15 en ander). Busa se projek word bespreek in verskeie bydraes in *Computers and the Humanities*. Dit was net te verwagte dat die rekenaar ook algou vir al hoe meer dergelike toepassings, asook vir ander toepassings op teksmateriaal, gebruik sou word.

Gebruikers op die gebied van taal- en teksmateriaal het die vinnig ontwikkelende moontlikhede aangegryp en in die belang van hulle eie navorsingsgebiede aangewend. Dit geld nie net van taalnavorsers nie, maar in toenemende mate ook van die gebruikers op ander terreine wat noodwendig met teksmateriaal moes omgaan omdat dit so universeel en strategies is en die kern vorm van toepassings op talle gebiede.

Oor die strategiese belang van teks, vergelyk Gill & Woll (1986, pp. 137-141):

Text is the symbolic representation of intelligence. Texts put in tangible form the content of different kinds of human expression: thoughts, ideas, concepts, facts, figures, and fiction. In other words, text makes it possible to store and disseminate human expression.

Op dié wyse het daar talle programme vir die verwerking van natuurlike taal tot stand gekom wat nie aan die eie aard van taal laat reg geskied nie en wat nie vanuit 'n taalkundige perspektief gekonsipieer is nie. Gebruikers het baie van hierdie programme intensief begin gebruik en deur die (beperkte) moontlikhede daarvan gekondisioneer geraak. Ander het in die oorskatting van hulle eie vermoë self probeer om die geskikte program vir hulle toepassings te skep – en het daardeur in die slaggat getrap dat hulle nie die gevorderde vlak van programmering kon bereik wat nodig is vir sommige van die gesofistikeerde vlakke van toepassings wat vandag nodig is nie.

Hoe positief die ontwikkelings met betrekking tot die gebruik van die rekenaar vir die verwerking van natuurlike taal gedurende die eerste paar dekades ook al aangeslaan moet word, neem dit nie weg daarvan nie dat, breedweg gesproke, baie gebruikers op sleeptou geneem is deur die ontwikkeling van die tegnologie en die rekenaarprodukte wat op die mark beskikbaar geraak het – sonder die nodige mate van besinning oor al die aspekte

betrokke by hierdie driehoek van (i) tegnologie, (ii) objek van verwerking en (iii) gebruikersgroep ten bate van wie die verwerking plaasvind.

Dit is eers in die afgelope dekade dat daar toenemend op 'n fundamentele vlak besin word oor vroe soos

- die rol van die rekenaar in die gemeenskap, en met betrekking tot die verwerking van inligting
- die implikasies van dergelike verwerking van taal vir die gebruikers
- die mate waarin die verwerking van taal wel laat reg geskied aan die eie aard van taal en inpas by die inligtingsoekende gedrag van die gebruiker (*information seeking behaviour*)
- die doeltreffendheid van die verwerking, gemeet aan verskillende standaarde
- die identifisering van die toekomstige moontlikhede (en behoeftes) met betrekking tot die verwerking van natuurlike taal
- die dinamika van verandering met betrekking tot nuwe en totaal andersoortige toepassings op byvoorbeeld die gebied van taal en die verwerking daarvan – wat temeer belangrik is in die lig van die baie vinnige verandering op dié gebied.

Dit blyk dus van die grootste belang te wees vir 'n verslag van dié aard dat die beoordeling van programme, toepassingsgebiede en moontlikhede nie op die losse sand van die ontwikkelings van die dag gebaseer sal wees nie, maar binne 'n raamwerk van die analise van tendense en patrone van ontwikkeling gegrond sal wees.

Ook hierdie tendense kan net weer binne 'n wyer raamwerk van analise of tipering van die moderne gemeenskap betekenis kry. Hierdie verslag gaan van die standpunt uit dat die verskynsel wat as "die inligtingsera" beskryf word, die raamwerk behoort te vorm van enige besinning oor die tagtigerjare en die jare daarna (vergelyk Naisbitt, 1984; Claassen, 1986). Wanneer daar gepraat word oor tendense met betrekking tot programmatuur en apparatuur, moet daar dus ook gepraat word oor tendense met betrekking tot die inligtingsera in die geheel, wat vanselfsprekend ook weer die behoeftes en persepsies van die gebruikers van inligting raak.

Hierdie aanname neem nie weg daaraan dat daar ook ander belangrike aspekte is wat die moderne era help bepaal en definieer nie. Die inligtingsera het ook nie in dieselfde mate deurgebreek na alle gemeenskappe nie – in baie gemeenskappe sou gesê kon word dat die "industriële era" eintlik nog die dominerende is. Ander tendensontleders skryf 'n groter betekenis toe aan fasette soos die nuwe ekologiese bewussyn en die sogenaamde *new age*-beweging; in Europa is die ekologiese bewussyn en beweging uiteraard van die grootste belang (Gerken, 1988). Dit blyk toenemend dat die hele moderne era nie deur een enkele aspek of beweging getipeer kan word nie en dat daar dus gereken moet word met 'n aantal

perspektiewe wat mekaar aanvul of soos sirkels wedersyds oorvleuel (Gerken, 1988). Tog is dit so dat die aspekte wat die daarstelling, bemiddeling en benutting van inligting en teoretiese kennis raak, dié aspekte is wat deurgaans op die spel is. Om van die huidige era as die "inligtingsera" te praat, hoef nie te geskied met uitsluiting van die erkenning van die ander bepalende faktore nie.

2.2 ALGEMENE TENDENSE

- a) Daar is 'n oormaat van inligting beskikbaar, wat – juis op grond van die oormaat daarvan – vir die gebruiker nie meer 'n bate is nie, maar 'n las en 'n bedreiging; dit skep onsekerheid en belemmer effektiewe besluitneming. Doeltreffende toegang tot en omgang met inligting beteken vir die gebruiker nie net hoeveel inligting daar oor 'n bepaalde saak ingewin kan word nie, maar eerder hoeveel tersaaklike/relevante inligting. Om die rekenaar net te gebruik as 'n meganisme om as't ware die bladsye van die boek om te blaai, help die gebruiker nie. Daar word dus toenemend meganismes benodig wat die gebruiker in staat stel om
 - by die regte inligting uit te kom
 - op 'n wyse wat by sy behoeftes aanpas.
- b) Daar is 'n duidelike aksentverskuiwing vanaf aanbod na aanvraag en seleksie. Daar is te veel inligting om alles te kan verwerk en die klem val al hoe meer op seleksie, op die verkryging van die relevante inligting. Die meganismes sal nou ontwikkel en verfyn moet word om dit moontlik te maak om op 'n meer gesofistikeerde wyse toegang tot die data te verkry en daarmee saam ook toegang tot net die gewenste data. As die inligting nou in taal gegiet is, ontstaan die vraag onmiddellik hoe die taal deur die rekenaar hanteer moet word om die beste toegang moontlik tot die relevante inligting te verkry. Trouens, dit is al hoe meer wenslik dat die gebruiker in staat gestel sal word om sy navraag in gewone omgangstaal te doen, eerder as dat hy 'n stel kunsmatige terminologie moet aanleer om by die gewenste resultaat uit te kom. Dit is juis die aksentverskuiwing vanaf aanbod na seleksie wat meebring dat ons nie maar kan klaarkom met 'n stelsel wat as't ware die bladsye van die teks elektronies omblaai nie, maar 'n stelsel moet hê wat die beste toegang tot die mees relevante materiaal bied.
- c) Die inligtingsera het sy eie momentum en dinamiek, wat meebring dat die groot aantal produkte van hierdie era net weer deur die vermoë van die rekenaar onder beheer gebring en gehou kan word. Die gebruiker se probleem om by die regte inligting uit te kom, kan slegs deur 'n rekenaaroplossing aangespreek word.
- d) Teks gaan toenemend op die rekenaar geprosesseer word deur 'n ander groep mense as dié wat dit tot nou toe hanteer het. In plaas daarvan dat – soos meesal vroeër – teks eers geskryf word en daarna deur byvoorbeeld sekretariële personeel ingetik word, gaan dit toenemend deur die gebruiker self op die rekenaar gegenereer word – en dan is dit juis belangrik dat al die middele op die rekenaar beskikbaar sal wees om die persoon wat teks genereer te ondersteun en natuurlik ook weer om die persoon wat inligting moet vind, te kan ondersteun om in terme van sy natuurliketaal-prosesse daarmee te kan omgaan. In plaas daarvan dat net die

besonder kundige mense teks op die rekenaar manipuleer, gaan baie meer "gewone" mense in alledaagse beroepe dié teks manipuleer en gaan hulle al hoe hoër eise stel in terme van die vlak van verwerking van die taalgewens.

- e) Die inligtingsera is 'n tyd van veelvuldige opsies. Dit was John Naisbitt (1984) wat aangedui het dat ons aan die begin van die dekade begin wegbeweeg het van die òf/òf situasie en al hoe verder begin beweeg het in die rigting van 'n omgewing van veelvuldige opsies. Ten opsigte van inligting beteken dit dat inligting in 'n groot verskeidenheid vorms of media (en nie net in verskillende boekformate nie) hanteer of verpak kan word, elkeen gerig op 'n bepaalde behoefte. Dit het implikasies vir die hele "inligtingverpakingsbedryf".

Daar is persone wat hierdie verskuiwing as so fundamenteel wil beskou dat hulle dit as 'n paradigmatverskuiwing in die uitgewersbedryf wil beskou. Die volgende aanhalings uit Zoelick (1987a) se inleidende hoofstuk tipeer dié standpunt:

Traditionally publishers have rushed to format information on a page, and in the process, fenced the information in.

Dan gee die outeur egter ook die oplossing:

The solution to these problems is at once simple and radical: Reverse the process. Build the database first and then extract the information needed for printing from the database. Reversing the process not only makes it easier to build databases, but makes it possible to publish using media other than paper.

Hy waarsku egter ook dat hierdie proses nie maar net 'n meganiese proses of 'n opdrag aan 'n operateur kan wees nie:

Like all revolutions, this paradigm shift to producing electronic databases rather than printed products requires new tools and new ways of thinking....

Die inligtingsgebied (en dus ook die tradisionele boekwêreld) is een van veelvuldige opsies, eerder as net die strenge letters-op-papier opsie wat as tipies van inligting in vorige dekades beskou kan word. Dit geld trouens ook op soveel ander gebiede, soos reeds in 1982 aangedui deur John Naisbitt in sy bekende werk *Megatrends* (1984, herpublikasie van die 1982-uitgawe), met name in die hoofstuk *From either/or to multiple option* (pp. 231-247). Die gewone verbruiker van inligting is soms onbewus van hoe dieselfde inligting telkens net op 'n ander wyse vorm kry in die medium waarin hy dit toevallig onder oë kry.

Daar is nou telkens ruimte vir 'n beslissing oor die mees geskikte formaat waarin inligting verpak kan word. Van persone betrokke in dié bedryf word nou genoegsame kundigheid verwag om keuses te doen wat aan die media, die inligting en die gebruikersbehoefte reg laat geskied.

- f) Alle rekenarisering is nie noodwendig tipies deel van die inligtingsera nie. Daar is baie toepassings wat eintlik maar net daarop neerkom dat ou perspektiewe gereke-

nariseer word, eerder as dat die kennisgebiede en die toepassings wat daarmee saamhang, grondig deurdink en inligtingkundig suiwer verantwoord word.

- g) By die beplanning van produkte, toepassings, ensovoorts in die inligtingsera moet daar baie duidelik beplan word in terme van die ontwikkelings wat gaan kom. Die leeftyd van baie produkte is maar kort en die kompetisie is sterk. Dit is nou meer as ooit in die verlede belangrik dat die tendense van ontwikkeling duidelik geïdentifiseer sal word en dat daar nou reeds gevra sal word na die aard van toekomstige gebruikersbehoefte, stelsels, toepassings, ensovoorts.
- h) Die navolging van standarde met betrekking tot toerusting en die verwerking van materiaal van enige aard het belangriker geword as ooit tevore. Dit hang saam met die vinnige ontwikkeling van rekenaartegnologie, die besef van die voorlopigheid van huidige toepassings en die besef van die noodsaaklikheid dat data gemaklik uitruilbaar moet kan wees. Met die oog op die grootste moontlike mate van oordraagbaarheid van toepassings en data, asook die uitskakeling van tydrowende herhalende menslike inset in die verwerking van die data, is dit baie belangrik om te hou by internasionaal erkende standarde.

Dit gaan hier oor standarde gestel deur amptelike, normerende liggame of de facto-standarde wat tot stand gekom het deur die eksplisiete of stilswyende aanvaarding van die gebruikerskorps. Voorbeelde is PostScript as bladsybeskrywingstandaard (vergelyk 5.4.1) en SQL as databasisstandaard (vergelyk Shaw, 1988).
- i) Ten spyte van die desentralisering van rekenaarverwerking (2.3a hieronder) gebeur dit nou weer baie dikwels dat daar 'n sentralisering ten opsigte van die konseptualisering en die logistiek in die geval van groter projekte nodig is. Die kompleksiteit en omvang van baie van dié projekte (hetsy programontwikkeling of die ontwikkeling van omvattende inligtingstelsels) word dikwels grootliks onderskat. Die insette van die afsonderlike persone wat betrokke is, moet vanuit 'n sentrale punt beplan en begelei word ten einde te verseker dat daar uiteindelik 'n geordende en werkende stelsel tot stand kom (vergelyk byvoorbeeld Hughes 1987d, 1988a oor omvangryke programme).
- j) Vir sinvolle en relevante projekte waarin van die rekenaar gebruik gemaak word, word dikwels baie omvattende befondsing benodig. Hierdie aspek hang saam met (i) hierbo. In die geval van teksmateriaal kan die koste vir die datavaslegging besonder hoog wees (vergelyk 2.3 en Hughes, 1987d; 1988a).
- k) Daar is 'n besonder goeie balans nodig tussen die gebruik van bestaande programmatuur en die ontwikkeling van nuwe en doelgerigte programmatuur. Programmatuur het te omvattend en kompleks geraak dat daar telkens uit die grond uit ontwikkel kan word. Dit is baie meer sinvol om eerder so ver moontlik van bestaande programmatuur gebruik te maak en dan net die ontbrekende komponente self te skep. So 'n benadering maak dit des te meer noodsaaklik om standarde met betrekking tot toerusting en die hantering van materiaal na te volg (vergelyk (h) hierbo).

2.3 TENDENSE MET BETREKKING TOT APPARATUUR

- a) Die verskuiwing van hoofraamrekenaars na kleiner rekenaars, veral na mikrorekenaars, kan sekerlik as die belangrikste faktor vir die vinnige ontwikkeling van die rekenaarmatige verwerking van natuurlike taal beskou word. Die rekenaartoneel is lank oorheers deur hoofraamrekenaars en die eerste groot toepassings het dus noodwendig op dié rekenaars plaasgevind. Namate meer persone toegang tot mikrorekenaars het en dié se verwerkingspoed en kapasiteit so dramaties toegeneem het, het daar ook 'n grootskaalse verskuiwing in die rigting van kleiner rekenaars plaasgevind.

Hiermee saam het die logistiese opset van die gebruik van rekenaars grootliks verander. Met hoofraamrekenaars gaan tipies (hoewel nie noodwendig nie) gepaard verskynsels soos:

- i) Die konsentrasie van die aktiwiteite aan sekere (groter) sentra waar die nodige toerusting en infrastruktuur beskikbaar is en die nodige ondersteuning aan die persone gegee kan word wat dié soort toepassings ontwikkel.
- ii) Die sentralisering van data-versamelings, uit hoofde van
 - die feit dat die materiaal in elk geval net op hierdie plekke sinvol verwerk kon word (weens die spesifieke formate van die materiaal, ensovoorts) en
 - die ywer om die voorsprong te behou, wat dikwels weer saamhang met die feit dat die data teen hoë koste vasgelê is en programmatuur teen hoë koste deur skaars programmeerders ontwikkel is.

Die motiewe vir sentralisering moet nie (in alle gevalle) beskou word as kleinlike of selfsugtige motiewe nie. Die algemene gees waarin die materiaal versamel en hanteer is, was veeleer mede gevorm deur die gees van rekenaar-implementering in die algemeen – in dié geval die sentraliserende benadering wat tipies saamgehang het met die gebruik van hoofraamrekenaars.

- b) Die proliferasie van rekenaars, wat in feite neerkom op 'n demokratisering van die rekenaar. Die miljoene mikrorekenaars wat in die plek van die groter rekenaars van vroeër gekom het, het daartoe gelei dat die potensiaal en vindingrykheid van 'n baie groot aantal gebruikers ontgin is. By baie gebruikers is daar 'n toenemende reaksie teen die komplekse gebruikers-onvriendelike kodeerstelsels van vroeër. Daar word veeleer gesoek na stelsels wat uiting gee aan die gewone, menslike prosesse van die soek na inligting.
- c) Nuwe moontlikhede het ontwikkel vir die rekenaarmatige verwerking van uitermate groot hoeveelhede data en selfs die mees opwindende ontwikkelings van vandag gaan nog deur dié van die toekoms oorskadu word. Dit geld vir harde skywe vir die werklike stoor van data tydens verwerking, asook vir die gebruik van optiese skywe wat die gemaklike berging en oordrag van baie groot dataversamelings moontlik maak. Dit hou voordele na twee kante toe in:
- i) Vrae kan nou gevra word ten opsigte van baie groot datakorpora.

- ii) Versamelings data wat betrekking het op komplekse en geïntegreerde kennisgebiede kan nou gemaklik op optiese skyf byeengebring word. Juis omdat sulke groot volumes data nou beskikbaar gestel kan word, is dit so belangrik dat dié soort programmatuur ontwikkel sal word wat nodig is om die gebruiker toegang tot die materiaal te gee op 'n wyse wat werklik inpas by sy inligtingsoekende gedrag.

Ontwikkelings met betrekking tot CD-ROM-skywe en (her)skryfbare optiese skywe geskied so vinnig dat die mees resente inligting net uit lopende rekenaartydskrifte verkry kan word. Die tydskrif **CD-ROM Review** is spesifiek gewy aan ontwikkelings op dié gebied. Dit verskaf ook inligting oor van die nuutste tegnieke, toerusting en beskikbare skywe. Vir 'n baie goeie oorsig oor die meeste tegniese aspekte betrokke by die tegnologie, vergelyk die verskillende bydraes in Lambert & Ropiequet (1986) (o.a. ook oor die moontlikhede van die medium CD-ROM op verskillende gebiede) en Ropiequet et al. (1987). 'n Nuttige resente lys CD-ROM-skywe op 'n verskeidenheid gebiede kan gevind word in Faxon (1988). 'n Skyf wat vir 'n groot verskeidenheid gebruikers baie nuttig kan wees, is **Microsoft Bookshelf**, wat tien groot naslaanbronne bevat, o.a. **American Heritage Dictionary**, **Roget's Electronic Thesaurus**, **The World Almanac** (Faxon, 1988). Helgerson (1988) gee 'n oorsig oor CD-ROM-skywe (deels in ontwikkeling) vir navorsing in die geesteswetenskappe. Hughes (1987d) verstrek inligting oor die besonder groot **Thesaurus Linguae Graecae**-projek: op een CD-ROM-skyf word die hele Griekse literatuur vanaf ongeveer 750 v.C. tot 600 n.C. beskikbaar gestel, ongeveer 60 miljoen woorde. 'n Ander skyf bevat 'n versameling data wat van besondere belang is vir navorsers op die gebiede Ou en Nuwe Testament, naamlik **PHI-CAT** (Hughes, 1987d).

- d) Die verwerkingspoed van selfs die gewone mikrorekenaars neem hand oor hand toe. Dit kan dus verwag word dat die gebruiker al hoe groter krag tot sy beskikking sal hê in al hoe kleiner rekenaars.
- e) Hoë resolusie grafika raak toenemend belangrik, nie net soos vroeër in spesiale werkstasies vir byvoorbeeld ingenieurstoepassings nie, maar ook vir rekenaars wat vir kantoor- of tuisgebruik bestem is. Die groter klem op grafika dra beslis ook daartoe by om rekenaars by meer mense populêr te maak (vergelyk verder 2.4g).

2.4 TENDENSE MET BETREKKING TOT PROGRAMMATUUR EN KUNDIGHEID

Baie tendense ten opsigte van programmatuur word bepaal deur die bogenoemde algemene tendense en dié ten opsigte van apparatuur. Die volgende kan uitgesonder word:

- a) Die ontwikkeling vanaf privaatprodukte na markprodukte. Die meeste aanvanklike programme wat spesifiek gerig is op die verwerking van natuurlike taal is ontwikkel vir die eie toepassings van persone of groepe en dié programme kon net deur spesiale reëlins aan ander persone beskikbaar gestel word. Die groter mikrorekenaarmark en die mate van standaardisering wat teweeggebring is deur die wye

gebruik van die MS-DOS-bedryfstelsel, het meegebring dat dit vir firmas lonend geword het om sinvolle produkte vir die verwerking van natuurlike taal op die mark te bring. WordCruncher is hiervan 'n goeie voorbeeld. Die gevaar verbonde aan dié tendens is dat sommige kleiner en baie beperkte programme wat werklik nie konseptueel of programmaties goed deurdag is nie, nou ook op die mark kom en dat die gebruiker se keuse moeiliker raak.

- b) Baie meer klem word in die ontwikkeling van programmatuur gelê op die gebruikersraakvlak (of: gebruikerskoppelvlak – *user interface*). Teenoor die gebruikers-onvriendelike raakvlak van baie programme in die eerste fases van rekenaar-verwerking, is daar nou 'n groter klem daarop dat die gebruiker op 'n gemaklike wyse met die program moet kan omgaan. Voorbeelde hiervan is die gebruik van die muis en ingebedde spyskaarte (wat voorkom dat die gebruiker al die opsies gelyktydig sien). Hierdie gemak kom nie net eenvoudig neer op 'n luukse nie. Dit gaan eerder daaroor dat die bevatlikheid van die gebruiker, asook sy tipiese wyse van omgang met inligting, in ag geneem word, maar nog baie ontwikkeling in dié verband is nodig (vergelyk verder 3.5 en 4.7.1).
- c) Die konseptuele basis by die verwerking van natuurlike taal word steeds belangriker. Die konsepte onderliggend aan sommige programme is so nuut (in terme daarvan dat dit op 'n rekenaar uitgevoer kan word) dat die gebruiker dit aanvanklik vreemd kan vind, terwyl dié konsepte moontlik op 'n baie meer diepliggende vlak met die eie aard van die kennisverwerkings- en kennisverwerwingsproses kan saamhang. Trouens, dit mag van die gebruiker vereis word om te breek met die praktyke en konsepte wat eintlik saamgehang het met of bepaal is deur uitgediende en beperkende tegnologie. Dit is eers wanneer die gebruiker die program begin gebruik en ervaar dat hy 'n nuwe greep op die inligting kry dat hy self daarvan bewus word dat sy vroeëre omgang met inligting deur verouderde tegnologie bepaal en beperk was. 'n Goeie voorbeeld in dié verband is die ontwikkeling en implementering van hiperteks, wat eers moontlik gemaak is deur die ontwikkeling en sameloop van bepaalde fasette van die tegnologie (vergelyk 4.7.3).
- d) By die beoordeling van nuwe programmatuur moet die gebruiker telkens die oog oophou vir verrassende nuwe moontlikhede, wat moontlik neerkom op 'n "kwantumsprong" in terme van dit wat vroeër gebruiklik was. Dié soort toepassings kan dan op hulle beurt weer perspektiewe open ten opsigte van verdere rigtings en moontlikhede van ontwikkeling. Die konsepte waarvan die haalbaarheid in die eerste hiperteks-toepassings getoon is, het inderdaad die veld geopen vir 'n totaal nuwe fase van besinning oor die inligtingkundige (en hermeneutiese) aspekte ten opsigte van alle ander media wat vir die verpakking van inligting gebruik kan word (vergelyk Dolan, 1988).
- e) Daar is 'n baie duidelike verskuiwing na intydse stelsels – teenoor die ouer stelsels waarby die gebruiker 'n proses aan die gang moes sit en moes wag op gedrukte resultate. WordCruncher is 'n intydse stelsel vir die opstelling van konkordansies, waarby die gebruiker die greep op die materiaal dinamies volgens sy veranderende behoeftes kan beïnvloed (vergelyk 4.4.3.3c). Die tegnologie is vandag beskikbaar wat hierdie soort omgang met die materiaal moontlik maak.

- f) Stelselontleding by taaltoepassings raak al hoe belangriker. Die verskuiwing vanaf die hoofraamvlak na mikrorekenaars (2.4a hierbo) en die feit dat al hoe meer persone self die benodigde programmeertale aanleer en hulle eie programme ontwikkel, bied sekerlik groot voordele. Dit hou egter ook die nadeel in dat die programmering nou uit die hande is van programmeerders wat (gewoonlik) goeie stelselontleding geleer het. Dit hou die gevaar in dat toepassings ontwikkel word waarin daar nie behoorlike stelselontleding plaasvind nie (vergelyk 5.2).
- g) Programme maak in toenemende mate en met groot verbeelding gebruik van grafika. Dit geld ook vir gewone toepassings waarin byvoorbeeld grotendeels met teks gewerk word. Die integrasie van teks en grafika in programme word toenemend belangrik, wat ook weer uit die oogpunt van die rekenaar-gebruiker-raakvlak baie belangrik is (vergelyk 3.5).

HOOFSTUK 3: VLAKKE VAN DIE HANTERING VAN TAAL MET BEHULP VAN DIE REKENAAR

3.1 INLEIDING

In die inleidende hoofstuk het dit geblyk dat een van die gesigspunte waaronder die verwerking van natuurlike taal georden kan word, die vlak van verwerking is, dit wil sê die mate van gesofistikeerdheid van die aksies wat met betrekking tot die taalgegewens uitgevoer word. By 'n bespreking van die aksies wat met die gegewens uitgevoer word, kom hierdie vlakke dan telkens weer ter sprake (vergelyk Hoofstuk 4).

Wanneer daar gepraat word van die rekenaarmatige verwerking van natuurlike taal word daar dikwels gebruik gemaak van baie vae en oorkoepelende benamings, byvoorbeeld "rekenaartoepassings op taal". Hierdie vae benamings verberg soms 'n grootskaalse onkunde met betrekking tot die werklike omvang en aard van die verwerking van taalgegewens wat telkens in die verskillende toepassings plaasvind.

Dit is die bedoeling met dié hoofstuk om 'n aanduiding te gee van die verskillende vlakke waarop die rekenaar by die verwerking van taalgegewens betrek kan word.

Die rekenaar kan die materiaal op talle verskillende vlakke verwerk, waarvan sommige baie meer gesofistikeerd as ander is. Trouens, dit kan gesê word dat die verwerking in sommige gevalle baie eenvoudig of naïef is. Dit gaan soms net oor die sortering van woorde, asof dié net sowel die name van produkte sou kon wees. In ander gevalle is die gebruiker van die rekenaar juis besig om deur taal heen as't ware in gesprek te tree met die rekenaar en vorm die woorde die raakpunte van dié "kommunikasie". In nog ander gevalle word die reaksie van die rekenaar bepaal deur afleidings wat gemaak word uit die taalgebruik van die gebruiker, wat op 'n baie gesofistikeerde vlak kan lê.

Die probleem ontstaan nou dat die werklike aard van die verwerking van taal op die rekenaar vir die gewone, nie-taalkundige gebruiker maklik verdoesel kan word deurdat daar 'n vinnige en behendige verwerking van die gegewens plaasvind wat die indruk laat dat daar verwerking op 'n gesofistikeerde vlak plaasgevind het. Ter illustrasie: indien iemand gebruik maak van 'n woordverwerkingsprogram waarin daar ook 'n spelkontrole-program of 'n ingeboude tesourus vervat is, kan dit maklik gebeur dat hy van mening is dat hy op 'n baie gesofistikeerde vlak met taalgegewens omgaan en kan hy onder die indruk verkeer dat die rekenaar die taalgegewens "intelligent" hanteer; immers, deur 'n ingeboude tesourus te hanteer, is hy tog besig om betekenisdraende elemente te manipuleer. In werklikheid lê hierdie toepassing egter op 'n totaal ander vlak as wanneer betekenis by wyse van 'n proses van die intelligente verwerking van die taalgegewens plaasvind.

Hiermee word nie geïmpliseer dat toepassings waarin dit gaan om die "meganiese" hantering van taalgegewens minder belangrik of minder nuttig is as dié waarin dit gaan oor die "intelligente" hantering van taal nie. Trouens, taal is so 'n komplekse stelsel dat dat

die meeste rekenaartoeepassings op taal wat pasklaar bruikbaar is, grotendeels gebaseer is op die meganiese verwerking van die gegewens op so 'n wyse dat dit aan die gebruiker die geleentheid gee om sy eie taalvermoë optimaal te gebruik. Die rekenaar speel net 'n ondersteunende rol, maar die gebruiker ervaar dit so asof die rekenaar die grootste deel van die verwerking doen.

Hierdie twee aspekte hang in die meeste gevalle baie nou met mekaar saam en vul mekaar aan: die gebruiker se vermoë tot die intelligente verwerking en interpretasie van taalgegewens kom eers dan werklik tot sy reg wanneer hy voorsien word van materiaal wat meganies op so 'n wyse verwerk en aangebied word dat hy dit gemaklik kan hanteer.

In dié hoofstuk word vervolgens aandag gegee aan 'n aantal vlakke met betrekking tot die hantering van taalgegewens. Die bedoeling is net om die toneel voor te berei vir die bespreking van aksies met betrekking tot taalgegewens wat in Hoofstuk 4 volg.

3.2 DIE GENERERING VAN TEKS

In die inleiding is daarop gewys dat daar noodwendig ook aandag gegee moet word aan die proses van die generering van taal, en nie net aan die produk nie. Dit is immers moontlik dat die feit dat rekenaarmatige prosesse betrek word, 'n invloed op die prosesse van die generering van teks kan hê of dit selfs wesenlik kan beïnvloed. Die rekenaarmatige prosesse waarvoor dit hier gaan, is hoofsaaklik dié van woordverwerking.

Wat 'n bespreking van die vlak van verwerking van natuurlike taal in verband met woordverwerking bemoedlik, is die feit dat dit (meestal) dieselfde programmatuur is wat gebruik word vir beide die meganiese oortik van materiaal wat alreeds geskep is en vir die generering van materiaal wat eers in en deur die gebruik van die betrokke program tot stand kom. Die bespreking kan slegs gaan oor die wyse waarop die program gebruik word deur die gebruikersgroep wat wel teks daarmee genereer.

Woordverwerking het gegroei uit die gebruik van die rekenaar op ander gebiede, met name die administratiewe gebied. Dit was aanvanklik bedoel vir sekretariële gebruik, maar mettertyd het dit die instrument by uitnemendheid van baie navorsers en skrywers geword. Die verdere ontwikkeling van woordverwerkingsprogrammatuur is aangedryf deur die kompetisie tussen verskeie programmatuurfirmas en baie meer klem het geval op die uitbou van die funksies van die programme as op die aard van die prosesse van teksgenerering en -verwerking. Dit het tot nadeel van die gebruikers plaasgevind. Hierby kom die feit dat gebruikers van woordverwerkingsprogramme gewoonlik binne 'n bepaalde program (of familie van programme) bly beweeg en nie maklik na totaal nuwe programme oorskakel nie.

In die bespreking van woordverwerkingsprogramme (4.2.1.1) sal dit blyk dat, alhoewel die funksies van hierdie programme grotendeels dieselfde is, daar tog ingrypende verskille is in die wyse waarop taal verwerk word en in verband met die prosesse van teksgenerering wat hulle daar afspeel.

3.3 DIE MANIPULASIE VAN DIE UITERLIKE TAALTEKENS

Die taalkundige stel dikwels daarin belang om die taalmateriaal wat hy ondersoek uit 'n ander gesigspunt georden te hê as dié waarin dit normaalweg in die algemene funksionering van taal voorkom, byvoorbeeld wanneer hy die materiaal in die vorm van 'n konkordansie georden wil hê of wanneer hy 'n woordeboek opstel. Die gewone gebruiker van taal stel uiteraard ook dikwels belang in dié soort ordening van die materiaal, byvoorbeeld wanneer hy 'n woordeboek of 'n tesourus gebruik. Die verskil is egter dat die feit dat die materiaal op dié wyse georden is, vir die gewone gebruiker net die geleentheid bied om die rekenaar "in werking te stel" ten opsigte van 'n bepaalde aksie wat hy uitgevoer wil hê. Die woord vorm in dié geval die skakelpunt of punt van interaksie.

Vanuit die woord as skakelpunt of punt van interaksie kan die inligting wat deur die gebruiker benodig word, maklik gevind word deur die rekenaar en dan vir die gebruiker aangebied word op so 'n wyse dat sy eie taalvermoë op die spel kom en hy self die nodige beslissings kan vel.

Hoewel hierdie skakelfunksie van die rekenaar eenvoudig mag voorkom, kan dit so vinnig en vindingryk plaasvind dat daarmee toepassings moontlik raak wat uit die oogpunt van die gebruiker as hoogs gesofistikeerd kan voorkom, byvoorbeeld in die geval van hiperteks (4.7.3). By die voorbereiding van hiperteksdokumente gaan dit oor die skakeling van verskillende brokke inligting met mekaar op 'n betreklik meganiese wyse. Wanneer die uiteindelijke gebruiker daarmee omgaan, word hierdie skakeling maar net in werking gestel of geoperasionaliseer (vergelyk 4.7.3.2). Dié wyse van omgang met teks is egter die naaste aan die prosesse wat hulle tipies afspeel wanneer die mens self (sonder die rekenaar) met inligting omgaan. Ons sou dus kon praat van die ondersteuning van die menslike prosesse van die intelligente omgang met sy wêreld (vergelyk 3.5, 4.7.3.5, 4.7.3.10h en 4.7.3.11g).

Hierdie relatief eenvoudige manipulerings van die uitwendige taalgegewens moet ten opsigte van die talige vermoë van die mens nie onderskat word nie en moet ook nie beskou word as 'n luukse waarsonder hy net sowel kan klaarkom nie. Op hierdie wyse vind daar 'n ondersteuning plaas van die menslike vermoë om met inligting om te gaan wat as essensieel beskou moet word in terme van die eise van die inligtingsera. Daar kom iets van die haalbaarheid en uitvoerbaarheid van die omgang met talige materiaal op die spel wat daarsonder net nie denkbaar sou gewees het nie.

Die manipulasie van die uiterlike taaltekens is miskien nie so interessant vir baie persone wat in toepassings op die gebied van "kunsmatige intelligensie" belangstel nie en dit kan ook verwag word dat ontwikkelings op dié gebied kan bydra tot 'n groter mate van gesofistikeerdheid in die wyse waarop die "meganiese" hantering van taalmateriaal op die rekenaar plaasvind. Tog kan met groot stelligheid gesê word dat dit juis dié relatief eenvoudige omgang met taalmateriaal is wat die gebruiker tot nuwe moontlikhede verhelp.

3.4 DIE HANTERING VAN BETEKENIS EN KONTEKS

Indien daar verbygekyk word by die "oëverblindery" wat soms plaasvind deurdat die rekenaar die taalgegewens vinnig en vindingryk op 'n baie meganiese wyse hanteer, blyk dit dat daar eintlik maar baie min wydverspreid geïmplementeerde programme is waarby met taal omgegaan word op die wyse wat die mens self taal prosesseer, byvoorbeeld deur ook betekenis daarby te betrek.

Soos blyk uit die bespreking van spelprogramme (4.2.1.2c) en grammatika- en stylkontroleprogramme (4.2.1.2d), gaan dit in die meeste programme van dié aard nog oor die uitwendige manipulasie van taalgegewens. Selfs op vlakke waar dit nog betreklik eenvoudig mag voorkom om rekenaarmatig daarvoor voorsiening te maak, kan dié programme nie werklik diepgaande kontrole meebring nie – hoofsaaklik omdat dié programme nie in staat is om die konteks in ag te neem nie. Deur brute krag (die vinnige verwerking van baie groot aantal moontlikhede wat vooraf gestoor word) sou dit nog moontlik kon wees om 'n effens hoër suksespersentasie te bereik. Só sou dit moontlik kon wees om byvoorbeeld voorsiening te maak vir woordgroepe (eerder as net losstaande woorde) of vir woorde wat met mekaar gekombineer kan word. Die probleem van die konteks kan egter nie werklik op dié wyse opgelos word nie.

Daar is gedurende die afgelope paar jaar baie vooruitgang gemaak in die ontwikkeling van programmatuur wat dit moontlik maak om betekenis te kan hanteer en die konteks in ag te kan neem. Dit kan verwag word dat ontwikkelings op die gebied van "kunsmatige intelligensie" toenemend sal bydra tot die vermoë om aan die probleem van betekenis en konteks te laat reg geskied (vergelyk Roux, 1988).

3.5 DIE REKENAAR-GEBRUIKER-RAAKVLAK

In die afgelope paar jaar het die rekenaar-gebruiker-raakvlak (*computer-user interface*) toenemend die aandag van navorsers en ontwikkelaars geniet. Daar word uiteraard al lank reeds aandag gegee aan dinge soos sinvolle en bevredigende skermuitleg, die gebruik van kleur, die helderheid van die skerm, ensovoorts. Baie van dié aspekte kan verstaan word binne die raamwerk van die ergonomika en het dus hoofsaaklik te make met die gemak waarmee die gebruiker die toerusting hanteer, die fisiese inspanning verwerk, ensovoorts. Sedert die middel van die tagtigerjare word daar al hoe meer aandag gegee aan simboliese aanduidings van funksies op die skerm van 'n rekenaar, die sogenaamde "ikone", byvoorbeeld 'n liasseerlaai as aanduiding van 'n lêer of 'n snippermandjie as aanduiding van 'n funksie "skrap/weggooi". Die groot voordeel van die gebruik van sulke simbole is dat dit die gebruiker help om die groot verskeidenheid benamings vir funksies in verskillende programme te sistap en eerder direk te reageer op die begrip wat deur die ikoon opgeroep word.

Met behulp van 'n meganisme vir die aanduiding van en reaksie op sulke simbole of spesiaal aangeduide areas (of ook op woorde), byvoorbeeld 'n lippen of muis, kan die

gebruiker dan baie gemaklik opdragte gee, ensovoorts. Die bedoeling met die ontwikkeling hiervan was om basiese wyses waarop die mens sy weg deur sy omgewing vind, oor te dra na die rekenaar-omgewing. Die ontwikkeling en wye aanvaarding van hierdie soort meganismes het oor die afgelope paar jaar gelei tot 'n totale verandering in die rekenaar-gebruiker-raakvlak.

Ook dié ontwikkelings kan egter nog beskou word as die verdere ontwikkeling van die uitwendige raakvlak (vergelyk 4.7.1). Die feit dat aspekte soos kleur en ruimtelike ordening verreken word, het uiteraard nie net betrekking op die gemak van die gebruiker by die hantering van 'n program nie en korrelleer ook op die een of ander wyse met die intellektuele prosesse betrokke by die verwerking van inligting.

Die term "gebruikersvriendelikheid" word dikwels as 'n omvattende term gebruik vir die soort aspekte wat hier ter sprake kom. In die afgelope paar jaar het dit egter duidelik geword dat dié term nie net betrekking het op die uitwendige raakvlak nie, maar ook op die intellektuele prosesse wat hulle afspeel wanneer die gebruiker besig is om inligting te verwerk. In selfs die mees gesofistikeerde programme moet die gebruiker hom nog steeds aanpas by die rekenaar (en wel by apparatuur en programmatuur). Dié aanpassing geskied al hoe makliker, aangesien gebruikers mettertyd in 'n groot mate gekondisioneer word in terme van hoe rekenaarprogramme funksioneer. In die afgelope paar jaar het dit egter toenemend duidelik geword dat die gesprek oor "gebruikersvriendelikheid" en die "rekenaar-gebruiker-raakvlak" impliseer dat die rekenaar en sy programme by die gebruiker en sy tipiese wyse van die omgang met inligting moet aanpas. Dit beteken dat die interaktiewe prosesse gerig moet word op die inligtingsoekende gedrag van die gebruiker, waarby die tipiese menslike verwerking van taalgegewens noodwendig ook betrek moet word. Dié kan egter nie meganies en met brute krag plaasvind nie. Die wyse waarop die konteks meehelp om betekenis te bepaal, waarop die gebruiker ervarings uit sy eie intellektuele verlede en uit die onmiddellik voorafgaande prosesse/stappe van die verwerking van inligting in daaropvolgende prosesse in werking kan stel, die beslissing oor wat hoegenaamd moontlik is en wat nie, ensovoorts, help alles mee in die wyse waarop die verskillende prosesse plaasvind.

In die mees resente gesprekke oor die rekenaar-gebruiker-raakvlak word dus toenemend aandag gegee aan aspekte wat semantiek, logika en kognisie raak. Met kognisie word 'n baie komplekse gebied aangesny, enersyds in die mate dat daar oor die prosesse van die verkryging en verwerking van indrukke en inligting nog soveel onduidelikheid is en andersyds in die mate dat dit nie duidelik is hoe die bestaande insigte met betrekking tot die prosesse van menslike kognisie in programfunksies omgesit kan word nie. Die gesprek oor hiperteks het baie daartoe bygedra om die besinning oor kognisie en oor die rekenaaromatige voorsiening vir sulke konsepte te stimuleer. Trouens, die gebied van hiperteks is die gebied in verband met rekenaar-toepassings op taalmateriaal waarin daar die meeste nagedink word oor kognisie.

Die kognitiewe kom in die gesprek oor hiperteks op verskeie vlakke ter sprake, byvoorbeeld: (a) die wyse waarop hiperteks gestruktureer moet word om aan die kognitiewe

prosesse uiting te gee (byvoorbeeld Baird & MacMorrow, 1988), (b) die feit dat hiperteks self 'n bykomstige kognitiewe lading (*overhead*) meebring (vergelyk Conklin, 1987, p. 40) en (c) die effek van die gebruik van hiperteks: Beeman et al. (1987, p. 81) praat van veranderinge by studente wat intensief van die Intermedia-stelsel gebruik gemaak het wat dui op *a new pluralistic cognitive style*. (Vergelyk ook Fiderio, 1988.)

Dit kan verwag word dat die rekenaar-gebruiker-raakvlak in toenemende mate sal laat reg geskied aan die eie aard van taal en aan die feit dat die mens 'n talige wese is.

HOOFSTUK 4: AKSIES TEN OPSIGTE VAN DIE VERWERKING VAN NATUURLIKE TAAL

4.1 INLEIDING

In hierdie hoofstuk word aandag gegee aan die aksies ten opsigte van taalgegewens wat telkens plaasvind. Die indeling van hierdie hoofstuk in verskillende hoofies reflekteer die verskillende soorte aksies met betrekking tot taalgegewens. Soos in die inleiding (1.1) aangedui, is die bedoeling nie om al die programme te bespreek wat in die ander verslae genoem is nie.

Dit is uiteraard moeilik om die aksies ten opsigte van die hantering van taal wat hieronder bespreek word, duidelik van mekaar af te grens. In sommige van die programme wat aan die gebruiker bekend is, gaan dit om meer as een aksie, loop die verskillende aksies deurmekaar, of is die aksies nie so duidelik sigbaar nie. Die onderskeidings wat wel getref word, is dus relatief en ruimte moet gelaat word vir 'n groot mate van wedersydse grensoorskryding.

4.2 DIE GENERERING VAN TEKS

Uiteraard gaan dit hier oor die generering van teks in elektroniese vorm, waarby daar 'n groot aantal moontlikhede is. In 3.2 is aangedui dat dit belangrik is dat daar meer aandag gegee sal word aan die proses van die generering van teks, en nie net aan die produk van dié proses nie.

4.2.1 Direkte (skeppende) generering van teks

In hierdie verband gaan dit hoofsaaklik oor woordverwerkingsprogramme. Programme vir kantoorpublikasie/lessenaarpublikasie (*desktop publishing* - *DTP*) kom ook in dié verband ter sprake, maar dié behels gewoonlik die verdere formattering van teks wat alreeds deur 'n proses van woordverwerking geskep is.

4.2.1.1 Woordverwerkingsprogramme en meegaande fasiliteite

a) *Basiese aard en uitgangspunte van enkele programme*

Baie woordverwerkingsprogramme leen hulle ideaal tot die oortik van teks vanaf 'n geskrewe of gedrukte vorm. Al die funksies wat 'n tikmasjien altyd gebied het, word as't ware in die oortreffende trap met betrekking tot spoed en gesofistikeerdheid aangebied. 'n Goeie voorbeeld van so 'n program is *WordStar*, wat een van die eerste woordverwerkingsprogramme was en wat telkens verder ontwikkel is (verskeie weergawes en selfs ook 'n verdere ontwikkeling soos *WordStar 2000*).

Met die totstandkoming van *WORD* (Microsoft Corporation) het daar 'n totaal nuwe fase in die ontwikkeling van woordverwerkingsprogramme aangebreek. Dit kom voor asof daar - in elk geval in 'n groter mate as voorheen - grondliggend nagedink is oor die

menslike prosesse betrokke by die hantering van teks. (Vergelyk verder 4.2.1.1c.) Die perspektiewe wat in die vorm van programfunksies neerslag gevind het, is veral belangrik vir die generering van teks – hoewel dit waarskynlik nie in die eerste plek met die oog daarop beplan is nie, maar eerder om woordverwerkingsprogramme nader te bring aan die werkwinkel van die professionele setter.

Die funksie van stylstelle (*style sheets*) blyk van groot nut te wees by die generering van teks. Dit stel die gebruiker in staat om 'n aantal kodes (of "style") te skep in terme waarvan die teks waaraan hierdie kodes gekoppel word, vorm kry op die skerm en op die drukker. Elkeen van hierdie kodes verteenwoordig 'n aantal attribute, byvoorbeeld lettertipe, formaat, plasing, ensovoorts. Hierdie attribute bepaal in die eerste plek die uitwendige voorkoms van die teks, maar aangesien die vorm en plasing van die teks gewoonlik ook struktuur en samehang impliseer, kan hierdie fasiliteit ook nuttig aangewend word in die proses van die generering van die teks. Deur vinnig die kode (of "styl") van 'n bepaalde paragraaf te verander, kan die onderlinge samehang van gedagtes ook duidelike uitdrukking kry en kan die prosesse betrokke by die generering van die teks ondersteun word. Byvoorbeeld: deur die koppeling aan paragrawe (wat ook net enkele woorde of frases kan wees) van die kode wat 'n inkeping op die derde vlak meebring, kan dit duidelik vorm kry dat hierdie ingekepte paragrawe aan mekaar newegeskik is en terselfdertyd ondergeskik is aan 'n paragraaf op die tweede vlak van inkeping. Deur slegs aan die kode te verander, verander die plasing van die paragrawe en kan 'n ander struktuur bewerkstellig word.

Die uitgangspunt van stylstelle verskil van eenvoudiger fasiliteite in ander programme waarby die outeur net 'n vasgestelde bladsyformaat of uitleg gebied word. Hierdie fasiliteite is ook nuttig, maar dien hoofsaaklik die tipografiese versorging en afronding van die teks. Die krag van die stylstelle van WORD lê daarin dat dit ook op die vlak van afsonderlike paragrawe van toepassing gemaak kan word en dat dit gemaklik en dinamies verander kan word in die proses van die generering van teks.

Dit is belangrik dat fasiliteite soos dié van stylstelle ook beskou sal word as 'n aspek van die verwerking van taal, al val dit miskien nie binne die gangbare beskrywing van dit wat aan die groot sentra gedoen word in verband met taalverwerking nie. Dit is eers deur verwerking met behulp van die rekenaar dat hierdie moontlikhede met betrekking tot die hantering van teks 'n werklikheid word. Daarteenoor geld egter ook dat programme fasiliteite kan bied wat nie vir hierdie beweeglikheid in taal voorsiening maak nie of wat juis dít bewerkstellig wat nie houdbaar is in terme van die eise van taal of taalgebruik nie.

Die funksie van stylstelle word in navolging van WORD tans ook gevind in verskillende ander programme, byvoorbeeld Ventura ('n program vir kantoorphublikasie) en WordPerfect (weergawe 5.0). Dit blyk dus dat die beginsel wat hier op die spel is, wel oortuig.

Dergelike vorme van beweeglikheid by die generering van teks word vandag ook gebied deur 'n verskeidenheid ander woordverwerkingsprogramme. Dit raak egter baie moeilik om 'n oordeel te vel, aangesien gebruikers dikwels so verknog raak aan bepaalde

programme dat hulle weë vind om by probleme verby te beweeg en die program tog sinvol aan te wend binne die raamwerk van hulle eie teksverwerkingstyl (vergelyk Heim, 1987, pp. 139-151).

b) *Raamwerkprogramme /-fasiliteite (outliners)*

Verskeie moderne woordverwerkingsprogramme sluit tans raamwerkfasiliteite in (verge-lyk Foster, 1985 en Hershey, 1985). Hierdie fasiliteite word met verskeie name aangedui, byvoorbeeld *outliners of idea processors*. Die eerste program van hierdie aard in die MS-DOS-mikrorekenaaromgewing was *ThinkTank* (van die firma Living Videotext, Mountain View, CA) (vergelyk Kamin, 1984). Daar was ook voorlopers van hierdie program in ander rekenaaromgewings, byvoorbeeld in die CP/M-omgewing. Dit is interessant dat hierdie program juis onder die invloed van die Chicago-skool van filosofie ontwikkel is, met die doel om die gebruiker te help om sy gedagtes in hiërargiese vlakke te struktureer (vergelyk Heim, 1987, p. 141). *ThinkTank* is 'n onafhanklike program wat in samehang met verskeie woordverwerkingsprogramme gebruik kon word, maar waarby die oorgang tussen die twee betrokke programme tog betreklik omslagtig was. Kort daarna is dieselfde fasiliteite ingebou in *WORD* en sedertdien ook in 'n verskeidenheid ander programme. Vir die doel van die beskrywing van hierdie soort fasiliteite word dié van *WORD* kortliks bespreek.

Die funksie van die raamwerkprogram kom op die volgende neer:

- i) Die program bied aan die gebruiker die moontlikheid om te beweeg of te spring tussen drie punte in of perspektiewe op die dokument
 - * die raamwerk of buitelyne van die hele dokument (in die vorm van hofies of opskrifte van paragrawe, wat dan veronderstel is om idees te verteenwoordig)
 - * 'n stel ondergeskikte hofies (wat ondergeskikte idees verteenwoordig)
 - * die teks op enige bepaalde punt in die dokument.

Hierdie beweging tussen verskeie perspektiewe op die dokument kan beskryf word in terme van drie prosesse:

- * ineenvou (*collapsing*)
- * uitsprei (*expanding*)
- * vertoon (*display*).

Die term *zooming* word dikwels ook gebruik om die eerste twee prosesse te beskryf.

- ii) Die beweging tussen enige van hierdie perspektiewe vereis die druk van hoogstens twee toetse op die sleutelbord.
- iii) Enige verskuiwing van 'n deel van die hofies (op enige vlak) bring terselfdertyd ook 'n hernumerering teweeg van die deel van die dokument wat daardeur geraak word.

- iv) Die koppeling van die fasiliteite van stylstelle aan die raamwerkfasiliteite bring mee dat die raamwerk ook wanneer dit uitgedruk word die struktuur van die dokument in gedrukte vorm (met alle tipografiese verskynsels) vertoon.

Dit is dus duidelik dat raamwerkfasiliteite veral nuttig is by die manipulasie en ordening van groter blokke van teks, spesifiek in die stadium dat die teks gegenereer word of bestaande teks tot 'n geordende geheel saamgevoeg word.

Die aansprake wat gemaak word met betrekking tot hierdie soort fasiliteite is soms oordrewe, as sou dit ook 'n uitwerking hê op baie ander terreine van die lewe, byvoorbeeld Kamin:

Children who grow up using ThinkTank, or a program like it, may develop the capacity to handle large amounts of information, and to structure it at an early age. Thus, they may be able to solve problems more effectively than their elders (1984, p. 218).

Die oordeelkundige gebruik van stylstelle ((i) hierbo) bied ook die geleentheid om 'n mate van strukturering te bewerkstellig, maar hoegenaamd nie so doeltreffend soos deur die gebruik van 'n raamwerkprogram nie.

Die werking van die raamwerkfasiliteite in WORD geskied gemaklik en glad. Dit is ten volle met die ander fasiliteite van WORD geïntegreerd en die gebruiker van die raamwerkfasiliteite kan ook op alle ander fasiliteite van WORD aanspraak maak.

Raamwerkfasiliteite dien ook die doel van die verwerking van taal, nie net met betrekking tot die uiterlike struktuur nie, maar ook met betrekking tot interne samehang van taal. Dit is veral op die vlak van die generering van teks dat hierdie soort programme van die grootste waarde is.

c) *Onlangse besinning oor woordverwerkingsprogramme*

Woordverwerkingsprogramme bied in beginsel die moontlikheid om teks op 'n ander wyse te genereer as in die vorm van liniêre teks. (Vir 'n bespreking van die nie-liniêre aard van die proses waarvolgens teks met woordverwerkingsprogramme gegenereer word, vergelyk Heim, 1987.) Raamwerkprogramme bied sekerlik 'n nog groter mate van beweeglikheid en oorsigtelikheid (vergelyk (b) hierbo), veral deurdat die gebruiker die dokument telkens gemaklik kan reduceer tot die raamwerk of kan *zoom* na die detail. Tog is dit so dat hierdie moontlikhede in 'n mate alreeds gebied word deur eenvoudige woordverwerkingsprogramme sonder raamwerkfasiliteite. Hierdie moontlikhede kom daarin tot uiting dat die gebruiker in staat is:

- om te begin met 'n raamwerk van gedagtes (wat vorm kry as trefwoorde, frases, opskrifte of sinne) en om enigeen daarvan algaande uit te brei na volwaardige teks
- om, namate die gedagtes ontwikkel, van een afdeling na die ander te beweeg en (eers) daar voort te gaan

- om die dokumentasie (en detail-argumentasie) op enige deel te kan doen sonder om die hele tyd die implikasies vir die geheel van die dokument in ag te neem (byvoorbeeld deurdat die meeste programme outomatiese voetnota-fasiliteite het).

By die ontwerp van woordverwerkingsprogramme het dit blykbaar selde gebeur dat daar fundamenteel besin is oor die aard van die intellektuele en kognitiewe prosesse wat hulle afspeel in die proses van die generering van teks. Die voortgesette uitbou van die funksies van die tradisionele tikmasjien het blykbaar die voorrang geniet. Aspekte soos die raamwerkfasiliteite is afsonderlik ontwikkel en het eers mettertyd ingang gevind in die tradisionele programme. ThinkTank is eers afsonderlik ontwikkel en bemark vir gebruik saam met 'n verskeidenheid programme. 'n Weergawe daarvan is mettertyd as geïntegreerde raamwerkfasiliteite ingebou in WORD.

Die ontwikkeling van hiperteks (vergelyk verder 4.7.3), wat op sy beurt saamgehang het met bepaalde ontwikkelings in die tegnologie (vergelyk 4.7.3.3), het daartoe gelei dat daar toenemend ook indringend aandag gegee word aan die proses van die generering van teks. Hiperteks impliseer – breedweg gesproke – dat daar meer erns gemaak word met die faset van nie-liniêre teks.

Aan die Department of Computer Science van die University of North Carolina is daar indringend aandag gegee aan die kognitiewe prosesse betrokke by die hantering van teks, en wel op die vlakke van die generering en die lees van teks (vergelyk Smith et al., 1987). Dit is veral die vlak van teksgenerering wat dié groep interesseer en wat die werk in die rigting van die ontwikkeling van 'n Writing Environment (WE) gestimuleer het.

Teksgenerering word beskou as 'n komplekse proses wat gebaseer is op verskillende kognitiewe vaardighede. Dit gaan nie net oor die vertaling van idees in woorde nie, maar oor inligtingsherwinning, die identifisering van assosiatiewe gedagtes, die opbou van hiërargiese strukture, woordkeuse, die rekursiewe keuse van woorde in die proses van redigering, ensovoorts (Smith et al., 1987, p. 197). Skrywers maak gebruik van verskillende kognitiewe modi om hierdie prosesse te laat plaasvind of om die oorgang van die een na die ander te bewerkstellig. Die volgende modi (wat iets anders is as temporeel op mekaar volgende stadia) kan onderskei word: die voor-skriftelike modus (*pre-writing*), die modus van organisasie (*organizing*) en dié van teksgenerering (*writing*).

In die ontwerp van die WE is die doel om meganismes daar te stel wat die kognitiewe prosesse in daardie betrokke modus sinvol kan ondersteun. Dit lei tot 'n ontwerp wat op die volgende neerkom:

- a) Die skerm – en 'n groter skerm as die gewone is wenslik – word verdeel in 'n aantal vensters wat elkeen vir 'n bepaalde modus voorsiening maak.
- b) Die gebruiker kan gemaklik tussen die vensters beweeg. Die inhoud van die vensters is egter op so 'n wyse (op die vlak van deel-inhoude) onderling verbind dat die gebruiker vanaf die betrokke stuk deel-inhoud van die een

venster direk by die alternatiewe realisering daarvan in 'n ander venster kan uitkom.

- c) Daar word voorsiening gemaak vir die volgende vensters en modi (vergelyk Smith et al., 1987, pp. 201-207):
- i) Die netwerkmodus ondersteun die verkennende fases van die ontwikkeling van 'n dokument of teks. Hierdie modus toon baie nou ooreenstemming met dit wat tipies in hipertekstelsels gevind word, naamlik die moontlikheid om een nodus aan 'n ander te verbind deur 'n skakel (*link*). (Vergelyk 4.7.3.6.)

Die konsepte word as nodi in grafiese vorm op die skerm voorgestel (byvoorbeeld blokkies of sirkels), kan in groepe saamgevoeg word en kan gemaklik (deur die gebruik van die muis) rondgeskuif word.
 - ii) Die boomstruktuur-modus help die skrywer om 'n enkele, geïntegreerde, hiërargiese struktuur vir die dokument op te bou. Die afsonderlike konsepte wat in die netwerkmodus geskep is – en daar nog op 'n vrye basis rondgeskuif is – word nou streng hiërargies georden. Nuwe nodi kan nou net in samehang met die gehele boomstruktuur geskep word.
 - iii) Die redigeermodus bied 'n gewone redigeerder, wat die gebruiker toelaat om die konsep (tot nog toe as nodus voorgestel) in teks te enkodeer.
 - iv) In die teksmodus word die dokument in liniêre vorm weergegee. Die komponente van die teks is egter gekoppel aan die nodi in die boomstruktuur en deur in die boomstruktuur rond te beweeg, word die ooreenstemmende teksgedeeltes in die teksvenster vertoon.
- d) Die krag en soepelheid van die program word verseker deur daaraan databasisfasiliteite te koppel, asook die moontlikheid om gemaklik tussen al die komponente deur te kan beweeg met behulp van 'n muis.

Dit is duidelik dat WE nog in 'n ontwikkelingsfase is. Daar kan egter verwag word dat die aktiwiteite van hierdie navorsingsgroep sal uitloop op die ontwikkeling van 'n nuwe soort woordverwerkingsprogram waarin dié belangrike gesigspunte tot uiting kom, of dat insette van dié groep in ander programme ingang sal vind.

Die navorsing van 'n ander groep aan die University of North Carolina het alreeds uiting gevind in die program *Storyspace* (vir die MacIntosh-rekenaar) wat "interaktiewe fiksie" moontlik maak, beide op die vlakke van skryf en lees (wat in werklikheid baie nou met mekaar saamhang): die outeur skep die teks as 'n reeks teksepisodes, wat dan aan mekaar gekoppel word deur 'n diagrammatiese redigeerder. Die leser is egter vry om episodes op 'n ander wyse met mekaar te verbind, wat dan lei tot 'n "interaktiewe leesproses".

Die werk het plaasgevind in die "Writing Group" en in samewerking met die groep wat die *Writing Environment* ontwikkel. (Vergelyk Bolter & Joyce, 1987.) Die program *Storyspace* word tans in die vorm van 'n beta-kopie uitgetoets; dit het nog eksperimentele karakter. Dit is egter ook duidelik dat die insigte wat in albei hierdie soorte programme neerslag vind, van 'n fundamentele aard is en – veral noudat die tegnologie daarvoor die

nodige ondersteuning bied – groot moontlikhede vir (onder andere) woordverwerkingsprogramme inhou. (Oor die noodsaaklikheid van fundamentele denke met betrekking tot sulke toepassings, vergelyk Ott, 1989b en McCarty, 1986, p. 62.)

d) *Gespesialiseerde woordverwerkingsfasiliteite vir akademiese doeleindes en tekspublikasie*

Afgesien van die gewone woordverwerkingsprogramme vir algemene gebruik, het daar ook 'n groot aantal programme tot stand gekom vir spesifieke doeleindes. Al die gewone programme leen hulle ook nie eweseer tot enige soort toepassing nie. Vergelykende besprekings van die verskillende programme in resente rekenaartydskrifte kan in dié verband geraadpleeg word. Die programme **WORD** en **WordPerfect** blyk byvoorbeeld baie gewild te wees vir toepassings in die geesteswetenskappe. Intussen raak **Nota Bene** ook baie gewild vir toepassings op dié gebied. Vergelyk McCarty (1986) en Hughes (1986b) vir omvattende resensies van **Nota Bene**.

Die vermoë van woordverwerkingsprogramme om spesiale karakters (wetenskaplike simbole of karakterstelle van vreemde tale) te kan hanteer, is ook 'n belangrike faktor by die keuse van bepaalde programme. Tog kan dié aspekte beskou word as grotendeels van belang vir die uiterlike versorging van die teks (beide op die skerm en in die gedrukte vorm).

Die besondere eise van woordverwerking in die geval van omvangryke tekste met besondere eise van gevorderde tipografie het aanleiding gegee tot die ontwikkeling van 'n program wat vandag op groot skaal gebruik word vir die publikasies van tekste – spesifiek vir die publikasie van tekste van historiese waarde. Die **Tübinger System von Textverarbeitungsprogrammen (TUSTEP)** is ontwikkel aan die Rekensentrum van die Universiteit van Tübingen, Wes-Duitsland, onder leiding van Wilhelm Ott. (Vergelyk Ott 1979; 1989a; 1989b; 1989c.) Die program (wat eintlik uit 'n aantal afsonderlike modules bestaan wat baie goed met mekaar geïntegreer is) is aanvanklik ontwikkel vir hoofraamrekenaars, maar daar is tans ook weergawes beskikbaar vir minirekenaars (VAX) en vir mikrorekenaars (MS-DOS). Op grond daarvan dat die program aanvanklik net 'n hoofraamprogram was, was dit lank slegs aan die Universiteit van Tübingen beskikbaar en moes gebruikers daarheen kom vir die gebruik daarvan. Mettertyd is die grootste deel van die program aan ander sentra beskikbaar gestel, wat tans in toenemende mate die geval is. Die deel van die program wat die finale voorbereiding vir die setwerk hanteer, kan weens bepaalde regte op lettertypes nie buite die Rekensentrum beskikbaar gestel word nie. TUSTEP word tans op groot skaal gebruik vir die herpublikasie van tekste vanaf die klassieke tot by moderne tekste, veral die kritiese uitgawes van tekste waarby die redakteur 'n baie groter mate van beweeglikheid moet hê as wat deur enige van die ander bestaande woordverwerkingsprogramme gebied kan word. Verreweg die meeste uitgawes van dergelike tekste wat tans in die sentraal-Europese gebied verskyn, is met behulp van hierdie program voorberei.

TUSTEP maak erns met die feit dat tekste van hierdie aard op 'n bepaalde wyse tot stand gekom het en oorgelewer is: daar het verskeie manuskripte bestaan, verskeie historiese

teksuitgawes het tot stand gekom en al hierdie weergawes en variante moet in 'n stel "tekskritiese apparate" aangedui word (gewoonlik in drie dergelike "tekskritiese apparate"). Die verwerking van hierdie gegewens moet op 'n kragtige rekenaar kan plaasvind en gereelde uitdrukke vir die doeleindes van proeflees moet kan plaasvind, op so 'n wyse dat die plasing van die teks (insluitende woordbreuke of afkapping) in die finale uitdruk (op bro-mide op 'n setmasjien) presies dieselfde sal wees as in die vroeëre stadia van verwerking. Hier gaan dit onder andere oor presisie in terme van tipografiese standarde, gemeet in punte (*points* = 1/72 van 'n duim) en dele van punte. Die gebruik van 'n bladsy-beskrywingstandaard soos PostScript maak dit tans in 'n verskeidenheid van programme en vir gebruik op 'n groot aantal tipes toerusting van uiteenlopende aard moontlik om presies eenderse tipografie te genereer (uiteraard binne die beperkings van die fisiese/tegniese moontlikhede van die betrokke drukker).

Die beskikbaarheid van 'n aantal "tekskritiese apparate" maak dit moontlik om aan die eie aard van hierdie soort tekste te laat reg geskied. Die moontlikhede vir die opstelling van indekse en inhoudsopgawes is ook baie verder ontwikkel as in die geval van die gewone woordverwerkingsprogramme (en selfs verder ontwikkel as die meeste professionele publikasiestelsels). Spesiale funksies toon vir die gebruiker by opeenvolgende uitdrukke aan watter dele redaksioneel ingevoeg is, ensovoorts.

In die ontwikkeling van TUSTEP is baie aandag gegee aan die basiese funksies of aksies wat onderneem word met betrekking tot die teks en programkomponente is ontwikkel om die uitvoering van dié funksies of aksies moontlik te maak. Ott (1989b, p. 17) praat van "Grundoperationen": die gebruiker moet self sy probleem kan terugvoer tot basiese "Grundoperationen", dan 'n strategie ter oplossing van sy probleem ontwerp en vervolgens die "Grundfunktionen" van TUSTEP in werking stel.

Afgesien van hierdie fasiliteite – wat grotendeels op die uiterlike ordening van die materiaal betrekking het – maak TUSTEP ook voorsiening vir die ordening van die taalgegewens op so 'n wyse dat die taalnavorsers dit uitstekend in sy navorsing kan gebruik, byvoorbeeld die moontlikheid van die opstelling van konkordansies (vergelyk Bader, 1989).

Dit is duidelik dat TUSTEP nie in die eerste plek gebruik word vir dié soort prosesse wat deur die meeste mense onder die verwerking van natuurlike taal beskou word nie. Tog maak hierdie programme met 'n belangrike aspek van die verwerking van taal – veral wanneer dit gaan oor die hantering van literêre tekste of tekste van historiese belang. Die gladde interaksie van die fasiliteite vir die ordening van gegewens met die teks in die formaat waarin dit vir publikasie voorberei word, bring mee dat die navorsers met die materiaal in die mees korrekte vorm kan werk.

4.2.1.2 Ondersteuning ten opsigte van teksgenerering

Die fasiliteite wat in enige woordverwerkingsprogram gevind word, bied aan die gebruiker ondersteuning in die proses van die generering van teks. Raamwerkfasiliteite (4.2.1.1b hierbo) bied dit in 'n besondere mate. Baie van die algemene fasiliteite kan nie juis

as spesifiek relevant vir die kwessie van die verwerking van natuurlike taal beskou word nie en sou ewe geskik wees vir die verwerking van numeriese gegewens. Hierdie fasiliteite word nie in dié verslag bespreek nie.

Daar bestaan egter ook sekere ander fasiliteite wat baie nou met die verwerking van taal saamhang en wat afsonderlik die aandag verdien in 'n verslag soos hierdie een. Die volgende kan as die belangrikste uitgesonder word: tesourusfasiliteite, vertalings-ondersteuning (onder sekere omstandighede), spelkontrole, die kontrole van grammatika en styl, ensovoorts. Afkapping speel uiteraard ook in dié verband 'n rol, maar hang in 'n groter mate met die uiterlike versorging van teks saam en verdien dus apart die aandag (4.3.2).

a) *Tesourusfasiliteite*

Die vind van die mees geskikte woord speel 'n baie belangrike rol op die stadium van teksgenerering (vergelyk ook Smith et al., 1987). Uiteenlopende faktore kan bepaal of die gebruiker 'n ander woord soek as een wat hy alreeds as keuse het, byvoorbeeld

- die vraag of daar nie 'n meer geskikte woord is om uitdrukking te gee aan die begrip nie
- die register van taal waarin die gebruiker beweeg
- die strewe om meer afwisseling in die woordkeuse te kry.

Tesourusprogramme is binne die dinamiek van woordverwerking net doeltreffend wanneer hulle goed geïntegreer is in die betrokke woordverwerkingsprogram. Daar bestaan egter ook verskeie tesourusprogramme wat onafhanklik is of net losweg met woordverwerkingsprogramme saamhang. Dié programme kan egter ook in ander toepassings gebruik word (vergelyk Raskin, 1987).

In werklikheid is 'n elektroniese tesourus slegs die elektroniese weergawe (soms effens verkort of vereenvoudig) van dit wat in 'n gewone tesourus gebied word. Al wat bykom, is programfunksies wat hierdie versameling materiaal operasionaliseer in terme van die proses van woordverwerking. Die vraag kan dus ontstaan of elektroniese tesouri nie 'n onnodige luukse is nie. Hierteenoor moet dit gestel word dat elektroniese tesouri een voordeel meebring wat nie gebied kan word deur die gedrukte weergawe nie, naamlik dié van onmiddellikheid: die funksie word as't ware onmiddellik deur knopdruk in werking gestel en die alternatiewe woorde verskyn op die skerm. Die druk van een verdere toets (na die keuse van die geskikte woord) bring die verandering in die dokument teweeg. Hierdie voordeel van onmiddellikheid is, gesien die dinamiek van woordverwerking, nie 'n luukse nie, maar eerder noodsaaklik, veral aangesien die funksie nie in dieselfde mate gebruik sal word indien dit nie so gemaklik geoperasionaliseer kan word nie. Daar is vroeër verwys na die psigodinamiese aspekte verbonde aan woordverwerking (1.3). Die onmiddellike beskikbaarheid van 'n tesourus wanneer teks gegenereer word, bring 'n totaal ander dimensie in die beweeglikheid waarmee natuurlike taal verwerk word in die proses van generering.

Vir persone wat in 'n veeltalige gemeenskap teks genereer, verkry 'n thesaurus 'n nog groter relevansie. Hier gebeur dit dikwels dat teks gegenereer moet word in die taal wat nie die moedertaal van die gebruiker is nie. Dit geld byvoorbeeld van die situasie in Suid-Afrika, waarin Afrikaanssprekende akademici (en insgelyks mense in talle ander beroepsrigtings) teks in Engels genereer. Die probleme in verband met die keuse van die regte woord is nou heelwat groter as vir mense wat slegs met een taal werk. In so 'n geval sou dit uiteraard nog beter wees indien daar aan die tesourus ook 'n vertalingswoordeboek gekoppel sou kon wees (hieroor verder (b) hieronder).

Tesourusprogramme (of -fasiliteite in woordverwerkingsprogramme) is van uiteenlopende karakter en gehalte. In die geval van fasiliteite wat in woordverwerkingsprogramme geïntegreerd is, kom die funksies van dié programme op die volgende neer:

- die gebruiker skuif die wyser (*cursor*) op die woord waarvoor hy die alternatiewe wil sien en druk 'n bepaalde kombinasie van toetse (gewoonlik twee)
- 'n venster word op die skerm oopgemaak en die ander woorde wat gekies kan word, word vertoon
- die gebruiker kies die mees geskikte woord deur met die wyser daarna te beweeg en druk 'n bepaalde toets
- die tesourus-venster verdwyn, die nuwe woord vervang die ou woord in die teks en die gebruiker moet hoogstens nog in sekere gevalle sekere grammatikale aanpassings maak.

Wat betref die gehalte en funksionaliteit van die voorstelle wat gemaak word, verskil programme in terme van:

- die getal woorde en woordkombinasies (wat kan wissel tussen 50 000 en meer as 200 000)
- die vraag of alle woorde ook as toegangsterme in die soektog benut kan word; byvoorbeeld: alhoewel Turbo Lightning 50 000 woorde in die tesourus bied, kan daar tot hulle toegang verkry word slegs deur 5 000 woorde (Raskin, 1987, p. 277). Hierdeur word die funksionaliteit van die program ingrypend beperk. In ander gevalle is daar tot 20 000 toegangswoorde (*root terms*) (p. 280).
- die indeling en rangskikking van woorde in die tesourus en die wyse waarop dit aan die gebruiker gebied word.

Nie alle programme bied aan die gebruiker 'n indeling van woorde wat die gebruik van die tesourus vergemaklik nie. Alhoewel die meeste programme die woorde indeel volgens woordsoorte (Raskin, 1987, p. 276), bied slegs enkele programme 'n indeling wat ook nog betekenskategorieë onderskei. In Word Finder word woorde in verskillende kategorieë, byvoorbeeld in horisontale groepe (lyne of groepe lyne), op die skerm aangedui. Raskin (1987, p. 278) gee 'n voorbeeld van die waarde van so 'n benadering: die tesourus-inskrywing *clear* bied 18 woordkategorieë (beide naamwoorde en werkwoorde). Uit die

oogpunt van die sinvolle gebruik van 'n tesourus moet hierdie ordening as baie belangrik beskou word.

Enkele funksies wat die gebruiker wat erns met taal maak, graag sou wou hê, is glad nie beskikbaar nie, byvoorbeeld die insluiting van woordgroepe. Raskin (1987, p. 280) noem slegs een tesourusprogram wat op 'n beperkte skaal vir frases voorsiening maak (naamlik *Webster's New World On-Line Thesaurus*). Trouens, dit kom voor asof die meeste elektroniese tesouri net gerig is op wat breedweg "hoofwoorde" genoem kan word.

'n Interessante voorbeeld van die wyse waarop die woorde in die tesourus georden kan word om die gebruik te vergemaklik, kan gevind word in die tesourus van *WORD*. Hierdie tesourus is 'n weergawe van *Word Finder*, wat ook as aparte tesourusprogram beskikbaar is (vergelyk Raskin, 1987, p. 276).

Daar is uiteraard nog baie ruimte vir die verdere ontwikkeling van tesourusfasiliteite. Die hoër verwerkingspoed van die nuwe generasie rekenaars en die groeiende skyfkapasiteit bring mee dat al hoe meer sinvolle funksies nog ingebou kan word. Die volgende sou byvoorbeeld wenslik kon wees:

- Die insluiting van woordrelasies anders as net dié met (ongeveer) sinonieme betekenis (byvoorbeeld ook antonieme, nouer en breër relasies). Dit is byvoorbeeld denkbaar dat die gebruiker 'n sin totaal anders wil formuleer.
- Die inagneming van woordgroepe (wat byvoorbeeld voorsetsels insluit).
- Die indeling van die woordvoorraad in verskeie registers (byvoorbeeld algemene gebruik en kombinasies van toepassing op spesifieke vakgebiede).
- Die inagneming van die grammatikale vorm by soektog en vervanging, wat vir 'n taal soos Afrikaans nog belangriker as vir Engels is. 'n Voorbeeld is werkwoorde met "ge-", wat eers "ontleed" moet word voordat die soektog in die tesourus kan begin. Hierdie probleem word selfs in Engels nog nie doeltreffend en konsekwent hanteer nie en tussen programme onderling is daar groot verskille (vergelyk Raskin, 1987, p. 280). In die tesourusfasiliteite van *WORD* neem die woord *playing* die gebruiker wel na die hofie *play* in die tesourus, maar *caught* en *went* neem hom nie verder nie.

By die gebruik van 'n tesourusprogram vir Engels in Suid-Afrika moet telkens in ag geneem word dat die meeste programme uit die VSA gelewer word met 'n tesourus vir Amerikaanse Engels, wat vir Suid-Afrikaanse Engels betreklik baie probleme oplewer.

b) *Vertalingsondersteuning*

Indien die gebruiker teks genereer in 'n taal waarin hy nie baie behendig is nie, het hy behalwe 'n tesourus en 'n spelkontroleprogram eintlik ook nog 'n vertalingswoordeboek nodig wat hom in staat stel om vanaf die woord in die taal waarin hy meer gemaklik beweeg die woord in die minder bekende taal te kan vind. Vir gebruik in Suid-Afrika sal so 'n funksie van groot waarde wees. Geeneen van die meer bekende woordverwerkingsprogramme maak egter vir so 'n funksie voorsiening nie. Die materiaal wat vir so 'n proses

gebruik kan word, raak in toenemende mate beskikbaar. In 'n onlangse lys CD-ROM-skywe (Faxon, 1988, pp. 42-43) word drie skywe vermeld wat tweetalige (of veeltalige) woordkombinasies bevat. In die meeste gevalle gaan dit blykbaar oor tegniese terminologie. Wat nog kortkom, is dat daar in die woordverwerkingsprogramme voorsiening gemaak sal word vir vertalingsfasiliteite, wat op dieselfde wyse in werking gestel kan word as die gebruik van 'n tesourus.

Dit sal ideaal wees indien die materiaal in 'n vertalingswoordeboek ook voorsiening kan maak vir die ordening van woorde in woordgroepe en vir die desiderata wat hierbo by tesourusfasiliteite genoem is ((a) hierbo), nou op meertalige vlak.

c) *Spelkontroleprogramme*

Die meeste woordverwerkingsprogramme het ingeboude fasiliteite vir spelkontrole en dié word deur baie gebruikers op 'n gereelde basis gebruik. Dié fasiliteite maak dit moontlik om spelfoute in 'n teks te identifiseer en korreksies aan bring op die basis van 'n ingeboude woordeboek met die korrekte spelling. Spelfoute is in baie gevalle eintlik maar net tikfoute, byvoorbeeld baie dikwels in gevalle waar die teks nie deur professionele tiksters getik word nie, dit wil sê die persoon besef dadelik dat 'n woord verkeerd is wanneer hy dit sien, maar ontdek die fout nie dadelik nie. As die fout egter eers gemaak is, word dit nie maklik raakgesien tydens proeflees nie, aangesien die konteks van die woord die betekenis van die geheel dra. Spelkontroleprogramme is gevolglik veral sinvol vir persone wat self teks op die woordverwerker genereer. Die keuse vir of die geïdentifiseerde fout wel 'n fout is en of die voorgestelde spelwyse wel aanvaar moet word, lê by die gebruiker. Daar word gewoonlik 'n stuk konteks saam met die woord vertoon om die gebruiker in staat te stel om sy keuse te doen. Die gebruiker kan dan ook 'n eie woordeboek (of meer as een) opbou waarin woorde vervat is wat nie in die woordeboek van die program voorkom nie, maar tog korrek is, byvoorbeeld eiename. Die gebruikerswoordeboeke kan agterna deur die gebruiker geredigeer word. Woorde wat uit twee of minder karakters bestaan, word gewoonlik nie gekontroleer nie.

Bogenoemde funksies is dié wat gemeen is aan die meeste spelkontroleprogramme. Sommige programme bied ander bykomende funksies: sommige wys woorde uit wat direk langs mekaar herhaal word, verkeerde hoofletters in woorde, ensovoorts, so byvoorbeeld die spelkontroleprogram van WORD. Dié soort funksie kan ook by grammatika- en stylkontroleprogramme gevind word (vergelyk (d) hieronder).

Afgesien van hierdie algemene funksies is daar gewoonlik ook verdere funksies wat tussen programme onderling kan verskil, byvoorbeeld in WORD word 'n korreksie net een keer aangebring in dieselfde dokument; alle ander gevalle word dan op die basis van dié korreksie hanteer.

'n Eerste onderskeiding wat tussen programme gemaak kan word, is of die program in bondelmodus (*batch mode*) of interaktief funksioneer. Indien die program in bondelmodus funksioneer, word die teks nadat dit gegeneer is, eers verwerk en het die gebruiker dan die geleentheid om die programvoorstelle te beoordeel en te aanvaar al dan

nie. In interaktiewe modus word daar dadelik (deur 'n geluid) vir die gebruiker aangetoon wanneer hy 'n spelfout begaan. Wanneer teks direk gegenereer word (eerder as oorgetik word), kan laasgenoemde vir die gebruiker steurend wees. In sommige programme kan die gebruiker kies watter een van die twee benaderings hy wil volg.

Die basiese werkverrigting van spelkontroleprogramme vind op een van die volgende wyses plaas (vergelyk Hughes, 1987f, pp. 2-3 vir dié onderskeiding en vir 'n oorsig oor dié soort programme):

Model A (brute krag):

- Die unieke woordvorme in die teks word geïdentifiseer.
- Die woorde word gesorteer en duplikate word tydelik uitgeskakel.
- Die woorde word teen die woordeboek gekontroleer: eers teen die basiese woordeboek en dan teen die gebruikerswoordeboek.
- Die interaksie met die gebruiker vind plaas (aanduiding van spelfoute, aanbieding van voorstelle, toevoegings tot die gebruikerswoordeboeke).
- Die veranderinge word in die dokument aangebring.

Model B (sekwensieel, met algoritmes):

- Die woorde van die teks word liniêr deur die teks beoordeel. Soms word daar gebruik gemaak van tegnieke wat as "intelligent" beskou sou kon word (vergelyk hieronder).
- Indien 'n woord nie in die woordeboeke voorkom nie en ook nie deur die deur een van die algoritmes hanteer kan word nie, word dit gemerk en word die gebruiker die geleentheid gegee om dit te ignoreer of dit in 'n gebruikerswoordeboek op te neem.

Hughes noem drie losstaande programme (dit wil sê nie ingebou in woordverwerkingsprogramme nie) van die tipe Model B.

Een van die voordele van die programme van Model B is dat die gebruiker gewoonlik die volle konteks van die woord kan sien.

Dit is juis by laasgenoemde soort spelkontroleprogramme dat daar dikwels ook 'n mate van intelligensie voorkom. Hiermee word bedoel dat woorde wat nie presies net so in die woordeboek voorkom nie, ontleed kan word in stam (*stem*) en affikse en daar dan beoordeel kan word of die woord korrek gespel is op grond van die feit dat die stam in die woordeboek voorkom en dit 'n toegelate kombinasie van stam en affiks is. Vergelyk Hughes (1987f, pp. 2-3) vir 'n oorsig oor die verskillende wyses waarop "intelligente" programme funksioneer. Een program wat vir Duits beskikbaar is, naamlik *Othocheck*, het ook nog 'n saamvoegingsfunksie: ook woorde wat saamgestel is uit twee of meer woorde wat afsonderlik in die woordeboek voorkom, word as korrek aanvaar (Uhlig-Döring, 1988, p. 41). So 'n funksie kan ook vir 'n taal soos Afrikaans nuttig wees.

Afgesien van die basiese werkverrigting, is die grootste verskille tussen spelkontroleprogramme die volgende (Hughes, 1987f, p. 2):

- a) die aantal inskrywings in die woordeboek,
- b) die verhouding van die aantal woordeboek-inskrywings tot die grootte van die woordeboek (wat die spoed bepaal),
- c) die grootte van die program,
- d) die spoed,
- e) die geheuebehoefte en
- f) die "intelligensie" van die program (vergelyk hierbo).

Die grootte van die woordeboek moet telkens geïnterpreteer word met inagneming van die taal waaroor dit gaan en die beginsel waarop die program funksioneer. 'n Woordeboek vir Duits wat op die beginsel werk dat alle woorde volledig in die woordeboek moet voorkom, kan byvoorbeeld baie groter wees as dié vir ander tale, ten einde vir die talle verboë vorme voorsiening te maak (Uhlig-Döring, 1988, p. 40).

Sommige programme maak ook voorsiening vir woorde wat foneties ooreenstem, byvoorbeeld dat alternatiewe uit die woordeboek voorgestel word op grond daarvan dat die woorde eenders uitgespreek sou word (wat dan moontlik die oorsaak van 'n spelfout sou kon gewees het). Dié benadering staan bekend as die *Soundex*-benadering (Salton, 1989, p. 431). Dit word in *Volkswriter* en *WordPerfect* geïnkorporeer (Uhlig-Döring, 1988, p. 43).

Die werkverrigting en gemak van hantering van spelkontroleprogramme kan ingrypend verskil en dit is raadsaam om van programresensies gebruik te maak wanneer 'n besluit geneem moet word. (Vergelyk Hughes, 1987f vir 'n bespreking van 'n aantal spelkontroleprogramme.) In rekenaartydskifte word daar dikwels sulke resensies aangetref.

Soos ook in die geval van tesourusfasiliteite, moet daar met die bevatlikheid en tipiese situasie van die gebruiker rekening gehou word. Indien daar te veel woorde in die spelkontrolewoordeboek opgeneem word, kan sekere spelfoute makliker deurglip, aangesien hulle as gespesialiseerde terminologie op skaars gebiede kan voorkom. Dit kan dus beter wees om gespesialiseerde, vakkundige terminologie in afsonderlike gebruikerswoordeboeke op te neem en dié dan in werking te stel wanneer die teks dit regverdig.

Daar bestaan ook outomatiese spelkorreksieprogramme, waarby die spelfout nie net uitgewys word vir die oordeel van die gebruiker nie, maar deur die program reggestel word (Salton, 1989, pp. 430vv.). Dié fasiliteite kom egter nie voor in die gebruiklike woordverwerkingsprogramme waarmee die gewone gebruiker te make kry nie.

d) *Grammatika- en stylkontroleprogramme*

Daar bestaan ook verskeie programme wat onder die benaming grammatika- en stylkontroleprogramme beskryf kan word (voortaan: GS-programme). Salton (1985, p. 436) praat van hierdie funksie as *syntax and style checking*. Dié programme kom nie voor as komponente van die gewone woordverwerkingsprogramme nie, maar kan afsonderlik aangekoop word. Die benaming "grammatika- en stylkontroleprogramme" is miskien ietwat hoogdrawend as in ag geneem word watter betreklik eenvoudige vlak van verwerking in werklikheid ter sprake is, maar dit is 'n gangbare benaming wat vir almal duidelik is.

Dié soort programme word hieronder die hofie "ondersteuning" ten opsigte van teksgenerering aan die orde gestel, aangesien die aspekte van taal wat hulle aanspreek wel die kommunikasie van die persoon wat teks genereer, beïnvloed en nie net op die uiterlike gerig is nie (soos byvoorbeeld afkappingprogramme).

GS-programme het 'n uiteenlopende graad van gesofistikeerdheid. Sommige beskik oor 'n baie meer gevorderde vlak van sintaktiese analise as ander. EPISTLE maak byvoorbeeld gebruik van 'n volledige sintaktiese analise (Salton, 1989, p. 438). Ander is in 'n groter mate ingestel op die meganiese verwerking van die teksmateriaal. Dit is dié soort programme wat gewoonlik in die mikrorekenaar-omgewing aangetref word. Alle programme van hierdie aard is daarop gebaseer dat hulle die tekslêers wat hanteer word, kontroleer teen 'n vooraf opgestelde lys grammatikale en stilistiese reëls en die gebruiker se aandag vestig op afwykings van dié reëls. In sommige gevalle word dan ook voorstelle gemaak om die teks te verbeter.

Die voorloper van die meeste GS-programme is *Writer's Workbench*, 'n omvattende program of stelsel wat vir die UNIX-omgewing ontwikkel is en wat uit 'n groot aantal afsonderlike programme bestaan. (Vergelyk WWB, 1985; Macdonald et al., 1982 en die bespreking by Salton, 1989, pp. 436-438.) Salton (1989, p. 437) gee 'n oorsig oor 'n aantal van dié kleiner programme, waarvan sommige gebaseer is op klein woordeboeke wat telkens geraadpleeg word). Baie van die meer algemene GS-programme behels dan 'n substel van die prosedures van dié program. Een van die mees gebruiklike programme in die MS-DOS-omgewing is *Grammatik II* (Grammatik, 1987).

Die funksies van *Writer's Workbench* gee 'n goeie aanduiding van die soort kontrole wat in programme van dié aard plaasvind (WWB, 1985 en Salton, 1989, pp. 436-438). Die teks word naamlik op die volgende vlakke hanteer:

- a) **Woordvlak:** Verdagte woorde of swak woordgebruik word onder die gebruiker se aandag gebring.
- b) **Frasevlak:** Hier gaan dit oor verdagte of leë frases, naamwoordelike frases met te veel naamwoorde wat mekaar bepaal, gesplete infinitiewe, ensovoorts.
- c) **Sinsvlak:** Sinne wat te lank of te kort is, word gemerk, asook dié wat te veel werkwoorde of negatiewe het. Sinne waarin die afstand tussen subjek en

werkwoord te groot is, word ook aangedui, asook dié met te veel afhanklike frases.

- d) Paragraafvlak: Die gebruiker word gewys op paragrawe wat te veel passiewe sinne het, gedeeltes met baie komplekse sinne en paragrawe wat 'n lae leesbaarheidsgraad het.

Die aantal kriteria waarvolgens die mikrorekenaar-GS-programme grammatika en styl kontroleer, verskil tussen programme onderling, maar daar is gewoonlik ongeveer twintig kriteria.

Die gewone GS-programme vir mikrorekenaars moet nie beskou word as hoogs gesofistikeerde programme wat die laaste woord oor grammatika en styl kan spreek nie. Dié programme is hoogstens hulpmiddels wat die gebruiker se aandag kan vestig op sekere van die meer opvallende ongerymdhede of oneffenhede in sy teks. Dit is juis belangrik dat die taalgebruiker se bewussyn opgeskerp sal word vir dit wat in die teks moontlik verkeerd kan wees. GS-programme laat ook nog baie foute deurglip – volgens berekening tot 75% van alle grammatikale- en stylfoute (Hüsmert, 1988, p. 60). Omgekeerd geld ook dat sekere van die voorstelle ter verbetering wat gemaak word, self foutief is.

Writer's Workbench is blykbaar baie meer suksesvol en bied die beste moontlikhede van dergelike programme. Salton (1989, p. 438) oordeel soos volg oor dié programme:

Although each component of the Writer's Workbench is conceptually simple, the complete collection of programs constitutes a powerful system that may be more useful, and certainly more cost effective, than alternative approaches based on more sophisticated syntactic and semantic analysis methods.

Die grootste probleem verbonde aan die huidige vlak van GS-programme is die feit dat die konteks nie tot sy reg kom nie (Hüsmert, 1980, p. 60). Die funksies wat wel uitgevoer word, is wel baie nuttig en behoort meer intensief gebruik te word, maar daar is in werklikheid 'n baie meer gevorderde vlak van die analise van grammatika en styl nodig. Dit kan verwag word dat die ontwikkeling van kunsmatige intelligensie (soos van toepassing op die verwerking van natuurlike taal) in dié verband beduidende vordering sal meebring.

In die geval van programme wat gebruik word in gemeenskappe waar mense dikwels in 'n ander taal as hulle moedertaal teks genereer, moet daar juis voorsiening gemaak word vir foute wat tipies in so 'n situasie voorkom. In sulke omstandighede is daar baie dikwels inmenging van die een taal by die ander. In die geval van Afrikaans is daar die probleem van anglisismes. Dié verskynsel kom nie net tot uiting in woordkeuse nie, maar veral in die hantering van woordgroepe (spesifiek die voorsetsels) en woordorde. 'n GS-program kan in dié verband 'n baie nuttige rol speel.

4.2.2 Meganiese generering van teks

Naas die skeppende generering van teks (4.2.1) word die rekenaar ook gebruik vir 'n meer meganiese generering van teks. Vir die verdere verwerking van teks wat op dié wyse

geskep is, kan daar ook grotendeels van dieselfde programmatuur gebruik gemaak word as in die geval van die skeppende generering van teks. Uiteraard verval sekere van die fasiliteite in dié funksie (byvoorbeeld die raamwerkfasiliteite).

Die volgende moontlikhede ten opsigte van die meganiese generering van teks kan onderskei word:

- a) Die oortik van teks vanaf teks wat in 'n ander vorm bestaan (hetsy vanaf geskrewe/ gedrukte teks of vanaf teks in 'n vorm waarin dit op 'n oudio-medium gestoor is).
- b) Die meganiese inlees van teks vanaf 'n getikte/gedrukte vorm (die "optiese lees van karakters" – *optical character reading*). Met betrekking tot hierdie soort programme het daar gedurende die afgelope paar jaar baie groot ontwikkelings plaasgevind. Dié programme is vandag vir gewone mikro-rekenaars (met eenvoudige skandeerders) beskikbaar en handhaaf 'n hoë vlak van akkuraatheid.

Spelkontroleprogramme behou uiteraard hulle funksionaliteit en kan aangewend word vir die kontrole van die teks wat op dié wyse geskep word. By sekere programme vir die optiese lees van karakters kan die sensitiwiteit van die program ingestel word om karakters sonder hulp te interpreteer slegs wanneer daar enersyds geen twyfel is oor die vorm van die karakter nie en andersyds wanneer die woord in 'n vooraf gespesifiseerde woordeboek of tesourus voorkom (bv. in SPOT).

4.3 DIE UITERLIKE VERSORGING VAN TEKS

4.3.1 Vorm en voorkoms

Tipografie het nie net te make met die uitwendige voorkoms van 'n dokument nie, maar kommunikeer ook betekenis. Die mens raak daaraan gewoond dat 'n sekere tipografiese verskynsel 'n sekere betekenis dra en koppel die betekenis dan aan die verskynsel. Daar is alreeds gewys op die belang van stylstelle in die proses van die generering van teks (4.2.1.1a). Dit het geblyk dat stylstelle 'n dubbele funksie vervul, naamlik dié van hulp ten opsigte van die logiese strukturering van teks, sowel as eenvormige formattering en tipografie (waarvoor dié funksie aanvanklik beplan is).

Programme vir woordverwerking en kantoorpublikasie wat vandag beskikbaar is, stel die gebruiker in staat om deur hierdie uiterlike versorging van teks ook die nodige betekenis te kan oordra. Programme vir kantoorpublikasie (asook enkele ander geïntegreerde woordverwerkingsprogramme) kan alreeds beeldmateriaal in die teks hanteer. Dit kom voor asof baie van die beter gehalte woordverwerkingsprogramme ook binnekort vir dié fasiliteite voorsiening sal maak.

4.3.2 Afkapping

Byna alle woordverwerkingprogramme het ook fasiliteite vir die afkapping van woorde (vir byvoorbeeld woordbreuke aan die einde van lyne). Hierdie funksie kan sekerlik ook

met betekenis te make hê— Salton (1989, p. 91) gee die volgende voorbeeld met betrekking tot Amerikaanse Engels: *rec-ord* (naamwoord) en *re-cord* (werkwoord) – hoewel dit in die meeste gevalle hoofsaaklik te make het met die uiterlike versorging van teks.

Daar is in alle tale gevalle waar die afkapping nie volgens reëls kan plaasvind nie. In die meeste fasiliteite vir afkapping word dus die volgende benadering gevolg:

- Die woorde in die teks word eers teen 'n uitsonderingswoordeboek gekontroleer. Indien dit daar gevind word, word die afkappingvoorskrifte van die woordeboek gevolg.
- Indien die woord nie in die uitsonderingswoordeboek gevind word nie, word 'n stel reëls toegepas. (Vergelyk Salton, 1989, pp. 91-92 vir die reëls wat gevolg word in die program TEX.)

Die reëls vir afkapping moet uiteraard telkens opgestel word vir 'n spesifieke taal. Die benadering ten opsigte van afkapping verskil selfs tussen Amerikaanse en Britse Engels. In Amerikaanse Engels vind afkapping plaas op grond van uitspraak (dit wil sê die lettergreepgrense speel 'n belangrike rol); in Britse Engels gee die herkoms van die woord die deurslag (Hughes, 1987f, p. 7). Hughes gee die volgende voorbeelde: Amerikaanse: *democ-racy* en *knowl-edge*; Engels: *demo-cracy* en *know-ledge*. By die gebruik van woordverwerkingsprogramme afkomstig uit die VSA moet hiermee rekening gehou word.

4.4 DIE ORDENING VAN TEKS

4.4.1 Berging – inleidende opmerkings

Uiteraard kom daar by enige rekenaarmatige hantering van data die vraag na vore na die mees geskikte formaat of datastrukture waarin teks (of die afsonderlike woorde en/of grammatikale gegewens) op die rekenaar gestoor moet word. In die meeste gevalle gaan dit oor 'n tegniese vraag, wat die beste deur rekenaarkundiges/-wetenskaplikes beantwoord kan word. Vergelyk Salton (1989) vir 'n bespreking van die lêer-/datastrukture waarin materiaal vir die verskillende soorte tekstoepassings op rekenaar hanteer kan word. Halasz (1988) maak verskeie opmerkings oor die feit dat die regte datastrukture nodig is vir die sinvolle funksionering van hipertekstelsels.

By baie prosesse in verband met die verwerking van natuurliketaal is daar die een of ander vorm van databasis (oftewel databasisbestuurstelsel) ter sprake. Sonder so 'n databasis bestuurstelsel is dit uiteraard ook moontlik om soektogte op data uit te voer (byvoorbeeld met behulp van 'n teksredigeerder) en kan ook sinvolle resultate behaal word. So 'n benadering is egter nie wenslik nie: dit maak nie op die mees doeltreffende wyse gebruik van die rekenaartegnologie nie, bied te veel beperkings en laat nie die soepelheid in die omgang met taalgewens toe wat wenslik is nie.

Tov (1986) beskryf 'n baie unieke versameling data waarin soektogte tot baie onlangs grotendeels eenvoudig deur 'n teksredigeerder plaasgevind het. Nieuwoudt (1989b) toon aan hoe hierdie data in databasisse hanteer kan word om 'n baie groter mate van soepel-

heid te verseker. Deur die materiaal op geskikte wyse in 'n databasis te orden, kan baie gesofistikeerde ondersoekes uitgevoer word, lyste uitgedruk word, ensovoorts. Nieuwoudt toon ook aan dat dié soort verwerking van die gegewens dit moontlik maak om standpunte in te neem of hipoteses te toets op 'n statisties meer verantwoordbare wyse -- 'n aspek wat baie dikwels in dié soort ondersoekes nie tot sy reg kom nie.

4.4.2 Die ordening van taalgegewens in databasisse

4.4.2.1 Inleiding

Met betrekking tot die keuse van 'n databasis is dit raadsaam dat gebruikers hulle sal laat adviseer deur persone wat kundig is oor databasisse -- en veral oor databasisse wat geskik blyk te wees vir rekenaartoeepassings op taal.

By die keuse van 'n databasis is dit toenemend belangrik dat gebruikers rekening sal hou met die belang van standarde wat die toekoms van bepaalde toepassings sowel as die moontlike toekomstige oordrag van data en toepassings tussen verskillende rekenaars kan bepaal. Die rasionaal agter die navolging van standarde is kortliks die volgende:

- Die belegging in programmatuur. Die rekenaargebied verander baie vinnig en daar moet verseker word dat die belegging in programmatuur en in die voorbereiding van die data nie binne die afsienbare toekoms in die gedrang kom nie. Dit geld wat betref beide apparatuur en programmatuur. Baie vervaardigers verdwyn van die toneel of beskik nie oor die kapitaal of vermoë om hulle produkte aan te pas by die nuutste ontwikkelings nie.
- Die oordrag van toepassings. Dit kan nodig wees om vir sekere toepassings oor te skuif na meer kragtige rekenaars en dan kan die probleem ontstaan dat die databasis van die kleiner rekenaar nie op dié rekenaar funksioneer nie of dat die data nie maklik uit die eerste databasis afgevoer kan word nie.
- Die uitruil van data. Dit kan belangrik wees om binne 'n groep (byvoorbeeld 'n akademiese groep) data uit te ruil en die waarskynlikheid is groter dat dié uitruiling sal kan plaasvind indien alle partye hulle in die grootste mate moontlik aan ooreengekome standarde hou.

Standaarde word nie net bepaal deur internasionale organisasies nie. Dikwels is dit net die dinamiek van die mark wat bepaal watter stel spesifikasies of watter produkte tot 'n standaard ontwikkel. Byvoorbeeld die feit dat 'n firma/produkt die eerste op die toneel was, dat 'n groot internasionale mark ontwikkel het, dat daar groot gebruikersvertroue opgebou het. Daar is talle voorbeelde van de facto "standaarde" wat op dié wyse tot stand gekom het, byvoorbeeld: PostScript as bladsybeskrywingstandaard (vergelyk 5.4.1), dBASE III Plus (as databasisbestuurstelsel), SQL as databasisstandaard (Shaw, 1988). Dit is juis hierdie standarde wat die moeilikste is om raak te sien of te beoordeel. Dit vereis 'n openheid ten opsigte van die jongste ontwikkelings en 'n vermoë om tendense te kan raaksien en in terme daarvan te kan beplan (vergelyk 2.2g). Naisbitt (1984) beklemtoon die belang daarvan dat mense in die inligtingsera sal leer om die tendense waar te neem en op

grond daarvan te beplan in terme van die ontwikkelings wat aan die kom is. Die gebruiker wat die eerste keer op dié gebied begin beweeg, moet hom in dié verband laat lei deur persone wat getoon het dat hulle in staat is om die tendense suiwer te kan lees.

Ten opsigte van die keuse van databasisse kan hierdie oorwegings onder andere inhou dat die databasis met die beste funksies nie noodwenig die beste keuse is en die beste toekomsmoontlikhede bied nie.

Die volgende voorbeelde van beplanning met die oog op die oordraagbaarheid van databasistoepassings kan genoem word:

- Die gebruik van 'n SQL-gebaseerde databasis (byvoorbeeld Oracle) op 'n MS-DOS-mikrorekenaar (sodat die data en die toepassingsprogrammatuur mettertyd betreklik gemaklik oorgedra kan word na 'n weergawe van dié databasis op 'n mini- of hoofraamrekenaar).
- Die gebruik van die programmeertaal "C" op die mikrorekenaar (sodat ook die programme mettertyd oorgedra kan word na 'n die groter rekenaar, wat na alle waarskynlikheid ook dié taal sal kan hanteer). Turbo Pascal is baie gewild in die mikrorekenaar-omgewing, maar Pascal verskil grootliks tussen verskillende rekenaar-argitekture. Dit kan meebring dat daar 'n groot mate van herprogrammering moet plaasvind by die oordrag van die programme na 'n groter rekenaar.

Die verwagting is dat dBASE IV ook mettertyd gemaklik op groter rekenaars sal kan funksioneer. Tussen dBASE III Plus en dBASE IV is daar uiteraard nie noemenswaardige oordragprobleme nie.

'n Ander benadering met betrekking tot die feit dat taalmateriaal op verskillende rekenaars geprosesseer moet word, is om die databasis op die groter rekenaar as uitgangspunt te neem en 'n program vir mikrorekenaars te ontwikkel waarop die versameling van die data kan plaasvind, asook voorlopige stadia van verwerking. Die groot verwerking vind dan plaas op die hoofraamrekenaar. Dan word daar van die begin af voorsiening gemaak vir die gemaklike oordrag van die data.

Dié benadering word byvoorbeeld gevolg in verband met vakterminologiedatabasis van die Nasionale Vakterminologiediens. Aan gebruikers wat op die grondvlak besig is met die insameling en verwerking van data word dan 'n mikrorekenaardatabasisprogram beskikbaar gestel wat net die nodige funksies het. Die voorbereiding van omvattende vakterminologielyste vind dan op die hoofraamrekenaar plaas.

4.4.2.2 Veeldoelige databasisse

Daar word van talle veeldoelige databasisbestuurstelsels gebruik gemaak by die verwerking van natuurlike taal. Aangesien die meeste van die groter toepassings (byvoorbeeld die opstel van woordeboeke) aanvanklik in die hoofraam-omgewing plaasgevind het,

word sulke databasisse (of die kleiner weergawes daarvan vir mini- en mikrorekenaars, waar van toepassing) tans ook nog gebruik vir baie van dié toepassings.

Daarnaas word daar in toenemende mate gebruik gemaak van tipiese databasisbestuurstelselprogramme vir mikrorekenaars. Die gewildste program in dié verband is dBASE III Plus (vergelyk Nieuwoudt, 1989a; 1989b). (Vergelyk ook 4.4.2.4 vir die voordele van dergelike databasisbestuurstelsels met hulle eie programmeertale.)

Volteksmateriaal (vergelyk 4.6.1. en 4.6.3) kan nie ewe doeltreffend hanteer word in alle soorte databasisbestuurstelsels nie. In dBASE III Plus kan verwerking nie sonder bykomstige vindingryke programmering plaasvind op die sogenaamde notavelde nie, terwyl die lengte van velde wat wel teks kan bevat waarin die data nog manipuleer kan word, beperk was. In dBASE IV is sommige van dié beperkings opgehef. In die meeste van die groter databasisstelsels kan volteksmateriaal ongeveer net so doeltreffend hanteer word soos veld-georiënteerde data.

4.4.2.3 Taaldatabasisse

Afgesien van die algemene en veeldoelige databasisbestuurstelsels het daar ook 'n aantal databasisprogramme ontwikkel wat spesifiek bestem is vir rekenaartoe toepassings op taal-materiaal. Die voordele van dié programme is dat hulle oor funksies beskik wat tipies deur taalwetenskaplikes benodig word. 'n Voorbeeld is Lbase, wat spesifiek vir dié doel ontwikkel is (SMS, 1987). Vir die klassici en persone wat die Bybeltale bestudeer, word in die besonder voorsiening gemaak: spesiale karakters, die vermoë om teks uit die bestaande teksversamelings gemaklik te kan opneem (byvoorbeeld die TLG-materiaal; vergelyk 2.3c en Hughes, 1987d), die integrasie met ander programme vir taalverwerking (byvoorbeeld met IT, vergelyk 4.5 en Simons & Versaw, 1988). Vir die persoon wat 'n program soek wat pasklaar vir gewone taaltoepassings (sonder baie uitsonderlike vereistes) gebruik kan word, kan dié programme baie nuttig wees.

Baie belangrike en hoogs gesofistikeerde programme (en databasisse) vir die interaktiewe analise van tekste uit die oogpunt van die sintaksis, asook vir die stoor en herwinning van die gegewens, is die Linguist's Workbench (LWB) en Linguistic Database (LDB), 'n program wat in die TOSCA-projek tot stand gekom het (vergelyk Webb et al., 1988 p. 29; Aarts & Van den Heuvel, 1985).

Die voordele van dié soort programme is opsigtelik. Dit is egter ook belangrik om kennis te neem van hul ernstige nadele en beperkings. Die mark op dié gebied is nie so dinamies soos dié vir databasisbestuurstelsels van 'n veeldoelige aard nie, wat meebring dat die verdere ontwikkeling van die programme nie so vinnig en doeltreffend kan plaasvind nie. Die grootste probleem is egter die gebrek aan 'n eie programmeertaal wat meebring dat die programme nie so maklik aangepas kan word by die behoeftes van taalkundiges wat nie deur die opstellers voorsien is nie. Hierdeur word daar ingrypend ingeboet met betrekking tot die soepelheid waarvoor die materiaal juis in 'n databasis gestoor word. Vir die persoon wat spesiale vereistes het, wat voorsiening wil maak vir die grootste mate van toekomstige groei en ontwikkeling in sy toepassings, vir die maksimale benutting van die

rekenaarmatige, verdere verwerking van gegewens, ensovoorts, is dit beter om 'n bekende, veeldoelige databasisstelsel te gebruik (waarvan daar verskeie beskikbaar is).

4.4.2.4 Toepassingsprogramme

Die voordeel van die gebruik van veeldoelige databasisse met 'n eie programmeringstaal is dat die gebruiker gemaklik op grond van so 'n databasis sy eie toepassingsprogramme kan ontwikkel (byvoorbeeld deur die gebruik van dBASE III Plus of dBASE IV). Alle databasisbestuurstelsels maak voorsiening vir sekere funksies (sortering, navraagfunksies of die voorbereiding van rapporte). Dié funksies kan egter slegs in enkele van die behoeftes van die gebruiker voorsien. Baie van dié funksies wat in tipiese taal- of tekstoepassings voorkom, kan nie hanteer word nie.

In baie gevalle word daar in databasisbestuurstelsels voorsiening gemaak vir die gebruik van algemene programmeertale (byvoorbeeld Turbo Pascal). Indien die databasisbestuurstelsel oor 'n eie programmeringstaal beskik, vind die ontwikkeling van toepassingsprogramme uiteraard baie gemaklik plaas, aangesien dié taal dan baie nou inskakel by die algemene werkverrigting van die program.

Dit het geblyk dat daar baie gesofistikeerde toepassingsprogramme op die basis van dBASE III Plus of IV (en met gebruikmaking van dié se programmeertaal) ontwikkel kan word (vergelyk Cook, 1989; Nieuwoudt, 1989a; 1989b).

Dit is nie moontlik om binne die ruimte van dié verslag alle moontlike toepassings op te noem nie. Elke moontlike gebied van die verwerking van taal/teks kan ter sprake kom, hetsy met die oog op navorsing, ontwikkeling of implementering. In sommige toepassings is die databasis meer sigbaar en is die gebruiker duideliker daarvan bewus dat hy met 'n databasis te make het, byvoorbeeld in die geval van leksikografiese prosesse (vergelyk Swanepoel & Morris, 1988). In die geval van hipertekststelsels ervaar die gebruiker nie so duidelik dat daar onderliggend aan die stelsel tog ook 'n gesofistikeerde databasis lê nie (vergelyk 4.7.3).

Dit is nie sinvol om hier 'n oorsig te gee oor die verskillende soorte toepassings wat op databasisbestuurstelsels ontwikkel is nie. Die moontlikhede van die implementering van databasisbestuurstelsels met die oog op die verwerking van natuurlike taal is so groot soos die verbeelding van die navorser!

4.4.3 Konkordansies, woordelyste en woordeskatstatistiek

4.4.3.1 Inleiding

Die opstel van konkordansies is een van die eerste toepassings op tekstmateriaal waarvoor die rekenaar gebruik is (vergelyk 2.1 vir besonderhede oor die konkordansieprojek van Roberto Busa). 'n Groot aantal gedrukte konkordansies wat met behulp van die rekenaar opgestel is, het sedertdien tot stand gekom. Nog baie meer konkordansies is nie gepubliseer nie, maar word voorberei in die vorm van gewone rekenaaruitdruk vir navorsingsdoeleindes aan 'n spesifieke instansie.

Dit is veral Bybelkonkordansies wat wydverspreid in gebruik is by die gewone, nie-akademiese gebruiker. Konkordansies word intensief gebruik deur talle persone in taalkundige en literêre rigtings, maar ook deur persone wat hoofsaaklik in die inhoud van 'n teks of tekskorpus belangstel.

Die opstel van 'n konkordansie is 'n besonder tydrowende taak. Voor die tyd van die rekenaar kon die opstel van 'n konkordansie op 'n omvattende taalkorpus soos die Bybel met handprosessê maklik 'n dekade of meer in beslag neem. Dit is egter ook 'n baie meganiese taak, wat ideaal deur die rekenaar uitgevoer kan word.

Met die koms van die rekenaar het daar vir die opstellers van konkordansies 'n nuwe tydperk aangebreek. Dit het nou moontlik geraak om die proses van verskeie jare in 'n paar uur (en later minute) te laat verloop, met daarby nog die voordeel dat die materiaal ná die aanvanklike proses van voorbereiding telkens opnuut weer in verskillende vorms verwerk kan word, afhange van die behoeftes van die gebruiker.

In dié afdeling word hoofsaaklik aandag gegee aan konkordansies, aangesien woordelyste en woordeskatstatistiek gewoonlik as 'n neweproduk van konkordansieprogramme verskaf kan word.

4.4.3.2 Konkordansies: basiese uitgangspunte en opsies

'n Konkordansie kan beskryf word as 'n toegangsmeganisme tot 'n teks of tekskorpus waarin die materiaal op so 'n wyse georden is (i) dat die gebruiker 'n goeie aanduiding van die voorkoms van 'n woord kan kry (met 'n stuk konteks en aanduiding van die posisie van die woord in die teks) en (ii) dat daar voldoen word aan die vereistes van navorsing/studie van die gebruiker. Hierdie meganisme kan in gedrukte vorm of in elektroniese vorm wees; in laasgenoemde geval is dit dan ook moontlik dat die konkordansie op 'n interaktiewe wyse op 'n rekenaar hanteer kan word.

Daar is verskil van mening of die konteks essensieel is. Sommige persone is van mening dat die eintlik belangrike aspek is dat die teks waar die woord voorkom, weer gevind moet kan word. Daar is egter talle aspekte van die gebruik van 'n woord wat net uit die konteks blyk en waarby die gebruiker dus juis die konteks benodig om verdere beslissings te vel oor die vraag of hy die betrokke teks moet gaan naslaan. (Vergelyk die bespreking by Oakman, 1980, pp. 74-76.)

'n Konkordansie is dus telkens 'n konkordansie vir 'n bepaalde toepassing. Daar is beswaarlik iets soos 'n ideale konkordansie vir alle toepassings; dit is 'n stuk gereedskap wat vir 'n bepaalde toepassing ontwerp word.

Die vereistes van die gebruiker kan uiteenlopend wees, afhange van wat hy met die teks wil doen of watter soort toegang hy tot die teks wil hê. Konkordansies vir die leë (byvoorbeeld Bybelkonkordansies) vorm eintlik 'n tema in eie reg (in terme van gebruikersbehoefte, ensovoorts) en word afsonderlik bespreek (4.4.3.4). Konkordansies vir akademiese gebruik (meer of minder intensief) word benodig deur die meeste persone

wat met teksmateriaal omgaan. Dit is 'n bykans algemene behoefte by navorsers wat met taalmateriaal omgaan om 'n oorsig te kan kry oor die gebruik van 'n bepaalde woord in 'n spesifieke korpus of tydperk, om alle gevalle te vind waar die woord voorkom, ensovoorts. Uiteraard verskil die gebruik van dié inligting by taal- en letterkundiges aan die een kant en persone in ander rigtings wat toegang tot 'n tekskorpus benodig slegs vir naslaandoeleindes, byvoorbeeld regsgeleerdes of historici. Die behoefte aan die orde waarin die materiaal gerangskik moet word, bly egter meesal dieselfde, naamlik op die basis van die leksikale voorraad.

Wat ander besonderhede betref, is daar egter ruimte vir groot variasie, byvoorbeeld met betrekking tot

- die plasing of rangskikking van die teks op die bladsy/uitdruk, insluitende die lengte van die konteks wat verskaf word
- die uitsluiting van sekere woorde (byvoorbeeld partikels)
- die vraag of daar ook statistiese gegewens oor die voorkoms van woorde verskaf moet word (byvoorbeeld by elke woord of opsommend aan die einde van die lys)
- die sorteerorde (byvoorbeeld die vraag of hoofletters en karakters met diakritiese tekens in die sorteerorde saam met die gewone kleinletters geplaas moet word).

Die wyse waarop die finale produk gerangskik word, word nie net deur estetiese oerwegings bepaal nie. Dit kan vir bepaalde gebruikers juis van groot belang wees om die uitdruk op 'n bepaalde wyse gerangskik te hê. Dit is byvoorbeeld moontlik om die trefwoord aan die linkerkant te plaas en die tekste daaronder, links gerangskik; of om die trefwoord in die middel van die bladsy te sentreer en die konteks aan weerskante daarvan te druk.

'n Eenvoudiger vorm van 'n konkordansie, wat egter ook baie nuttig kan wees, is 'n woord-indeks. Dié benaming word gewoonlik gebruik vir 'n eenvoudige, alfabetiese lys woorde wat in 'n teks voorkom, saam met 'n aanduiding van waar die woord voorkom (Hockey, 1980, p. 41).

Voordat 'n konkordansieprogram gebruik word, is dit gewoonlik nodig om die teks op die een of ander wyse voor te berei vir die verwerking daarvan. Die volgende moet of kan byvoorbeeld aangedui word:

- Die verdeling van die teks in onderafdelings (byvoorbeeld boeke, hoofstukke en verse of aktes en tonele), met die oog daarop dat die aanduidings van waar 'n woord voorkom later deur die program aangebring kan word.
- Die onderskeiding van homografe, hetsy wanneer dit gaan oor woorde in verskillende tale in die teks of woorde in dieselfde taal. Dié proses vind

dikwels op rekursiewe wyse plaas, deurdat daar eers 'n voorlopige konkordansie geproduseer word en dan op grond daarvan verdere onderskeidings gemaak word wat in 'n volgende fase korrek hanteer word.

- Aanhalings, indien dié afsonderlik hanteer moet word.
- Die hantering van die morfologiese verskeidenheid in taal: dit is moontlik om woorde vooraf te lemmatiseer of aan te dui dat sekere woorde onder daarmee samehangende woorde gerangskik moet word. So byvoorbeeld in die geval van wisselende spelling of die onreëlmatige werkwoorde in Engels (byvoorbeeld die verskillende vorme van die werkwoord *to go*, wat kan wees *go, went, gone, goes, going*).

Die feit dat dit soveel tyd in beslag kan neem om 'n teks reg voor te berei vir byvoorbeeld konkordansieprogramme – en dat verskillende toepassings verskillende kodering vereis – het daartoe gelei dat daar mettertyd gesoek is na tekstkoderingstandaarde wat deur 'n verskeidenheid toepassings geïnterpreteer kan word (vergelyk 5.4).

Afgesien van die basies alfabetiese, leksikale ordening van die materiaal in 'n gewone konkordansie, is dit ook moontlik om 'n ander orde te kies, afhangende van die behoefte. 'n Retrograde of omgekeerde konkordansie (*reverse concordance*) kan byvoorbeeld van groot waarde wees vir die bestudering van die morfologie van sekere tale of vir poësie. Oakman (1980, p. 77) bespreek toepassings van dié soort konkordansie en verskaf verdere bibliografiese inligting. (Vergelyk ook Hockey, 1980, p. 49.)

Die benaming "grammatikale konkordansie" word ook gebruik, en wel as 'n aanduiding van 'n ordening van die tekstmateriaal uit die oogpunt van grammatikale verskynsels (hetsy morfologies of sintakties). Die benaming "grammatikale konkordansie" is veel-seggend, maar binne die raamwerk van hierdie verslag moet dié soort elektroniese ordening van die materiaal of die uitdruk daarvan eerder beskou word as 'n databasis of uitdrukke daarvan (vergelyk 4.4.2.2). Dit vereis in elk geval 'n mindere of meerdere mate van grammatikale kodering van die materiaal.

Burton (1968) gebruik die term "grammatical concordance" as 'n aanduiding van 'n konkordansie wat ook die klein funksiewoorde insluit (byvoorbeeld *a, the, this*) wat sy intensief benut in haar stilistiese studie van die werke van Shakespeare. Dié woorde word normaalweg uitgesluit uit gewone konkordansies.

Daar is inderdaad 'n groot verskeidenheid moontlikhede wat betref konkordansies, en wel met betrekking tot die teksvoorbereiding, die uitleg van die finale produk, die implementering, die aanpassing by die verwagte gebruik, ensovoorts. Hockey (1980, pp. 41-78) en Oakman (1980, pp. 68-87) bied 'n volledige oorsig in dié verband en bespreek sekere van die meer bekende konkordansies.

Dit moet in gedagte gehou word dat die afvoer wat deur 'n konkordansieprogram gegenereer word baie groot is. Hierdie aspek het implikasies vir die spoed van verwerking, die benodigde ruimte op 'n harde skyf en die hoeveelheid drukwerk. Hughes

toon byvoorbeeld aan dat 'n konkordansielêer wat met **Micro-OCP** gegeneer is 12 keer so groot soos die toevoerlêer was, terwyl die woordindeks 1,9 so groot was en die woordfrekwensielys slegs 0,9 keer so groot (1988b, p. 17).

4.4.3.3 Verskillende konkordansieprogramme

In die volgende bespreking word kortliks aandag gegee aan twee belangrike konkordansieprogramme wat (i) die moontlikhede van sulke programme aantoon en (ii) die verskil tussen gewone bondelverwerking- en interaktiewe konkordansieprogramme uitwys. Verder sal ook aandag gegee word aan die implikasies van toepassings wat besonder eie-soortig is.

a) *Oxford Concordance Package (OCP)*

Die **Oxford Concordance Package (OCP)** is die bekendste konkordansieprogram (Hockey & Marriot, 1984). Dit was baie jare lank die enigste kommersieel beskikbare konkordansieprogram. Dit is ontwikkel deur Ian Marriot en Susan Hockey van die Oxford University Computing Service in 1979 en 1980. OCP het ontwikkel uit 'n vorige program van die OCS, naamlik **COCOA (Word COunt and COncordance Generation on Atlas)**. In 1986-1987 is die eerste program volledig opgedateer en het weergawe 2.0 tot stand gekom.

OCP word wêreldwyd in 22 lande gebruik aan meer as 230 opvoedkundige instellings. Die eerste weergawe was daarvoor bekend dat dit besonder stadig was en groot aansprake op die rekenaar gemaak het. (Vergelyk O'Brien, 1986 vir 'n oorsig oor die ervaring van gebruikers van OCP.)

Vir die doel van 'n bespreking van die funksies en werkverrigting van OCP word die mikrorekenaar-weergawe bespreek.

b) *Micro-OCP*

Micro-OCP (OEP, 1988) is 'n uitgebreide weergawe van weergawe 2.0 van OCP, wat al die funksies van die hoofraamprogram bevat. Die gebruikersraakvlak is egter baie beter. Teks wat vir die een program voorberei is, kan gemaklik na die ander oorgedra word.

Die volgende aspekte van **Micro-OCP** kan uitgesonder word (vergelijk Hughes, 1988b):

- i) Dit vereis 'n mikrorekenaar met minstens 512K geheue en 'n harde skyf.
- ii) Dit is 'n bondelverwerkingsprogram.
- iii) Tekslêers wat verwerk word, moet in suiwer ASCII-formaat wees, wat gemaklik vanuit die gewone woordverwerkingsprogramme voorberei kan word. Die teks moet verder gemerk wees met een van drie verwysingsformate ("verwysing" het betrekking op posisie van die woord in die teks, byvoorbeeld hoofstuk of vers): **COCOA**, vaste formaat of in begin string-formaat (vergelijk Hughes, 1988b, pp. 14-15; Hockey, 1980, pp. 41-75; Oakman, 1980).

- iv) Die gebruiker het die geleentheid om die meeste prosesse van verwerking en sortering te beïnvloed, byvoorbeeld
 - die sorteerorde (enersyds met die oog op spesiale karakters of karakters met diakritiese tekens en andersyds met betrekking tot die vraag of die gewone sorteerorde gevolg moet word of 'n omgekeerde orde)
 - die plasing van sekere woorde onder ander woorde (vergelyk by 4.4.3.2 hierbo); die outomatiese hantering van grammatikale kategorieë of lemmatisering is uiteraard nie moontlik nie.
 - die bepaling dat net woorde met 'n sekere frekwensie hanteer moet word.
- v) **Micro-OCP** skep konkordansies, indekse, woordfrekwensielyste en woordeskatstatistiek en kan dié inligting op die skerm vertoon, uitdruk of in 'n gebruikerslêer plaas vir verdere verwerking.
 - In die geval van 'n konkordansie word die trefwoord en die frekwensie op een lyn geplaas en die tekslyne (verwysing en konteks) op daaropvolgende lyne, met die trefwoord in konteks onder mekaar geplaas. Aan die einde van die konkordansie word volledige inligting verstrekk oor die aantal woorde wat hanteer is, ensovoorts.
 - In die geval van woordeskatstatistiek word 'n groot verskeidenheid inligting verskaf, byvoorbeeld werklike tellings, frekwensies en persentasies (vergelyk Hughes, 1988b, p. 118).

Micro-OCP blyk 'n uitstekende stuk gereedskap te wees vir die voorbereiding van konkordansies, woordlyste en woordeskatstatistiek in 'n verskeidenheid tale. Dit kan sonder meer voorsien in die meeste behoeftes van gebruikers.

c) **WordCruncher**

Namate al hoe meer van die eindgebruikers van die konkordansie ook toegang verkry het tot (eers) terminale en (metertyd) mikrorekenaars met voldoende skyfruimte, het daar in die ontwikkeling van konkordansies 'n toenemende behoefte ontstaan na interaksie met die gegewens. Hiermee is die weg gebaan vir interaktiewe konkordansies. Alreeds by sekere van die vroeëre konkordansies was dit uiteraard moontlik om elektronies toegang tot die materiaal te verkry vanaf 'n terminaal – in 'n sekere sin ook interaktiewe toegang. Die resultate van **Micro-OCP** kan byvoorbeeld agterna op die skerm verder gemanipuleer word. Interaktiewe konkordansies behels egter dat die gebruiker die aard van sy navraag kan verander en as't ware onmiddellik die materiaal uit 'n ander perspektief kan sien. Die daarstelling van 'n gedrukte konkordansie kan nog steeds plaasvind, maar dit word nou 'n ondergeskikte deel van die hele proses van omgang met die teks – aangesien die interaktiewe omgang met die teks soveel meer moontlikhede vir die gebruiker inhou.

WordCruncher is die beste voorbeeld van 'n interaktiewe konkordansie (ETC, 1987). Trouens, dit is in so 'n mate interaktief en gerig op veranderende behoeftes van die

gebruiker dat dit ook beskou kan word as programmatuur vir dokumentherwinning (vergelyk 4.6.3). Dit bied in werklikheid meer as baie van die gewone programme vir dokumentherwinning (vergelyk byvoorbeeld 4.6.3 hieronder) en kan ook nog gedrukte konkordansies voorberei. WordCruncher toon juis aan dat die onderskeidings tussen verskillende soorte programmatuur wat soms gemaak word, kunsmatig is en dikwels bepaal word deur die stand van die tegnologie of deur die perspektiewe, ondernemingsgees en moed van die opstellers.

'n Interaktiewe konkordansie maak dit verder ook moontlik om dié soort navrae te formuleer wat nie in die geval van 'n gedrukte konkordansie moontlik is nie. Byvoorbeeld: Micro-OCP maak dit wel moontlik om Boolse operatore te gebruik in die beheertaal (Hughes, 1988b, p. 16), maar nie in die formulering van 'n navraag gebaseer op bepaalde woord nie; in WordCruncher is dit egter wel moontlik (Hughes, 1987c, p. 2). Hierdeur word totaal nuwe dimensies in die omgang met teks moontlik gemaak.

WordCruncher is die verdere ontwikkeling van 'n konkordansieprogram wat lank reeds deur akademiëci gebruik word, naamlik die BYU (= Brigham Young University) Concordance. Dit word bemark deur die diensmaatskappy Electronic Text Corporation (ETC), wat ook rekenaar-leesbare tekste verskaf wat alreeds voorberei is vir gebruik in WordCruncher. Trouens, 'n seleksie van voorbereide tekste word saam met WordCruncher gelewer.

WordCruncher kan konkordansies skep van uitermate groot tekslêers, wat gesamentlik vier gigagreep groot kan wees. Ten einde 'n geskikte spoed vir interaktiewe werk te kan verseker, is dit nodig dat die teks vooraf geïndekseer sal word. WordCruncher bestaan dan ook uit twee afsonderlike dele of programme (wat ook afsonderlik aangekoop kan word, afhangende van die gebruiker se behoefte). IndexETC word gebruik om indekse te skep uit vryevorm-tekste; ViewETC word gebruik om soektogte op dié geïndekseerde tekste te kan uitvoer.

IndexETC laat die gebruiker toe om die verwysingstelsel te definieer (byvoorbeeld hoofstuk, verse – tot op drie vlakke), die soorteerorde te bepaal, uitsluitingswoorde te definieer en om woordeboek- en tesouruslêers te indekseer (wat later by 'n dokument betrek kan word vanuit ViewETC).

ViewETC maak dit moontlik om bepaalde tekste te kies – in die taal van WordCruncher: *books* vanuit 'n *bookshelf*, alfabetiese lysse van woorde en hulle frekwensie te vertoon, te soek vir woorde en woordkombinasies met behulp van tipiese tegnieke betrokke by inligtingherwinning (byvoorbeeld Boolse operatore, soektogte volgens nabyheid of opeenvolging, wisselkarakters, frases/ woordgroepe; vergelyk 4.6.3a vir die ooreenstemmende Engelse terminologie), woorde en frases te vertoon, verwysings op te volg, verskillende soorte konkordansies te genereer, gebruiker-gedefinieerde woordeboeke of tesouri te implementeer, verslae of enige deel van die teks of van 'n voorbereide konkordansie uit te druk, woordfrekwensie-inligting te verstrek, ensovoorts. ViewETC is die vertoonvenster

van **WordCruncher**: soos hierbo gesien, bied dit baie kragtige fasiliteite vir die kombinerings van woorde om 'n navraag te vorm.

Die volgende aspekte van **WordCruncher** kan verder uitgesonder word (Hughes, 1987c):

- i) **WordCruncher** vereis 'n mikrorekenaar met minstens 512K geheue en 'n harde skyf (1,2 mg skyfruimte word slegs vir die program benodig). Dit is moontlik om van slapskywe gebruik te maak, maar dit belemmer die spoed en gemak van hantering.
- ii) **WordCruncher** is 'n volledig interaktiewe program. Die tekste moet uiteraard vooraf geïndekseer word, wat betreklik baie tyd in beslag kan neem.
- iii) Die program is ten volle spyskaart-aangedrewe. Daar word ongelukkig nog nie voorsiening gemaak vir 'n muis nie. Daar is help-fasiliteite wat telkens gerig is op die funksie waarin die gebruiker hom bevind. Dit is ook moontlik om die funksiesleutels te personaliseer.
- iv) Daar word ryklik gebruik gemaak van vensters, sodat die gebruiker materiaal in verskillende formate, asook verskillende instruksiestelle, gemaklik kan hanteer.
- v) Die naatlose integrasie met ander programme verdien spesiale vermelding:
 - Daar word voorsiening gemaak vir die gebruik van die **Duke Language Toolkit**, wat dit moontlik maak om Griekse, Hebreeuse en Cirilliese tekste te kan hanteer, op die skerm te vertoon en uit te druk; daar is enkele beperkings in **WordCruncher** wanneer nie-Romeinse karakters gebruik word (Hughes, 1987c, p. 6).
 - Daar is 'n baie nuwe interaksie moontlik met **WordPerfect**: die funksies van **WordPerfect** se *Library* kan gebruik word en 'n gedeelte van 'n voorbereide konkordansie kan deur die *clipboard*-fasiliteite van dié program oorgedra word vir insluiting in 'n woordverwerkingsdokument.
- vi) Omvattende statistiese gegewens kan verstrekkend word. eensyds deur **IndexETC** ten opsigte van die indeksering wat plaasgevind het (Hughes, 1987c, p. 5) en andersyds deur **ViewETC** ten opsigte van soektogte wat plaasgevind het of konkordansies wat voorberei is (Hughes, 1987c, pp. 6-7).
- vii) Daar word voorsiening gemaak vir 'n aantal konkordansiestyle op die stadium van vertoon of uitdruk, asook vir allerlei opsies ten opsigte van die finale afwerking.
- viii) Die moontlikheid om woordeboeke en tesouri te kan skep en daartoe toegang te kan kry vanuit **ViewETC** maak dit vir die gebruiker moontlik om sy tipiese werkswyse in die omgang met tekste ook op interaktiewe wyse met **WordCruncher** te volg.

WordCruncher bied op interaktiewe wyse 'n veelvuldigheid van wyses in die hantering van taalgewens wat baie meer is as dit wat uiteindelik in die vorm van 'n konkordansie vorm kan kry. Die moontlikheid van die kombinasie van verskeie kriteria in die soekproses of die vinnige afwisseling van alternatiewe perspektiewe en invalshoeke (*views*) op die korpus, kan net gebied word deur 'n interaktiewe proses. Hiermee verskuif die klem in die implementering van die "konkordansie"-begrip vanaf die daarstelling van 'n gesorteerde lys woorde-in-konteks na die gebruik van 'n meganisme vir die manipulasie van taal wat, as een van vele moontlikhede, die skepping van so 'n lys moontlik maak. In die navorsingsgemeenskap is hierdie 'n bevrydende konsep, aangesien daar nou met 'n korpus materiaal omgegaan kan word op verskeie wyses – beperk slegs deur die verbeelding van die gebruiker.

WordCruncher is 'n besonder kragtige en soepel stuk gereedskap wat baie wyer gebruik behoort te word. Die feit dat een enkele program die funksies van dokumentherwinning/teksherwinning en die opstel van konkordansies kan uitvoer, bring mee dat dit vir baie gebruikers die moeite en koste werd kan wees om dit aan te skaf en te leer gebruik.

d) Konkordansies vir spesiale doeleindes

Een van die grootste probleme verbonde aan baie konkordansieprogramme is dat die gebruiker nie kan ingryp in sekere van die prosesse nie. Dit is meesal nog moontlik om die lengte van die konteks, die plasing van die uitgedrukte produk, die sorteerorde, ensovoorts te bepaal. Vir spesifieke doeleindes kan dit egter nodig wees om ook nog ander parameters te wysig of ander funksies beskikbaar te hê, byvoorbeeld vir die voorbereiding van gedrukte konkordansies wat aan bepaalde vereistes moet voldoen. Aangesien die bronkode nie beskikbaar is nie, is die enigste uitweg om self die nodige programme te skryf vir die manipulering van die data. Die mees sinvolle benadering is om sulke programme te skryf as toepassingsprogramme wat in samehang met 'n algemene en veel-doelige databasis funksioneer.

Die volgende twee voorbeelde toon aan vir watter soort funksies dit nodig kan wees om afsonderlike programme te skryf:

- i) Met die oog op die voorbereiding van 'n gedrukte konkordansie vir Siries was dit nodig om spesiale roetines as toepassingsprogramme van dBASE III Plus te skryf, wat dit moontlik maak om vir die besondere van Siries voorsiening te maak: die keuse van die juiste karakterstelle, die uitdruk van regs na links, ensovoorts. Dié projek word onderneem in die Navorsingeneheid vir Rekenaartoe toepassings op die Taal en Teks van die Ou Testament, Universiteit van Stellenbosch (vergelyk Nieuwoudt, 1989a).
- ii) Vir die voorbereiding van 'n konkordansie op die Nuwe Afrikaanse Vertaling van die Bybel vir algemene gebruik (vergelyk 4.4.3.4) is dit nodig om voorsiening te maak vir funksies wat nie in gewone konkordansieprogramme vervat is nie: die insluiting van woordgroepe, die massering van die konteks ten einde die maksimale sinvolle konteks, met inagneming van karakterwydtes

op die setstadium, te bereik. Die programme hiervoor moet ongelukkig spesiaal geskryf word. Dié projek word eweneens onderneem in bogenoemde Navorsingseenheid.

'n Ander geval wat die aandag verdien van persone op die gebied van die Bybeltale of die Klassieke (en wat sekerlik ook vir ander persone van groot waarde kan wees) is die IBYCUS-rekenaar – 'n gededikeerde rekenaar met die nodige programmatuur vir die navorser op dié gebiede (vergelyk Hughes, 1986a). Daar word ook programmatuur ingesluit vir die opstel van konkordansies wat volledig rekening hou met die eiendommelikhede van Grieks, Hebreeus en Siries. IBYCUS is egter 'n geslote stelsel en gewone MS-DOS-programmatuur kan nie gebruik word nie. Daar bestaan gevolglik 'n groot gevaar dat die gebruiker uitgesluit word van die dinamiese ontwikkelings wat tipies meegebring word deur 'n vrye mark en 'n algemeen aanvaarde bedryfstelsel wat programme vir telkens nuut ontwikkelende gebiede vinnig en betreklik goedkoop ontwikkel.

4.4.3.4 Konkordansies vir die leek

Bybelkonkordansies blyk baie gewild te wees by professionele persone op die gebied van die teologie, sowel as by die leek. Die aard van dié tekskorpus en die gebruik wat tipies daarvan gemaak word, dra by tot dié gewildheid.

Vir die kundige gebruiker van 'n konkordansie – kundig in terme van die eiendommlike van taal en die verwagtinge wat aan die meganisme "konkordansie" gestel kan word – is 'n konkordansie 'n baie nuttige en kragtige instrument. Dit blyk egter dat die konkordansie totaal oorskakel word deur die gewone gebruiker van 'n konkordansie in die alledaagse lewe – veral deur die gebruikers van 'n Bybelkonkordansie. Die meeste gebruikers wil toegang tot die teks verkry, ten einde daar verder na te lees oor 'n bepaalde begrip. Aan die begrip wat die doel van die soektog is, word egter in die meeste gevalle een enkele woord gekoppel en dan word die ongeoorloofde sprong gemaak om te dink dat wanneer alle gevalle van die woord gevind is, ook alle gevalle van die betrokke begrip gevind is. Byvoorbeeld: die gebruiker soek inligting oor die begrip "iemand van die toestand van lewe na die toestand van dood verskuif" en koppel dan daaraan die woord "doodmaak". Die soektog in die konkordansie lewer dan byvoorbeeld twintig tekste waar dié woord voorkom en die gebruiker verkeer onder die indruk dat hy daarmee al die gevalle opgespoor het waar dit gaan om die begrip wat hy eintlik gesoek het. In werklikheid moes hy ook gaan soek het onder woorde of woordgroepe soos: moord pleeg, ombring, iemand se lewe neem, iemand van kant maak, doodslaan, doodsteek, ensovoorts.

Wat in dié geval van die oorskakeling van die meganisme konkordansie gebeur, is dat dit uit die oog verloor word dat die konkordansie eintlik maar net die herrangskikking van die woord van 'n bepaalde tekskorpus volgens 'n bepaalde gesigspunt is. Op meer as dit kan 'n konkordansie nie aanspraak maak nie.

Dit kom voor asof dié gesigspunt nie eers deurgedring het tot die opstellers van die konkordansies nie, te oordeel daaraan dat die een konkordansie na die ander eenvoudig

maar op die ou trant voortgegaan het sonder om eers die gebruikers in 'n voorwoord in te lig oor die beperkings van 'n konkordansie of oor bogenoemde gevaar.

Een manier om hierdie probleem in 'n beperkte mate te omseil, is om die gebruiker by die betrokke woorde daarop te wys dat daar ook ander woorde is wat vir (ongeveer) dieselfde begrip gebruik word en dat dié woorde ook geraadpleeg kan word.

Die beste voorbeeld van hierdie soort hulp is die konkordansie op die *Good News Bible*, wat in die geval van sommige sleutelwoorde vir die gebruiker ander woorde aantoon waaronder hy kan gaan soek. Verder maak dié konkordansie ook voorsiening vir die feit dat die "woordeskat" van die gebruiker dikwels gekondisioneer is deur ander (gewoonlik ouer) Bybelvertalings. Op dié wyse word daar dan verwysings gemaak na woorde in die nuwe vertaling vanaf woorde in die ouer vertaling. Ongelukkig laat dié konkordansie ook nog nie reg geskied aan woordgroepe nie en die gebruiker moet 'n bepaalde woordgroep maar dikwels onder die mees prominente woord van die woordgroep gaan soek tussen ander alleenstaande woorde.

Konkordansies behoort rekening te hou met die basiese onderskeiding tussen begrip en woord, op so 'n wyse dat die gebruiker gehelp word om reg te laat geskied aan die eie aard van taal. *WordCruncher* maak dit byvoorbeeld vir 'n gebruiker moontlik om vanuit 'n dokument toegang te kry tot 'n woordeboek of tesourus wat hy self opgestel het (vergelyk 4.4.3.3c en ETC, 1987, p. 103).

Hierdie probleem word nog baie vererger deur die gebruik van interaktiewe, elektroniese Bybelkonkordansies wat toenemend beskikbaar raak. In dié konkordansies is dit nog baie makliker om voorsiening te maak vir 'n ingeboude tesourus wat afgestem is op die terminologie wat tipies in Bybelvertalings voorkom.

4.5 DIE KOMMENTARIËRING OF ANNOTASIE VAN TEKS

Dit is uiteraard moontlik om teks op so 'n wyse te kommentariseer dat daar 'n nuwe teks tot stand kom waarin die teks wat gekommentariseer word dan betrek word deur die aanhaling van frases of deur 'n stelsel van verwysingsnommers. Dikwels word verskillende soorte of vlakke van kommentaar op een bladsy gevind, byvoorbeeld in die *Interpreter's Bible Commentary*. 'n Sinvolle ordening van die verskillende vlakke van inligting kan dan bereik word deur van die een of ander vorm van tipografiese aanduidings gebruik te maak (byvoorbeeld verskillende paragraafformate of lettertipes). Bybelkommentare bied voorbeelde van die mees uitgebreide soorte kommentaar: in sommige gevalle is daar verskillende vlakke van kommentariëring, byvoorbeeld onderskeidelik oor die teks, die literêre tipe, die historiese agtergrond, die toeligting van die afsonderlike verse in opeenvolgende orde, die boodskap, ensovoorts. Juridiese tekste bied dikwels ook dieselfde omvang en verskeidenheid van kommentariëring. Vir die gebruiker (= leser) van die kommentaar bied hierdie wyse van ordening van die gegewens gewoonlik alles wat hy benodig. In hierdie proses kan daar dan van al die moontlikhede

van woordverwerkingsprogramme gebruik gemaak word (4.2.1.1 hierbo). Wat tot stand kom, is telkens 'n nuwe teks.

In die geval van taalkundige, letterkundige of antropologies-kulturele navorsing wil die navorser dikwels kommentaar op die teks lewer op so 'n wyse dat hy ook weer agterna toegang tot die betrokke kommentaar kan kry, dit kan orden (byvoorbeeld sorteer), ensovoorts. Hierdie kommentaar kan gelewer word ten opsigte van 'n hele sin, 'n frase, 'n woord of selfs op suprasegmentele vlak. Dit is ook moontlik om ten opsigte van 'n enkele woord kommentaar op verskeie vlakke te lewer, byvoorbeeld: grammatikale vorm, betekenis, kombinerbaarheid, die verhouding tot ander woorde, vertalingsekwivalente. Hierdie soort kommentaar staan bekend onder verskeie name, byvoorbeeld "annotasie" of "kommentaar". Die gebruik van dié terme is nie konsekwent nie, maar dit kom voor asof die term "annotasie" gebruik word wanneer die inligting wat op dié wyse aan woorde of teks gekoppel word langerig raak en "kommentaar" wanneer dit omvangryk raak, aldus Van der Merwe (1988, pp. 242-243). Hy bespreek die agtergrond en praktyk van teks-annotasie breedvoerig.

Uiteraard kan al die verskillende soorte annotasies wat op dié wyse aan woorde of teks gekoppel word, in 'n databasis gestoor en weer herwin word. Die gebruik van 'n gewone databasis vir dié soort proses kan egter baie omslagtig wees.

Die program IT (Interlinear Text Processing) is geskep om hierdie soort kommentariëring van teks, en die benutting van die materiaal agterna, moontlik te maak. Dit is geskep vir die gebruik van taalkundiges en Bybelvertalers verbonde aan die Summer Institute of Linguistics (SIL) (Dallas, Texas). Vergelyk 4.8 vir meer besonderhede oor die tipes toepassings waarvoor daar by die SIL voorsiening gemaak word.

Die skeppers van IT wil reg laat geskied aan die multidimensionaliteit van teks en die gebruiker (die taalontleder) die geleentheid gee om op alle dimensies kommentaar te lewer.

Text is a multidimensional object, and IT is a tool which allows the analyst to treat it as such. A conventional text editing program views text as a simple sequence of characters. But from a linguist's perspective, text is far more than that. We may write down a text as a one-dimensional sequence of characters, but the object we are thereby representing is in fact a multidimensional one. A given stretch of speech can be simultaneously viewed in terms of its phonetic interpretation, its phonemic interpretation, its morphophonemic interpretation, its morphemic interpretation, or its lexemic interpretation. The meaning of the text operates in many dimensions as well (Simons & Versaw, 1988, p. 2-2).

IT laat interliniëre annotasie toe – na analogie van die tipiese interliniëre teksuitgawes of die handgeskrewe notas of annotasies van taalkundiges in gedrukte tekste: die gebruiker sien 'n lyn teks op die skerm en kan dan 'n aantal lyne annotasie daar onder aanbring.

Daar word voorsiening gemaak vir die volgende soorte inligting:

- a) Identifikasie van die eenheid onder behandeling (*unit identification*): dit definieer die korpus en die teksgrense (byvoorbeeld hoofstuk en verse).
- b) Basislyn-tekse (*baseline text*): die teks wat geannoteer word.
- c) Opgelynde annotasies (*aligning annotations*): vertikaal opgelynde annotasies (morfeem-tot-morfeem of woord-tot-woord) ten opsigte van die basislyn-tekse.
- d) Vryevorm-annotasie (*freeform annotations*): annotasies van 'n hele eenheid van basislyn-tekse of annotasies ten opsigte van ander annotasies.
- e) Buite-velde (*extraneous fields*): enige ander velde van inligting.

In die geval van opgelynde annotasies op die basis van morfeem-tot-morfeem of woord-tot-woord sorg die program dat die annotasies vertikaal opgelyn word en bly (Simons & Versaw, 1988, p. 2-16). In dié vorm word dit ook in die databasis gestoor vir toekomstige verwerking. Die hantering van die program word vergemaklik deurdat daar van die "sneeubal-effek" gebruik gemaak word: indien 'n woord wat alreeds een keer geannoteer is, weer voorkom, word die vorige annotasie uit die opgeboude databasis geneem en as 'n voorstel aan die gebruiker aangebied. Dit staan die gebruiker vry om dié voorstel te gebruik (wat dan deur die druk van 'n enkele toets plaasvind) of om 'n ander voorstel te maak. Op dié wyse kan die verwerking van die materiaal al hoe vinniger verloop namate die proses vorder, terwyl dit tog nie oorgelaat word aan outomatiese prosesse nie.

Die materiaal wat in IT geskep word, kan agterna gemaklik verder verwerk word:

- a) gesorteerde lysse daarvan kan voorberei word,
- b) dit kan afgevoer word na die taalkundige databasis Lbase (4.4.2.3), waarin dit sonder meer opgeneem word, of
- c) dit kan verwerk word vir opname in 'n algemene databasis.

Dit blyk dus dat IT daarmee erns wil maak dat teks iets meer is as net die liniêre opeenvolging van woorde. IT wil die geleentheid skep om dié teks te kommentarieer op 'n wyse wat inpas by die tipiese aktiwiteite van die navorser. IT is 'n baie nuttige program wat vir 'n groot aantal navorsers baie goeie moontlikhede bied. Dit is duidelik dat IT 'n program-in-ontwikkeling is en dat daar in die toekoms vir nog meer sinvolle funksies voorsiening gemaak sal word. Vergelyk 4.8 oor die CELLAR-omgewing waarin daar mettertyd ook vir die funksies van IT voorsiening gemaak sou kon word.

Die gegewens wat deur annotasie of kommentariëring geskep word, kan op 'n baie sinvolle wyse interaktief aan die gebruiker beskikbaar gestel word deur van hipertekse gebruik te maak (vergelyk 4.7.3, en Van der Merwe, 1988). Uiteraard kan annotasie ook gedoen word in 'n hipertekseprogram, maar dan net in die vorm van vrye brokke teks en nie op so 'n wyse dat die annotasies opgelyn word vir latere ordening soos in IT nie.

4.6 TOEGANG TOT DATABASISSE

4.6.1 Inleiding

In 4.4 is aandag gegee aan meganismes vir die ordening van teks, hetsy in die vorm van konkordansies of in databasisse. Dit het geblyk dat dit moontlik is om die teks in haas enige vorm op te breek en te stoor, wat agterna die moontlikheid van toegang tot dié materiaal bied volgens 'n groot verskeidenheid kriteria en kombinasies.

Dit gebeur tans in toenemende mate dat ook die volle teks van dokumente in 'n databasis gestoor word. Voltekstdatabasisse of dokument-databasisse (vergelyk hieronder vir terminologie) het gedurende die afgelope paar jaar 'n al hoe meer bekende en gewilde rekenaartoevoeging geword. Daar was uiteraard al geruime tyd toepassings van dié aard. Die ontwikkeling van die tegnologie gedurende die afgelope dekade, met name die feit dat die skyfbergingskapasiteit van rekenaars al hoe groter geword het en die volle teks van dokumente geredeliker in reenaar-leesbare vorm beskikbaar was, het daartoe gelei dat hierdie toepassings so gewild geraak het. Die koms van CD-ROM het dié ontwikkeling versnel en weer opnuut die aandag daarop gevestig dat die mees doeltreffende tegnieke aangewend moet word om toegang tot data in die verskillende soorte databasisse te verkry. In die geval van persoonlike rekenaars was die grootste beperkings voor die koms van CD-ROM die onbetroubaarheid en stadige spoed van kommunikasielyne. Deur 'n CD-ROM-leser aan 'n persoonlike reenaar te koppel, kan 'n baie hoë oordragspoed van data behaal word. Vir die belang van CD-ROM, vergelyk die talle artikels in Lambert & Ropiequet (1986) en Ropiequet et al. (1987).

Databasisse van die aard wat hier bespreek word, word lank reeds gebruik in inligtingkundige toepassings. Dit is nie die bedoeling met die bespreking in dié afdeling om die hele terrein van databasisse, hetsy uit 'n inligtingkundige oogpunt of uit die oogpunt van die rekenaarmatige prosesse, betrokke daarby, in besonderhede te bespreek nie. Soos egter elders aangedui (4.7.3.2), speel die aspek van die rekenaarmatige hantering van taal 'n al hoe groter rol in die inligtingkunde. Dit is dus noodsaaklik om dié aspekte ook hier ter sprake te bring.

Terminologies is dit belangrik om op die volgende te let:

- * Verskeie Engelse terme word gebruik met verwysing na databasisse waarin tekste volledig gestoor word: *full-text databases* (Tenopir, 1984; 1988), *document databases*, *document retrieval systems* (Zoellick, 1987b), *text bases*, *text base programs*, *text retrieval programs* (Hughes, 1987c). Voorkeur word in dié verslag gegee aan die term "voltekstdatabasisse", bo "dokument-databasisse".
- * Daar moet duidelik onderskei word tussen voltekstdatabasisse en "dokument-as-beeld-databasisse" (*document image databases - DIP*). In eersgenoemde geval word die volle teks van die dokument as elektroniese teks (in karakterwaardes) gestoor en kan die teks byvoorbeeld vir 'n bepaalde woord deursoek

word; in laasgenoemde geval word die teks as beeld digitaal gestoor (die beeld sou net sowel 'n illustrasie kon gewees het) en kan die teks nie sonder meer elektronies deursoek word nie. Laasgenoemde soort databasisse is nie ter sprake in dié verslag nie, hoewel dit nie as onbelangrik beskou moet word nie.

Die prosesse van inligtingherwinning kan ingrypend verskil tussen gevalle waar dit gaan oor dataherwinning enersyds en dokumentherwinning andersyds. Dié twee prosesse kan uiteraard nie só afgegrens word dat die een slegs van toepassing is op "gewone", veldgeoriënteerde databasisse en die ander op voltekst-databasisse nie. Die prosesse kan immers ook in een databasis verbind word en in samehang met mekaar plaasvind. Tog is dit so dat dié onderskeiding breedweg geld.

Die prosesse van dokumentherwinning verskil inderdaad ingrypend van dié van dataherwinning. Dié verskille word baie goed aangedui deur Blair (1984). Dit kom breedweg daarop neer dat dataherwinningstelsels die inligting (kan) bied wat gesoek word, terwyl dokumentherwinningstelsels 'n lys dokumente verskaf (of ook die teks van toepassing) waar die inligting moontlik gevind kan word. In die een geval gaan dit oor data wat in die vorm van vaste waardes voorgestel word (syfers, name, adresse, deskriptore, ensovoorts); in die ander geval gaan dit oor 'n resultaat wat in 'n meerdere of mindere mate van waarskynlikheid relevant of bruikbaar kan wees. Die vereiste van bruikbaarheid bring egter weer mee (i) dat meganismes in werking gestel moet word om 'n waarde te heg aan die graad van bruikbaarheid van die dokumente wat gevind word, (ii) dat die konteks betrek moet word en (iii) dat die gebruiker (= soeker van inligting) in staat gestel moet word om self 'n oordeel te vel.

Tussen dié twee prosesse (en die daarmee gepaard gaande soorte databasisse) is daar uiteraard baie raakpunte en dit is moontlik om van dieselfde soort meganismes gebruik te maak, byvoorbeeld van tesouri wat dit moontlik maak om die nodige terminologiese ooreenstemming tussen die gebruiker en die stelsel teweeg te bring of 'n navraag te presiseer (vergelyk 4.6.4 oor tesouri).

Dit is veral met betrekking tot dokumentherwinning dat daar 'n verdere onderskeiding van belang getref moet word: dit gaan in dié geval hoofsaaklik daaroor dat die gebruiker die dokumente moet vind wat waarskynlik relevant is ten opsigte van sy navraag, sodat hy verder gemaklik al lesende kan "rondblaai" (*browse*) en die werklik relevante dokumente kan vind. Die twee prosesse van soek (*searching*) en rondblaaier (*browsing*) word dan onderskei. Daar is verskillende soorte rondblaaier deur dokumente, waarvan sommige vir sekere toepassings meer of minder geskik is, byvoorbeeld sekwensiële rondblaaier of rondblaaier gebaseer op trefwoordtoegang (*keyed-access browsing*). Vir 'n goeie oorsig oor die verskillende soorte rondblaaier, vergelyk Zoellick (1987b) en Fand (1986).

'n Besondere soort rondblaaier wat gedurende die afgelope paar jaar baie aandag geniet het, is "geskakelde rondblaaier" (*linked browsing*). Vanuit die oogpunt van die toegang tot databasisse kan hiperteks as 'n soort "geskakelde rondblaaier" beskou word. Omdat

die term "rondblaaier" die karakter dra van op-pad-wees-na-die-relevante-inligting, terwyl hiperteks (in die geval van 'n behoorlik gestruktureerde hiperteks-databasis) veronderstel dat die nuwe inligting wat gevind word, telkens alreeds relevant is en dat die gebruiker net inligtingverkenning al hoe verder deur die inligting wil beweeg, word hiperteks in dié verslag as iets anders beskou as gewone geskakelde rondblaaier. Hiperteks word veeleer beskou as 'n interaktiewe stelsel vir inligtingverkenning en onder die afdeling "teks-aanbieding" verder bespreek (vergelyk 4.7 en veral 4.7.3).

Dit is vanselfsprekend dat dit die doel van enige goeie inligtingherwinningstelsel is dat die werklik relevante dokumente (of onderdele van dokumente) gevind sal word, en juis ook nêr dié. In die ontwikkeling van stelsels word intensief gesoek na die optimale oplossing vir dié uitdaging, onder andere deur die implementering van die meer gesofistikeerde tegnieke, veral ook dié wat op 'n intelligente wyse met die taalgegewens wil omgaan. Solank dié stelsels nog nie die optimale graad van doeltreffendheid bereik het nie – en dit kan inderdaad nog betreklik lank duur – moet die gebruiker aanvaar dat 'n tegniek soos rondblaaier ook alreeds 'n baie doeltreffende wyse van die omgang met teks-inligting is. Dit is juis deurdat die rekenaar dit moontlik maak om (i) vinnig toegang tot waarskynlik relevante dokumente te kry en dan (ii) gemaklik die tegniek van rondblaaier te kan uitvoer, dat die rekenaar die proses van toegang tot databasisse ondersteun en op 'n nuwe vlak plaas. Ook hierdie oënskynlik beperkte moontlikhede van die rekenaar moet intensief benut word.

Dit is interessant dat ook op die gebied van toegang tot visuele databasisse dit belangrik blyk te wees om die menslike beeldherkenningsmoontlikhede ten volle te ontgin, met die maksimale mate van rekenaar-ondersteuning in die proses (Batley, 1988). Beskrywing van voorwerpe op beelde met behulp van trefwoorde is nuttig, maar nog nie voldoende nie. Dit is moontlik dat daar mettertyd 'n gevorderde graad van patroonherkenning kan ontwikkel, maar tot tyd-en-wyl kan die rekenaar ook wel ingespan word om die baie doeltreffende menslike beeldherkenningsvermoë optimaal te ondersteun.

In dié afdeling sal kortliks aandag gegee word aan die verskillende soorte toegang tot databasisse, met spesifieke verwysing na aspekte waar taalverwerking op die spel is. Noodwendig moet dan ook aan inligtingkundige tesouri aandag gegee word. 'n Voltekst-databasisstelsel vir mikrorekenaars sal as voorbeeld bespreek word. Daar sal nie aandag gegee word aan al die rekenaarwetenskaplike tegnieke wat ter sprake kom met betrekking tot die opstel van indekslêers, ensovoorts nie.

4.6.2 Dataherwinning

So uiteenlopend soos die struktuur van data is waarvolgens dit in 'n databasis gestoor kan word (4.4.2), so uiteenlopend kan ook die toepassings wees wat daarop gebou word, en die soorte toegang tot die data.

In eenvoudige databasisse gerig op dataherwinning gaan dit gewoonlik om die herwinning van spesifieke data, byvoorbeeld die bevolking van 'n land, die naam van 'n voorwerp wat korrelleer met 'n itemnommer, die prys van 'n item, ensovoorts. In die geval

van 'n taaldatabasis gaan navrae oor die teksaanduiding waar 'n bepaalde woord voorkom, die etimologie van 'n woord, die grammatikale vorm van 'n woord, ensovoorts. Die items wat in die databasis toeganklik gemaak word, word in dié geval beskryf met betreklik eenduidige kodes of vaste waardes. In baie gevalle kan items egter nie anders beskryf word nie as deur terminologie wat nie so eenduidig is nie, byvoorbeeld in die geval van bibliografiese items, beelde, dokumente, ensovoorts. Dan word deskriptore of trefwoorde aan die dokument toegeken as beskrywing van die inhoud. Dié terminologie kan baie minder eenduidig wees en daar kan groter kans wees vir verwarring. Daar is nou naamlik drie fasette/instansies op die spel, naamlik (i) die indekseerder, (ii) die soeker van inligting en (iii) die stel moontlike terminologie en die komplekse verhouding tussen die terme onderling, asook die verhouding tussen term en begrip (vergelyk 4.4.3.4). Dit is hierdie kompleks van verhoudinge wat meebring dat dit baie dikwels noodsaaklik is om by die herwinning van inligting in databasisse van dié aard gebruik te maak van 'n gekontroleerde en gestruktureerde stel terminologie, naamlik 'n tesourus (vergelyk 4.6.4 vir besonderhede oor tesouri).

In baie databasisse wat vir die algemene gebruik beskikbaar is, word daar hoegenaamd nie (of nie voldoende) voorsiening gemaak vir sulke tesouri nie.

4.6.3 Dokumentherwinning

a) *Algemeen*

Voltekstatabasisse word vir 'n wyd uiteenlopende reeks toepassings gebruik. Tenopir (1988, p. 266) stel dit soos volg:

Full text databases are being used for: document location (like bibliographic databases are used), fact retrieval, paragraph or peripheral material retrieval, citation searching, and known-article retrieval.

In dié artikel gee Tenopir 'n baie duidelike en omvattende oorsig oor die gebruike van voltekstatabasisse en die gebruikerspatrone en -praktyke. (Meer inligting kan verkry word in Tenopir, 1984.)

Gedurende die afgelope drie jaar het daar ook 'n groot aantal voltekstatabasisstelsels beskikbaar geraak, waarvan sommige baie gewild blyk te wees. Vir besonderhede, vergelyk die gereelde bespreking van hierdie programme/stelsels in die rekenaar-tydskrifte. (Vergelyk ook Hughes, 1987e vir 'n algemene oorsig oor dié programme.)

Dit het hierbo geblyk dat konkordansies in 'n sekere mate ook beskou kan word as 'n stelsel vir teksherwinning. Dit is veral die interaktiewe konkordansies wat dieselfde soort funksies as dokumentherwinningstelsels bied en wat dan ook vir dié doel gebruik kan word – en wat dit in sekere gevalle op uitmuntende wyse doen (vergelyk 4.4.3.3c). Die grootste verskil lê daarin dat die meeste konkordansieprogramme ook nog 'n ander soort manipulasie van die teks toelaat as wat tipies in dokumentherwinningstelsels plaasvind.

Twee vlakke van die omgang met dokumente in 'n dokumentherwinningstelsel kan onderskei word, naamlik dié van die indeksering en dié van die uiteindelijke herwinning. In die meeste gevalle word die volledige korpus van dokumente nie eers tydens die proses van soek en herwinning deurgesoeek nie, maar vind daar 'n vorm van indeksering plaas in die stadium dat die dokumente in die databasis opgeneem word. Vir die resultate wat later behaal word in die stadium van soek/herwinning, is dit dus van deurslaggewende betekenis dat die dokument op so 'n wyse in die databasis opgeneem word, dat dit in die latere stadium die mees doeltreffende vlak van herwinning moontlik maak. Doeltreffendheid beteken nie net die grootste aantal dokumente nie, maar die ideale balans (telkens in terme van die behoeftes van die betrokke soeker van inligting) tussen noukeurigheid en herwinning. Noukeurigheid (*precision*) is die mate waarin 'n soektog slegs die relevante items vind; herwinning (*recall*) is die mate van deeglikheid of uitputtendheid waarmee alle relevante items in 'n databasis wel gevind word.

Daar is al baie aandag gegee aan die vraag na die geskikste metode van indeksering: per hand of outomaties. In die proses van indeksering en herwinning word daar uiteraard telkens gebruik gemaak van talle tegnieke om te verseker dat die grootste mate van doeltreffendheid bereik word, byvoorbeeld gekontroleerde woordelyste, uitsluitende woordelyste (*stoplists*), tekskompresie, ensovoorts. Ook tydens die stadium van soek/herwinning word aan die gebruiker 'n aantal meganismes ter beskikking gestel om die verloop van dié uiters komplekse proses te beheer en aan die talige karakter van die soekproses te laat reg geskied, byvoorbeeld die soektog volgens frases, volgens nabyheid (*proximity searching*), Boolese operatore (*Boolean operators*), die gebruik van wisselkarakters (*wildcard operators*), tesouri (vergelyk 4.6.4), terugverwysingsmoontlikhede (*back referencing*); vir besonderhede, vergelyk Fand (1987).

Daar is al baie ernstige bedenkinge uitgespreek oor die doeltreffendheid van soektogte op voltekstatabasisse. In die meeste gevalle het dit gegaan oor die doeltreffendheid van outomatiese indeksering op die meer bekende hoofraamstelsels. Oor verskeie jare is bevindinge in dié verband voorgelê, met resultate wat in beide rigtings dui. In 'n belangrike artikel het Blair & Maron (1985) aangedui dat in soektogte wat hulle gemoniteer het slegs 20% van die relevante dokumente gevind is. Hulle stel voor dat indeksering per hand meer doeltreffend kan wees. Die resultate van dié ondersoek word weer deur ander navorsers betwis (Zoellick, 1987b, p. 71) en voorstelle word gemaak oor hoe die moontlikhede van outomatiese indeksering verbeter kan word. Zoellick meen dat deur die outomatiese indeksering te beperk tot terme uit die titels en opsommings (*abstracts*), kan die doeltreffendheid oor die algemeen verhoog kan word. Vergelyk ook Feinberg (1973) oor die indeksering uit titels. In die geesteswetenskappe is dit egter juis problematics om van die woorde in die titel gebruik te maak. Salton (1989, pp. 275-312) bied die mees resente oorsig oor outomatiese indeksering.

Die bedenkinge oor die doeltreffendheid van soektogte op voltekstatabasisse geld nie net ten opsigte van outomatiese indeksering nie. O'Neill & Morris (1988, p. 287) dui aan dat hierdie beperkte mate van doeltreffendheid geld ten opsigte van bykans alle intydse

stelsels; die probleem lê by die feit dat trefwoorde in sigself (wat dan gemanipuleer word deur Boolse operatore) nie baie goed verteenwoordigend kan wees van die inligting in 'n databasis nie. Wat nodig is, volgens O'Neill & Morris, is dat daar in 'n groter mate gebruik gemaak moet word van

- die verwerking van natuurlike taal (*natural language processing*), waardeur enersyds die navraag van die gebruiker en andersyds die volle teks in die databasis op so 'n wyse verwerk kan word dat die beperkinge verbonde aan die indeksering op die basis van enkele trefwoorde omseil kan word en
- deskundige stelsels (*expert systems*), waardeur afleidings gemaak kan word oor die gebruiker se kundigheid en behoeftes.

Daar is reeds talle eksperimentele stelsels ontwikkel waarin daar gepoog word om hierdie moontlikhede te ontgin. Op mikrorekenaargebied is die stelsel Q & A 'n goeie voorbeeld hiervan (Roux, 1988, p. 32). Besonder baie ontwikkeling op dié gebied is egter nog nodig. Dit neem nie weg nie dat daar intensief gebruik gemaak behoort te word van die programmatuur wat reeds beskikbaar is en dat dié stelsels reeds 'n baie bruikbare vlak van inligtingherwinning bied.

Daar word voortdurend aandag gegee aan moontlikhede om dié programme meer doeltreffend te maak. 'n Belangrike aspek waarmee in dié verband toenemend rekening gehou word, is die behoefte van die gebruiker. Die gebruiker van of soeker na inligting moet naamlik in staat gestel word om die mees relevante inligting – en nét dié – te vind. Daar moet rekening gehou word met die kundigheidsvlak van die gebruiker en die proses van inligtingsoek moet geoptimeer word om hierby aan te pas. Blair (1984) praat van die gebruiker se punt van oorlaaiing (*futility point*), wat intree wanneer 'n oormaat van inligting gevind word (*output overload*).

Dit is dus duidelik dat daar weer eens baie aandag gegee word aan die rekenaargebruiker-raakvlak! (Vergelyk hieroor verder 3.5.) Dié raakvlak kom in verskillende toepassings of aksies met die data telkens op verskillende wyses ter sprake. Wat egter opval, is die oriëntering aan die inligtingsbehoefte en inligtingsverwerkingsmoontlikhede van die gebruiker.

Die doeltreffendheid van voltekstatabasisse kan egter verhoog word deur in 'n groter mate gebruik te maak van die meganisme van rondblaaiery.

b) AskSam – 'n mikrorekenaarprogram

Voltekstatabasisstelsels vir mikrorekenaars raak intussen al hoe meer gewild. Die moontlikhede van dié stelsels moet nie oorskat word nie. Tog bied dit in baie gevalle baie nuttige oplossings vir die behoeftes van die gebruiker. AskSam (Seaside Software, 1987) het gedurende die afgelope paar jaar geblyk een van die gewildste stelsels in dié verband te wees. In besprekings in rekenaartydskrifte is dit ook deurgaans baie positief beoordeel (meesal as die doeltreffendste program) (vergeelyk Puglia, 1986 en Pruett, 1987).

AskSam is 'n spyskaart-aangedrewe program wat beide vryevorm-teks en teks in veld-formaat kan hanteer wat nie vooraf geïndekseer is nie. Die feit dat die teks nie vooraf geïndekseer word nie, bring mee dat nuwe inligting maklik en dinamies opgeneem kan word sonder dat daar lang indeksering moet plaasvind.

Die volgende aspekte van **AskSam** verdien aandag:

- a) Tekste wat in **AskSam** opgeneem word, kan enige lengte hê. 'n Rekord bestaan uit 1600 karakters en 'n aantal van dié saam vorm dan 'n dokument of teks.
- b) Toevoer van teks geskied gemaklik, mits dit suiwer ASCII-teks is.
- c) **AskSam** bied 'n verskeidenheid soekmeganismes: Boolse operatore (*AND, OR, NOT*), soektogte volgens nabyheid (*proximity operators*), vier posisionele operatore, agt teks-operatore en wisselkarakters in alle posisies.
- d) Alhoewel die soek van inligting sekwensieel deur die nie-geïndekseerde data plaasvind, is die program kragtig en vinnig (Hughes, 1987b, p. 15). In 'n interaktiewe stelsel van dié aard is die aspek van spoed nie 'n kritieke beperking nie, aangesien die gebruiker gewoonlik eers wil kyk na die rekords wat gevind word.
- e) **AskSam** beskik oor hiperteksfasiliteite, vergelyk 4.7.3.10g, wat dit moontlik maak om gemaklik na ander gevalle te beweeg waar 'n bepaalde woord voorkom.

Daar is egter ook sekere nadele verbonde aan **AskSam**:

- a) **AskSam** kan nie die spoed van sommige indeks-gebaseerde programme ewenaar nie.
- b) In die geval van baie groot lêers (byvoorbeeld honderde megagreep) sal die beginsel van sekwensiële soektog nie sinvol wees nie.
- c) **AskSam** beskik nie oor tesourusfasiliteite nie.

Vir baie toepassings sal **AskSam** blyk 'n baie doeltreffende stelsel te wees. Daar kan verwag word dat dié stelsel vinnig en dinamies verder ontwikkel sal word.

4.6.4 Tesouri (inligtingkundig)

Die term "tesourus" word gebruik vir 'n gestruktureerde versameling van die algemene woordeskat van 'n taal – hetsy in gedrukte vorm of as 'n elektroniese databasis wat dan kan dien as fasiliteite vir die ondersteuning van die proses van teksgenerering (vergeelyk 4.2.1.2 hierbo).

Afgesien van hierdie gebruik van die term, word dit ook lank reeds op inligtingkundige terrein gebruik ter aanduiding van 'n gestruktureerde versameling terminologie op 'n (dikwels gespesialiseerde) vakgebied, byvoorbeeld die ingenieurswese, die drukkers-

bedryf, die farmaseutiese bedryf, ensovoorts. Dit het vroeg reeds in die geskiedenis van inligtingsontsluiting – lank voor die koms van die rekenaar – geblyk dat die een of ander vorm van die ordening van die terminologie nodig was met die oog op suksesvolle soeke na inligting. Dit was immers duidelik

- dat verskillende mense verskillende terminologie vir dieselfde entiteit gebruik,
- dat daar voorkeurterme is, terwyl ander terme hoegenaamd nie gebruik moet word nie en
- dat mense wat inligting soek op verskillende vlakke met die terminologie kan omgaan, byvoorbeeld: die een kan soek op die basis van 'n spesifieke term (byvoorbeeld "armband"), terwyl die ander op die basis van 'n meer algemene term (byvoorbeeld "juweliersware") kan soek, aangesien hy onder die indruk is dat die dokument onder 'n meer algemene term geïndekseer is.

Dit het gevolglik duidelik geword dat die terminologie op so 'n wyse gestruktureer moet word dat dit as't ware die eenbaan-bruggie moet vorm waardeur verseker kan word dat daar eenvormigheid sal wees op die vlak van die beskrywing/indeksering en op dié van inligtingsoek. Daar moet 'n versameling terminologie wees in terme waarvan die indekseerder 'n dokument moet beskryf, terwyl die persoon wat inligting soek ook weer van dieselfde terminologie gebruik kan maak en daarvan seker kan wees dat hy op dié wyse die dokument sal kan vind.

'n Lys toegelate terme is egter nog nie voldoende nie. Selfs deur voorsiening te maak vir 'n mate van kruisverwysings (horisontale strukturering), is die probleem egter nie opgelos nie. In die proses van indeksering vel die indekseerder naamlik 'n beslissing oor die mate waarin 'n spesifieke term op die mees doeltreffende wyse 'n item of die inhoud van 'n dokument kan beskryf. Indien die dokument verskeie spesifieke terme bevat, kan dit beter wees om 'n meer algemene of omvattende term te gebruik. Die persoon wat inligting soek, moet dan egter in staat wees om te kan agterkom hoe die terme met mekaar saamhang. Dit het dus ook nodig geraak om voorsiening te maak vir 'n hiërargiese strukturering in die terminologie.

In tesouri in die inligtingkundige sin van die woord word daar dus tipies voorsiening gemaak vir 'n strukturering van die terminologie op die basis van die volgende relasies (net die belangrikste word aangedui, vir verdere besonderhede vergelyk Aitchison & Gilchrist, 1972; Soergel, 1974 en Wersig, 1978). Die Engelse aanduidings van die relasies (*relators*) word hier gebruik, aangesien dit gewoonlik só in databasisse geïmplementeer word:

- USE x: gebruik term x (en nie dié een nie)
- UF y: (*used for*) gebruik dié term (en nie y, wat jy as gebruiker dink 'n moontlik relevante een is nie)
- SAL x: (*see also*) albei terme kan gebruik word; raadpleeg dus albei

- BT: (broader term) (byvoorbeeld x BT y) y is 'n meer algemene term ten opsigte van x
- NT: (narrower term) (byvoorbeeld x NT y) y is 'n meer spesifieke term ten opsigte van x
- SYN x: x is 'n sinoniem van die term ter sprake
- ANT x: x is 'n antoniem van die term ter sprake.

In baie inligtingstelsels word daar voorsiening gemaak vir tesourusfasiliteite in die soekproses, byvoorbeeld in die bekende hoofraamprogramme STAIRS en BASIS. Die tesourus word dan op so 'n wyse in die proses geïntegreer dat die gebruiker in die soekproses begelei word om vanaf sy eie terminologie te beweeg na die terminologie wat by die beskrywing van die item gebruik is, byvoorbeeld:

- a) Die gebruiker soek 'n spesifieke term A en stel aan die stelsel die vraag of daar dokumente (of items) is wat deur dié term beskryf of geïndekseer is.
- b) Die stelsel dui aan dat daar twee dokumente is, maar dat daar in die tesourus ook nog 'n meer algemene of omvattende term B is en 'n nog meer spesifieke term ten opsigte van A, naamlik C. Daar is byvoorbeeld 12 dokumente wat deur B beskryf is en 4 wat deur C beskryf is.
- c) Die gebruiker kan nou sy soekproses voortsit en daarop aandrang dat hy wel net belangstel in die dokumente beskryf deur term A – of hy kan sy soekstrategie verander om vir die meer algemene of vir die meer spesifieke terminologie voorsiening te maak.

Die ontwikkeling van 'n tesourus is 'n moeisame en tydrowende taak wat baie 'n baie groot mate van kundigheid van die opstellers vereis. Verskeie soorte kundigheid is ter sprake: kundigheid met betrekking tot die spesifieke vakgebied, met betrekking tot inligtingsontsluiting in die algemeen en met betrekking tot die onderhawige taal. Tog is dit vir sekere soorte inligtingsontsluiting die enigste sinvolle uitweg.

Ongelukkig word daar nog nie in die meer gebruiklike databasisse op mikrorekenaars voorsiening gemaak vir geïntegreerde tesourusfasiliteite nie. Die enigste uitsondering is **WordCruncher**, (vergelyk 4.4.3.3c) wat dit egter op 'n ietwat eiesoortige wyse gebruik wat nie dieselfde funksionaliteit het as dergelike fasiliteite in byvoorbeeld **BASIS** nie.

4.7 TEKS-AANBIEDING

4.7.1 Inleiding

In 'n sekere sin kan alle prosesse waarby teks met die oog op die lees daarvan aangebied word op die skerm van 'n rekenaar (of in 'n drukstuk), beskou word as teks-aanbieding. Die moontlikheid bestaan dan dat dit so 'n algemene en vae term word dat dit ineens ook 'n ondoeltreffende benaming word. Daar het in die laaste paar jaar egter juis op die gebied

van die interaksie tussen die gebruiker en die rekenaar 'n aantal van die mees sinvolle ontwikkelings plaasgevind wat nie anders omskryf kan word nie as "teks-aanbieding" (of tekspresentering – *text presentation*). Die eintlik opwindende ontwikkelings het plaasgevind met betrekking tot die meer intelligente en leser-gebaseerde aanbieding van teks op die skerm, hoofsaaklik op die basis van 'n nie-liniêre greep op die teksmateriaal.

Die enigste manier om hierdie soort toepassing – wat moontlik tot van die mees strategiese rekenaartoepassings met betrekking tot teks kan ontwikkel – as 'n aspek van teks-aanbieding te omskryf, is om ander soorte teks-aanbieding uit te sluit, byvoorbeeld die volgende:

- a) die aanbieding van teks op papier (hetsy as gewone uitdrukke of as tipografies hoogs versorgde uitleg met die hulp van kantoorpublikasie-programmatuur);
- b) die aanbieding van teks op die skerm as die resultaat van 'n proses van woordverwerking (of redigering) of van 'n soektog op die basis van indekseringsmeganismes (byvoorbeeld die resultaat van 'n volteksoektog, vergelyk 4.6.3); die resultaat op die skerm kan dan wel 'n stuk volteks wees, maar die soekprosedure het nou op so 'n wyse verloop dat die gebruiker juis die volle teks wil sien;
- c) die aanbieding van die resultate van 'n proses van teksmanipulering (byvoorbeeld met behulp van konkordansieprogrammatuur, vergelyk 4.4.3.3, of 'n elektroniese woordeboek), waarby die gebruiker die materiaal juis in die vorm van rekord-georiënteerde data wil sien;
- d) die aanbieding van teksmateriaal (en ander materiaal, byvoorbeeld beeldmateriaal) aan gehore tydens (of ter vervanging van) voordragte, wat saamgevat kan word onder die term "aanbiedingstechnologie(ë)"; die interaksie verloop hierby meesal buite die beheer van die kyker om. Hierdie gebied is (veral sedert 1987/1988) een van die mees dinamiese groeigebiede met betrekking tot mikrorekenaartoepassings. Soms gaan dit om omvattende aanbiedings op die skerm, byvoorbeeld in die geval van die sogenaamde *slideshow*-benadering.

Wat oorbly, is 'n toepassingsarea waarin die teks op 'n spesiale wyse verpak word om voorsiening te maak vir 'n soort interaktiewe toegang daartoe op die basis van 'n behoefte of versoek van die gebruiker. Die klem val daarop dat die proses op 'n interaktiewe wyse verloop. Ook hierdie soort toepassing kan op 'n baie meganiese en ongesofistikeerde wyse hanteer word (vergeelyk 4.7.2). Dit kan egter ook op 'n baie gevorderde vlak hanteer word waarin daar taalverwerking plaasvind in ooreenstemming met gebruikersbehoefes.

Daar moet gelet word op die feit dat die prosesse van die omgang met inligting via die skerm van 'n rekenaar ingrypend verskil van dié betrokke by die gedrukte woord. Hierdie verskille moet nou verreken word in die wyse waarop die meganismes geskep word waarmee die materiaal hanteer word, asook met betrekking tot die wyse waarop die

materiaal vir interaksie met die rekenaar gestruktureer of verpak word. Kerr (1986, p. 321, vergelyk ook Brown, 1986) beskryf die probleem as een wat hoofsaaklik op twee vlakke lê:

- i) oppervlak-ontwerp (*surface design*), waarin dit gaan oor beslissings in verband met tipografie, uitleg, die gebruik van kleur, beeldmateriaal, ensovoorts en
- ii) raakvlak-ontwerp (*interface design*), waarin dit by elektroniese teks veral die navigasieprobleem (*navigation of die wayfinding problem*) is wat die meeste aandag verdien en wat weer op drie verskillende vlakke ter sprake kom:
 - die direkte struktuur (hoe inligting oor navigasie op die vlak van die bladsy of die skerm verskaf word, dit wil sê die instruksies en hoe dit aangedui word)
 - die interne struktuur (hoe die inligting binne-in 'n bepaalde dokument gestruktureer word)
 - die eksterne struktuur (die navigeringshulpmiddele waarvan die "leser" gebruik kan maak om van een dokument na 'n ander (of van een deel van 'n dokument na 'n ander deel) te beweeg.

In die gesprek oor die wyse waarop teks in die driehoekige verhouding (i) leser/gebruiker, (ii) rekenaar en (iii) teks hanteer moet word, word dus inderdaad 'n groot verskeidenheid aspekte aangeraak. In die vakliteratuur oor hierdie tema blyk dit dat dit hier om interdisiplinêre navorsing en ontwikkeling in die volle sin van die woord gaan. (Vir 'n baie nuttige literatuuroorsig tot 1985, vergelyk Kerr, 1986. Vergelyk ook 1.2 en hoofstuk 3 oor die verskillende aspekte/gebiede ter sprake.) Die aspek van die verandering in die moontlikheid van die mens om inligting in hierdie ander dimensies te kan hanteer, het tot nog toe nie genoeg aandag geniet nie. Vergelyk hieroor Heim (1987), wat verkies om dit onder die trefwoord "psigodinamika" te bespreek.

4.7.2 Rekenaargesteuende onderrig (RGO)

In rekenaargesteuende onderrig gaan dit gewoonlik oor die een of ander vorm van die opbreek van teks in 'n aantal stukke teks wat onder verskillende voorwaardes vir die student op die skerm aangebied word.

In baie gevalle is rekenaargesteuende onderrig niks meer nie as 'n stel eenvoudige vrae en antwoorde, soms darem met 'n sekere hoeveelheid terugvoer aan die student (byvoorbeeld om die regte antwoord te gee of om bykomstige inligting te verskaf). In sulke gevalle is dit te bevraagteken of daar noemenswaardige voordele daarin is om materiaal wat andersins ook in die formaat van boek-programmering aangebied word spesifiek in die vorm van 'n program vir RGO aan te bied. In ander gevalle is daar 'n baie sinvolle stratifikasie van die liniêre teks van byvoorbeeld die handboek in nie-liniêre teks en word die aanbieding van die materiaal binne die konteks van die RGO-opset op so 'n wyse aangebied dat daar tog van sinvolle interaksie gepraat kan word. Verskeie moderne outeurstelsels (*authoring*

languages) laat byvoorbeeld 'n baie goeie strukturering van die materiaal toe, byvoorbeeld QUEST.

'n Aantal ontwikkelings van die afgelope paar jaar het aan die konsep van die interaktiewe aanbieding van teks egter 'n totaal ander dimensie gegee. Hierdie ontwikkelings en die programmatuur en toepassings wat daaruit voortvloei, kan ten beste onder die benaming "hiperteks" aan die orde gestel word.

4.7.3 Hiperteks

4.7.3.1 Inleiding

Hiperteks en hipermedia (oor die terminologie, vergelyk 4.7.3.4) kan beskou word as een van die mees belowende areas van rekenaartoepassing ten opsigte van die betekenisvolle hantering van teks. Baie rekenaartoepassings op taal is hoofsaaklik van belang vir mense wat op die een of ander wyse op die gebied van navorsing oor taal/teks betrokke is (deur byvoorbeeld die opstel van interaktiewe konkordansies). Hierteenoor is hiperteks 'n toepassing wat wel vir hierdie groep van belang is, maar ook – en miskien in 'n nog groter mate – vir mense op 'n wyer gebied wat meen dat hulle buite die gebied van taalnavoring en -ontwikkeling staan. Hiperteks kan van belang wees vir enigiemand wat op enige wyse hoegenaamd met die hantering of bemiddeling van inligting in teksvorm te make het.

Tog is dit so dat die ontwikkeling van hierdie nuwe gebied grootliks verbygegaan het by 'n groot aantal persone en instansies wat betrokke is by gebiede wat eintlik belang behoort te hê by hierdie ontwikkeling. Die redes vir hierdie aparte ontwikkeling is velerlei:

- Daar is so baie programme wat potensieel relevant is vir die hantering van teks dat die ontwikkeling van hiperteks as maar net nog 'n toepassing in lyn met die ander beskou word terwyl dit hier gaan oor 'n radikaal anderssoortige ontwikkeling wat as 'n sprong in die geskiedenis van rekenaarimplementering beskou moet word.

Die ontwikkelings in verband met hiperteks word deur baie persone beskou as een van die mees sinvolle ontwikkelings wat nog op die vlak van die gebruik van die rekenaar deur die gewone gebruiker plaasgevind het. Dit word deur baie mense beskou as die rigting wat die moontlikheid bied om met teks op dieselfde behendige wyse om te gaan as wat 'n sigbladverwerker (*spreadsheet*) die gebruiker toelaat om met syfers om te gaan. Vergelyk in dié verband Dyson (1988):

... the big advances over the next few years won't come from better spreadsheets or even from prettier output, whether of text or data. The excitement will come from a profusion of tools to manipulate text. ... Until now it's been tougher to automate the meaningful manipulation of text than of numbers, and the industry has hardly even tried. ... Now we're about to deal with text the way spreadsheets deal with numbers.... The business of manipulating text is just starting, and HyperTEXT '87 was its christening.

- Mense het in so 'n mate gewoon geraak aan die karakter en die krag van teks in sy konvensionele vorme dat dit baie moeilik is om te dink aan 'n ander en kwansuis meer doeltreffende wyse om teks te kan hanteer. Die boekkultuur het byvoorbeeld in Duitsland tot so 'n vlak ontwikkel dat dit vir baie mense moeilik is om 'n begrip te ontwikkel van die moontlikhede van ander inligtingsdraers (selfs vir sekere toepassings waarvoor dié inligtingsdraers meer geskik blyk te wees). Wat egter nie deurgedring het na baie voorstanders van die boek in sy tradisionele vorm en posisie nie, is die feit dat die volume materiaal in gedrukte vorm alreeds die perke van bruikbaarheid oorskry het en dat – minstens op sekere gebiede – 'n alternatiewe en meer doeltreffende soort toegang tot inligting nodig is (vergelyk verder 2.2a).
- Die onrealistiese voorstellings oor die "elektroniese boek" wat dan kwansuis die tradisionele boek heeltemal sou vervang – en wat opsigtelik baie ver van die werklikheid en die realistiese toekoms af lê – het ook daartoe bygedra dat enige beweging in dié rigting nie as ernstig beskou is nie.
- Hiperteks is 'n rekenaartoeëpassing wat eers onlangs ontwikkel is. Die programme is grotendeels onbekend en net die mees resente literatuur is van belang.

4.7.3.2 'n Kort omskrywing van hiperteks

Daar bestaan geen eenvoudige en eenduidige definisie van hiperteks nie. Verskeie van die belangrike artikels oor hiperteks probeer nie 'n definisie gee nie, maar eerder 'n omskrywing van die essensiële eienskappe, byvoorbeeld Conklin (1987, pp. 17-19). Die rede hiervoor is dat hiperteks eerder 'n konsep ten opsigte van die operasionalisering van teks is as 'n bepaalde vorm van teks of 'n bepaalde program. Verder is daar 'n aantal programme of programfasiliteite wat elkeen bepaalde kenmerke in gemeen het met hiperteks en wat 'n kort en bondige definisie bemoeilik. Hierby kom ook nog dat die term oor 'n tydperk van twintig jaar in verskillende betekenisse gebruik is (vergelyk 4.7.3.3).

Hiperteks kan omskryf word as 'n rekenaar-gebaseerde medium vir die operasionalisering van nie-liniêre teks (of ander inligtingskomponente), wat die gebruiker in staat stel om rekenaargestesteunde inter-dokument- en intra-dokument-skakels te maak en wat daardeur die gebruiker se vermoëns aanvul met betrekking tot die prosesse van lees, skryf en dink.

Die struktuur van hiperteks-dokumente is eintlik baie eenvoudig. Dit bestaan naamlik net uit twee komponente, naamlik nodi (*nodes*) en skakels (*links*), met 'n meganisme wat die beweging tussen die nodi toelaat (vergelyk hieroor verder 4.7.3.6).

Wat baie belangrik is, is dat die heen en weer beweeg tussen nodi in hiperteks gemaklik en teen hoë spoed moet plaasvind. Slegs die kleinste vertraging is aanvaarbaar; daar word gedink in terme van hoogstens een of twee sekondes (Conklin, 1987, p. 33). Indien dit enigsins langer gaan duur om data te gaan haal (byvoorbeeld 'n beeld vanaf 'n videoskyf) moet dit onmiddellik vir die gebruiker aangedui word dat dit tyd in beslag gaan neem. Hierdie aspek van die hoë spoed van beweeg tussen nodi hang daarmee saam dat

hiperteks juis die kognitiewe prosesse op die voet wil volg – en in dié verband is spoed van die grootste belang.

Onderliggend aan hiperteks is die een of ander soort databasisprogram, wat uiteraard 'n algemeen kommersieel beskikbare databasis kan wees.

By die omskrywing van hiperteks is dit ook belangrik om aan te dui wat dit nie is nie. Dit gebeur naamlik telkens dat persone hiperteks beskou as maar net die verdere ontwikkeling, of as die toevallige saambestaan in een program, van 'n paar eienskappe wat in elk geval ook op rekenaars beskikbaar is. Sommige van dié eienskappe kan moontlik die "gevoel" van hiperteks gee, maar is dit nie werklik nie. Hiperteks moet nie verwar word met, of beskou word as slegs 'n verdere ontwikkeling van, die volgende nie (Conklin, 1987, p. 18):

- Vensterstelsels, wat die gebruiker ook toelaat om gemaklik tussen verskillende soorte data te beweeg. Tekort: die gemaklike en veelvlakkige heen en weer beweeg binne dieselfde databasis is gewoonlik nie moontlik nie.
- Lêerstelsels. Tekort: die konsep van nodi en skakels is nie voldoende ontwikkel nie.
- Raamwerkprosesseerders (vergelyk 4.2.1.1b). Tekort: Daar is te min ondersteuning vir die opvolg van skakels op 'n ander vlak as die raamwerk, met ander woorde binne die teks onder 'n bepaalde punt in die raamwerk.

Ook beeldmateriaal word in hiperteks ingesluit (vergelyk 4.7.3.4 oor terminologie). Dit sou gesê kon word dat hiperteks eerder op die terrein van inligtingshantering (of -ontsluiting) funksioneer as op dié van teks en die verwerking van natuurliketaal. Dat hiperteks baie te make het met inligtingshantering, is uiteraard korrek (vergelyk Smith & Weiss, 1988, p. 816), maar dan beteken dit nie dat hiperteks nie binne die gebied van die verwerking van natuurlike taal lê nie; dit beteken juis eerder dat die verwerking van natuurlike taal nou 'n meer prominente plek moet kry in die hantering van inligting!

Tydens die groot internasionale kongres van die inligtingsbedryf in 1988, Online Information 88, is 'n groot deel van die program gewy aan hiperteks. Dit het duidelik geword dat die faset van die verwerking van taal (en daarmee saam ook hiperteks) 'n fundamentele faset van die ontsluiting van inligting op bykans enige gebied is. (Vergelyk Claassen & Bothma, 1988; Van der Merwe, 1988; Dolan, 1988; Kesselman & Trapasso, 1988 en ander bydraes in die bundel **Online Information 88**.)

Die hele ontwikkelingsgeskiedenis van hiperteks bevestig net weer eens baie duidelik dat die tradisionele grense van baie vakdissiplines kunsmatig is en lei tot 'n beoefening van die wetenskap in isolasie van die werklikheid: die perspektief van inligting is integrale deel van enige wetenskap en die bewussyn van wat taal is en meebring, is weer deel van enige poging om inligting in belang van die gebruiker en ooreenkomstig sy werklike

inligtingsbehoefte en -praktyke te operasionaliseer – en daarmee het ons noodwendig ook by hiperteks uitgekóm.

Anders as baie ander benaderings tot die verkryging van inligting wat die inligtingsoeker tot by 'n bepaalde punt bring, wat hoogstens deur 'n nuwe soektog opgevolg kan word, is hiperteks 'n kennisverkennde stelsel, wat dit moontlik maak om (meestal) van die resultaat van die soekproses al verkennde nog verder te beweeg.

Hiperteks is nie 'n luukse wat ontwerp is deur en bedoel is vir persone wat nie die moeite wil doen om 'n gewone, tradisionele boek te hanteer nie. Dit is veeleer 'n noodwendige ontwikkeling om die gebruiker in staat te stel om die groot volumes teksmateriaal wat in die moderne lewe en met die ondersteuning van die tegnologie gegenereer word, sinvol te kan hanteer. Meer nog: dit is 'n medium wat miskien in 'n baie groter mate met die mens se intellektuele prosesse korreleer as die gedrukte boek wat 'n "plat" medium (*flat medium*) is. Die boek het uiteraard wel uit 'n uiterlike- en nuttigheidsperspektief sy voordele, maar dit is beslis nie die mees geskikte inligtingsdraer vir alle soorte inligting, gebruikers en toepassings nie (vergelyk Fabian, 1983).

4.7.3.3 'n Kort historiese oorsig oor hiperteks

Vannevar Bush, eertydse wetenskaplike adviseur van president Roosevelt, word allerweë beskou as die vader van hiperteks. Hy het self nie die term "hiperteks" gebruik nie, maar het in 1945 alreeds die inligtingsontploffing sien kom, die belang van skakeling (*linking*) tussen items of brokke inligting raakgesien en die moontlikheid van ondersteuning van die intellektuele prosesse deur toerusting voorsien (Bush, 1945). Dit kan dus wel gesê word dat die konsep hiperteks van Bush afkomstig is.

Sedert die sestigerjare het Douglas Engelbart die moontlikhede van die rekenaar probeer ontgin vir wat hy noem *the augmentation of man's intellect* (Engelbart, 1963; Conklin, 1987, p. 22).

Die term "hiperteks" is egter afkomstig van Ted Nelson, wat die term in die sestigerjare gemunt het en sedertdien deurlopend daarby betrokke was. Vir die publikasies van Nelson, vergelyk die literatuur by Conklin (1987). Nelson se idees is in besonderhede en vir 'n wyer gehoor uitgewerk in sy boek *Literary Machines* (1981). Vir besonderhede, vergelyk Conklin (1987, pp. 23-24) en Nelson (1981). Nelson word allerweë as 'n visionêre persoon (*visionary*) beskou: hy het baie belangrike dinge raakgesien, baie ontwikkelings voorsien en baie persone betrokke by hiperteks gestimuleer (vergelyk byvoorbeeld Van Dam, 1988). Vir Nelson is twee fasette belangrik:

- Die rekenaar maak dit moontlik om 'n vorm van nie-liniêre teks te skep en te operasionaliseer, waarin daar geleentheid is vir interaktiewe vertakking (*branching*).
- Daar moet 'n eenvormige literêre omgewing op 'n globale basis geskep word, waarin inligting van oor die hele wêreld (!) gestoor kan word en vanwaar dit

gemaklik (met skakels tussen dokumente) toeganklik kan wees vir die gebruiker vanaf sy eie rekenaar.

Een van die eerste kragtige hipertekstelsels was NoteCards (ontwikkel aan Xerox PARC onder leiding van Trigg, wat self ook van die mees fundamentele werk in verband met hipertekst gedoen het). Die beginsel van hipertekst is deur dié stelsel baie duidelik gedemonstreer en dit het die uitgangspunt gevorm van verskeie latere ontwikkelings. Die groot bydrae van hierdie program was dat dit gebruikers oortuig het dat 'n nuwe era ten opsigte van die intelligente, interaktiewe, rekenaar-gesteunde omgang met teks aangebreek het. Die stelsel is nie aktief bemark nie en daar is slegs enkele verkoop of vir ontwikkelingsdoeleindes beskikbaar gestel. (Vergelyk Halasz, 1987; Halasz et al., 1987; Conklin, 1987 en Trigg & Irish, 1987.)

Die Intermedia-stelsel van die IRIS-groep (Institute for Research on Information and Scholarship, Brown University) is grotendeels onafhanklik van die ander stelsels ontwikkel en is veral belangrik op grond daarvan dat dit dadelik aangewend is vir groter toepassings, waaruit die werklike waarde van hipertekst vir toepassings in die onderrig (in die breedste sin van die woord, naamlik as kennisbemiddeling) geblyk het. (Vir 'n kort opsomming, vergelyk Conklin, 1987, pp. 28-29. Vir verdere besonderhede, vergelyk Meyrowitz, 1986; Yankelovich et al., 1985; 1987a; 1988; Yankelovich & Van Dam, 1987.)

Agter die stelsel lê grondige inligtingkundige denke en die rekenaar-gebruiker-raakvlak het deurgaans 'n belangrike rol gespeel. Daar was telkens baie noue samewerking tussen die IRIS-groep en die Department of Computer Science van Brown University. Vir besonderhede oor die dinamiek van van dié samewerking en ontwikkeling, vergelyk Van Dam (1988). Die stelsel is sedert die middel van 1989 beskikbaar vir die MacIntosh-rekenaar.

Intussen het die ontwikkelings voortgegaan, deels in aansluiting by die NoteCards-stelsel en deels heeltemal onafhanklik daarvan – wat weer 'n aanduiding is dat hipertekst nie net in die verbeelding van 'n individu ontstaan het nie, maar in werklikheid 'n noodwendige ontwikkeling was in terme van die behoeftes van die tyd en die moontlikhede van die tegnologie. Trouens, verskeie stelsels was alreeds in gebruik toe die kongres Hypertext '87 plaasgevind het. Let veral op die stelsels Thoth-II van Bell Communications Research (Collier, 1987) en Document Examiner van Symbolics (Walker, 1987), asook Conklin se bespreking van verskeie stelsels.

Die eerste hipertekstprogramme en -dokumente is ontwikkel op sogenaamde "werkstasie"-rekenaars (*workstations* of *workstation computers*). Dié rekenaars was in daardie stadium kragtiger as die gewone mikrorekenaars en het veral baie goeie grafiese fasiliteite gebied – wat vir die ontwikkeling van hipertekst essensieel was.

Daar moet op gelet word dat vir die ontwikkeling van hipertekst dit noodsaaklik was dat sekere aspekte in die ontwikkeling van die tegnologie by mekaar sou uitkom, naamlik

- 'n kragtige verwerker (*chip*), met die moontlikheid om groot geheue te kan hanteer

- baie goeie (en vinnige) grafika-fasiliteite (met gesofistikeerde programmatuur vir vensters)
- die muis as meganisme wat vinnige interaksie moontlik maak.

Vir hipertekstoepassings is gewoonlik ook groot hardeskyfruimte nodig, maar dié het in elk geval met 'n eie momentum en parallel daaraan ontwikkel.

Geskikte tegnologie vir hiperteks was aanvanklik net in die werkstasie-omgewing beskikbaar. Die groot deurbraak in beide die ontwikkeling en die benutting van hiperteks het gekom toe hierdie fasiliteite ook op mikrorekenaars aangebied is. Die ontwikkeling van gesofistikeerde, hoëvlak hipertekstoepassings het uiteraard bly voortgaan en dit kan gesê word dat die omgewing van die kragtige en gesofistikeerde werkstasies die beste ontwikkelingsomgewing bied, vanwaar 'n bepaalde program mettertyd oorgedra kan word na die gewone mikrorekenaaromgewing. Die bekendste hiperteksprogram vir die MS-DOS-omgewing, naamlik *Guide*, is aanvanklik vir UNIX-rekenaars ontwikkel en eers daarna oorgedra na die MacIntosh-omgewing en toe na die MS-DOS-omgewing (Brown, 1987). Die *Intermedia*-stelsel is eweseer op werkstasie-rekenaars ontwikkel en ontwikkelings het reeds ver gevorder om dit na die MacIntosh-omgewing oor te dra.

Die mees wydverspreid gebruikte hiperteksstelsel is *HyperCard*, wat net vir MacIntosh-rekenaars beskikbaar is (vergelyk Oren, 1987 en Goodman, 1987). Die grafiese fasiliteite van die MacIntosh is uitstekend geskik vir hiperteks. *HyperCard* is aanvanklik verkoop as 'n aparte program, maar Apple het gou besef dat dié program in sigself aanloklik is. Mettertyd is die program as standaardkomponent ingesluit by alle MacIntosh-rekenaars. *HyperCard* word selfs beskou as deel van die sogenaamde stelselprogrammatuur van die MacIntosh. Dit is dan ook *HyperCard* wat meegebring het dat hiperteks by gewone gebruikers ingang gevind het en vir 'n groot verskeidenheid gebruikers die ideale stelsel vir 'n aantal uiteenlopende toepassings geword het: databasis, bibliografieë, onderrig, die vinnige ordening van persoonlike data, ensovoorts. Hiermee het ook 'n nuwe term ingang gevind, naamlik "stapels" (*stacks*). Die woord verwys na 'n hipertekstoepassing in *HyperCard* en die analogie met die versameling of stapel (*stack*) kaartjies is baie duidelik. Tans word sulke stapels in die handel aangebied vir verskeie gebiede en toepassings.

HyperCard het ook die grootste verskil tussen baie van die huidige hiperteksstelsels na vore gebring. Soos die naam aandui, is *HyperCard* gemodelleer op die metafoor van die kaartjie (soos in 'n persoonlike kaartjie-versameling), wat dan vinnig en gemaklik deur die rekenaar hanteer word. Inligting van groter omvang (byvoorbeeld 'n stuk lopende teks) moet nou in 'n sekere mate afgebreek word tot – by wyse van spreke – kaartformaat. Ook *NoteCards* (vergelyk hierbo) is op dié model gebaseer. Daarteenoor is ander stelsels gemodelleer op die metafoor van die teks of dokument, as die basiese draer van inligting. Dit is byvoorbeeld die uitgangspunt van *Intermedia* en *Guide*. Hierdie onderskeiding blyk baie belangrik te wees vir toepassings wat gebaseer is op materiaal in teksvorm.

Terwyl die MacIntosh-weergawe van *Guide* aanvanklik baie gewild was, moes dit gou dié plek afstaan aan *HyperCard*, wat teen die middel van 1987 aan die publiek beskikbaar

gestel is. Kort daarna is die eerste hipertekstelsel vir die MS-DOS-omgewing bekend gestel, naamlik *Guide*, en dit is opgevolg deur 'n paar ander programme (byvoorbeeld *Hyperties*, *KnowledgePro*). MS-DOS-rekenaars moes (aanvanklik) telkens eers met spesiale apparatuur toegerus word om hiperteks te kon hanteer, teenoor die MacIntosh wat ook in sy basiese vorm alreeds daarvoor gebruik kon word. Dit het die ontwikkeling van hiperteks in die MS-DOS-omgewing aanvanklik effens aan bande gelê.

Sedert November 1987 was die geskiedenis van hiperteks 'n geskiedenis van (i) toepassings op 'n groot verskeidenheid gebiede en (ii) kongresse waarby dié toepassings bekend gestel is. Die Hypertext '87-kongres (November 1987) aan die University of North Carolina (Chapel Hill, N.C.) staan bekend as 'n unieke geleentheid en het 'n ongekende belangstelling uitgelok. Die referate het werklik sentrale onderwerpe met betrekking tot hiperteks behandel. (Die referate kan gevind word in die kongresbundel *Hypertext '87 Papers*. Sommige daarvan is herpubliseer in *Communications of the ACM*, volume 31, nommer 7, Julie 1988.) Verskeie ander kongresse het hierop gevolg. In Engeland het alreeds 'n aparte hipertekskongres plaasgevind en 'n tweede het plaasgevind in Junie 1989. Tydens die Online Information '88-kongres (Desember 1988) is 'n spesiale deel gewy aan hiperteks. Die tweede groot hipertekskongres (Hypertext '89) vind in November 1989 plaas.

Nelson se aanvanklike idee van 'n universele stelsel van dokumente wat aan mekaar verbind is, het in die geskiedenis van die ontwikkeling van hiperteks nie werklikheid geword nie. Tydens die Online Information 88-kongres het Nelson weer sy misnoë met die ontwikkeling van hiperteks uitgedruk en aangedui dat 'n nuwe meganisme in sy *Xanadu*-stelsel en groot steun van 'n bekende programmatuurfirma tog die moontlikheid bied om 'n *universal server mechanism for linked materials* te skep (Nelson, 1988): Nelson se konsep hou egter nie tred met die algemene ontwikkeling van hiperteks en die dinamiek verbonde aan die ontwikkeling van geslote stelsels vir plaaslike toepassings nie. In die plek van 'n universele stelsel het daar verskeie stelsels ontwikkel wat Nelson se konsep van nie-liniêre teks verder geneem het, maar dit probeer bereik binne die raamwerk van geslote hiperteksstelsels. Hiermee word bedoel dat die stelsel nie (gemaklik) oop is na ander hiperteksstelsels nie. Daar is geen poort na 'n universele of globale stelsel nie en dit kan ook nie so maklik daargestel word nie, aangesien die interne werking en struktuur (vergelyk 4.7.3.6) van die stelsels verskil. Die IDEX-stelsel is 'n hipertekstelsel in netwerkverband, wat (binne 'n geslote, of minstens vasgestelde, gebruikersgemeenskap) tog sekere van die perspektiewe van Nelson werklikheid laat word.

4.7.3.4 Terminologie in verband met hiperteks

Met die keuse van die term hiperteks is daar doelbewus probeer om die nuwe soort strukturering van en toegang tot teks (naamlik teks met die moontlikheid van skakels wat deur die rekenaar geoperasionaliseer word) te plaas teenoor die tradisionele konsep van teks soos dié noodwendig deur die tradisionele inligtingsdraers van teks gekondisioneer is. Die bedoeling met dié naam was om aan te dui dat, teenoor tradisionele teks, hiperteks die perspektiewe van gemak van hantering, ensovoorts, in 'n oortreffende trap bied.

Vanselfsprekend het dit hierby ook gegaan oor ander nie-teks komponente wat in tekste kan voorkom, byvoorbeeld illustrasies.

Sedert die middel van die tagtigerjare het die rekenaartegnologie in so 'n mate ontwikkel dat dit nou ook moontlik geword het om ander media in die opbou en "lees" van 'n hiperteksdokument te betrek, byvoorbeeld videoskyf en klank. Die naam *multimedia* het nou al hoe meer dikwels voorgekom. In een groep word die term *intermedia* gebruik, naamlik die IRIS-groep van Brown University, Rhode Island. Die benaming word egter hoofsaaklik gebruik as benaming vir die spesifieke stelsel wat aan hierdie sentrum ontwikkel is (naamlik *Intermedia*).

Dit kom egter voor asof die term "hiperteks" in toenemende mate as generiese term gebruik word vir dokumente wat in die gees van hiperteks (4.7.3.2 hierbo) gestruktureer is, selfs wanneer beeldmateriaal ingesluit word of daar ook van ander media gebruik gemaak word. Onderliggend aan dié uitgangspunt is die feit dat teks tog in die meeste van hierdie dokumente sentraal bly. In hierdie verslag word die benaming "hiperteks" in hierdie gangbare betekenis gebruik (vergelyk Conklin, 1987, p. 18).

Die term "hiperteks" self is egter so algemeen, dat dit nou ook weer nodig is om duidelik te onderskei tussen (onder andere) die volgende:

- Die konsep hiperteks, wat in programme van uiteenlopende aard vorm kan kry. Dit gaan hier dus eintlik oor programme wat op nie-liniêre teks gebaseer is waarby die skakels deur die rekenaar geoperasionaliseer kan word.
- Hiperteksstelsels, of -programme, wat die gebruiker in staat stel om hiperteksdokumente te skep, aan mekaar te koppel en weer te lees. Soms word dié benaming gebruik slegs vir die program, sonder inagneming van die materiaal wat daarmee geskep word; soms word dit egter ook gebruik vir die program saam met die materiaal, dit wil sê die hele werkende stelsel.

Dit is jammer dat die terminologie so onnoukeurig gebruik word. Die terminologie word egter wel so gebruik in die vakliteratuur. Die keuse vir die term "stelsel" wanneer die program bedoel word, hang onder andere daarmee saam dat die "program" soms uit 'n aantal nou met mekaar geïntegreerde programme of modules bestaan.

- Hiperteksdokumente, wat die materiaal behels wat deur hiperteksprogramme opgestel is en wat weer met behulp van so 'n program gelees kan word. Die benaming vir hiperteksdokumente verskil tussen verskillende programme/stelsels: in die geval van Guide (vergelyk 4.7.3.9) word daar gepraat van *guidelines* en in die geval van HyperCard van *HyperCard stacks* of net *stacks*. In baie gevalle word daar egter net van (elektroniese) dokumente gepraat.

Die dokument (as makro-eenheid) bestaan dan telkens weer uit verskillende mikro-eenhede. In die meeste gevalle gaan dit oor 'n skerm vol inligting, wat in

verskillende stelsels onderskeidelik aangedui word as raam (*frame*), bladsy (*page*) of kaart (*card*) (*HyperCard*). In *Guide* is die verhouding tussen skerm en dokument ietwat meer ingewikkeld: die inhoud van 'n venster kan 'n dokument wees, waarbinne dan beweeg kan word deur 'n rolbalk (*scrolling bar*). In verband met die terminologie, vergelyk Brown (1988, p. 48).

Die term "hipertekstoepassing" (*hypertext application*) is nog meer verwarrend en kan dui op enige van die bogenoemde fasette.

(Hierdie onderskeidings moet byvoorbeeld duidelik getref word wanneer die verskillende toepassings bespreek word. Vergelyk 4.7.3.10 hieronder.)

4.7.3.5 Belangrike aspekte in verband met hiperteks

- a) Hiperteks bring 'n wye, interdisiplinêre verskeidenheid aspekte op die tafel. Aan die een kant is daar verskeie aspekte van die tegnologie wat indringende aandag verdien en ten opsigte waarvan baanbrekerswerk nodig is. Aan die ander kant is 'n groot inset van kundiges op die gebied van taal en teks nodig, asook van kundiges op gebiede soos die sielkunde en die ergonomika.
- b) Hiperteks het besonder baie te make met die hele proses van menslike kennis (persepsie, assosiasie, inligtingsoek en inligtingverwerking, leer, ensovoorts). Die aanspraak word gemaak dat hiperteks baie naby aan die proses van menslike kennis te staan kom en dat dit ingrypende moontlikhede inhou vir die ondersteuning van dié prosesse deur die rekenaar (vergelyk ook 3.5).
- c) Hiperteks is so 'n kragtige meganisme dat dit noodsaaklik is om groter klarigheid te kry oor die epistemologiese en hermeneutiese aspekte wat by die gebruik daarvan ter sprake kom (Dolan 1988). Deurdat brokke inligting op 'n bepaalde wyse aan mekaar verbind of geskakel word (*linking*), kan dit baie maklik gebeur dat die inligting op 'n spesifieke (bedoelde of onbedoelde) wyse ideologies skeefgetrek word. Dit beteken natuurlik nie dat inligting wat in die tradisionele media (byvoorbeeld die boek) verpak word vry is van hierdie gevaar nie. Daar is te weinig oor die meer bekende media en hulle eie epistemologiese implikasies besin. Dit is juis die krag van hiperteks wat die oë geopen het vir die feit dat die implikasies van die media inderdaad by alle media in ag geneem moet word.

In die geval van 'n boek speel gewoon die feit dat dit nie beskikbaar is nie kennis-sosiologies gesproke ook 'n rol in die proses van kennisbemiddeling. Die sosiologiese en logistiese aspekte van die inligtingsdraers waarmee soms onnadenkend omgegaan word, verdien dus ook die aandag. Oor die posisie van die boek as inligtingsdraer in die geesteswetenskappe, vergelyk Fabian (1983).

- d) Hiperteks is 'n konsep en nie 'n finaal ontwikkelde produk nie. Trouens, vir baie toepassings is dit nog totaal onvoldoende en is nog baie ontwikkeling en verfyning nodig (vergelyk 4.7.3.11). Dit is egter belangrik om die konsep en die

spektrum van nuwe moontlikhede te skei van die werklike produkte en elkeen apart te beoordeel.

4.7.3.6 Die struktuur van hiperteksdokumente

Die struktuur van hiperteksdokumente is eintlik baie eenvoudig. Dit bestaan naamlik net uit twee komponente, naamlik nodi (*nodes*) en skakels (*links*). Hierdie nodi word nou deur die skakels op so 'n wyse met mekaar verbind dat die leser van 'n hiperteksdokument in staat is om van die een nodus na die volgende te beweeg op grond van sy inligtingsbehoefte en binne die raamwerk van moontlikhede wat die betrokke hiperteksstelsel/-program vir hom toelaat.

'n Aantal skakelareas (*linking areas*) binne 'n nodus (wat 'n woord of 'n simbool kan wees) word as die aktiewe areas gedefinieer en aan dié word 'n verdere nodus gekoppel deur 'n skakel.

Saam hiermee is 'n bepaalde program nodig om die gebruiker in staat te stel om die skakels te volg en daardeur na ander nodi te beweeg of terug te beweeg na die plek vanwaar hy begin het.

Skakels word op uiteenlopende wyse gebruik in die verskillende stelsels. Tog kan die skakels as een van twee soort beskryf word. Die terminologie verskil grootliks: Brown (1988, pp. 45-46) praat van die tipes *hierarchical* en *cross-referencing*; Conklin (1987, pp. 33-35) noem die ooreenstemmende skakels *organizational* en *referential*.

- i) 'n Hiërargiese skakel lei tot verdere vlakke van inligting.
- ii) 'n Kruisverwysingskakel lei tot ander inligting in dieselfde of in 'n ander dokument (naamlik inligting wat assosiatief verband hou).

Die aard en die funksie van skakels in die verskillende programme/stelsels is een van die belangrikste aspekte van hiperteks en vorm die voorwerp van indringende bespreking in die mees resente literatuur. Sommige baie belowende moontlikhede in dié verband is byvoorbeeld nog nie behoorlik ontgin nie (vergelyk 4.7.3.11).

Nodi is in die onderskeie stelsels ook van uiteenlopende aard en word deur verskillende benamings aangedui (vergelyk bespreking by 4.7.3.4).

4.7.3.7 Die lees van hiperteksdokumente

Breedweg gestel, funksioneer die proses van die lees van 'n hiperteksdokument soos volg:

- i) Die leser het voor hom op die skerm 'n stuk teks, waarbinne hy die moontlikheid het om op en af te kan rondbeweeg (*scrolling*). Hy kan met die wyser (*cursor*) (wat ook deur die muis geaktiveer kan word) vinnig enige posisie op die skerm aandui en daarop reageer.
- ii) Sekere woorde in die teks is op 'n manier (byvoorbeeld deur tipografie) aangedui as woorde ten opsigte waarvan die leser nog meer inligting kan

aanvra deur slegs 'n knoppie te druk. Enige area op die skerm (byvoorbeeld 'n deel van 'n voorwerp of 'n simbool in die dokument) kan aangedui word as 'n area ten opsigte waarvan die leser bykomstige inligting kan aanvra deur vanuit daardie area te reageer.

- iii) Deur die wyser met die muis op die betrokke area te plaas en die knoppie van die muis te druk, kan die leser dan onmiddellik meer inligting (hetsy in die vorm van teks of beeldmateriaal) oor die betrokke woord/item kry. Gewoonlik gaan 'n venster op die skerm oop, waarbinne die nuwe materiaal dan vertoon word en waarbinne daar ook weer rondbeweeg kan word. Waaroor dit in die betrokke venster gaan, word deur 'n naam of opskrif in die boonste raam van 'n venster aangedui sodat die leser die band na vorige vensters kan behou. Die nuwe inligting kan op verskillende wyses aangebied word, wat van program tot program verskil.
- iv) Die nuwe stuk inligting wat nou vir die leser gebied word, bied moontlik ook weer die geleentheid om verdere inligting aan te vra. In so 'n geval word 'n verdere venster oopgemaak waarbinne die inligting vir die leser vertoon word.
- v) Nadat die leser na al hoe meer gevorderde vlakke deurgedring het, kan hy ook weer deur die verskillende vlakke terugbeweeg na waar die soektog begin het, deur telkens 'n venster toe te maak deur die muis te aktiveer op 'n hoekie van die venster. In sommige stelsels is dit moontlik om vanaf laer vlakke direk na hoër vlakke te beweeg, sonder om eers al die vensters tussen-in toe te maak.

4.7.3.8 Die ontwikkeling van hiperteksdokumente

Vir die ontwikkeling van 'n hiperteksdokument moet groter gedeeltes of tekste opgebreek word in kleiner modules (wat self uiteraard ook die karakter van langerige dokumente kan hê). Hierdie proses van die opbreek van bestaande dokumente in modulêre eenhede wat as nodi kan funksioneer, is 'n kuns en vereis 'n baie goeie kundigheid van die samehang van die elemente van inligting binne die teks, van die verwagte of moontlike begrip van die betrokke tema aan die kant van die waarskynlike of moontlike leser en van die struktuur en moontlikhede van die betrokke program.

Die konsep van nie-liniêre teks moet deurlopend in gedagte gehou word, wat beteken dat die verdeling in modules – Conklin praat van 'n *modularization of ideas* (1987, pp. 35-36) – moet geskied op so 'n wyse dat verdere vlakke van inligting ook inligting op 'n verdere vlak (gewoonlik 'n laer vlak van bekendheid) moet verteenwoordig. Indien dié inligting nie werklik op 'n ander vlak van kennis lê nie, is dit natuurlik ook moontlik om dit in 'n vlak agter die boonste (of vorige) vlak te verpak, maar dan moet dit op so 'n wyse in verband gebring word met die boonste vlak en moet die skakelpunte so duidelik aangedui word dat die leser nie basiese inligting misloop nie.

Claassen & Bothma (1988, p. 86) toon aan dat hierdie beweging na ander vlakke van inligting nie so eenvoudig is soos in sommige stelsels aangebied nie en dat die behoefte aan verdere inligting nie net as 'n vertakking op hiërargiese vlak beskou kan word nie

(vergeelyk 4.7.3.6 oor die aard van nodi). By inligtingstelsels op komplekse gebiede is die behoefte nie net om meer inligting te kry nie, maar dikwels ook om meer van 'n spesifieke soort inligting te kry.

Wat baie belangrik is, is dat die nodige skakels en skakelareas saam met die opbreek van die teks beplan moet word. Indien trefwoorde as skakelareas gaan dien, is dit belangrik dat daar in elke hoër vlak van die dokument die regte trefwoorde sal wees wat die leser by die volgende vlakke sal uitbring.

Om gewone, liniêre teks voor te berei vir verpakking in 'n hiperteksdokument, kan dus 'n baie ingewikkelde proses wees. Om teks uit die staanspoor te skep, vereis dat die skrywers eers opgelei moet word om te begryp wat nie-liniêre teks is en hoe dit in 'n hiperteksstelsel funksioneer.

Die grootste probleem by die skep van hiperteksdokumente is die feit dat daar 'n bepaalde vorming (opbreking, formattering, kommentariëring) en skakeling van die materiaal plaasvind wat alreeds met die moontlikhede van 'n spesifieke hiperteksstelsel saamhang (en wat op hulle beurt weer saamhang met gegewenhede in die apparatuur en stelselprogrammatuur). Dit kan gladnie voorsien word hoe die hiperteksstelsels van oor vyf jaar daar gaan uitsien nie. Die enigste wyse om hierdie probleem – minstens in 'n baie groot mate – te omseil, is om die teks (en meegaande illustrasiemateriaal) in sy kleinste elemente te ontwikkel in 'n "medium-onafhanklike" vorm, vanwaar dit dan weer betreklik gemaklik in verskeie vorms (onder andere die mees sinvolle hiperteksstelsel) herverpak kan word. Dan is dit belangrik dat daar gehou sal word by internasionaal aanvaarde standaarde, waarvan 'n mens weet dat daarvoor in toekomstige stelsels voorsiening gemaak sal word. Die materiaal kan dan uiteraard ook in ander vorms (of stelsels of media) verpak word. SGML is so 'n standaard vir tekstkodering en daar word alreeds in 'n paar hiperteksstelsels daarvoor voorsiening-gemaak, byvoorbeeld in die IDEX-stelsel deur OWL International Inc. (OWL, 1988). Ook in die omvattende Perseus-stelsel vir die Griekse kultuur word voorsiening gemaak vir die verpakking van die materiaal in 'n medium-onafhanklike vorm, vandaar dit dan, ooreenkomstig die ontwikkeling van die behoeftes, in 'n ander stelsel verpak kan word (Hughes, 1988a, pp. 3, 6). Tekstkoderingstandaarde word in 5.4 verder bespreek.

4.7.3.9 Hiperteksprogramme

Binne die raamwerk van hierdie verslag is dit nie moontlik om indringend aandag selfs aan 'n paar van die wye verskeidenheid hiperteksprogramme te gee nie. Die meer algemene of generiese aspekte is alreeds behandel en die spektrum van toepassings sal hieronder die aandag geniet (4.7.3.10). Daar sal kortliks aandag gegee word aan een baie populêre hiperteksprogram in die MS-DOS-omgewing, naamlik Guide. Verder sal daar gewys word op sekere aspekte van 'n ander baie belowende hiperteksprogram in die MS-DOS-omgewing, naamlik KnowledgePro.

Guide was een van die eerste kommersieel beskikbare programme en het die konsep van hiperteks na die gewone gebruiker van mikrorekenaars gebring. Guide is aanvanklik vir

die UNIX-omgewing ontwikkel en is vandaar oorgedra na die MacIntosh-omgewing en die MS-DOS-omgewing onderskeidelik (Brown, 1987). Die eerste weergawe (Guide 1.0) vir algemene gebruik is ontwikkel vir die MacIntosh (Hershey, 1987). Guide vir die MS-DOS-omgewing was ook die eerste hipertekstprogram vir dié omgewing (OWL, 1988a). Guide word intussen gebruik vir talle toepassings van groter en kleiner omvang, vanaf omvattende inligtingstelsels (byvoorbeeld in 'n diagnostiese stelsel vir die versiening van motors, vergelyk Markoff, 1988) tot 'n vorm van rekenaargestuende onderrig wat nie moontlik is met behulp van tradisionele RGO-programme nie (byvoorbeeld in die Departement Duits van die Universiteit van Stellenbosch, vergelyk Kussler, 1989).

In die handleiding van Guide word hipertekst beskryf as *a high-speed, automated way of browsing through information* (OWL, 1988, p. 3). Dit word gestel dat elektroniese dokumente as medium soveel meer bied as papier. Guide wil 'n stuk gereedskap wees vir die skryf en die lees van elektroniese dokumente: die dokument kan nou gestruktureer word om die inligting te pas, eerder as andersom in die geval van dokumente op papier. Die dokumente wat op dié wyse geskep word, kan dinamies en interaktief wees: inligting wat nie van belang is vir sekere lesers nie, kan verbygegaan word sonder dat dit eers vlugtig deurgelees hoef te word; gespesialiseerde vlakke van inligting kan egter wel gemaklik verkry word deur dié wat dit nodig het. Omdat sommige van dié funksies verwantskap toon met funksies in ander programme, word Guide spesifiek afgegrens van ander programme/funksies, byvoorbeeld van woordverwerkingsprogramme, raamwerk-fasiliteite en intydse help-fasiliteite (OWL, 1988, pp. 3-6). Die dokumente wat in Guide geskep word, heet *guidelines*.

Hierdie voordele van Guide word gerealiseer deur die gebruik van knoppies (*buttons*). 'n Knoppie in Guide is 'n aktiewe (of "lewendige") area wat op die skerm geskep kan word. Deur die wyser met die muis op die knoppie (= aktiewe area) te plaas en die knoppie van die muis te druk, word nuwe inligting te voorskyn gebring, op 'n wyse wat afhang van die soort knoppie of skerm-area wat geaktiveer is. Daar is in Guide drie soorte knoppies wat dien om die vlakke van inligting te beheer:

- Uitbreidingsknoppies (*expansion buttons*)
- Verwysingsknoppies (*reference buttons*)
- Notaknoppies (*note buttons*).

'n Verdere knoppie is 'n opdragknoppie (*command button*) wat gebruik word om toegang te verkry tot inligting buite Guide.

Die knoppies funksioneer soos volg:

- Uitbreidingsknoppies: Deur die aktiewe area te aktiveer, word die knoppie uitgebrei of vervang deur teks, grafika of deur meer knoppies. Uitbreidingsknoppies word in die geval van 'n woord tipografies aangedui deur vet letters (*bold*). Die leser kan vooraf agterkom dat die knoppie wel bykomstige inligting

kan bied, aangesien die vorm van die wyser verander wanneer dit op die knoppie geplaas word.

- Verwysingsknoppies: Dié knoppies word gebruik om skakels na inligting in ander dele van dieselfde hiperteksdokument (*guideline*) of in ander hiperteksdokumente te aktiveer (gewoonlik in 'n afsonderlike venster). Dit word dus gebruik vir kruisverwysings. Verwysingsknoppies word tipografies aangedui deur kursivering. Sodra die inligting op hierdie ander plek verwerk is, kan die leser teruggaan (*backtrack*) na die plek waar hy vroeër was deur met die muis die teruggaan-simbool te aktiveer. Elke aktivering van die muis op dié simbool neem die leser 'n verdere vlak van verwysing terug.
- Notaknoppies: Dié knoppies word gebruik vir annotasie. Deur die muis te aktiveer op die knoppie of aktiewe area, verskyn die inligting in 'n venster wat net vir dié doel geskep word. Die inligting (en venster) bly op die skerm net solank as wat die knoppie van die muis ingedruk word.

In plaas van moeisaam (deur die aktivering van die teruggaan-simbool) terug te beweeg deur die verskillende vlakke van inligting wat deur die verwysingsknoppies oopgevlak is, kan die leser ook met een enkele opdrag terugbeweeg na die hiperteksdokument in sy eenvoudigste vorm. Hiervoor word die opdrag *Top Level* onder die opdraggroep *Display* gebruik. In dié opdraggroep is daar ook ander opsies wat 'n gemaklike manipulerings van die hiperteksdokument moontlik maak (OWL 1988, pp. 28-35).

Die program maak dit ook moontlik om self hiperteksdokumente te skep. Teks kan van buite af ingevoer word, asook grafika vanaf bekende grafika-programme.

Die volgende verdere opmerkings kan oor *Guide* gemaak word:

- a) *Guide* is dokument-gebaseerd (anders as *HyperCard* wat kaartjie-gebaseerd is) (vergelyk 4.7.3.3).
- b) Die leser kan op en af deur die dokument beweeg deur van die rolbalk (*scrolling bar*) gebruik te maak.
- c) Daar word in *Guide* baie dinamies en verbeeldingryk gebruik gemaak van kleur.

In die geheel beskou is *Guide* 'n baie bruikbare program, wat vir die belangrikste aspekte van hiperteksvoorsiening maak. Die toekomstige ontwikkeling van *Guide* in die MS-DOS-omgewing word sekerlik effens aan bande gelê deurdat die grafika in dié omgewing nie so soepel is soos in die Macintosh-omgewing nie. Tog is die MS-DOS-omgewing 'n baie belangrike omgewing vir die toekomstige implementering van hiperteks. Daar is aangekondig dat die ontwikkelaars van *Guide* in hulle ander program *IDEX* voorsiening gaan maak vir netwerkfasiliteite en SGML as koderingstandaard (OWL, 1988, p. 8). Dit sal baie sinvol wees indien dié fasiliteite ook in toekomstige weergawes van *Guide* geïnkorporeer kan word.

KnowledgePro dien hom aan as 'n program wat hiperteks kan verbind met kunsmatige intelligensie en deskundige stelsels. Dit is byvoorbeeld moontlik om die heen en weer beweeg tussen verskillende vlakke van inligting parameter-aangedrewe te laat verloop, dit wil sê dat daar nie telkens net méér inligting verskaf word nie, maar dat die program die keuse vir verdere inligting of vir spesifieke vlakke van inligting dinamies kan modifiseer op grond van reaksie van die leser/gebruiker en 'n manipulerings van die data. KnowledgePro bied volle ondersteuning vir die skryf van reëls vir deskundige stelsels. Hierdie benadering hou baie belofte in, aangesien dit 'n baie groter mate van soepelheid en vindingrykheid as ander stelsels toelaat.

Shepard (1987) sê van KnowledgePro:

It provides expert system development and consultation capabilities, a list processing language, hypertext, windows, access to external graphics and other specialized software, and seamless data exchange with spreadsheets, databases, and popular artificial intelligence application tools.

Die volle funksionaliteit van die kombinasie van hiperteks, kunsmatige intelligensie en deskundige stelsels kan verkry word deur saam met KnowledgePro die program KnowledgeMaker (van dieselfde firma) te gebruik. Dié program kan weer baie sinvol gebruik maak van TurboPROLOG-programme. (Vergelyk ook Thompson, 1988.)

4.7.3.10 Toepassings van hiperteks

Die toepassingsmoontlikhede van die konsep van hiperteks is byna onbeperk. Soos vroeër aangetoon, moet daar egter onderskei word tussen die konsep hiperteks, hiperteksstelsels of -programme en hiperteksdokumente.

a) *Ensiklopediese werke*

Die tipiese aksie van die gebruiker in 'n ensiklopedie kan ook beskryf word as 'n nie-liniêre beweeg tussen verskeie inskrywings, waarby die leser telkens nuwe leidrade optel en opvolg deur onder 'n verdere trefwoord te gaan naslaan. Die vinnige toegang tot teks in 'n elektroniese ensiklopedie (byvoorbeeld die *American Academic Encyclopedia* van Grolier) kan die indruk wek dat dit wel om hiperteksmatige toegang tot die materiaal gaan. In werklikheid word die vinnige toegang egter bewerkstellig deur omvangryke indekse waarin alle (moontlik relevante) woorde in die teks georden word.

Ensiklopedieë kan op twee wyses deur hiperteks geoperasionaliseer word:

- i) deurdat die materiaal van die ensiklopedie self gestruktureer word deur 'n veelvlakkigheid in te bou, of
- ii) deurdat 'n ensiklopedie as 't ware ingebou word in 'n meer omvattende stelsel en vanuit die stelsel daar gemaklik toegang verkry kan word tot die verskillende inskrywings.

Dit kan verwag word dat daar op die gebied van elektroniese ensiklopedieë interessante ontwikkelings gaan plaasvind in die volgende paar jaar. Vergelyk Shneiderman (1987) vir die wyse waarop 'n elektroniese ensiklopedie in hiperteks (spesifiek in die Hyperties-stelsel) hanteer kan word.

b) Inligtingstelsels vir omvattende kennisgebiede

Deur sy vermoë om die gebruiker gemaklik tussen inligting op verskillende vlakke en bestaande uit verskillende datatipes te kan beweeg, het hiperteks hom baie gou aangedien as die ideale stelsel vir omvattende kennisgebiede, waarin gebruikers met uiteenlopende kundigheid gemaklik inligting kan vind. Uiteraard sou gewone elektroniese databasisse ook sekere moontlikhede bied, maar dit was duidelik dat daarmee nie 'n inligtingstelsel met dieselfde soepelheid en groot mate van gebruikersvriendelikheid van hiperteks tot stand sou kon kom nie.

Die volgende gebiede toon aan watter verwagtinge van hiperteks gekoester word as dié medium wat sekere skaakmatposisies in kennisbemiddeling in die geval van omvattende kennisgebiede kan ophef:

- i) Die Perseusprojek is die mees omvangryke projek wat nog as hiperteksprojek aangepak is. Die doelstelling met die stelsel is om 'n stel teks- en beeldmateriaal oor antieke Griekeland te ontwikkel met die oog op die gebruik daarvan in kursusse (van uiteenlopende aard) op universiteitsvlak. Die stelsel wil egter nie slegs op didaktiese vlak 'n rol speel nie; dit wil ook help om brûe te skep tussen gebiede wat deur toenemende spesialisasie al hoe verder van mekaar afgegrens word. Ook navorsers moet dus in staat gestel word om gemaklik oor die grense van eng gebiede te beweeg. Die gebruiker sou byvoorbeeld by 'n Griekse teks kon begin en dan deur knopdruk na die woordeboek-beweeg-of-na-'n-vertaling-'n-Sprong-na-die-kultuurhistoriese materiaal sou net so maklik kon plaasvind.

Daar word gebruik gemaak van HyperCard op die Macintosh-rekenaar, alhoewel die materiaal van die begin af medium-onafhanklik kodeer word met die oog op die gebruik van enige ander sinvolle hiperteksstelsel in die toekoms.

Die stelsel sal uiteindelik op CD-ROM beskikbaar wees. Die Perseusprojek ontvang baie groot finansiële ondersteuning van groot stigtings in die VSA en geniet die samewerking van 'n groot aantal kenners op die vakgebied (vergelyk Hughes, 1988a; Crane, 1987).

Dit is belangrik om daarop te let dat die Perseusprojek 'n uitermate groot versameling teksmateriaal behels wat op 'n bepaalde wyse verpak is om in die inligtingsbehoefte van 'n uiteenlopende verskeidenheid gebruikers te voorsien.

- ii) Tydens die eerste hiperteks-kongres is 'n ander vakgebied voorgestel waar die probleem ervaar word van 'n omvangryke versameling teksmateriaal wat nie

voldoende toeganklik is vir die gebruiker nie, naamlik die mediese vakgebied. Dit het geblyk dat die meganisme van hiperteks baie geskik is vir die prosesse van inligtingsoek wat tipies in die mediese praktyk plaasvind. Hierdie inligting bestaan ook uit 'n groot hoeveelheid teksmateriaal. 'n Deeglike analise is gemaak van die tipiese "inligtingstelsel" wat die medikus vir homself skep om die daaglikse situasies te kan hanteer en 'n sekere hoeveelheid kennis sy eie te maak – terwyl hy dan tog ook soms toegang tot 'n stel ander, meer gevorderde inligting benodig. Op die basis van hierdie ondersoek word daar tans aan die Washington School of Medicine 'n omvattende hiperteksstelsel vir die medisyne ontwikkel (vergelyk Frisse, 1988).

- iii) 'n Teksgebaseerde vakgebied soos die eksegetiese wetenskappe (en daarmee saam die Oriëntalistiek) bied ook baie interessante moontlikhede vir die implementering van hiperteks. In hierdie geval – in 'n mate vergelykbaar met die Perseusprojek – speel taal op twee vlakke 'n rol: enersyds kan die groot volume teksmateriaal net deur die intelligente omgang met taal toeganklik gemaak word; andersyds is die taal in die eksegetiese prosesse ook die voorwerp van bestudering en word nuwe perspektiewe net verkry uit 'n noukeurige omgang met die taalgegewens. Die oorspronklike teks van die Bybel vorm telkens die oriëntasiepunt van die verstaansproses (vergelyk bronnelys in Claassen & Bothma, 1988).

c) *Die operasionalisering van literêre tekste*

In al die gevalle hierbo (b) is daar tekste wat 'n basiese kern vorm en waarop daar telkens weer teruggegryp word, of, andersom gestel, tekste van waaruit 'n proses van ontginning van die betekenis van die teks, die agtergrond daarvan, ensovoorts, plaasvind. Die tekste word gekommentareer, verduidelikende aantekeninge ten opsigte van hierdie tekste word gemaak, ensovoorts. Dit geld veral ten opsigte van (i) en (iii), maar sekerlik ook (ii) waar daar ook basiese tekste is wat fundamentele insigte (of behandelings) weergee en waarop daar in veranderende omstandighede verdere kommentaar gelewer word. Aan hierdie proses van die kommentariëring van teks is alreeds aandag gegee (4.5). Die teks wat in die vorm van kommentariëring of as toeliggende teks aangebied word, kan op verskillende vlakke van inligting lê.

Hiperteks vorm nou egter 'n nuttige meganisme om hierdie kommentaar op die teks te operasionaliseer. Dit is juis wanneer daar 'n baie groot hoeveelheid teks is en dit op verskillende vlakke van inligting lê dat hiperteks 'n baie doeltreffende meganisme vir die operasionalisering vorm. Die feit dat die inligting as't ware op aanvraag beskikbaar kan wees, verhoog die moontlikhede van die sinvolle omgang met teks (vergelyk Van der Merwe, 1988).

d) *Handleidings*

Die verskaffers van produkte waarvoor omvattende handleidings benodig word, het baie gou groot belangstelling in hiperteks getoon. Die probleem is naamlik dat die gedrukte produk te omvangryk kan wees en die gebruiker kan oorlaai met inligting. D.m.v. 'n

hipertekstelsel kan die groot hoeveelheid teks- en beeldmateriaal egter baie sinvol in 'n inligtingstelsel verpak word, op 'n wyse wat aan die gebruiker se direkte inligtingsbehoefte voldoen. In verskeie bydraes by die Hypertext '87-kongres is die moontlikhede van toepassings van dié aard uitgelig – ook in enkele voorstelle vir referate wat uiteindelik nie aanvaar is nie (vergelyk die kongresbundels *Hypertext '87 Papers*, *Hypertext '87: Information en Online Information 88*).

Daar kan verwag word dat hipertekst in toenemende mate gebruik sal word vir toepassings van dié aard.

e) Woordverwerkingsprogramme

Daar is alreeds aandag gegee aan 'n woordverwerkingsprogram wat as uitgangspunt neem die konsep van hipertekst (4.2.1.1c). In *Writing Environment* word die verskillende kognitiewe modi betrokke by die prosesse van lees en teksgenerering onderskei en meganismes geskep wat vir elkeen die mees doeltreffende blyk te wees. Die tipering van dié program as 'n hipertekst-omgewing hang daarmee saam dat die dokument vanuit verskillende perspektiewe beskou kan word, dat dit uit verskillende nodi bestaan en dat die beweging tussen dié nodi eintlik net deur die skakeling daarvan op die rekenaar geoperasionaliseer kan word.

f) Rekenaar-gebruiker-raakvlakke by databasisse

Baie van die groot konvensionele databasisse is bekend daarvoor dat die gebruikers-raakvlak nie baie gemaklik is nie: die gebruiker moet verskillende stelle instruksies leer vir die verskillende programme, ervaar probleme met verskillende soekstrategieë en ervaar die intydse aktiwiteit tog as ietwat verwyderd van die werklikheid van sy tipiese opset van waar hy die inligting benodig. Die toenemende aantal databasisse op CD-ROM wat intussen beskikbaar raak, bring mee dat die probleem van moontlike verwarring en oneffenhede tussen verskillende stelsels nog groter gaan word. Hierby kom ook nog dat meer en meer gebruikers met onvoldoende kennis van databasisraakvlakke (en opleiding in die inligtingkunde) toegang tot databasisse sal wil soek. In die lig van die algemene bekendheid van 'n hipertekstprogram soos *HyperCard*, met die moontlikheid om te dien as 'n meganisme vir die sinvolle integrasie van 'n verskeidenheid inligting, word daar tans baie aandag gegee aan die moontlikheid om bekende programme van dié aard aan te wend as gebruikersraakvlakke tot die minder bekende databasisse. Die krag van hipertekst om 'n gebruikersvriendelike omgewing te bied, 'n verskeidenheid tipes inligting naatloos aan mekaar te koppel en die gemaklike integrasie van taalkomponente soos thesouri moontlik te maak, maak hipertekst hiervoor die ideale stelsel. Op dié wyse kan dit moontlik raak om knopdruk-vinnig deur algemene of hoogs gespesialiseerde tesouri te beweeg of onbekende terme vanuit 'n intydse ensiklopedie te belig, terwyl die ontwikkeling van KI-tegnieke en natuurlike taal-soekmeganismes die moontlikhede nog verder kan uitbrei (vergelyk Kesselman & Trapasso, 1988). (Vergelyk Davis, 1989 in verband met hipertekst-raakvlakke by bestuursinligtingstelsels.)

g) Databasisse

Die konsep van hiperteks het so belowend geblyk te wees dat dit nie lank geduur het nie voordat dit ook uiting gevind het as 'n funksie in databasisse. Dit is veral in die geval van volteks-databasisse dat die konsep van skakels blyk baie belangrik te wees: skakels tussen woorde (wat as die skakelarea/-punt dien) en nodi (die brokke teks of ander materiaal waarheen dit lei). Hierdie funksie is byvoorbeeld ingebou in AskSam, 'n hoogs suksesvolle volteks-databasis. Dit mag baie ondernemend voorkom om die term hiperteks te gebruik vir iets wat net deels ooreenstemming vertoon met werklike hipertekstelsels. Hughes beskou dit as 'n *hypertextlike search utility* (1987a, p. 14). Dit kan egter verwag word dat hierdie soort funksie toenemend geïmplementeer sal word en toenemend die volle omvang van hiperteksfasiliteite sal behels.

h) 'n Nuwe konsep van rekenaargesteunde kennisbemiddeling en kennisverkenning

Besonder baie van die toepassings waarin hiperteks vorm kry, het op dié een of ander wyse te make met kennisbemiddeling. Om egter al hierdie toepassings te beskou as 'n moderne of gesofistikeerde vorm van rekenaargesteunde onderrig, kom neer op 'n growwe vereenvoudiging. Baie toepassings is uiteraard nie meer as net 'n gesofistikeerde vorm van RGO nie – veral HyperCard word intensief hiervoor gebruik.

Soos hierbo geblyk het (4.7.1), kan beide rekenaargesteunde onderrig (RGO) en hiperteks onder 'n gemeenskaplike hofie van teksaanbieding geplaas word. Van tradisionele RGO verskil hiperteks egter aansienlik, in die mate dat dit hier nie net om 'n oor-en-weer aanbieding van vrae en antwoorde gaan nie, maar om 'n totaal anderssoortige koppeling van elemente aan mekaar, wat veral vanuit die perspektief en behoeftes van die gebruiker self gestuur word.

Dit is eers wanneer hiperteks binne die raamwerk van ander inligtingsdraers van teks of beeldmateriaal gesien word, en wanneer in ag geneem word dat die meeste vorms van ernstige teks in elk geval 'n kennisbemiddelende karakter het, dat die rol van hiperteks ten opsigte van kennisverwerwingsprosesse in perspektief gesien kan word. 'n Handleiding wil byvoorbeeld kennis bemiddel, maar dit is ook duidelik dat baie handleidings nie die proses van die soek van inligting aanhelp nie, aangesien die feit dat alles op 'n "plat oppervlak" lê, die soektog juis kan bemoeilik.

Die mees oortuigende bewyse van die waarde van hiperteks in 'n onderrigopset is die **Intermedia**-stelsel van die IRIS-groep, een van die eerste, maar ook een van die mees doeltreffende stelsels (vergelyk 4.7.3.3 hierbo). Kursusse in uiteenlopende vakke soos Engels en Biologie het aangetoon dat hiperteksmatige, kennisverkennde omgang met die vakinhoud inderdaad lei tot die ontwikkeling en bevordering van nie-liniêre denke (*non-linear thought*) (vergelyk Beeman et al., 1987; Landow, 1988; Meyrowitz, 1986; Yankelovich et al., 1985; 1987a; 1988; Yankelovich & Van Dam, 1987).

Die Perseusprojek (b (i) hierbo) het ook 'n baie duidelike didaktiese inslag en beloof groot moontlikhede in te hou. Verdere verwysings na projekte met in die eerste plek 'n didaktiese doel word gevind in vele onlangse publikasies oor hiperteks.

Dit blyk dus dat hiperteks en hipertekstoepassings tog ook weer iets anders is as rekenaargesteuende onderrig (RGO). Hiperteks kan wel gebruik word in die proses wat algemeen bekend staan as rekenaargesteuende onderrig, maar dit bied ook veel meer, onder andere omdat dit 'n baie meer soepel en gebruiker-gestuurde proses van kennisverkenning toelaat.

Dit kom egter voor asof die ontwikkeling en wyer gebruik van hiperteks kan meebring dat die tradisionele konsep van RGO aangepas sal moet word. Dit kan verwag word dat tradisionele RGO-programme toenemend komponente van hiperteks sal begin akkommodeer. Die vraag is egter of die proses nie eerder fundamenteel omgekeer moet word en daar eerder in die ryke, kennisverkennende omgewing van hiperteks meer toetsingskomponente ingebou moet word nie. Soos hierbo aangetoon (4.7.3.9), maak KnowledgePro dit reeds moontlik om die kennisverkennende proses parameter-aangedrewe te laat verloop. Hierdie parameters kan net sowel dié wees wat tipies in die RGO-proses die vertakking, terugvoer en bepunting bepaal.

As aanvaar word dat die meeste gevalle van teks wat in elektroniese vorm beskikbaar raak of behoort te wees, tog in elk geval 'n kennisbemiddelingskarakter het, behoort dit duidelik te wees dat die grense van wat RGO is en wat nie, nou baie fyner geraak het. Kerr stel dit so dat die meeste gevalle van elektroniese teks beskou moet word as teks wat op die oog het dat die gebruiker ingelig of opgelei moet word (*instructional text*) (1986, pp. 325-327).

Die vraag moet selfs gestel word of die ontwikkeling van hiperteks nie selfs vereis dat daar indringende vrae gestel moet word oor die didaktiese terrein in die breë of selfs oor die hele terrein van die verpakking van kennis in die algemeen nie.

i) Onbeperkte toepassings?

Die verrassend nuwe aan hiperteks, asook die feit dat dit opsigtelik 'n nuwe omgang met inligting moontlik maak, het daartoe gelei dat 'n aantal skrywers al verby konkrete toepassings kyk en 'n nuwe visie van die toekoms sien. Vergelyk byvoorbeeld Drexler (1986). Die talle publikasies van die visionêre Ted Nelson verdien ook vermelding in dié verband. Dié aansprake mag onrealisties en te ondernemend voorkom, maar dit toon minstens 'n optimisme ten opsigte van die gebruik van die rekenaar ten opsigte van taal- en teksmateriaal, wat 'n groot stukrag aan die ontwikkelings kan gee.

4.7.3.11 Probleme in verband met hiperteks

Daar is alreeds daarop gewys dat die beperkinge verbonde aan bestaande hiperteksprogramme nie daartoe moet lei dat die konsep van hiperteks ook noodwendig in die gedrang kom nie. Baie van die huidige beperkinge hang nog saam met die onvermoë van die persone betrokke by die ontwikkelings om te ontsnap aan die beperkinge van die

medium van liniêre teks en die beperkende tegnologiese moontlikhede van vroeër. Ander word nog bepaal deur die huidige stand van die tegnologie.

Die volgende probleme en beperkinge kan uitgesonder word (vergelyk Claassen & Bothma, 1988, pp. 87-88 vir 'n lys van die probleme wat nog die ontwikkeling van 'n stelsel vir 'n gebied soos Bybelse inligting in die weg staan):

- a) Die moontlikheid van verlore te kan raak tussen die talle skakels en nodi (met telkens ondergeskikte vlakke van inligting) is in die meeste stelsels 'n werklikheid. In die hiperteks-jargon word gepraat van *getting lost in hyperspace*. In sommige stelsels word 'n grafiese oorsig oor die samehang van die nodi verskaf, maar in die meeste gevalle is dié nog onvoldoende. Daar word alreeds indringend aandag gegee aan dié probleem.
- b) Die spesifieke woord as skakelarea wat 'n direkte skakel na ander brokke inligting aktiveer (4.7.3.6), is nog te star en beperkend. Die interaksie tussen woord en begrip vind so dinamies plaas dat 'n stelsel wat met die funksionering van taal by die gebruiker rekening wil hou, voorsiening behoort te maak vir 'n ingeboude tesourus. In so 'n geval kan die beweging na verdere inligting dan plaasvind nie net op grond van 'n spesifieke woord nie, maar ook op die basis van verwante woorde. Dit kom voor asof dié soort funksie in KnowledgePro hanteer sal kan word, aangesien dié program oor kragtige opdragte beskik om woorde of lyste woorde te kan manipuleer (vergelyk Shepard, 1987). In die geval van persone wat tipies met toepassings te make het wat nie in hulle moedertaal is nie, kan 'n ingeboude tweetalige woordeboek die proses nog verder vergemaklik.
- c) Dit is nog nie duidelik op watter wyse die insigte van kunsmatige intelligensie en tegnieke wat daaruit voortvloei, geïntegreer kan word in 'n hiperteks-omgewing nie. In 'n baie groot mate gaan dit in bestaande hiperteksstelsels oor die meganiese verpakking van die teks (en beeldmateriaal) in 'n nuwe omgewing. "Meganies" moet in dié verband nie negatief geïnterpreteer word nie; die verpakking kan juis baie ingewikkeld wees en 'n groot mate van vindingrykheid vereis. Die koppeling van kunsmatige intelligensie en hiperteks toon baie belofte, aangesien dit hierdeur moontlik word om die gehalte van die interaksie met die rekenaar te beheer en sinvoller by die behoeftes van die gebruiker aan te pas. Shepard (1987) praat (in aansluiting by N. Negroponte) van *qualities of interaction*. Deur die "intelligente" rekenaaromatige toegang tot en omgang met tekstmateriaal buite die hiperteks-omgewing (byvoorbeeld databasisse) op die basis van kunsmatige intelligensie kan die hiperteksstelsel 'n venster tot 'n baie wye wêreld van inligting vorm (vergelyk Halasz, 1988, pp. 846-847 en Shepard, 1987).
- d) In bestaande hiperteksstelsels word daar nog nie voldoende voorsiening gemaak vir die koöperatiewe aspek nie. In omgewings waar skeppende werk plaasvind, is dit egter baie belangrik. Dit is juis die konsep van hiperteks wat uitstekende moontlikhede bied vir die ondersteuning van dié prosesse van die koöperatiewe hantering van teks (vergelyk 5.3). Ongelukkig laat bestaande stelsels nog nie reg

geskied aan dié prosesse nie (vergelyk Halasz (1988, pp. 848-850) vir 'n oorsig oor die mate waarin bestaande stelsels alreeds voorsiening maak vir koöperatiewe werk).

- e) Die geslotenheid van bestaande hiperteksstelsels (vergelyk 4.7.3.3) bring mee dat (potensiële) gebruikers soms huiwer om bestaande materiaal in hiperteksstelsels te verpak. Indien daar mettertyd beter stelsels ontwikkel, kan dit meebring dat die materiaal dan weer verpak moet word, wat 'n reuse-taak kan wees. Dit is belangrik dat daar stelsels of databasisse (of aanpassings ten opsigte van bestaande databasisse) tot stand sal kom waarin die veelvlakkige hiperteksmateriaal (met gemerkte skakelareas en skakels) as't ware stelsel-onafhanklik ontwikkel kan word, as hiperteksdokumente geoperasionaliseer kan word (al is dit ook voorlopig en eenvoudig) en vanwaar dit mettertyd weer met betreklik min moeite na ander stelsels oorgedra kan word, indien nodig. Dit is vanselfsprekend dat daar by so 'n ontwikkeling ook rekening gehou moet word met die tekskoderingstandaard SGML (vergelyk 5.4).
- f) Daar is 'n gebrek aan riglyne oor presies hoe teks op 'n veelvlakkige wyse gestruktureer moet word. Indien die skakeling deur woorde (as skakelareas) plaasvind, moet daar gesorg word dat die regte woorde en woord-omgewing telkens gekies word om te verseker dat die leser dit as relevant beskou om na 'n volgende vlak van inligting te wil beweeg. Die streng hiërargiese onderskeid tussen die hoofteks en die annotasie (soos gebruiklik in tradisionele geleerde publikasies) gaan nie meer op nie. Landow (1987) dui aan dat materiaal wat vir nie-liniêre aanbieding in hiperteksstelsels ontwikkel word, spesiale diskoersteknieke vereis om voorsiening te maak vir die proses van "skakelspronge" (*link traversal*). Volgens hom is 'n *rhetoric of arrival* en 'n *rhetoric of departure* nodig, wat aan die leser inligting verskaf oor die verhouding tussen brokke inligting wat deur die skakel teweeggebring word (vergelyk ook Halasz, 1988, p. 850). Hiermee hang saam die probleem van die wyse waarop bestaande teks wat nie met die oog op 'n veelvlakkige inligtingstelsel geskep is nie, agterna na so 'n omgewing oorgedra kan word sonder om volledig oorgeskryf te moet word. Frisse (1988) gee aandag aan moontlikhede vir die oordrag van gewone teks (*flat text*) na die hiperteksomgewing.
- g) Daar word in verband met hiperteksbare aansprake gemaak met betrekking tot die wyse waarop dit inskakel by hele proses van menslike kognisie. Sommige van die aansprake oortuig wel, maar die proses van menslike kognisie en die wyse waarop dit deur die tussenkoms van die rekenaar beïnvloed word, vereis nog verdere aandag, onder andere ook van die kant van kognitiewe sielkundiges. (Vergelyk 3.5, 4.7.3.5 en 4.7.3.10h. Vergelyk ook Beeman et al., 1987; Baird & MacMorrow, 1988; Fiderio, 1988.)

Halasz (1988) gee aandag aan die hele spektrum van probleme wat die aandag verdien in die ontwikkeling van hiperteksprogramme en -toepassings.

Daar is nog baie ander fasette met betrekking tot hiperteksbare wat intensiewe ondersoek vereis; in baie opsigte kan gesê word dat dié tegnologie nog in sy kinderskoene staan. Dit

geld beide die gebruikersvlak en die datamodelle en -strukture onderliggend aan die stelsel. Die feit dat hiperteks soveel fasette van die mens en sy omgang met inligting aanraak, asook dat soveel eise aan die toerusting gestel word, bring mee dat daar nog baie navorsing en ontwikkeling gedoen moet word. Die basiese konsepte van hiperteks is egter baie belowend en sal nog lank die toneel van die rekenaaromatige verwerking van teks bepaal (vergelyk Fiderio, 1988, p. 244 en Halasz, 1988).

4.8 'N GEÏNTEGREERDE OMGEWING VIR DIE VERWERKING VAN TAAL

Die toenemende gebruik van die rekenaar met betrekking tot die bestudering van taal en teks, die groot verskeidenheid van toepassings wat in toenemende mate ontwikkel, en die probleem van die fragmentêre karakter van sommige van die programme het daartoe gelei dat 'n navorsingspan van die Summer Institute of Linguistics (Dallas, Texas) op 'n baie fundamentele wyse begin aandag gee het aan die prosesse wat onderliggend aan al dié toepassings is (Simons et al., 1988). Hulle het aandag gegee aan die redes vir die gebruik van die rekenaar vir navorsing oor taal, teks en kultuur, asook aan die inligtingkundige prosesse wat telkens plaasvind en die behoeftes van die mense wat tipies met sodanige prosesse werk.

Dit moet in gedagte gehou word dat die Summer Institute of Linguistics (SIL) die "akademiese"- en navorsingsarm is van die Wycliffe Bible Translators en op 'n deurlopende basis opleiding en ondersteuning te gee aan 'n groot aantal Bybelvertalers wat in bykans alle dele van die wêreld besig is met verskeie aspekte verbonde aan Bybelvertaling. Dit behels onder andere die optekening en beskrywing van tale waarvoor daar geen ontwikkelde skrifstelsel is nie, die ontwikkeling van woordeboeke en proefvertalings, asook die prosesse van die redaksionele en uiteindelik tipografiese voorbereiding van vertalings. Die program IT is byvoorbeeld juis ontwikkel om hierdie prosesse van die beskrywing van tale met die oog op die opbou van 'n grammatikale beskrywing te ondersteun (vergelyk 4.5).

Dit het vir die span duidelik geword dat die grootste probleem die feit is dat 'n groot verskeidenheid programme gebruik word wat uiteindelik nie by mekaar aanpasbaar is nie. Die materiaal word telkens omgeskakel van die een formaat na die ander, met die daarmee gepaard gaande verlies aan doeltreffendheid en soepelheid. Die analise van die verskillende prosesse het aan die lig gebring dat dit wel moontlik is om die verskillende soorte inligting te verenig. Al die aksies ten opsigte van taal wat hierbo beskryf is (4.2 tot 4.7) kan inderdaad in een omvattende stelsel verenig word, waarin die verskillende komponente of prosesse met mekaar geïntegreerd is. Hulle stel 'n stelsel voor wat genoem word CELLAR (Computing Environment for Linguistic, Literary, and Anthropological Research) (Simons et al., 1988). In dié taalverwerkingsomgewing sal ook intensief gebruik gemaak word van konsepte van die gebied van kunsmatige intelligensie en hiperteks. Die tekstkoderingstandaard SGML sal aan die grondslag van alle formate van data lê, wat die

grootste moontlike mate van beweeglikheid in die aanpasbaarheid met ander stelsels, sowel as die grootste moontlike mate van oordraagbaarheid sal verseker.

Daar word tans gewerk aan die ontwikkeling van die CELLAR-stelsel. Bestaande programme wat bruikbaar is, word so ver soos moontlik in die omvattende stelsel geïntegreer.

Hierdie konsep van 'n geïntegreerde omgewing vir die verwerking van taal is die mees ondernemende konsep wat nog op die gebied van die rekenaarverwerking van taal voorgestel of aangepak is. Dit hou baie belowende moontlikhede in vir navorsing op 'n verskeidenheid gebiede.

HOOFTUK 5: DIE LOGISTIEK VAN DIE VERWERKING VAN NATUURLIKE TAAL

5.1 INLEIDING

Dit het geblyk dat daar 'n aantal fundamentele verskuiwings plaasgevind het met betrekking tot die totale omgewing waarin taal/teks verwerk word. Enkele van dié verskuiwings is as tendense beskryf in Hoofstuk 2. Een van die belangrikste verskuiwings is die feit dat die verwerking wegbeweeg het van die hoofraamomgewing en toenemend in 'n gedesentraliseerde omgewing plaasvind. Hiermee saam het egter ook 'n ander belangrike verandering plaasgevind: die ontwikkeling en implementering van programme het in die hande beland van persone wat nie so vertrouwd is met stelselanalise as dié wat voorheen in die hoofraamomgewing daarmee gewerk het nie. Dit is natuurlik 'n vraag of dié persone in elk geval die nuwe dinamiek van stelselanalise (vergelyk 5.2 hieronder) sou kon hanteer.

Hierdie en ander redes maak dit noodsaaklik om ook aandag te gee aan die logistiek van die verwerking van natuurlike taal en om die vraag te stel of die veranderinge in die omgewing van rekenaarverwerking nie van so 'n aard is dat stelselanalise vir die verwerking van natuurlike taal nou 'n nuwe inhoud gekry het nie.

5.2 STELSELANALISE

Stelselanalise was nog altyd 'n integrale deel van enige programmering, indien nie in die praktyk nie, dan ten minste in die teorie. In stelselanalise gaan dit gewoonlik oor die analise van die probleem of beoogde toepassing, die opbreek daarvan in verskillende stappe of prosedures, die ondersoek na die mees geskikte programfunksies (en programme) om die aksie te kan uitvoer en die uiteindelijke aaneenskakeling van die prosedures of programme om 'n werkende stelsel te vorm. Die opset van die gebruiker van die stelsel is uiteraard ook op die spel en daarvoor word dan gepas voorsiening gemaak.

Die feit dat dit taal/teks is wat verwerk word, bring mee dat daar vir die eiendomlike van taal voorsiening gemaak moet word – wat heelwat anders kan wees as in die geval van die verwerking van numeriese gegewens of eenvoudige administratiewe data. Die wyse waarop die verwerking van teks moet plaasvind, is nie vanselfsprekend nie en daar kan nie sonder meer aangesluit word by die wyse waarop stelselanalise in die geval van ander toepassings plaasvind nie. In verskeie publikasies oor die verwerking van taal word uitvoerig aan dié aspekte aandag gegee (byvoorbeeld Oakman, 1980). Tog kom dit voor asof daar ook nog nie werklik erns gemaak word met sekere aspekte van taal nie, byvoorbeeld die aspek van betekenis, die verband tussen woord en begrip, die hantering van woordgroepe, die belang van die konteks (vergelyk 4.2 en 4.4). Daar is dus nog steeds 'n behoefte aan riglyne vir stelselanalise ten opsigte van die eiendomlike van taal.

Indien dit nou ook nog gaan oor teks, wat meer is as net losstaande taalgegewens of taalgegewens van beperkte omvang (byvoorbeeld losstaande sinne of uitings) en wat op 'n eiesoortige wyse in die samelewing funksioneer, moet daar ook rekening gehou word met hierdie besondere rol van teks, die posisie van teks in sy verskillende uitingsvorme in die samelewing en die hele spelruimte van rekenaarverwerking van teks (funksies en programme).

Hieruit volg dat stelselanalise in verband met die verwerking van natuurlike taal nou 'n ander karakter kry as net dié van die moontlikhede van program-aksies ten opsigte van bepaalde data, sy dit ook talige data. Dit kry enersyds 'n omvattende en andersyds 'n eksplisiet interdissiplinêre karakter. Uiteraard sou dieselfde ook gesê kon word van ander soorte data en die rol daarvan in die samelewing (byvoorbeeld gegewens oor strategiese minerale, oor swak debiteure, ensovoorts). Tog is dit so dat teks op so baie wyses en in so baie vorms geoperasionaliseer kan word en so nou verknoop is met die menslike vermoë om inligting te kan verwerk, dat dit hier 'n besondere posisie inneem.

Stelselanalise met betrekking tot tekstmateriaal kan net sinvol plaasvind indien daar ook rekening gehou word met die totale omgewing waarbinne teks hanteer word en die moontlike aksies wat in die verwerking daarvan uitgevoer kan word en dus ook moontlik op die een of ander stadium uitgevoer gaan word. Dit raak dus belangriker om 'n toepassing te kan plaas op 'n punt tussen die hele spektrum van aksies wat in verband met die verwerking van teks ter sprake kom en terselfdertyd rekening te hou met die implikasies indien die teks ook mettertyd in 'n ander vorm verwerk moet word.

Anders as in die geval van die meeste administratiewe toepassings, is die vlak van programmering wat agter baie van die sinvolle en bruikbare programme vir taalverwerking lê (en die vlak van stelselanalise wat telkens daarmee saamhang) van so 'n aard dat dit nie maklik geëwenaar kan word deur beginners of selfs deur programmeerders wat vir 'n spesifieke toepassing programme moet ontwikkel en nie voltyds by dié soort toepassings betrokke is nie. Hierby kan gevoeg word dat die fondse vir programmering beskikbaar vir persone betrokke by die verwerking van natuurlike taal meesal nie so ruim is as in baie besigheids- of tegniese rigtings nie. Dié aspek is gewoonweg 'n werklikheid wat in ag geneem moet word by stelselanalise vir die verwerking van natuurlike taal. Dit kan dus byvoorbeeld inhou dat daar meer intensief as in ander rigtings gebruik gemaak moet word van bestaande programme of programfasiliteite en dat daar eerder gesoek moet word na dié mees sinvolle integrasie van sulke programme in 'n omvattende proses of stelsel. So 'n benadering bring egter noodwendig weer die kwessie van aanpasbaarheid van programme en data ter sprake.

Ook die keuse van die rekenaaromgewing waarin natuurlike taal verwerk word, is van die grootste belang (vergelyk hoofstuk 6). Dit is naamlik moontlik dat 'n program vir gebruik op een rekenaar (byvoorbeeld 'n mikrorekenaar met drukker wat ideaal kan wees vir die intensiewe eksperimentering met formattering en tipografie) ontwikkel word, maar dat die hoeveelheid materiaal wat uiteindelik geprosesseer moet word te veel is om op so 'n rekenaar te hanteer en daar dan 'n oordrag na 'n ander rekenaar moet plaasvind. So 'n stap

kan ingrypende implikasies inhou. Indien die totale omgewing van verwerking egter reg gekies is, kan dit maklik wees om die programme oor te dra en die aanpassings te maak vir die databasisse op 'n groter stelsel. Vergelyk Shaw (1988) oor die belang van SQL as standaard met betrekking tot die oordraagbaarheid van databasisse. Ook dié aspek hang baie nou saam met stelselanalise.

Sedert daar alternatiewe ten opsigte van die tradisionele boek ontwikkel het, wat vir sommige toepassings baie meer doeltreffend as die boek blyk te wees, het die vraag na die mees geskikte wyse vir die operasionalisering van teks (wat heelwat meer behels as net die publikasie van teks) 'n nuwe dimensie gekry. Die teks self vorm nou die kern; die verpakking is opsioneel, afhange van die behoefte. Ook in dié opsig moet dus 'n aantal beslissings geneem word oor die mees geskikte wyse/vorm van die verpakking van tekstmateriaal.

Die volgende aspekte wat by stelselanalise met die oog op die verwerking van tekstmateriaal op die spel moet kom, kan dus byvoorbeeld onderskei word:

- a) Die eie aard van talige materiaal: Dit is nodig om rekening te hou met alle aspekte van taal, insluitende betekenis. Besondere aandag moet gegee word aan die eie aard van teks.
- b) Die onmiddellike en uiteindelijke doel met of bestemming van die tekstmateriaal: Die ervaring het geleer dat dit dikwels nodig is om tekstmateriaal wat in een vorm of vir een toepassing verpak is, ook mettertyd vir ander toepassings of in ander vorms te verpak. Hiermee kom 'n sosiologiese dimensie van teks in die samelewing ter sprake. Dit kan nou nodig wees om rekening te hou met aspekte soos:
 - i) Die vlakke van kennis, mediageletterdheid, tegnologiese bewussyn of kundigheid, en die inligtingsoekende praktyke van die gebruiker of waarskynlike gebruiker, asook die beskikbaarheid van toerusting.
 - ii) Die vraag of die materiaal liniêr of veelvlakkig verpak moet word.
 - iii) Die vraag of 'n herverpakking van die materiaal vir ander doeleindes moontlik of sinvol is.
- c) Die menslike ervaring en verwerking van tekstmateriaal: Dit kan nodig wees om rekening te hou met aspekte soos:
 - i) Die implikasies van die oormaat van tekstmateriaal in die inligtingsera, die vermoë van die mens om die oormaat te kan hanteer en sy tipiese reaksie ten opsigte van die oormaat.
 - ii) Die kognitiewe aspekte ter sprake (alle soorte verpakking van inligting is nie ewe geskik vir alle groepe nie).

- d) Die sosiologiese dimensies van die groep of samelewing waarin die teksmateriaal moet funksioneer, byvoorbeeld
- i) Die mate van eenvormigheid al dan nie van die groep of groepe waarin die materiaal gebruik gaan word.
 - ii) Die vraag of 'n teks op so 'n wyse verwerk of verpak moet word dat dit (mettertyd) in verskeie vorms geoperasionaliseer kan word, byvoorbeeld in 'n meer populêre weergawe of in terme van die behoeftes van verskeie groepe gebruikers.

Die onsamehangende aard van die uiteensetting van aspekte wat hierbo gebied is, asook die veelvuldige oorvleuelings enersyds en skynbare leemtes andersyds, is 'n duidelike aanduiding dat dié aspekte nie op een vlak lê of net hiërargies gelys kan word nie. Dit gaan veeleer om 'n multidimensionele verhouding van relevante aspekte, wat moontlik in die vorm van die drie dimensies of vlakke van 'n blok voorgestel kan word, waarby die drie vlakke van wydte, hoogte en diepte (elkeen met sy eie vier kante) gedui kan word as:

- die tegnologiese (met as kante:- apparatuur; programmatuur [bestaande en te ontwikkel]; vlak van kundigheid en middele beskikbaar vir die ontwikkeling van programmatuur; standaarde)
- die soort aksie wat met die taalgegewens uitgevoer word (met as kante:- teks-generering; teksordening; toegang tot teksmateriaal; teks-aanbieding – en by elkeen telkens alle aspekte van taal wat daarby ter sprake is)
- die menslike aspek (met as kante:- die inligtingsbehoefte van die individu of groep; die vlak van mediageletterdheid [kundigheid, ensovoorts]; die sosiologiese aspek; die kognitiewe aspek).

~~Die benadering by stelselanalise moet dan wees om 'n stelsel of toepassing te kan plaas op 'n punt tussen al dié vlakke en kante (of moontlik op 'n lyn wat die huidige stand en die groei of toekomstige ontwikkeling verteenwoordig). By komplekse stelsels kan dit uiteraard moontlik wees om verskeie punte (of lyne) tussen dié vlakke en kante te bepaal.~~

Bogenoemde uiteensetting moet slegs as 'n voorlopige oorsig oor en model met betrekking tot die spektrum van fasette betrokke by die verwerking van natuurlike taal beskou word. Daar is so 'n groot verskeidenheid aspekte op die spel dat die volle implikasies daarvan slegs oor 'n langer tydperk in groepverband deurdink en uitgewerk kan word. Omdat teks en die funksionering daarvan so nou verknoop is aan die samelewingsverbande waarin dit funksioneer, is dit waarskynlik so dat daar nie 'n finale lys van aspekte opgestel kan word nie en dat daar dus voortdurend dinamies aangepas sal moet word by ontwikkelings.

5.3 NETWERKVORMING

Die verskuiwing vanaf hoofraamrekenaars na mikrorekenaars kan as 'n vorm van desentralisasie beskou word. Die probleem wat egter hiermee ontstaan het, is dat toepassings/gebruikers in 'n toenemende mate ongekoppeld geraak het: die gebruiker werk op sy eie rekenaar, verwerk 'n sekere hoeveelheid taal-/teksmateriaal en stel die resultate (meestal in gedrukte vorm) aan ander beskikbaar.

Teks funksioneer egter nie net op dié wyse nie. Vir die navorser wat 'n ondersoek met betrekking tot 'n bepaalde teks of tekskorpus wil onderneem en daaruit bepaalde bevindinge wil maak, is dit uiteraard die normale patroon dat die teks uitgehaal word uit die omgewing waarin dit dinamies funksioneer en as't ware in laboratoriumtoestande hanteer word. Baie van die aksies ten opsigte van teks wat in dié verslag bespreek is, het dan ook in 'n groot mate met hierdie soort omgang met teks te make. Dit het egter ook geblyk dat sekere aksies met betrekking tot die rekenaarmatige verwerking van teks juis gaan oor die dinamiese funksionering van teks in die samelewing. Ook ten opsigte van dié aksies is dit belangrik dat die verwerking op so 'n wyse moet geskied dat dit aan taal/teks op daardie gebied laat reg geskied. In sulke gevalle funksioneer teks nie ongekoppeld nie, maar juis in die dinamiese interaksie tussen die lede van die kleiner of groter taalgemeenskap. Trouens, as simboliese representering van kennis en tipiese wyse van menslike kommunikasie speel teks hier 'n strategiese rol.

In verskeie omgewings waarin met teks gewerk word, word die teks deur verskeie persone hanteer en het verskeie van hulle die geleentheid om die dokument te lees, verbetering of veranderinge voor te stel, die veranderinge van ander te aanvaar of te verwerp, ensovoorts. Dit geld van die omgewing van kantoororganisasie, maar word veral dikwels gevind in omgewings waar daar skeppende werk met teks plaasvind, byvoorbeeld in die akademiese omgewing, in die sakewêreld, in ontwerp- en beplanningsafdelings, ensovoorts. Halasz (1988, p. 848) sê hieroor:

Most idea processing and information management tasks are inherently collaborative, with groups of two to ten people working on a common project.

Terminologie wat dikwels voorkom, is: *work group (workgroup) computing* en *collaborative environment*. In die natuurwetenskaplike en tegniese omgewings vind daar gewoonlik baie meer koöperatiewe werk plaas as in die geesteswetenskappe (met sommige van die sosiale wetenskappe daar tussen-in). Die toenemende spesialisasie aan die een kant en die opsigtelike noodsaaklikheid aan die ander kant om toenemend interdisiplinêre samewerking te ontwikkel, kan meebring dat persone in die geesteswetenskappe ook al hoe meer van dié soort koöperatiewe omgewing gebruik gaan maak. In baie gevalle waar projekte van groter omvang aangepak word, raak dié benadering al hoe meer gebruiklik. Die koöperatiewe aspek kom nie net ter sprake in die stadium wanneer die teks alreeds voorberei is nie, maar ook juis in die stadium wanneer dit nog net gaan oor idees wat rondgegooi en algaande uitgebou word, totdat daar mettertyd 'n struktuur daaruit te

voorskyn kom. Halasz (1988, p. 849) gee 'n baie goeie oorsig oor die aard van aktiwiteite in so 'n koöperatiewe omgewing. Afgesien van prosedurale aktiwiteite noem hy ook annotasie-aktiwiteite, wat juis in die hiperteks-omgewing baie goed hanteer kan word. Hierdie soort koöperatiewe werk vind toenemend in elektroniese vorm plaas. Daar word al lank reeds gebruik gemaak van elektroniese posstelsels, maar dié het meesal hoogstens 'n boodskap-karakter en het geblyk te beperkend te wees vir die dinamiese ondersteuning van koöperatiewe werk (Halasz, 1988, pp. 848-849).

Verskeie programme het intussen voorsiening begin maak vir 'n gesofistikeerde elektroniese omgewing waarbinne koöperatiewe werk kan plaasvind. Vergelyk Begeman & Conklin (1988) vir 'n uiteensetting van die belang van die koöperatiewe omgewing in die proses van stelselontwerp en vir 'n bespreking van die program gIBIS (Graphical IBIS) wat gebaseer is op die IBIS-metode (Issue-Based Information Systems). Hyperdoc maak vir netwerke voorsiening en OWL (1988) is besig met die afronding aan 'n netwerkgebaseerde hiperteksstelsel IDEX (wat verwantskap toon met Guide). Die program For Comment stel 'n groep van tot 16 persone in staat om saam te werk in die generering, beoordeling en hersiening van dokumente (Marcus, 1988, pp. 13-14). Dit maak voorsiening vir tot vyf vlakke kommentaar, asook kommentaar op kommentaar. Die funksionaliteit van dié programme verskil grootliks van dié van gewone netwerkprogramme wat dit maar net moontlik maak dat persone in 'n organisasie met dieselfde dokumente kan werk (byvoorbeeld die ATEX-stelsel vir elektroniese redigering in 'n uitgewersbedryf). (Vergelyk ook Meyrowitz, 1985.)

Die hiperteks-omgewing blyk in dié verband ideale moontlikhede te bied. Alle hiperteksprogramme maak nog nie voorsiening daarvoor nie en dit word juis as een van die groot desiderata ten opsigte van die ontwikkeling van programmatuur beskou. Die verwagting is dat daar toenemend vir hierdie soort koöperatiewe werk voorsiening gemaak sal word.

Afgesien van hierdie vlak van geïntegreerde, koöperatiewe werk is daar ook nog 'n ander aspek ten opsigte waarvan daar deesdae vir netwerkvorming voorsiening gemaak moet word, naamlik dat alle programme in staat moet wees om in netwerkverband te funksioneer. Netwerkvorming het gedurende die afgelope paar jaar geblyk 'n sleutelwoord van die rekenaarbedryf te wees. Die deling van duur, gesofistikeerde randapparatuur (byvoorbeeld drukkers, stoorruimte [harde skyf of optiese skyf]) en programmatuur maak dit noodsaaklik dat programme ook in netwerkverband moet kan funksioneer. Dit verg dikwels heelwat aanpassing aan programme om te verseker dat programme wel foutvry in 'n netwerk kan funksioneer.

5.4 TEKSKODERINGSTANDAARDE

5.4.1 Inleiding en agtergrond

Dit het hierbo geblyk dat teks in 'n groot verskeidenheid toepassings verwerk en in 'n groot verskeidenheid formate verpak kan word. Die rekenaar leen hom ideaal tot die verwerking van 'n groot volume teksmateriaal. Dit is egter belangrik dat die verskillende

elemente in die teks gemerk sal word op so 'n wyse dat die teks deur verskillende persone op verskillende plekke op dieselfde wyse hanteer of geïnterpreteer sal kan word. Die teksmerkers wat te make het met punktuasie en tipografie (lettertipes, paragraaf-indelings en inkeping) is algemeen bekend en word as vanselfsprekend verwerk deur die leser wat die stelsel alreeds sy eie gemaak het. Die rekenaar kan met sekere van hierdie teksmerkers gemaklik omgaan, aangesien hulle eenduidig is, byvoorbeeld

- woorde word voorafgegaan deur en gevolg deur 'n spasie (of ook gevolg deur 'n bepaalde punktuasieteken)
- sinne word deur 'n punt (of vraagteken of uitroepeteken) afgesluit
- 'n bepaalde kode in die teks dui die begin en die einde van 'n lettertipe aan
- 'n kode aan die einde van 'n aantal karakters (sinne) dui 'n paragraaf aan:

Wat egter nie ewe gemaklik deur die rekenaar hanteer kan word nie, is onderskeidings soos die volgende:

- die verhouding tussen paragrawe onderling (byvoorbeeld die feit dat 'n bepaalde paragraaf of 'n aantal paragrawe ondergeskik is aan 'n ander, byvoorbeeld as ingekeepte paragrawe)
- dat 'n bepaalde paragraaf as opskrif ten opsigte van 'n hele gedeelte funksioneer
- die verhouding tussen afdelings, hoofstukke, seksies en paragrawe (byvoorbeeld)
- die wisseling tussen sprekers.

Op die eenvoudige en uiterlike vlak van woordverwerking is teksmerkers van hierdie aard nie nodig nie; enkele van die onderskeidings kan hoogstens grafies aangedui word (byvoorbeeld deur kursivering). Tensy hierdie onderskeidings gemerk word, sal materiaal by ander soorte verwerking van die teks net as 'n lang string karakters beskou word.

Dit is ongelukkig die geval dat die merk of kodering van tekste vir elke toepassing op 'n ander wyse geskied. Daar is aanvanklik nie veel nagedink oor die stelsel van kodering wat gebruik word nie en besluite is geneem op grond van die behoeftes van 'n bepaalde toepassing, byvoorbeeld woordverwerking, databasisse, konkordansies, setwerk, hiperteks. Op hierdie wyse het daar 'n groot aantal tekstkoderingstandaarde ontwikkel, elkeen van toepassing op 'n enkele of op 'n paar toepassings. Wanneer dieselfde teks nou in verskillende toepassings verwerk moet word (byvoorbeeld selfs net wanneer dit vir setwerk voorberei word en dan weer in 'n databasis vir skermhantering gestoor word), moet die kodering telkens aangepas word by die nuwe stelsel of behoeftes.

Dit is om hierdie rede dat daar oor die afgelope paar jaar heelwat aandag gegee is aan die formulering van standaarde wat die moontlikheid bied van die eenvormige beskrywing

van teks op so 'n wyse dat, as dit eenmaal gedefinieer/beskryf is, die teks met betreklik min moeite (of selfs met volledig meganiese prosesse) in 'n ander vorm verpak kan word. Hierdie standaarde of formate staan bekend as tekstkoderingstandaarde (*text markup languages*).

Tekstkoderingstandaarde moet duidelik onderskei word van twee ander standaarde of soorte kodering van aspekte wat ook teks, die verwerking en die vormgewing daarvan raak:

- Grammatikale kodering (of *tagging*), wat gewoonlik die kleiner elemente van die grammatika tot op sinsvlak (maar soms tot op die vlak van styl) op die oog het (vergelyk Van der Merwe, 1988, pp. 242-3).
- Bladsybeskrywingstandaarde (*page description languages*), wat op die oog het om 'n eenvormige standaard te definieer vir die afvoer van die rekenaar na 'n drukker, en wat dus 'n eenvormige raakvlak tussen 'n verskeidenheid programme en die drukker daarstel (PostScript is 'n goeie voorbeeld).

Die omvang van hierdie probleem word duidelik indien dit in ag geneem word dat dit kan wees dat dieselfde teks (wat moontlik besonder omvangryk is en verskeie megagreep data behels) in verskillende omstandighede en volgens verskillende behoeftes van gebruikers op, onder andere, die volgende wyses verpak of geoperasionaliseer moet kan word:

- in 'n woordverwerkingspakket (tydens die skepping of vir die doel van redigering)
- in 'n pakket vir kantoorpublikasie (byvoorbeeld Ventura) (met die oog op die voorbereiding van meesters vir drukwerk en waarin die teks op gevorderde wyse geformateer is); van hier af kan die teks ook weer na 'n setmasjien gestuur word waar dit in presies dieselfde uitleg as op 'n laserdrukker sal uitkom (indien dit in die DTP-pakket in PostScript voorberei was).
- in 'n stelsel vir kantooroutomatisasie (waarin dit nie net gaan oor die berging en meganiese herwinning van materiaal nie, maar ook oor assosiatiewe beweeg tussen verwante materiaal)
- in 'n volteks-databasis (met die oog op inligtingsherwinning op die basis van trefwoorde of 'n kombinasie van trefwoorde)
- in 'n hiperteksstelsel (waarin die materiaal op so 'n wyse gestruktureer is dat die inligtingsoeker op intelligente wyse en betekenismatig inligting kan herwin en van meer algemene vlakke na verdere en meer gespesialiseerde vlakke van inligting kan beweeg, of hom kan laat lei deur assosiatiewe denkproesse)
- in 'n onderrigstelsel (waar die materiaal, soos in die vorige punt, in hiperteksformaat gestruktureer is en die kennisverwerwingsproses as't ware deur die rekenaarproesse gesimuleer kan word).

Om die inligting telkens per hand in 'n ander wyse te gaan merk vanaf rou, ongeformateerde teks (of vanaf ASCII-teks), is net nie haalbaar nie. In dié opsig is teks eintlik 'n baie meer problematiese item as byvoorbeeld numeriese gegewens en dit verg baie gevorderde werk om dit telkens in 'n ander vorm of medium te verpak. Die rasionaal agter die bestaan van en noodsaaklikheid van die implementering van tekskoderingstandaarde is dus soos volg:

- Teks is vandag in al hoe groter volumes en in 'n al hoe groter verskeidenheid vorms beskikbaar.
- Daar is al hoe meer uitingsvorme van teks, elkeen met sy eie moontlikhede en toepassingsgebiede, waarin teks verpak moet (en kan) word.
- Om teks in verskillende formate te kan verpak, vereis baie tyd, energie en koste (en soms ook hoëvlakkundigheid).
- Die teks moet dus van die begin af op so 'n wyse gemerk word dat dit vandaar maklik herverpak kan word in 'n verskeidenheid formate, soos wat dit vir elke gebruik en omstandigheid benodig mag word.
- Daarvoor is 'n eenvormige standaard nodig in terme waarvan die verskillende komponente van teks beskryf word.

Die konsep van tekskoderingstandaarde het ongetwyfeld sy agtergrond in die setbedryf en die proses van merk of kodering (*markup*) wat daar plaasvind (Fraser, 1986, pp. 1-4, 26-27). Pogings om hierdie faset op die vlak van woordverwerking te betrek (soos die stylstelsel van WORD, vergelyk 4.2.1.1), het die konsep in die een rigting verder laat ontwikkel, maar uiteraard nog binne die raamwerk van 'n enkele program en vir die oordrag tussen programme is nie voorsiening gemaak nie.

Hierdie aspek van die "beweeglikheid" met teks is een van die belangrikste aspekte van die wyse waarop teks in die inligtingsera hanteer gaan word. Tekskoderingstandaarde gaan die kern vorm van enige sinvolle hantering van teksmateriaal in die neëntigerjare. In verskeie tydskrifte op die gebied van rekenaartoe toepassings op taal word aandag gegee aan die noodsaaklikheid van 'n *single, rational set of guidelines for encoding machine-readable texts* (Sperberg-McQueen, 1987, pp. 1-2).

5.4.2 Verskillende tekskoderingstandaarde

Verskeie tekskoderingstandaarde het (soms onafhanklik van mekaar) deur die jare tot stand gekom. Die wyer gebruik van rekenaars, die groei van gedesentraliseerde verwerking van teks (vergelyk 2.3a), die groot toename in die volume teksmateriaal wat deur die rekenaar geprosesseer word, en veral die feit dat die skepper/redakteur van die teks en die drukkersbedryf al hoe nader aan mekaar beweeg, het daartoe gelei dat die kwessie van die uitruilbaarheid van tekste vir verskeie toepassings en 'n eenvormige standaard al hoe meer die aandag geniet het.

5.4.2.1 Eerste pogings

a) *Die Oxford Concordance Package en die COCOA-formaat*

Die Oxford Concordance Package (OCP) (vergelyk 4.4.3.3 (a) en (b) hierbo) vereis slegs 'n minimale kodering of merk van die teks, en wel slegs om te verseker dat die sortering op die korrekte wyse sal geskied (Hockey & Marriott, 1984). Die mate van kodering wat verder nog benodig word, is afhanklik van die behoeftes en verwagtings van die gebruiker. OCP aanvaar kodering in drie formate, waarvan COCOA die belangrikste is. Spesifieke merkers (*tags*) word egter nie voorgeskryf nie.

b) *Die Lancaster-Oslo/Bergen (LOB) Engelse korpus*

Die kodering wat hier gevolg word, is hoofsaaklik gerig op taalkundige navorsing en het nie wye ingang gevind vir ander doeleindes nie. (Vir besonderhede, vergelyk Barnard et al., 1988a.)

c) *Textcode*

Hierdie stelsel word gebruik aan die universiteite van Liverpool en Edinburgh. Alle gegewens wat relevant is vir grammatikale navorsing word aangedui. Textcode is egter nie baie geskik vir die kodering van groter eenhede soos sinne, paragrawe of hoofstukke nie, aangesien dit van die uitgangspunt uitgaan dat alle inligting op die vlak van die woord gekodeer moet word. Vir groter strukture moet die betrokke inligting oor die samehorigheid daarvan dus telkens by elke woord aangedui word, wat te omslagtig is (Barnard et al., 1988a, p. 27).

5.4.2.2 Moderne tekstkoderingstandaarde

Slegs die belangrikste standaard word hier bespreek, met name dié waaraan daar op internasionale vlak gewerk word en wat ook as internasionale standaard goedgekeur is of daarvoor voorberei word. Dit gaan uiteraard oor liggame of komitees wat amptelik op internasionale vlak verantwoordelik is vir die formulering van standaarde, byvoorbeeld ISO.

a) *Office Document Architecture (ODA)*

Verskeie instansies het aan hierdie standaard gewerk: die European Computer Manufacturers Association (ECMA), die International Organization for Standardization (ISO) en die International Telegraph and Telephone Consultive Committee (CCITT). Die uitgangspunt van die opstellers van hierdie standaard was om die mees gemaklike oordrag van dokumente binne die tipiese kantooromgewing moontlik te maak. Die standaard bestaan eintlik uit twee dele:

- i) Die ODA-deel beskryf 'n abstrakte dokument-argitektuur-model wat gebruik kan word vir die beskrywing van dokumente, asook die model van dokumentverwerking onderliggend daaraan.
- ii) Die ODIF-deel gee aandag aan die beginsels vir die formaat van oordrag.

Die dokument-argitektuur-model van ODA/ODIF kan inderdaad komplekse strukture doeltreffend beskryf. Die metode waarop hierdie beskrywing moet plaasvind, is egter betreklik ondoeltreffend en omslagtig (Fraser, 1986, pp. 18-22). Uit die tegniese oogpunt is dit baie goed, maar uit die oogpunt van die mens as gebruiker nie in dieselfde mate suksesvol nie.

Die ODA/ODIF-standaard is in die omgewing van modelle van kantooroutomatisasie baie gewild en kan beslis as een die van die twee mees gebruiklike modelle beskou word, saam met SGML-tipe formate ((c) hieronder).

b) *Office Information Architecture (OIA)*

Hierdie standaard is deur IBM ontwikkel en geniet nie wye steun buite IBM om nie.

c) *SGML en "generalized markup"*

Op die gebied van die verwerking van natuurlike taal het SGML (Standard Generalized Markup Language) as standaard tot nog toe die meeste belangstelling geniet en blyk dit die mees belowende te wees. SGML bied die mees omvattende en sinvolle vergestaltung van die konsep van *generalized markup* – 'n tekskoderingsformaat waarvolgens die kodering van die teks generies is eerder as op die prosesse gerig. (Oor SGML, vergelyk Barnard et al., 1988a, pp. 28-30; Fraser, 1986, pp. 26vv.; Barnard et al., 1988b; Smith, 1987a; 1987b; Smith & Stutely, 1988; Bryan, 1988. Die besonderhede van koderingskema van SGML word in besonderhede uitgewerk in die tesis van Fraser, 1986.)

By die operasionalisering van 'n dokument wat met 'n stelsel van "algemene kodering" (*generalized markup*) gekodeer is, verloop die verwerking soos volg:

- i) Herkenning: 'n attribuut van die dokument word herken (byvoorbeeld 'n element met 'n identifikator wat dié element as 'n gewone paragraaf merk); die ooreenkoms met die beginsel van die stylstelle van WORD (4.2.1.1a) is opvallend.
- ii) Afbeelding (*mapping*): die attribuut word afgeteken op, of geassosieer met, 'n bepaalde prosedure, wat uiteraard van toepassing tot toepassing kan verskil.
- iii) Uitvoering: die prosedure word uitgevoer.

Die onderskeiding van die generiese beskrywing van die elemente van 'n dokument, afsonderlik van die prosedure, bring mee dat die dokument gemaklik in verskeie toepassings geoperasionaliseer kan word, byvoorbeeld: in 'n druktoepassing kan 'n indentering geskep word, in 'n skermttoepassing kan 'n verandering van karaktertipe bewerkstellig word, terwyl in 'n inligtingstelsel dit toegelaat word (of nie) dat die woorde van die betrokke paragraaf as trefwoorde in 'n indeks opgeneem word. Vir elke toepassing is daar telkens net 'n SGML-interpreteerder nodig wat die brug slaan tussen die SGML-dokument en die toepassing.

Dit behoort duidelik te wees dat hierdie soort benadering in die kodering van teks uitstekende moontlikhede inhou vir die feit dat teks in 'n verskeidenheid van toepassings geoperasionaliseer kan word – wat juis ook weer tipies is van die ryke verskeidenheid aanwendings wat met teks en met die gebruiker se omgang met teks geassosieer kan word. Dit is eweseer van belang wanneer toepassings (of herpublikasie) plaasvind lank nadat die dokument vir die eerste keer gepubliseer of in SGML-standaard voorberei is.

Daar is dikwels verwarring oor die aard van SGML. SGML is nie self 'n koderingstelsel nie; dit wil dus nie 'n stel generiese identifikatore of kodes voorskryf nie. Dit wil veeleer 'n basis en 'n metodologie daarstel vir die skepping van koderingstelsels.

Die volgende kan verder oor die aard van SGML gesê word:

- Dit is rekenaar-onafhanklik (*machine-independent*).
- Dit dui die logiese struktuur van 'n teks aan (Smith, 1987a, p. 1).
- Dit word aanvaar dat die struktuur hiërargies is.

SGML behels die volgende (vergelyk Fraser, 1986, pp. 28-39 en Barnard et al., 1988, pp. 28-29):

- i) 'n Abstrakte sintaksis (in die rekenaarwetenskaplike sin) vir die beskrywing van die struktuur en attribute van 'n dokument. Dié aspek hoort by alle implementerings van SGML.
- ii) 'n Konkrete sintaksis wat die abstrakte reëls in verband bring met 'n ander stel elemente (om byvoorbeeld voorsiening te maak vir bepaalde stelsels of apparatuur).
- iii) Verklarings oor kodering (*markup declarations*), wat die gebruiker in staat stel om 'n spesifieke "woordeskaf" van generiese identifikatore en ander attribute te definieer wat in verskillende soorte dokumente gebruik kan word.
- iv) Entiteitverwysings (*entity references*), wat stelsel-onafhanklik is en wat verwys na dele wat buite die dokumentlêer lê (en wat in SGML "entiteite" genoem word). Hierdie funksie kan ook gebruik word soos die *glossary*-funksie van woordverwerkingsprogramme (byvoorbeeld WORD), dit wil sê om kodes (wat ook 'n afkorting kan wees) binne die teks te koppel aan inhoude daarbuite of in 'n ander lêer of om dié kodes in bepaalde toepassings onder programbeheer met die inhoude te vervang.
- v) Entiteitsinhoude (*arbitrary data content*): voorsiening vir die data of inhoude (in samehang met die entiteitsverwysings ((iv) hierbo), wat nie self deur die tekstkoderingstandaard gemerk word nie (byvoorbeeld die relasie tot sekere beeldmateriaal of wetenskaplike formules).
- vi) 'n Metode om verwerkingsopdragte te kan onderskei.

- vii) 'n Funksie om opsionele eienskappe te kan aktiveer (byvoorbeeld die vermoë om kodering te vereenvoudig of kodes te kan herdefinieer).

In ondersoek wat uitgegaan het van 'n algemene stel vereistes vir tekstkoderingstandaarde (op die basis van die behoeftes van gebruikers in die humaniora) en waarteen alle bestaande standaarde getoets is, het dit geblyk dat SGML aan al die vereistes voldoen (Fraser, 1986, pp. 42-46). Dit beteken nie dat daar nie ook nog probleme is met betrekking tot die besonderhede van die standaard en die aanvaarding daarvan nie. (Vergelyk Coombs et al., 1987 oor die stelsel van *descriptive markup* en die kritiek daarop van Makram et al., 1988.) Die probleem van Makram et al. is die feit dat dit tog nie so maklik is om dokumente tussen verskillende stelsels en tekstkoderingstelsels oor te dra indien daar nie 'n presiese ooreenstemming tussen die twee betrokke stelsels is nie. Dit kom egter voor asof dié probleme opgelos kan word binne die algemene raamwerk van SGML en dus eerder as die normale ontwikkeling van 'n standaard en sy toepassing beskou moet word. Barnard et al. (1988b) gee aandag aan twee van hierdie probleme: (i) Die behoefte aan veelvuldige perspektiewe (*views*) op 'n dokument kan opgelos word deur oorvloedig te kodeer en dan filters in werking te stel. (ii) Die probleem van die omslagtigheid van die inkodering van gegewens kan opgelos word deur die intelligente gebruik van makros.

SGML is intussen as ISO-standaard gepubliseer: ISO 8879. **Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML)**. Genève. 1986. Dit word tans wydverspreid gebruik in verskeie toepassings in die humaniora. By kongresse kom temas oor SGML toenemend op die agenda.

Ook by 'n aantal kommersiële toepassings word alreeds voorsiening gemaak vir SGML of word aangedui dat dié standaard binne die afsienbare toekoms geakkommodeer sal kan word. So byvoorbeeld die hiperteks-georiënteerde dokumentverwerkingstelsel (op netwerkbasis) IDEX, wat aangekondig is deur OWL International Inc., wat ook Guide ontwikkel (OWL, 1988). Ook die geïntegreerde rekenaaromgewing vir linguïstiese navorsing wat deur Simons et al. voorgestel word, sal op SGML gebaseer wees (1988, p. 29). (Vergelyk 4.8.) Daar kan verwag word dat hierdie tendens sal toeneem.

d) *FORMEX*

Parallel met die ontwikkeling van ander tekstkoderingstandaarde het die Office for Official Publications van die Europese Gemeenskap ook aandag gegee aan die ontwikkeling van 'n standaard om die groot hoeveelheid gegewens van die deelnemende state te kan stoor, verwerk en oordra. In hierdie ontwikkeling is daar so ver moontlik rekening gehou met internasionale standaarde, hoewel daar deels ook eie paaie geloop is. Dié werk het uitgeloop op 'n verdere stelsel vir tekstkodering – FORMEX (Formalized Exchange of Electronic Publications) (Guittet, 1985), wat op amptelike vlak binne die Europese Gemeenskap moontlik 'n belangrike faktor kan word. Opsteller van hierdie verslag het egter nog geen ander verwysings na die implementering van FORMEX gevind nie.

Die besondere aan FORMEX is dat dit aan die een kant hou by internasionale standaarde, maar tog ook probeer om twee verskillende benaderings tot die oordrag van tekste te

verenig, naamlik die *common communication format* (CCF), wat grotendeels op die oordrag van data gerig is (Guittet, 1985, p. 5), en SGML. Die bedoeling is om dit met FORMEX moontlik te maak om enigen van hierdie twee benaderings te kan volg. FORMEX is eweseer gerig op die logiese struktuur van teks, aangesien die fisiese struktuur op grond daarvan programmaties gestalte kan kry.

Dit is belangrik om daarop te let dat FORMEX nie breek met die SGML-standaard nie (Guittet, 1985, p. 26). Daar is egter wel 'n aantal versuimwaardes wat by verwerking van 'n FORMEX-dokument deur 'n SGML-ontleder eksplisiet gespesifiseer moet word. Die belangrikste aanpassings ten opsigte van SGML lê op die volgende vlakke: 'n meer uitgebreide karakterstel word gebruik, die opsionele eienskappe verskil en daar is verskille met betrekking tot die "konkrete sintaksis" (Guittet, 1985, p. 26).

5.4.3 Opsommende opmerkings

Tekskoderingstandaarde is volledig deel van die terrein van die verwerking van natuurlike taal en moet saam met programmatuur en apparatuur as deel van die toekomstige toneel van rekenaarverwerking in die humaniora (onder andere) beskou word. Dit funksioneer op die gebied tussen apparatuur, programmatuur en gebruiker en bied die moontlikheid van beweeglikheid en soepelheid. SGML het sigself geprofileer as die mees geskikte standaard op dié gebied.

HOOFSTUK 6: DIE KEUSE VAN APPARATUUR

6.1 ALGEMEEN

Gebruikers wat betrokke is by die reknaarmatige verwerking van natuurlike taal ervaar dikwels ernstige probleme en onsekerhede in verband met hulle toepassings of beoogde toepassings. Dié probleme kan talle aspekte van hulle toepassings raak; die meeste het egter te make met programmatuur. Indien die probleme met betrekking tot programmatuur ervaar word op die heel basiese vlak van die funksionering van die rekenaar, gaan dit baie dikwels oor probleme wat met die keuse van apparatuur, oftewel met die rekenaaromgewing, saamhang.

Die woord "omgewing" word gebruik as vertaling vir (*computer*) *environment*. Dié term word meesal gebruik as aanduiding van die gemeenskaplikheid op die vlak van bedryfstelsel (*operating system*) tussen 'n aantal rekenartipes of -fabrikate. Byvoorbeeld: UNIX word as so 'n omgewing beskryf, aangesien dit 'n bedryfstelsel is wat 'n vlak van gemeenskaplikheid tussen 'n aantal fabrikate skep, ook van rekenaars wat in terme van werkverrigting en prys in totaal verskillende klasse val. Die belangrikste is dat UNIX-programmatuur op al sulke rekenaars kan funksioneer. Insgelyks kan van MS-DOS as so 'n rekenaaromgewing gepraat word. Dit sou meer korrek gewees het om te praat van 'n programmatuur-omgewing, maar die benamings word oor die algemeen baie onwetenskaplik gebruik. (Dikwels word die woord "omgewing" in die omgangstaal selfs gebruik van apparatuur: daar word byvoorbeeld van die "VAX-omgewing" gepraat, waarby dit in werklikheid gaan oor 'n bepaalde tipe apparatuur waarop verskeie bedryfstelsels gebruik kan word, onder andere UNIX.)

Die probleme ten opsigte van die rekenaaromgewing het gewoonlik te make met die volgende:

- a) Die vermoë van die rekenaar om die betrokke toepassing te kan hanteer (verwerkingspoed, geheue en skyfkapasiteit).
- b) Die oordraagbaarheid van programmatuur en toepassings tussen verskeie rekenaars en rekenaaromgewings.
- c) Die opdatering van programmatuur en die probleme om dan nog altyd seker te wees dat programmatuur doeltreffend sal kan funksioneer saam met ander programmatuur.

Die onsekerhede raak gewoonlik ook bogenoemde probleme, maar verder ook die volgende:

- i) Die vraag of die regte toerusting gebruik word (byvoorbeeld of 'n toepassing nie eintlik op 'n hoofraamrekenaar uitgevoer moet word nie).

- ii) Die vraag na die ontwikkelingspad van die bepaalde rekenaartipe waarmee of omgewing waarin gewerk word.
- iii) Die kwessie van standarde (met betrekking tot apparatuur, algemene programmatuur en programmatuur met die oog op rekenaartoepassings op taalmateriaal).
- iv) Die doen van 'n keuse tussen die advies van welmenende handelaars van rekenaartoerusting aan die een kant en persone betrokke by toepassings op taal aan die ander kant.

Die meeste van dié onsekerhede en probleme van gebruikers hang saam met die feit dat baie van hulle slegs met programmatuur werk wat algemeen beskikbaar is en dat hulle dus afhanklik is van markontwikkelings en van die verskaffers van programmatuur.

6.2 DIE KEUSE VAN 'N APPARATUUR-OMGEWING

6.2.1 Betrokkenheid

As heel fundamentele oorweging by die keuse van apparatuur moet dit duidelik gestel word dat die belangrikste sekerlik is dat die persoon wat daarin belangstel om betrokke te raak by die rekenaarmatige verwerking van natuurlike taal wel betrokke sal raak – al beteken dit ook dat hy aanvanklik nie van die mees ideale toerusting gebruik kan maak nie.

Wie dus nie 'n groot keuse het nie en moet besluit om gebruik te maak van dit wat beskikbaar is of hoegenaamd nie op dié gebied betrokke te raak nie, moet eerder kies om wel betrokke te raak. Die ervaring wat opgedoen word, is soms baie werd en kan verseker dat dié persoon in die toekoms sinvol by toepassings op ander toerusting (of in ander omgewings) betrokke kan raak. Dit is altyd weer moontlik om data van die een rekenaar-omgewing na 'n ander oor te dra. Dit kan uiteraard betreklik moeilik wees, hoë onkoste meebring en 'n groot mate van herformattering vereis, maar dit kan wel gedoen word.

6.2.2 Rekenaaromgewings

Die meeste gebruikers het egter altyd 'n keuse met betrekking tot die omgewing van rekenaartoerusting waarbinne hulle gaan werk. Dan is dit belangrik dat 'n besluit geneem word wat die beste moontlikhede bied vir die eie aard van die verwerking van taal – nie net die mees geskikte toerusting vir die onmiddellike toepassings nie, maar ook dié wat die beste moontlikhede bied vir toekomstige uitbreiding en ontwikkeling.

Dit is vroeër (2.3a) aangedui dat die wegbeweeg van die hoofraamomgewing na 'n gedesentraliseerde omgewing beskou moet word as 'n fundamentele verskuiwings met betrekking tot die totale omgewing waarin taal/teks verwerk word. Die gevolg hiervan was (onder andere) dat die verwerking van taal in die hande van 'n groot aantal gebruikers van mikrorekenaars beland het. Vir baie gebruikers bly dit egter nog 'n moeilike keuse, aangesien daar by sommige mense 'n algemene voorstelling is dat die hoofraam-

omgewing tog nog steeds die mees betroubare is, terwyl ander baie sterk daarvoor voel dat dinamiese en vindingryke programmering en implementering vir dié doel juis die beste in die mikrorekenaaromgewing kan plaasvind. Tussen dié twee omgewings bestaan daar vandag ook al hoe meer moontlikhede, enersyds deur rekenaars van middelgrootte en andersyds deurdat die omgewings nie meer net tot een enkele klas van werkverrigting of prys afgegrens kan word nie. Die verandering op die hele rekenaartoneel het die (potensiële) gebruiker se keuse dus eintlik net verder bemoeilik.

Die keuse van 'n omgewing van rekenaarapparatuur het gelukkig nie altyd dwingende implikasies vir die keuse van programmatuur nie. Dié stelling kan deur die volgende gestaaf word:

- a) Dit gebeur toenemend dat die ontwikkelaars van programmatuur daarvoor voorsiening maak dat hulle programme in meer as een rekenaaromgewing bedryf kan word, byvoorbeeld 'n IBM-hoofraamomgewing, UNIX, MS-DOS, MacIntosh. Voorbeelde kan gevind word met betrekking tot onder meer woordverwerkingsprogramme en databasisbestuurstelsels: die woordverwerkingsprogram WORD wat beskikbaar is vir MS-DOS- en UNIX-rekenaars, sowel as vir die MacIntosh; of die databasisbestuurstelsel Oracle waarvan weergawes beskikbaar is vir die gebruik op hoofraamrekenaars, minirekenaars (onder andere onder UNIX as bedryfstelsel) en MS-DOS-rekenaars.
- b) Dit gebeur toenemend dat daar voorsiening gemaak word daarvoor dat 'n bepaalde bedryfstelsel "onder" 'n ander een kan loop. Dit hou byvoorbeeld in dat MS-DOS-programme gebruik kan word op 'n rekenaar wat eintlik in die eerste plek op UNIX-gebaseer is. In dié geval gaan dit daaroor dat die een bedryfstelsel as die gasheer optree en die ander een as gas akkommodeer.
- c) Daar is ook al verskeie soorte aanpassingsapparatuur en -programmatuur ontwikkel wat dit moontlik maak om op 'n rekenaar voorsiening te maak vir 'n ander bedryfstelsel as dié wat as die basiese van die betrokke rekenaar beskou kan word (verwant aan (c) hierbo). So is dit byvoorbeeld moontlik om met 'n (betreklik duur) uitbreidingsbord MS-DOS-programme op 'n MacIntosh te kan hanteer.
- d) Daar word toenemend voorsiening gemaak vir netwerkkoppeling tussen rekenaars met verskillende bedryfstelsels. Dit beteken nie noodwendig dat die data ook gebruik kan word in die programme van die ander rekenaar nie; indien dit om anderssoortige rekenaars gaan, kan daar gewoonlik hoogstens van die randapparatuur van die ander rekenaar gebruik gemaak word. Die data sal nog telkens oorvertaal moet word om in programme op die ander rekenaar gebruik te kan word.
- e) Daar het ook al enkele poortprogramme (*gateway programs*) ontwikkel wat dit byvoorbeeld moontlik maak om 'n databasis op 'n groter rekenaar te hê, maar dit dan toeganklik te maak vanaf 'n mikrorekenaar wat wel 'n substel van die volle moontlikhede hanteer, maar op so 'n wyse dat daar 'n volledige

interaksie tussen die twee programme kan plaasvind. Dit is moeilik om met sekerheid te sê of hierdie benadering in groter mate gaan ontwikkel en of dit vir die rekenaarmatige verwerking van natuurlike taal enige moontlikhede inhou.

Daartéén spreek die feit dat die krag van mikrorekenaars baie vinnig toeneem, baie groot hoeveelhede data baie vinnig oorgedra moet word (wat meer doeltreffend kan geskied op 'n plaaslike rekenaar met vinnige aandrywerkaarte (vir die skerm of vir skyfaandrywers) wat direk op die moederbord aangebring word) en die grafika al hoe nouer met die werkverrigting van die betrokke program geïntegreer word.

6.2.3 Oorwegings by die keuse van 'n rekenaaromgewing

Die keuse van 'n rekenaaromgewing word egter nie net bepaal deur dit wat uit die oogpunt van apparatuur (en programmatuur) die beste blyk te wees nie. Die opset waarin die gebruiker funksioneer, speel ook 'n belangrike rol, byvoorbeeld die vraag na die mate van kundigheid en tegniese ondersteuning wat hy kan verwag in sy omgewing, die mate waarin hy gebruik kan maak van kragtiger toerusting of spesiale randapparatuur by kollegas, die gemaklike beskikbaarheid van programmatuur om te kan uittoets, ensovoorts. Byvoorbeeld: in die akademiese omgewing in die VSA word daar op al hoe groter skaal gebruik gemaak van MacIntosh-rekenaars eerder as van MS-DOS-rekenaars – in elk geval op baie groter skaal as in Suid-Afrika (of in Europese lande). In die eerste twee jaar na die ontwikkeling van die MacIntosh was dit egter 'n baie geslote omgewing (dit wil sê, data kon nie maklik na ander soorte rekenaars oorgedra word nie) en daar was beperkte programmatuur beskikbaar. Intussen is die situasie omgekeerd: vir die MacIntosh is baie goeie programmatuur beskikbaar (vergelyk hieronder in verband met grafika op die MacIntosh) en die data kan baie makliker as voorheen na ander omgewings oorgedra word, byvoorbeeld deur netwerke of deur 'n spesiale aanpassingskaart wat dit moontlik maak om diskette in MacIntosh-formaat op 'n MS-DOS-rekenaar te lees of te skryf.

Die belangrikste oorweging by die keuse van 'n rekenaaromgewing is sekerlik dat die gebruiker daarvoor voorsiening maak dat hy die volle voordeel van die markdinamiek kan geniet. 'n Gevestigde bedryfstelsel bied aan die ontwikkelaars van programmatuur die stabiliteit en sekerheid wat onontbeerlik is vir die langtermyn-belegging wat die ontwikkeling van programmatuur sekerlik is. Binne so 'n stabiele mark vind daar dan tipies 'n baie vinnige en dinamiese ontwikkeling plaas en die gebruiker deel in die voordeel daarvan.

Die inskakeling by 'n stabiele mark uit die oogpunt van die bedryfstelsel bied aan die gebruiker egter ook die verdere voordeel van die moontlikheid om weer betreklik gemaklik uit dié bedryfstelsel te kan uitbeweeg. Hy kan byvoorbeeld daarvan verseker wees dat hy nie alleen agtergelaat sal word wanneer daar wel groot veranderinge met betrekking tot tot bedryfstelsels plaasvind nie. Indien die mark groot genoeg is – wat die geval sal wees by 'n wydverspreid aanvaarde bedryfstelsel – sal daar altyd 'n onder-

nemende instansie wees wat voorsiening maak vir programmatuur wat die oordrag na die nuwe omgewing moontlik maak.

Daar is uiteraard voortdurend ontwikkeling met betrekking tot die oorwegings wat die keuse van 'n bedryfstelsel bepaal en die gebruiker moet leer om dié goed dop te hou of hom te laat adviseer deur iemand wat 'n goeie oorsig daaroor het. Byvoorbeeld:

- Die MS-DOS-omgewing was aanvanklik nie baie aantreklik vir toepassings soos die verwerking van taal nie, aangesien die ontwikkelaars van programmatuur hoofsaaklik op die besigheidsmark ingestel was. Tans is daar 'n besonder wye verskeidenheid programmatuur, met 'n baie hoë graad van gesofistikeerdheid, beskikbaar vir dié omgewing.
- Die MacIntosh-omgewing blyk tans 'n meer belowende omgewing te wees vir persone vir wie grafika baie belangrik is. Dié moontlikhede word ook in die MS-DOS-omgewing ontwikkel, maar is dáár telkens meer as 'n jaar agter dié van die MacIntosh.
- Die UNIX-omgewing is verskeie jare alreeds intensief benut en aangemoedig deur rekenaarwetenskaplikes en deur die ontwikkelaars van programme, waarna daar 'n fase van grootskaalse benutting vir besigheidsdoeleindes gevolg het. Die feit dat UNIX ook in die werkstasie-omgewing gebruik is – waar baie goeie grafika beskikbaar was – het meegebring dat baie van die goeie hipertekstelsels ook in dié omgewing ontwikkel het. UNIX raak tans 'n al hoe meer interessante omgewing vir kragtige rekenaar-toepassings op natuurlike taal. Een groot voordeel is dat daar in dié omgewing rekenaars van bykans elke grootte en prysklas beskikbaar is.

Dit kan as 'n algemene riglyn geld dat dit die veiligste is om 'n omgewing te kies waarin die moontlikhede vir groei en ontwikkeling oop is.

In die lig van bostaande oorwegings, en as die volgende gegewens in ag geneem word -

- watter rekenaars vir die meeste gebruikers in Suid-Afrika die mees geredelik beskikbaar is en
 - die feit dat die meeste gebruikers vir die wydste moontlike spektrum toepassings voorsiening wil maak en die moontlikhede vir toekomstige ontwikkeling wil oophou
- kan gesê word dat die MS-DOS-omgewing tans nog die mees geskikte omgewing is vir die verwerking van natuurlike taal.

Dit is moontlik dat die belang van dié omgewing ook verder kan toeneem op grond van die ontwikkeling van kragtige draagbare MS-DOS-rekenaars met groot skyfkapasiteit, wat meebring dat die gebruiker nou 'n kragtige meganisme vir die verwerking en verskaffing van taalgegewens by hom kan hê waar hy ook al gaan. Daar is twee soorte rekenaars ter sprake, naamlik draagbare kofferrekenaars (onder andere die tipe *lunchbox computers*) en

die skootrekenaars (*laptop computers*) van verskillende verskaffers. In die geval van hipertekstelsels (vergelyk 4.7.3) beteken dit ook dat hy 'n hoogs gesofistikeerde, omvattende kennisstelsel tot sy beskikking kan hê.

Daar bestaan een kombinasie van apparatuur en programmatuur wat spesiaal met die oog op die verwerking van natuurlike taal ontwikkel is, naamlik die bekende IBYCUS-stelsel. Dit is deur David Packard (seun van die bekende rekenaar-entrepreneur en self klassikus) ontwikkel en word bemark deur die Packard Humanities Institute (Los Angeles). (Oor die IBYCUS-stelsel, vergelyk Hughes, 1986a.) IBYCUS is veral geskik vir die hantering van die Bybeltale en ander tale met spesiale karakters. Die stelsel het sy eie bedryfstelsel en moet as 'n geslote stelsel beskou word. Daar word vir 'n sekere hoeveelheid programmatuur voorsiening gemaak, maar die gebruiker kan self ook verdere programmatuur ontwikkel. Die gebruiker wat nie self programmatuur wil ontwikkel nie, is egter uitgelewer aan die verskaffer vir die beskikbaarstelling van nuwe programmatuur, aangesien daar nie 'n wye mark is waarop die dinamiek van programmatuurontwikkeling kan gedy nie. Hoewel dié stelsel dus baie goed kan wees vir navorsers met spesiale behoeftes, is dit – soos ook ander geslote stelsels – nie die ideale stelsel vir die toekoms van die rekenaar-matige verwerking van taal nie.

6.3 MEEGAANDE TOERUSTING

In dié afdeling word net aandag gegee aan twee stukke toerusting wat van belang kan wees vir die rekenaar-matige verwerking van natuurlike taal. Sommige toerusting is in so 'n mate van algemene belang dat dit nie hier bespreek hoef te word nie (byvoorbeeld drukkers). Die ondergenoemde toerusting is egter wel van belang vir toepassings van dié aard.

6.3.1 Kleurgrafika

Dit is belangrik om ook rekening te hou met die verskillende soorte en standarde grafika wat beskikbaar is. Die gebruiker het gewoonlik die keuse tussen verskillende grafika-standarde op dieselfde rekenaar.

Die vraag met betrekking tot die mees geskikte grafika kan nie veralgemenend beantwoord word nie; die antwoord hang telkens af van die behoeftes van die gebruiker en van die toepassings waarmee hy tipies te make het of in die toekoms te make sal hê. Die volgende algemene opmerkings kan in ag geneem word met betrekking tot kleurgrafika:

- Kleurgrafika is vir sommige toepassings hoegenaamd nie nodig nie en die gebruiker kan met monochroom-grafika alles gedoen kry.
- Vir ander toepassings is kleur wenslik: daar word in baie programme toenemend gebruik gemaak van kleur ten einde die verskeidenheid perspektiewe op tekstmateriaal gemaklik te kan onderskei (en selfs om die groot aantal opsies in 'n spyskaart behoorlik te kan onderskei).

- Vir sommige programme is hoë gehalte kleurgrafika van die allergrootste belang, byvoorbeeld vir hipertekstprogramme.

Kleurgrafika word dikwels verkeerdlik as 'n luukse beskou. In baie gevalle is dit egter nie 'n luukse nie, maar hang dit veeleer baie nou saam met die rekenaar-gebruiker-raakvlak: die inligting wat die gebruiker moet sien en duidelik moet kan onderskei, is nie net die elemente van die karakterstel of 'n bepaalde formattering van die teks nie, maar ook 'n verskeidenheid ander perspektiewe op die data en dié kan baie goed aangedui word deur die gebruik van kleur.

In alle gevalle waar kleurgrafika oorweeg word, is dit belangrik om VGA-grafika ernstig te oorweeg. EGA-grafika kan ook sinvol gebruik word, maar moet tog beskou word as 'n tussenfase in die ontwikkeling van grafika-standaarde.

6.3.2 Optiese skyf

Enige soort harde skyf kan uiteraard ook gebruik word vir die rekenaarmatige verwerking van taal. Vir baie toepassings op taal word baie groot hoeveelhede data benodig. Die optiese skyf of laserskyf het geblyk die ideale medium vir die berging en verspreiding van dié groot hoeveelhede data te wees, veral in die vorm van CD-ROM-tegnologie, 'n optiese skyf waarop daar nie deur die gebruiker self "geskryf" kan word nie. (Vergelyk 2.3c oor CD-ROM-tegnologie. Vergelyk ook die verskillende bydraes oor dié tegnologie in Lambert & Ropiequet, 1986 en Ropiequet et al., 1987.) Daar bestaan ook ander soorte optiese skyf, byvoorbeeld herskryfbare optiese skywe, maar op dié gebied het daar nog nie 'n duidelike en algemeen aanvaarde standaard te voorskyn gekom nie. Daar bestaan trouens baie goeie produkte van die tipe "skryfbare optiese skyf": die tipe "eenmaal-skryf" (*write once*) is al goed gevestig (en wel volgens verskeie standaarde deur verskeie vervaardigers), terwyl dit voorkom of die tegniese hindernisse uit die weg geruim is met betrekking tot die tipe "herhaaldelik-skryf" (*write many*). In die geval van CD-ROM het die ontwikkeling egter wel gestabiliseer en geld 'n internasionale ISO-standaard. Daar moet egter in ag geneem word dat CD-ROM vir die MacIntosh-rekenaar nie dieselfde is as vir ander rekenaars nie. Die fisiese formaat is dieselfde, maar die rekordstruktuur verskil.

Dit is belangrik dat persone wat op dié gebied betrokke is, sal besef dat daar wel 'n standaardformaat vir CD-ROM bestaan en dat, al is daar tegnologies gesproke ander standaarde en 'n digter verpakking van data op optiese skyf moontlik, hierdie standaard die hede en die onmiddellike toekoms van die hantering van groot hoeveelhede data bepaal. Alle groot databasisse en dataversamelings wat vir wye gebruik bestem is, word op CD-ROM versprei. Daar is talle CD-ROM-skyfaandrywers beskikbaar wat gebruik kan word om dié optiese skywe te lees. Die data kan uiteraard weer vanaf so 'n CD-ROM-skyf oorgekopieer word na 'n harde skyf of na 'n gewone disket vir verdere gebruik.

HOOFSTUK 7: DIE BESKIKBAARHEID VAN PROGRAMMATUUR

7.1 DIE OPDRAG

Toe die opdrag van hierdie ondersoek uiteengesit is, het die vraag na die beskikbaarheid van geskikte programmatuur vir die verwerking van natuurlike taal baie prominent gefunksioneer. Die vraag is gestel of die LEXINET-ondersoek nie ook die prosesse aan die gang kon sit of kon stimuleer wat daartoe sou kon lei dat die nodige programmatuur teen realistiese koste plaaslik ontwikkel sou kon word of beskikbaar gestel sou kon word nie.

Een van die moontlikhede waaraan tydens die eerste gesprekke gedink is, was 'n databasis vir die opstelling van vakterminologielyste en woordeboeke.

Die vraag na die sinvolheid en haalbaarheid van die plaaslike ontwikkeling van programmatuur moet inderdaad die aandag geniet ten opsigte van die hele verskeidenheid van programmatuur wat in dié verslag ter sprake kom. Dit is ook belangrik dat dié probleem nie net vanuit die perspektief van 'n enkele sentrum of toepassing beskou sal word nie, maar binne 'n breëer, nasionale raamwerk. By die stelselanalise van spesifieke toepassings word die aandag eerder gevestig op die wyse waarop 'n spesifieke probleem opgelos kan word. Dan gaan dit ook oor die vraag hoe alle beskikbare programmatuur moontlik gesamentlik benut en geïntegreer kan word om die probleem op te los. (Vergelyk 5.2 oor stelselanalise.)

7.2 DIE PROBLEEM

Een van die grootste probleme met betrekking tot die verwerking van natuurlike taal, is inderdaad die beskikbaarheid van geskikte programmatuur. Die meeste persone wat hoegenaamd betrokke is of wil raak by dié gebied, het gewoonlik self 'n persoonlike rekenaar of kan toegang tot een verkry. Dit beteken nie noodwendig dat dié rekenaar oor die vermoë beskik om dié soort toepassings te kan hanteer wat met gesofistikeerde toepassings gepaard gaan nie.

Goeie programmatuur vir mikrorekenaars is in Suid-Afrika besonder duur op grond van die swak wisselkoerse en die hoë invoerbelasting daarop. Hierdie faktore is egter nie die enigste rede waarom die koste van programmatuur as buitengewoon hoog ervaar word nie. Dit is ook so dat fondse beskikbaar vir baie persone betrokke by dié soort werk (byvoorbeeld akademici) nie gestyg het in verhouding tot die prysstyging as gevolg van bogenoemde faktore nie. Die pryse van programmatuur word dikwels daardeur opgestoot dat programme (weens sanksies) nie direk van die verskaffers verkry kan word nie (of nie teen spesiale universiteitsafslag nie) en net via tussengangers.

In die privaatsektor is die hoë koste van programmatuur nie die belangrikste rede daarvoor dat geskikte programmatuur vir die verwerking van natuurlike taal nie beskik-

baar is nie. Dié programmatuur kos gewoonlik heelwat minder as algemeen administratiewe of gespesialiseerde professionele programmatuur. Die feit dat programmatuur vir die verwerking van natuurlike taal nie beskikbaar is nie, hang eerder saam met die feit dat mense in dié rigtings nie begryp dat taal eiesoortig is en dat spesiale programmatuur benodig word om daaraan reg te laat geskied nie. Die aankoop van die rekenaar geniet dan voorkeur en ander weë word gevolg met betrekking tot die verkryging van programmatuur.

In Suid-Afrika het die praktyk van die onwettige kopiëring van programmatuur ongelukkig baie wye ingang gevind. Dit vind plaas op beide die persoonlike (private) vlak en die amptelike vlak (dit wil sê in die uitvoering van 'n beroep).

In baie ander lande vind daar uiteraard ook kopiëring plaas, maar die probleem word daardeur getemper dat

- (i) daar dikwels 'n baie meer wydverspreide moraliteit ten opsigte van kopiëring is, wat meebring dat kopiëring nie as 'n heldedaad beskou word nie en in die gemeenskap oor die algemeen afgekeur word en
- (ii) die koste van programmatuur as baie laer ervaar word.

Daar behoort by gebruikers geen twyfel wees oor die feit dat kopiëring onwettig is nie. Dit geld ook van pogings om die kopieregwet te omseil deur interpretasies waarvolgens net een kopie aangekoop word vir gebruik deur 'n hele groep gebruikers. Die kontrak wat tot stand kom wanneer 'n nuwe program oopgemaak word, bepaal gewoonlik dat die betrokke program net op een rekenaar gebruik mag word. Kopiëring kan nooit goedgepraat word nie, wat ook al die omstandighede waaronder dit plaasvind. Gebruikers ervaar egter in dié verband 'n baie ernstige probleem, veral akademiese gebruikers en dié wat met taalnavorsing besig is, maar op hulleself aangewese is vir toerusting vir navorsing: sonder die gepaste programmatuur kan hulle gewoonweg nie voortgaan met hulle navorsing en ontwikkeling en byhou by die vinnige ontwikkelings op dié gebied nie.

Daar is ook 'n verdere probleem, naamlik dat gebruikers nie altyd bewus is van die beskikbare programmatuur nie of die verkeerde programmatuur gebruik vir hulle toepassings. Sodoende kry hulle nie waarde vir hulle geld nie.

7.3 DIE PLAASLIKE ONTWIKKELING VAN PROGRAMMATUUR

Die voor die hand liggende oplossing is dat daar toenemend plaaslik ontwikkel word aan programmatuur. Wisselkoersprobleme word uitgeskakel en invoerbelasting is dan ook nie van toepassing nie. Deur plaaslike montering en vervaardiging kon die pryse van apparaatuur byvoorbeeld relatief laag gehou word.

Daar is egter 'n aantal faktore wat die ontwikkeling van programmatuur vir die verwerking van natuurlike taal in Suid-Afrika bemoeilik en wat onder oë gesien moet word.

7.3.1 Belemmerende faktore

Die volgende faktore kan as enigsins belemmerende faktore met betrekking tot die plaaslike ontwikkeling van programmatuur beskou word.

- a) Die mark vir sommige van dié gespesialiseerde soort programmatuur in Suid-Afrika is nie groot genoeg om die ontwikkeling van sulke programmatuur aantreklik te maak vir ondernemers nie. Programmatuur vir gewone administratiewe of kommersiële toepassings vorm 'n baie meer aantreklike mark.
- b) Uit die PROLANG-onderzoek het dit geblyk dat daar inderdaad 'n baie groot verskeidenheid programmatuur benodig word, vir 'n bepaalde fase van die verwerking van natuurlike taal. Op sekere gebiede het die programmatuur in die afgelope paar jaar baie gesofistikeerd (en uit die oogpunt van programontwikkeling baie kompleks) geraak (byvoorbeeld 'n konkordansieprogram soos WordCruncher). Ander gebiede – wat juis sommige van die mees stimulerende gebiede blyk te wees – het eers werklik tot wasdom ontwikkel sedert daar met die ondersoek begin is (byvoorbeeld hiperteks). Op laasgenoemde gebied is die programontwikkeling só kompleks dat dit te bevraagteken is of dit vinnig deur 'n plaaslike ontwikkelingspan geëwenaar sou kon word.
- c) Die opdatering van programmatuur. By die ontwikkeling van programmatuur gaan dit nie net oor die eerste ontwikkelingswerk nie, maar ook oor die voortdurende instandhouding en opdatering daarvan. In dié verband moet gewys word op die besondere dinamiek van kommersiële programontwikkeling:
 - i) Programme word dikwels so gou soos moontlik bemark ten einde die voorsprong van 'n nuwe konsep of funksie te kan geniet. Dit gebeur dikwels dat alle denkbare funksies doelbewus nie in 'n program ingebou word nie, ten einde die moontlikheid vir opdatering te skep. Die waarde van ander konsepte of funksies word egter dikwels eers ingesien wanneer dit in die programme van kompeterende firmas raakgesien word.
 - ii) In strawwe kompetisie tussen firmas word dan gepoog om nuwe kopers te wen, bestaande kliënte gelukkig te hou of ander te oortreë om oor te skakel na hulle program. Hierdie oorwegings van programmatuurfirmas moet nie negatief beoordeel word nie; dit is eenvoudig deel van die dinamiek van hierdie sektor, en daarom juis ook deel van die dinamiek wat meebring het dat programmatuur van die gehalte wat vandag beskikbaar is, wel hoëgenaamd ontwikkel is.

Gebruikers het dikwels baie naïewe voorstellings oor die dinamiek van die ontwikkeling van hoëvlakprogrammatuur. Die inset aan kundigheid (en dus kapitaal) wat nodig is om van een weergawe van 'n program na 'n volgende een te kan beweeg, word dikwels onderskat. Dit gaan meesal oor 'n eksponensiële sprong. Die ontwikkeling van nuwe weergawes van programmatuur op 'n gereelde basis dra uiteraard ook daartoe by om die nodige kapitaalvloei te verseker.

- iii) Die gereelde opdatering van programmatuur het ook ten doel om telkens die nuut ontwikkelende moontlikhede in verband met apparatuur te kan ontgin. Deur volgende weergawes word dit eenvoudig moontlik om funksies in te bou wat baie direk saamhang met sulke nuwe moontlikhede.
- iv) In nuwe weergawes word gewoonlik ook voorsiening gemaak vir die akkommodering van nuwe standaarde in programmatuur, byvoorbeeld **Postscript** as bladsybeskrywingstandaard (vergelyk 5.4.1) en vir die onderlinge aanpasbaarheid van programmatuur, byvoorbeeld die feit dat programmatuur binne 'n tydperk van twee jaar voorsiening moes maak vir aanpasbaarheid met die netwerkprogrammatuur **Novell** – om hoegenaamd nog in tel te kon bly.

Programmatuur is in dié opsig anders as ander gebruiksvoorwerpe wat selfs na jare nog hulle funksionaliteit kan behou. 'n Program wat vandag goed is, kan binnekort hoegenaamd nie meer bruikbaar wees nie, aangesien dit nie meer aanpasbaar is met ander belangrike programmatuur wat ook gebruik moet word nie.

By plaaslike ontwikkeling sal daar rekening gehou moet word met hierdie dinamiek. Daar moet nie net voorsiening gemaak word vir die eerste ontwikkeling van programmatuur nie, maar ook vir die voortdurende opdatering daarvan.

7.3.2 Die strategiese belang van plaaslike ontwikkeling

By die plaaslike ontwikkeling van programmatuur moet daar nie net gedink word aan die besparing ten opsigte van die invoer van programmatuur nie. Dit gaan ook oor die strategiese belang van die vestiging van 'n plaaslike programmatuurnywerheid wat sy eie dinamiek kan genereer en wat telkens in die toekoms nuwe uitdagings sal kan hanteer. As in gedagte gehou word dat ons in die inligtingsera leef waarin daar telkens nuwe behoeftes of toepassingsgebiede te voorskyn gaan kom wat baie vinnig aangespreek moet word, is dit duidelik dat dit by die plaaslike ontwikkeling van programmatuur nie oor 'n luukse gaan nie, maar oor 'n noodsaaklikheid.

Dit moet ook in gedagte gehou word dat baie ingevoerde programme in elk geval nie (of nie voldoende) voorsiening maak vir die hantering van ander tale as Engels nie. Dit is dus juis baie belangrik dat daar plaaslik programmatuur ontwikkel sal word wat vir al die tale van Suid-Afrika voorsiening maak. Programmatuur vir baie besigheids- en tegniese toepassings kan nog maklik ingevoer word en net so bruikbaar wees, maar waar dit gaan oor die verwerking van natuurlike taal is daar telkens 'n baie noue band tussen die taal en die program. Dié soort programme sal plaaslik ontwikkel moet word (vergelyk byvoorbeeld die aanbevelings van Roux, 1988 oor die noodsaaklikheid van plaaslike ontwikkeling).

As dit dan blyk dat ontwikkelings op dié gebied so strategies is, is dit belangrik dat liggame/instansies wat op 'n oorkoepelende vlak by die beplanning van navorsing en ontwikkeling betrokke is, sal meewerk daartoe dat die nodige plaaslike ontwikkeling van programmatuur wel kan plaasvind.

'n Goeie voorbeeld van 'n plaaslik ontwikkelde program wat 'n baie nuttige funksie vervul, is die vakterminologiesdatabasis van die Nasionale Vakterminologiesdiens.

7.3.3 Alternatiewe vir ontwikkeling

Afgesien van plaaslike ontwikkeling van programmatuur moet die moontlikheid ook ondersoek word van die lisensiëring van programmatuur deur plaaslike instansies op so 'n wyse dat die koste vir die eindgebruiker haalbaar is.

7.4 KUNDIGHEID EN ADVIES

Gebruikers kan ook baie meer waarde vir die beskikbare geld kry indien hulle vooraf deeglik ondersoek sal instel en hulle deur kundige persone laat adviseer of 'n bepaalde program wel geskik is vir hulle toepassing. Dit is 'n bekende feit dat gebruikers 'n groot aantal programme opgaar en net een of 'n paar daarvan werklik gebruik. Daar kan egter met beperkte middele binne die raamwerk van die kopiereg beweeg word indien die keuse van programme op 'n meer oordeelkundige wyse plaasvind. Dan is dit egter ook nodig dat daar sentra van kundigheid sal wees waar persone met 'n belangstelling en behoefte op dié gebied hulle kan laat adviseer. As dié soort sentra dan 'n rol op nasionale vlak vervul, behoort hulle ook van sentrale vlak die nodige ondersteuning te ontvang om dié rol sinvol en met die nodige mate van permanensie te kan vervul.

Die ondervinding aan een so 'n sentrum, Die Navorsingseenheid vir Rekenaartoeepassings op die Taal en Teks van die Ou Testament, Universiteit van Stellenbosch, het onder andere die volgende geleer:

- i) Die meeste persone wat van die rekenaar gebruik wil maak vir die verwerking van natuurlike taal, is baie oningelig oor wat hulle kan doen en watter programmatuur beskikbaar is.
- ii) Daar kan aan so 'n persoon nie net 'n program voorgestel word nie, maar 'n diepgaande gesprek is nodig waarin aandag gegee word aan die inligtingkundige en metodologiese aspekte van die probleem.
- iii) Die gebruiker moet soms lank begelei word ten einde sy probleem met behulp van die beskikbare programmatuur te kan oplos.
- iv) Die oplossing vir 'n probleem op dié gebied kan nie oor die toonbank gekoop word nie; trouens, die meeste verskaffers van programmatuur is totaal onkundig oor dié soort programmatuur en het gewoonlik min begrip vir die aard van die probleem. (Dieselfde geld van baie rekenaarkundiges.)

Dit is selfs moontlik dat dié soort sentrum in 'n groter mate betrokke kan wees by die ontwikkeling van die nodige programmatuur vir 'n wyer gebruikersgroep.

7.5 SAMEVATTING

Die probleem in verband met die beskikbaarheid van geskikte programmatuur is inderdaad ernstig en daar is geen aanduiding dat dié probleem vanself opgelos sal word nie.

Sommige programmatuur sal net eenvoudig aangekoop moet word, aangesien die plaaslike mark nie plaaslike ontwikkeling regverdig nie, die plaaslike ontwikkeling te duur is of die kundigheid nie plaaslik beskikbaar is nie.

Dit is terselfdertyd ook van strategiese belang dat daar soveel plaaslike ontwikkelingswerk soos moontlik gedoen sal word.

Strategiese liggame/instansies sal daartoe moet meewerk dat die nodige ondersteuning vir die ontwikkeling van programmatuur gegee kan word.

Instansies waar persone betrokke is by die rekenaarmatige verwerking van natuurlike taal (byvoorbeeld universiteite) sal in 'n groter mate as voorheen moet toesien dat die middele beskikbaar gestel word vir die aankoop en/of ontwikkeling van programmatuur. Dit geld trouens vir programmatuur op alle gebiede. Programmatuur is deel van die gereedskap van die persone wat dinamies die moontlikhede van die inligtingsera wil ontgin en wil bydra tot die oplossing van die probleme daarvan.

Dit kan van groot belang wees dat plaaslike sentra op so 'n vlak ondersteuning ontvang dat hulle die funksie van advies aan en begeleiding van navorsers en ontwikkelaars kan onderneem, asook moontlik self by die ontwikkeling van programmatuur betrokke kan raak. Die aard van die toepassings maak dit belangrik dat daar telkens langtermyn-ondersteuning gegee sal word.

HOOFSTUK 8. SAMEVATTENDE OPMERKINGS EN AANBEVELINGS

By enkele van die afdelings van die verslag is 'n samevatting gebied. Die volgende opmerkings het 'n meer algemene en oorkoepelende karakter en het ten doel enersyds om die tema van die verslag in die geheel in 'n wyer perspektief te stel en andersyds om aanbevelings te maak met betrekking tot die implementering van dié perspektiewe. Uiteraard sluit baie van dié opmerkings ten nouste aan by die algemene tendense wat in Hoofstuk 2 uiteengesit is.

- a) 'n Baie wye verskeidenheid toepassingsmoontlikhede en programmatuur kom ter sprake onder die tema "die rekenaarmatige verwerking van natuurlike taal": taal kan op verskeie vlakke hantéer word, verskillende soorte aksies kan met taalgegewens uitgevoer word en 'n wye verskeidenheid programmatuur is beskikbaar, beide vir akademiese navorsing en vir toepassings in die praktyk.

Die toepassings wat ter sprake kom, raak alle fasette vanaf die generering van teks tot die ordening daarvan en die interaktiewe hantering van teks in databasisse en kennisstelsels, onder andere in die nuutste programmatuur wat in 'n baie groter mate as sekere toepassings van vroeër reg laat geskied aan die eie aard van taal/teks en die gebruiker se omgang daarmee. Dit is belangrik dat persone wat betrokke is by die rekenaarmatige verwerking van taal, kennis sal neem van die volle spektrum van toepassings en 'n openheid daarvoor sal hê dat die materiaal waarmee hulle werk, moontlik ook in ander vorms en op ander wyses geoperasionaliseer moet kan word. Die unieke aard van teks as draer van inligting kan meebring dat dit soms nodig kan wees om dieselfde teks op verskeie wyses, in verskeie vorms en op verskeie vlakke te operasionaliseer.

- b) Dit is nie net die hoogs gesofistikeerde toepassings wat op 'n gevorderde vlak van outomatiese verwerking gebaseer is wat sinvol is nie. Daar is ook ander toepassings waarin 'n "eenvoudiger" vorm van die verwerking van taal plaasvind wat vir die gebruiker (hetsy akademies of in die sakesektor) van groot waarde kan wees.

Persone wat daarin belangstel om die rekenaar te benut met betrekking tot die verwerking van natuurlike taal, moet nie altyd dink aan hoogs gesofistikeerde toepassings wat op kunsmatige intelligensie-tegnieke gebaseer is nie. Daar is talle ander toepassings (asook programme wat daarvoor aangewend kan word) wat dit moontlik maak om die potensiaal van die rekenaar te ontgin en om die verwerking van taal op 'n ander vlak te plaas. Dikwels word daar neergesien op toepassings van hierdie aard en word hulle as minderwaardig beskou (onder andere deur rekenaar-wetenskaplikes wat dikwels in 'n groter mate in kunsmatige intelligensie belangstel). In werklikheid bring baie semi-outomatiese of "meganiese" toepassings die taalkundige of die persoon wat met teks werk, baie meer in die sak. Trouens, die vermoë om mettertyd by sommige van die meer gesofistikeerde toepassings betrokke te raak, hang dikwels af van die perspektiewe wat ontwikkel word in hierdie eerste vlakke van rekenaar-implementering.

- c) Vir die gebruik van die rekenaar vir al hierdie verskillende toepassings en op al die vlakke is daar telkens 'n wye verskeidenheid programmatuur beskikbaar wat so intensief moontlik benut moet word vir navorsing en ontwikkeling.

Dit het uit die ondersoek geblyk dat daar 'n wye verskeidenheid baie goeie en bruikbare programmatuur beskikbaar is waarvan beide die taalnavorsers en die persoon wat by navorsing en ontwikkeling op heeltemal ander terreine betrokke is, gebruik kan maak. Een van die groot probleme is dat daar nie altyd genoeg kundigheid is oor wat wel beskikbaar is nie. In die verslag is 'n aanduiding gegee van die hele spektrum van aksies wat met taal uitgevoer kan word en van die meegaande programmatuur, wat die gebruiker kan help om sy behoefte êrens op dié spektrum van moontlikhede en beskikbare apparatuur te plaas. Alle programme is vanselfsprekend nie ewe geskik vir die aanwending op alle gebiede nie. Trouens, deur van programme gebruik te maak wat nie vir die betrokke toepassing geskik is nie, bestaan daar juis die moontlikheid dat die data skeefgetrek kan word en die taal nie tot sy reg kom nie. In stelselanalise op dié gebied is dit belangrik dat die bestaande programmatuur so goed moontlik verreken sal word.

- d) 'n Nuwe konsep van stelselanalise ten opsigte van die rekenaarmatige verwerking van natuurlike taal is nodig.

Hierdie konsep van stelselanalise moet rekening hou met aspekte wat nie gewoonlik in stelselanalise ter sprake kom nie en wat tot nou toe nie in rekenaartoepassings op taal gefunksioneer het nie. Daar sal aandag gegee moet word aan die volgende aspekte: die tegnologiese aspek (insluitende programmatuur en standaarde), die soort aksie wat met taalgegewens uitgevoer word en die menslike aspekte.

- e) Om sinvol en dinamies betrokke te wees by die rekenaarmatige verwerking van taal, word van die gebruiker 'n bepaalde instelling van openheid met betrekking tot verandering en met betrekking tot nuwe tegnologiese moontlikhede vereis.

Die tempo van verandering in die inligtingsera is heelwat vinniger as vroeër. Dit geld ook vir die verskillende moontlikhede met betrekking tot die verwerking of operasionalisering van taal, asook met betrekking tot die programmatuur wat beskikbaar word. Dit is belangrik dat gebruikers nie sal vassteek by die stand van denke en programmatuurmoontlikhede van enkele jare gelede nie, maar sal saambeweeg en ook die nuwere moontlikhede sal benut. Dit verg 'n bepaalde instelling en openheid om nuwere ontwikkelings dinamies te kan hanteer, die belangrikste tendense te kan identifiseer en die meganismes daar te stel vir die verwerking van taal op dié vlakke.

Vir baie persone betrokke by die verwerking van taal vorm programmatuur vir woordverwerking en (elementêre) databasisse die hele spektrum van moontlikhede en hulle vind dit moeilik om ruimte te skep vir ander fasette.

- f) Alhoewel daar baie goeie programmatuur beskikbaar is, is dit maar nog net die begin van die ontwikkeling en daar moet dinamies verder ontwikkel word aan programmatuur op dié gebied.

Hoe positief die ontwikkelings met betrekking tot die gebied "rekenaar en taal" tot nou toe plaasgevind het, hoe opwindend die deurbrake ook al is wat gemaak is m.b.t die benutting van die tegnologie vir die verwerking van natuurlike taal en hoe nuttig die programmatuur ook al is wat tot nou toe tot stand gekom het, blyk dit uit hierdie verslag dat daar nog baie ontwikkeling nodig is om programmatuur tot stand te bring wat werklik aan die eie aard van taal, aan die uiteenlopende verskeidenheid toepassings en aan die menslike verwerking daarvan laat reg geskied.

Sommige programme het die verskillende taaldissiplines waarin dit gebruik word en die verskillende toepassingsgebiede waarvoor dit aangewend word 'n baie stewige hupstoot gegee en selfs op 'n nuwe grondslag geplaas. Verandering is egter nodig met betrekking tot die instelling van die ontwikkelaars van programme. Daar moet meer erns gemaak word met die eie aard van taal en die persone met die nodige kundigheid ten opsigte van taal moet betrek word om daarmee behulpsaam te wees.

As daar programme vir stylkontrole ontwikkel word waarby aansprake gemaak word met betrekking tot grammatika en styl, sal daar erns gemaak moet word met wat sintaksis en styl werklik behels; as 'n tesourus daarop aanspraak maak dat dit die gebruiker wil help om die regte woord te vind, sal die ontwikkelaars moet erns maak met die volle kwessie van woordkeuse en semantiek; as 'n hipertekstprogram daarop aanspraak maak dat dit kan help om die kognitiewe prosesse van die gebruiker op die voet te volg, sal daar baie meer aandag gegee moet word aan die aard van kognitiewe prosesse en sal meer eksperimentele toetsing ten opsigte van die effek van dié soort programme moet plaasvind.

- g) Taal/teks het met die mens in al sy fasette te make. Alle navorsing en ontwikkeling wat hierop betrekking het, behoort dus ook 'n interdisiplinêre karakter te hê.

In baie programmatuur vir die hantering van taal is tot nou toe net voorsiening gemaak vir die manipulering van die uitwendige taaltekens, sonder om voldoende rekening te hou met die feit dat taal baie nou met die mens saamhang en 'n bepaalde rol in die gemeenskap vervul. Die ontwikkeling van doeltreffende toegangsmeganismes tot die oormaat van inligting in teksvorm wat vandag beskikbaar is, kan byvoorbeeld daartoe bydra dat die gebruiker se omgang met dié materiaal op 'n haalbare vlak gebring word. Dit is ook dringend noodsaaklik dat daar navorsing gedoen moet word met betrekking tot die kognitiewe aspek van die menslike verwerking van talige gegewens.

- h) Gebruikers en potensiële gebruikers moet opgelei word om rekenaarfasiliteite vir die verwerking van taal te begryp en te kan gebruik.

In die opleiding van studente oor 'n baie breë spektrum van vakgebiede (en persone in talle beroepe) moet voorsiening gemaak word vir 'n algemene oriëntering met betrekking tot die eie aard van taal en die moontlikhede waarop dit deur die implementering van die rekenaar geoperasionaliseer kan word. Hiermee word nie noodwendig bedoel kursusse in Rekenaarwetenskap nie, maar eerder kursusse wat toegespits is op die hantering van taalgegewens (en alles wat daarmee saamhang).

- i) Taal- en letterkundiges het 'n unieke geleentheid om betrokke te raak op dié dinamies en vinnig ontwikkelende gebied.

Die belang van die inset van taal- en letterkundiges ten opsigte van dié ontwikkelings word toenemend duidelik – veral daar waar taal- en letterkundiges bereid is om betrokke te raak en die werklik belangrike probleemgebiede uit die oogpunt van die ontwikkelaars van programmatuur aan te spreek. Dié persone moet beseft dat hulle 'n baie strategiese rol te speel het met betrekking tot dié ontwikkelings. Dit is hulle wat die kundigheid het oor wat taal op die verskillende vlakke behels, wat op die vlakke van toepassing en evaluering betrokke kan raak en wat moontlik self toenemend betrokke kan raak by die ontwikkeling van die nodige programme vir sinvolle toepassings in die toekoms.

- j) Daar moet rekening gehou word met die internasionale karakter van dié gebied en met die vinnige tempo van ontwikkeling wat daardeur teweeggebring word.

Die ontwikkelings op dié gebied 'n sterk internasionale karakter: persone uit verskeie lande is by die ontwikkelings betrokke en deur hulle gesamentlike inset word die tempo van ontwikkeling versnel. Die tempo waarteen die ontwikkelings plaasvind, is verstommend: hiperteks het in minder as twee jaar tot 'n belangrike wyse van die rekenaarmatige hantering van teks ontwikkel en die tekscoderingstandaard SGML het binne drie jaar tot 'n standaard ontwikkel wat deur niemand op dié gebied geïgnoreer kan word nie. Dit is nie voldoende om 'n oorsig net uit publikasies te probeer kry nie. Persone wat op dié gebied wil saambeweeg met die oog op die oordrag van kundigheid na Suid-Afrika, moet in die posisie gestel word om die belangrike ontwikkelings te kan meemaak deur dinamies internasionaal te kan saambeweeg. Dit kan verder nodig wees om spesiaal voorsiening te maak vir meganismes vir die hantering van innovasie.

Op baie gebiede in die sakewêreld waar daar vinnige ontwikkeling en innovasie plaasvind, word daar van die sogenaamde "hekwag"-funksie (*gatekeeper*) gebruik gemaak: 'n persoon van wie dit reeds duidelik geblyk het dat hy die vinnige en dinamiese ontwikkelings kan opvang, interpreteer en oordra, ontvang die opdrag om dié rol te speel vir 'n bepaalde groep gebruikers (byvoorbeeld 'n firma of 'n organisasie) en word in staat gestel om hierdie rol baie doeltreffend te kan vervul. Hierdie benadering blyk baie suksesvol te wees en verseker dat die kundigheid vinnig en doeltreffend by die regte persone uitkom.

9. BIBLIOGRAFIE

- AARTS, J. & T. VAN DEN HEUVEL, 1985, Computational Tools for the Syntactic Analysis of Corpora, *Linguistics* 23: 303-335.
- ABERCROMBIE, J. R. , 1984, *Computer Programs for Literary Analysis*. Philadelphia: University of Pennsylvania Press.
- AITCHISON, J. & A. GILCHRIST, 1972, *Thesaurus Construction. A Practical Manual*. London: Aslib.
- BADER, W., 1989, Textual Analysis with the Help of TUSTEP: Findings from a Study in Syntax, *Proceedings of the Second International Conference on Computers and Biblical Studies* (Jerusalem June 5-13, 1988).
- BAIRD, P. & N. MACMORROW, 1988, Cognitive Aspects of Constructing Non-linear Documents: HyperCard and Glasgow Online, In: *Online Information* (1988): 207-218.
- BARNARD, D. T. et al. (C. A. Fraser & G. M. Logan), 1988a, Generalized Markup for Literary Texts, *Literary and Linguistic Computing* 3/1: 26-31.
- BARNARD, D. T. et al., [1988b], *SGML-Based Markup for Literary Texts: Two Problems and Some Solutions*. (Technical Report #204, Department and Computing and Information Science, Queen's University, Kingston, Ontario, Canada.)
- BATLEY, S., 1988, Visual Information Retrieval: Browsing Strategies in Pictorial Databases, In: *Online Information* (1988): 373-381.
- BEEMAN, W. O. et al. (K. T. Anderson, G. Bader, J. Larkin, A. P. McClard, P. J. McQuilla & M. Shields), 1987, Hypertext and Pluralism: From Lineal to Non-lineal Thinking, *Hypertext '87 Papers*, Chapel Hill, NC (November 13-15): 67-88.
- BEGEMAN, M. L. & J. CONKLIN, 1988, The Right Tool for the Right Job, *Byte* October: 255-266.
- BLAIR, D. C., 1984, The Data-Document Distinction in Information Retrieval, *Communications of the ACM* 27/4: 369-374.
- BLAIR, D. C. & M. E. MARON, 1985, An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System, *Communications of the ACM* 28/3: 289-299.
- BOLTER, J. D. & M. JOYCE, 1987, Hypertext and Creative Writing, *Hypertext '87 Papers*, Chapel Hill, NC (November 13-15): 41-50.
- BROWN, P. J., 1986, Viewing Documents on a Screen, In: Lambert & Ropiequet (1986): 175-184.
- BROWN, P. J., 1987, Turning Ideas into Products: the Guide System, *Hypertext '87 Papers*, Chapel Hill, NC (November 13-15): 51-55.

- BROWN, P. J., 1988, Linking and Searching with Hypertext, *Electronic Publishing* 1/1: 45-53.
- BRYAN, M., 1988, *SGML: An Author's Guide to the Standard Generalized Markup Language*. Workingham: Addison-Wesley Publishing Company.
- BURTON, D.M., 1968, Some Uses of a Grammatical Concordance, *Computers and the Humanities* 2: 145-154.
- BUSH, V., 1945, As We May Think, *Atlantic Monthly* 176/1 (July): 101-108. [Herpubliseer in Lambert & Ropiequet (1986): 3-20.]
- BYERS, T. J., 1987, Built by Association, *Byte* April: 244-251.
- CAPURRO, R., 1986, *Hermeneutik der Fachinformation*. München: Verlag Karl Alber.
- CHARNEY, D., 1987, Comprehending Non-Linear Text: The Role of Discourse Cues and Reading Strategies, *Hypertext '87 Papers*, Chapel Hill, NC (November 13-15): 109-120.
- CLAASSEN, W. T., 1986, Die inligtingsprobleem en die beoefening van die wetenskappe, In: *Prioriteite tot die Jaar 2000*. Deel I. (Verrigtinge van die jaarvergadering van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, Stellenbosch, 1985): 1-18.
- CLAASSEN, W. T., 1987, A Research Unit for Computer Applications to the Language and Text of the Old Testament, *Journal of Northwest Semitic Languages* 13: 11-22.
- CLAASSEN, W. T., te verskyn, User Needs and Information Systems in Biblical Studies, referaat gelewer by die **Second International Conference on Computers and Biblical Studies** (Jerusalem June 5-13, 1988); te verskyn.
- CLAASSEN, W. T. & T. J. D. BOTHMA, 1988, Structuring diverse types of information in hypertext: the case of Biblical information, In: *Online Information* (1988): 83-90.
- COLLIER, G. H., 1987, Thoth-II: Hypertext with Explicit Semantics, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 269-289.
- CONKLIN, J., 1987, Hypertext: a Survey and Introduction, *IEEE Computer*, 20/9 (September): 17-41.
- COOK, J., 1989, A Computerized Data Base for the Qumran Biblical Scrolls, *Proceedings of the Second International Conference on Computers and Biblical Studies* (Jerusalem June 5-13, 1988).
- COOMBS, J. H. et al., 1987, Markup Systems and the Future of Scholarly Text Processing, *Communications of the ACM* 30: 933-947.
- CRANE, G., 1987, From the Old to the New: Integrating Hypertexts into Traditional Scholarship, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 51-55.
- DAVIS, S. G., 1989, Can Stand-Alone EISs Stand Up?, *Datamation* July 1, pp. 41-44.

- DOLAND, V. M. , 1988, *The Hermeneutics of Hypertext*, In: *Online Information* (1988): 75-82.
- DREXLER, K. E., 1987, *Engines of Creation*. New York: Anchor Books.
- DYSON, E., 1987, *Release 1.0* (Nuusbrief oor programmatuur). [Uitgegee deur E. Dyson].
- ENGELBART, D. C., 1963, *A Conceptual Framework for the Augmentation of Man's Intellect*, In: *Vistas in Information Handling, Volume I*, (reds.) P.D. Howerton & D.C. Weeks. 1-29, Washington, D.C.: Spartan Books.
- ETC, 1987, *WordCruncher Version 4.0*. Provo: Electronic Text Corporation.
- FABIAN, B., 1983, *Buch, Bibliothek und geisteswissenschaftliche Forschung*. Göttingen: Vandenhoeck & Rupprecht. (Schriftenreihe der Stiftung Volkswagenwerk, Band 24).
- FAND, J., 1987, *Full-Text Retrieval and Indexing*, In: Ropiequet et al. (1987): 83-102.
- FAXON, 1988, *Comprehensive Directory of CD-ROM Information*, *Access Faxon, Information Access Services 1/1*: 41-44.
- FEINBERG, H., 1973, *Title Derivative Indexing Techniques: A Comparative Study*. Metuchen, N.J.: Scarecrow Press.
- FIDERIO, J., 1988, *A Grand Vision*, *Byte* October: 237-244.
- FINKELSTEIN, R., 1987, *Lingua Franca for Databases*, *PC Tech Journal* December: 53-68.
- FINKELSTEIN, R. & F. PASCAL, 1988, *SQL Database Management Systems*, *Byte* January: 111-118.
- FOSTER, E., 1985, *Outliners: A New Way of Thinking*, *Personal Computing* (May): 74.
- FOX, E.A., 1986, *Information Retrieval: Research into New Capabilities*, In: Lambert & Ropiequet (1986): 143-174.
- FRASER, C. A., 1986, *An Encoding Standard for Literary Documents*. (Master of Science Thesis, Department of Computing and Information Science, Queen's University, Kingston, Ontario).
- FRISSE, M. E., 1987, *Searching for Information in a Hypertext Medical Handbook*, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 57-67. (Herpubliseer: 1988, *Communications of the ACM*: 880-886.)
- FRISSE, M. E., 1988, *From Text to Hypertext*, *Byte* October: 247-253.
- GARRETT, L.N. et al. (K. E. Smith & N. Meyrowitz), 1986, *Intermedia: Issues, Strategies, and Tactics in the Design of a Hypermedia Document System*, *Computer-Supported Cooperative Work (CSCW '86) Proceedings*, Austin, TX, December 3-5, 1986.

- GERKEN, G., 1988, *Die Geburt der neuen Kultur. Vom Industrialismus zum Light Age*. Düsseldorf: ECON Verlag.
- GILL, J. & T. WOLL, 1986, Full Text Management, In: Lambert & Ropiequet (1986): 137-141.
- GOODMAN, D., 1987, *The Complete HyperCard Handbook*. New York: Bantam Books.
- GRAMMATIK, 1987, *Grammatik II, Version 2.0*. Walnut Creek, CA: Digital Marketing
- GUTTET, C., 1985, (ed.) *FORMEX. Formalized Exchange of Electronic Publications*. Brussels and Luxembourg: Office for Official Publications of the European Communities 'New Technologies - Project Management' Department.
- HALASZ, F. G., 1987, Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 345-365. (Herpubliseer: 1988, *Communications of the ACM* 31/7: 836-852).
- HAMMWOHNER, R. & U. THEL, 1987, Context Oriented Relations Between Text Units - A Structural Model for Hypertexts, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 155-174.
- HEIM, M., 1987, *Electric Language: A Philosophical Study of Word Processing*. New Haven: Yale University Press. 1987.
- HELGERSON, L. W., 1988, CD-ROM and Scholarly Research in the Humanities, *Computers and the Humanities* 22: 111-116.
- HERSHEY, W., 1985, Idea Processors *BYTE*, June: 337.
- HERSHEY, W., 1987, Guide (Review) *BYTE*, October: 244-246.
- HOCKEY, S., 1980, *A Guide to Computer Applications in the Humanities*. London: Duckworth.
- HOCKEY, S. & I. MARRIOTT, 1984, *Oxford Concordance Program Version 1.0 Users' Manual*. Oxford University Computing Service.
- HUGHES, J. J., 1986a, The IBYCUS SC: A Multilingual Computer System for Scholars, *Bits & Bytes Review* 1/1: 1-8.
- HUGHES, J. J., 1986b, Nota Bene: State-of-the-Art Academic Word Processing, *Bits & Bytes Review* 1/1: 9-15.
- HUGHES, J.J., 1987a, *Bits, Bytes and Biblical Studies*. (A Resource Guide for the Use of Computers in Biblical and Classical Studies). Grand Rapids: Academie Books.
- HUGHES, J. J., 1987b, AskSam 4.0 - Fast, Flexible Text Retrieval Program (Product Review: Text Retrieval Program), *Bits & Bytes Review* 1/1: 12-15.

- HUGHES, J. J., 1987c, WordCruncher: High-Powered Text-Retrieval Software (Product Review: Text Retrieval Program), *Bits & Bytes Review* 1/3: 1-8.
- HUGHES, J. J., 1987d, From Homer to Hesychius - The Thesaurus Linguae Graecae Project (Data Base Review: Greek Texts), *Bits & Bytes Review* 1/7: 1-6.
- HUGHES, J. J., 1987e, Text Retrieval Programs - A Brief Introduction, *Bits & Bytes Review* 1/8: 1-4.
- HUGHES, J. J., 1987f, Four MS-DOS, Stand-Alone Spelling Programs (Product Review: Spelling Programs), *Bits & Bytes Review* 1/5: 1-16.
- HUGHES, J. J., 1988a, Studying Ancient Greek Civilization Interactively - the Perseus Project (Project Review: Hypertext/ Hypermedia), *Bits & Bytes Review* 2/1: 1-12.
- HUGHES, J. J., 1988b, Micro-OCP. Mainframe Power in a Micro-computer Text Analysis Program (Product Review), *Bits & Bytes Review* 2/1: 13-18.
- HÜSMERT, W., 1988, Auf dem Weg zum elektronischen Lektor, *Micro Extra Heft* 1: 59-61.
- JONASSEN, D.H., 1986, Hypertext Principles for Text and Courseware Design, *Educational Psychologist* 21: 269-292.
-
- KAMIN, J., 1984, *The ThinkTank Book*. Berkeley: Sybex.
- KERR, S. T., 1986, Transition from Page to Screen, In: Lambert & Ropiequet (1986): 321-344.
- KESSELMAN, M. & L. TRAPASSO, 1988, Hypertext and the End-user, In: Online Information (1988): 219-225.
- KUSSLER, R., 1989, Literatuuronderrig deur middel van "Hiperteks", *Tydskrif vir Taalonderrig* 23, 1: 72-88. —
-
- LAMBERT, S. & S. ROPIEQUET, 1986, (reds.) *CD ROM: The New Papyrus*. Redmond WA: Microsoft Press.
- LANDOW, G. P., 1987, Context 32: Using Hypermedia to Teach Literature, *Proceedings of the 1987 IBM Academic Information Systems University AEP Conference*.
- LANDOW, G. P., 1987, Relationally Encoded Links and the Rhetoric of Hypertext, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15: 331-343.
- MACDONALD, N. H. et al. (L. T. Frase, P. Gingrich & S. A. Keenan), 1982, The Writer's Workbench: Computer Aids for Text Analysis, *IEEE Transactions on Communications* 30/1: 105-110.
- MAKRAM, S. et al., 1988, Descriptive Markup - The Best Approach?, *Communications of the ACM* 31/7: 810-811.
- MARCUS, S., 1988, Reading, Writing, and Hypertext, *College Literature* 15/1: 9-18.

- MARKOFF, J., 1988, HP and Ford Motors, *Windows*: 11-13.
- MCCARTY, W., 1986, Nota Bene, an Academic Word Processing and Text Retrieval System for the IBM PC, *Computers and the Humanities* 20: 62-71.
- MEYROWITZ, N., 1985, Networks of Scholar's Workstations: End-User Computing in a University Community, Brown University, IRIS, Providence, RI (IRIS Technical Report: 85-3. June 1985).
- MEYROWITZ, N., 1986, Intermedia: The Architecture and Construction of an Object-Oriented Hypertext/Hypermedia System and Applications Framework, *Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA '86)*, Portland, Oregon (September 29 - October 2, 1986).
- MICHEL, S., 1987, Guide - A Hypertext Solution, *CD-ROM Review* (July/August): 22-24.
- MORRIS, R., 1988, LEXINET en die rekenarisering van taal (Hoofverslag van die LEXINET-program) (Verslag LEXI-3). Pretoria: RGN.
- NAISBITT, J., 1984, *Megatrends. Ten New Directions Transforming our Lives*. London: Macdonald & Co.
- NELSON, T. H., 1981, *Literary Machines*. Swathmore, PA: T.H. Nelson.
- NIEUWOUDT, B. A., 1989a, The Peshitta Project in Stellenbosch: Structure, Computer Programs and Usability of Data, *Proceedings of the Second International Conference on Computers and Biblical Studies* (Jerusalem June 5-13, 1988).
- NIEUWOUDT, B. A., 1989b, The CATSS Database: Progress in Research using Database Programs on Personal Computers and Mainframes, *Proceedings of the Second International Conference on Computers and Biblical Studies* (Jerusalem June 5-13, 1988).
- OAKMAN, R. L., 1980, *Computer Methods for Literary Research*. University of South Carolina Press.
- O'BRIEN, F., 1986, Oxford Concordance Program, *Computers and the Humanities* 20: 138-141.
- O'NEILL, M. & A. MORRIS, 1988, Database and Expert Systems - The Way Forward, In: *Online Information* (1988): 279-290.
- OEP, 1988, *Micro-OCP - Oxford Concordance Program*. Oxford: Oxford Electronic Publishing.
- OLSEN, M., 1987, Textbase for Humanities Applications: WordCruncher, *Computers and the Humanities* 21: 255-260.

- ONLINE INFORMATION, 1988, *Online Information 88. 12th International Online Information Meeting. Proceedings.* (6-8 December 1988, London) Oxford: Learned Information.
- OREN, T., 1987, *The Architecture of Hypertexts, Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 291-306.
- OTT, W., 1979, *A Text Processing System for the Preparation of Critical Editions, Computers and the Humanities* 13: 29-35.
- OTT, W., 1989a, *Vom Manuskript zur Edition. Das Programm SATZ als Baustein in TUSTEP*, In: (reds.) A. Schwob, K. Kranich-Hofbauer & D. Suntinger, *Historische Editionen und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung.* Graz: Leykan-Verlag: 153-176.
- OTT, W., 1989b, *Flexibilität ist Grenzlos*, *DUZ* 1-2/1989: 16-17.
- OTT, W., 1989c, *Transcription Errors, Variant Readings, Scholarly Emendations: Software Tools to Master them, Proceedings of the Second International Conference on Computers and Biblical Studies* (Jerusalem June 5-13, 1988).
- OWL, 1988a, *Guide. Hypertext for the PC.* Seattle: Owl International Inc.
- OWL, 1988b, *OWL Announces IDEX Hypertext-base Document-Management Software*, (brochure) 1 Maart.
- PRUETT, N. J., *Using askSam to Manage Files of Bibliographic References*, *Online* July: 46-52.
- PUGLIA, V., 1986, *TBMS (= Text-based Management Systems, WTC): Database Power Unleashed*, *PC Magazine*, November 25: 211-230.
- RASKIN, R., 1987, *Electronic Thesauri: Four Ways to find the Perfect Word*, *PC Magazine*, January 13: 275-283.
- ROPIQUET, S. et al. (J. Einberger & B. Zoellick), 1987, (reds.) *CD ROM Vol. 2: Optical Publishing. A Practical Approach to Developing CD ROM Applications.* Redmond WA: Microsoft Press.
- ROUX, J. C., 1988, *Kunsmatige intelligensie en die prosessering van natuurlike taal (AILANG-projek van die LEXINET-program)* (Verslag LEXI-8). Pretoria: RGN.
- SALTON, G., 1989, *Automatic Text Processing. The Transformation, Analysis, and Retrieval of Information by Computer.* Reading, Mass.: Addison-Wesley Publishing Company.
- SEASIDE SOFTWARE, 1987, *AskSam 4.0.* Perry, FL: Seaside Software.
- SHASHA, D., 1986, *When Does Non-linear Text Help?*, *Proceedings of the Expert Database Systems Conference.*

- SHAW, R. H., 1988, Project Database 3: SQL: An Emerging Database Standard for PCs, *PC Magazine* (May 17): 275-306.
- SHEPARD, S. J., 1987, AI Meets Hypertext: KnowledgePro & KnowledgeMaker, *Language Technology* (November/December) [aangehaal uit oordruk sonder bladsynommers]
- SHNEIDERMAN, 1987, User Interface Design for the Hyperties Electronic Encyclopedia, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15: 189-194.
- SIMONS, G. F. & L. VERSAW, 1988, *How to Use IT: A Guide to Interlinear Text Processing*. (Revised Edition, Version 1.1) Dallas: Summer Institute of Linguistics.
- SIMONS, G. F. et al. (J. V. Thomson & S. J. DeRose), 1988, *Computing Environment for Linguistic, Literary, and Anthropological Research. A Project Proposal*. Dallas: Summer Institute of Linguistics.
- SMITH, J. B. et al. (S. F. Weiss, G. J. Ferguson, J. D. Bolter, M. Landam & D.V. Beard), 1986, *WE: A Writing Environment for Professional*, University of North Carolina, Department of Computer Science, Chapel Hill, NC, August, 1986: 86-125.
- SMITH, J. B. et al. (S. P. Weiss & G. J. Ferguson), 1987, A Hypertext Writing Environment and its Cognitive Basis, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15: 195-214.
- SMITH, J. B. & S. F. WEISS, 1988, An Overview of Hypertext, *Communications of the ACM* 31/7: 816-819.
- SMITH, J. M., 1987a, *The Standard Generalized Markup Language (SGML): Guidelines for Authors*. (British National Bibliography Research Fund Report 27). London: British Library.
- SMITH, J. M., 1987b, *The Standard Generalized Markup Language (SGML): Guidelines for Editors and Publishers*. (British National Bibliography Research Fund Report 27.) London: British Library.
- SMITH, J. M. & R. STUTELY, 1988, *SGML: The User's Guide to ISO 8879*. New York: Ellis Horwood Limited.
- SMS, 1987, *LBASE*, Silver Mountain Software.
- SOERGEL, D., 1974, *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville.
- SPERBERG-MCQUEEN, M., 1987, ACH Project for Text Encoding Standards, *ACH Newsletter* 9/4: 1-2.
- SWANEPOEL, P. H. & R. MORRIS, 1988, Die rekenarisering van leksikografiese prosesse (WORDNET-projek van die LEXINET-program) (Verslag LEXI-4). Pretoria: RGN.

- TENOPIR, C., 1984, Full-Text Databases, *Annual Review of Information Science and Technology* 19: 215-246.
- TENOPIR, C., 1988, Users and Uses of Full Text Databases, In: *Online Information* (1988): 263-270.
- THOMPSON, B. & B., 1988, *KnowledgePro 1.0* (IBM PC Version). Nassau, N.Y.: Knowledge Garden Inc.
- TOV, E., 1986, *A Computerized Data Base for Septuagint Studies. The Parallel Aligned Text of the Greek and Hebrew Bible*. (Computer Assisted Tools for Septuagint Studies [CATSS], Volume 2) (= *Journal of Northwest Semitic Languages*, Supplement 1). Stellenbosch.
- TRIGG, R. H. & P. M. IRISH, 1987, Hypertext Habitats: Experiences of Writers in NoteCards, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 89-108.
- TRIVETTE, D., 1988, Webster's Thesaurus Pops up 4 Million Synonyms, No Antonyms, *PC Magazine* (April 22): 56.
- UHLIG-DÖRING, C., 1988, Dem Fehlerteufel an den Kragen, *Micro Extra Heft 1*: 40-43.
- VAN DAM, A., 1988, Hypertext '87: Keynote Address, *Communications of the ACM* 31/7: 887-895.
- VAN DER MERWE, D. P., 1988, Annotating literary texts with hypertext, In: *Online Information* (1988); 239-247.
- WALKER, J. H., 1987, Document Examiner: Delivery Interface for Hypertext Documents, *Hypertext '87 Papers*, Chapel Hill, NC, November 13-15, 1987: 307-323.
- WEBB, V. N. et al. (T. J. D. Bothma & R. Morris), 1988, *Rekenaartoepassings in die taalwetenskap* (TEXTNET-projek van die LEXINET-program) (Verslag LEXI-7). Pretoria: RGN.
- WERSIG, G., 1978, *Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis*. München: Dokumentation.
- WOODS, W. A., 1975, What's in a Link? Foundations for Semantic Networks, In: *Readings in Knowledge Representation*, (eds.) R. J. Brachman & H. J. Levesque. Los Altos, CA: Morgan Kaufmann.
- WWB, 1985, *Writer's Workbench Software, User's Guide*. Redport.
- YANKELOVICH, N. et al. (N. Meyrowitz & A. van Dam), 1985, Reading and Writing the Electronic Book, *IEEE Computer*, 18/10 (October): 15-30.
- YANKELOVICH, N. & A. VAN DAM, 1987, *Spinning Scholarly Webs, The Annenberg/CPB Project Report to Higher Education, The Annenberg/CPB Project*, Washington, D.C.

YANKELOVICH, N. et al. (G. P. Landow & P. Heywood), 1987a, Designing Hypermedia 'Ideabases'- The Intermedia Experience, IRIS Technical Report 87-4, Brown University, Providence, RI.

YANKELOVICH, N. et al. (B. Haan, N. Meyrowitz & S. Drucker), 1988, Intermedia: The Concept and the Construction of a Seamless Information Environment, IEEE Computer, January.

ZOELLICK, B., 1987a, Changing the Publishing Paradigm, In: Ropiequet et al. (1987): 1-7.

ZOELLICK, B., 1987b, Selecting an Approach to Document Retrieval, In: Ropiequet et al. (1987): 63-82.

10. BYKOMENDE LITERATUURLYS

- BARRETT, E. (ed.), 1988, *Text, ConText, and HyperText. Writing with and for the Computer*. Cambr., Mass.: The MIT Press.
- BARRETT, E., 1988, Introduction: A New Paradigm for Writing with and for the Computer, In: Barrett (1988): xiii-xxv.
- BROCKMANN, R. J., 1988, Exploring the Connections Between Improved Technology – Workstation and Desktop Publishing and Improved Methodology – Document Databases, In: Barrett (1988): 25-49.
- CARLSON, P.A., 1988, A Way of Incorporating User Feedback into Online Documentation, In: Barrett (1988), 93-110.
- CRANE, G. & E. MYLONAS, 1988, The Perseus Project: An Interactive Curriculum on Classical Greek Civilization, *Educational Technology* 28/11: 25.
- FISCHER, R., 1989, Sterben Bücher aus? Hypertext, Hypermedia: Die totale Informationsintegration, *PC Magazin* 34 (16 August): 28-32.
- HARRIS, M. & M. CADY, 1988, The Dynamic Process of Creating Hypertext Literature, *Educational Technology* 28/11: p. 33.
- HASELKORN, M. P., 1988, The Future of "Writing" for the Computer Industry, In: Barrett (1988): 3-13.
- JAMES, G., 1988a, Artificial Intelligence and Automated Publishing Systems, In: Barrett (1988): 15-23.
-
- JAMES, G., 1988b, The Ethics of Automated Publishing Systems (A Response to Dr. Brockmann), In: Barrett (1988): 50-54.
- JONASSEN, D.H., 1988, Designing Structured Hypertext and Structuring Access to Hypertext, *Educational Technology* 28/11: 13.
- KATZ, B., 1988, Text Processing with the START Natural Language System, In: Barrett (1988): 55-76.
- KEARSLEY, G., 1988, Authoring Considerations for Hypertext, *Educational Technology* 28/11: 21.
- LIN, X., 1988, A Selected Hypertext Bibliography, *Educational Technology* 28/11: 41.
- MARCHIONINI, G., 1988a, Introduction to Special Issue on Hypermedia, *Educational Technology* 28/11: 7.

- MARCHIONINI, G., 1988b, Hypermedia and Learning: Freedom and Chaos, *Educational Technology* 28/11: 8-12.
- MORARIU, J., 1988, Hypermedia in Instruction and Training: The Power and the Promise, *Educational Technology* 28/11: 17.
- RUBENS, P., & R. KRULL, 1988, Designing Online Information, In: Barrett (1988): 291-309.
- SHNEIDERMAN, B. & G. KEARSLEY, 1989, *Hypertext Hands-On! An Introduction to New Way of Organizing and Accessing Information*. Reading, Mass.: Addison-Wesley Publishing Company.
- SLATIN, J.M., 1988, Hypertext and the Teaching of Writing, In: Barrett (1988): 111-129.
- YOUNGGREN, G., 1988, Using an Object-Orientated Programming Language to Create Audience-Driven Hypermedia Environments, In: Barrett (1988): 77-92.

11. SAKEREGISTER

Aanbiedingstechnologie 67

Afkapping 31, 38-41

American Academic Encyclopedia 83

Annotasie van teks 56-57, 82, 90, 98

AskSam 63-64, 87

BASIS 66

Bedryfstelsel 16, 54, 107-112

Bladsybeskrywingstandaarde 13, 31, 42,
100, 117

Boolse operatore 51, 62-64

"Browsing" vgl. Rondblaaier

Bybelkonkordansie 46, 54-55

Bybelse inligting 89

"C" 43

CCF 106

CD-ROM 15, 35, 58, 84, 86, 113

CELLAR 57, 91-97

COCOA 49, 102

Databasisse 1, 41-45, 58-66, 84-90, 95, 99,
113, 120-121

Databasisstandaard 13, 42

Dataherwinning 59, 60

Dataherwinningstelsel 59

dBASE III Plus 42, 43, 44, 45, 53

dBASE IV 43-45

Desentralisering 13, 93

"Desktop publishing" 24, 100

Deskundige stelsels 63, 83

Document Examiner 73

Dokument-as-beeld-databasisse 58

Dokumentdatabasisse 58

Dokumentherwinning 51, 53, 59

Dokumentverwerkingstelsel (IDEX) 105

DTP vgl. "Desktop publishing"

Duke Language Toolkit 52

Eksegetiese wetenskappe 85

Elektroniese boek 70

Elektroniese ensiklopedie 83-84

Epistemologiese 5-6, 77

EPISTLE 38

Ergonomika 21, 77

For Comment 98

FORMEX 105, 106

Gebruikersraakvlak 16, 49, 86, 91

Gebruikersvriendelikheid 16, 22, 86

Geesteswetenskappe 15, 30, 62, 77, 97

"Generalized markup" 103

Generering van teks 6, 19, 24-28, 31, 39,
40, 120

gIBIS 98

Grafika 15-17, 74, 81, 82, 110-113

Grafika-programme 82

Grafika-standaarde 112

Grammatik II 38

Grammatika- en stylkontroleprogramme
21, 32, 35, 38, 39

- Grammatikale konkordansie 48
- Guide** 74-77, 80-82, 98, 105
- "Guideline" 76, 81-82
- Hermeneutiese 6, 16, 77
- Hipermedia 69
- Hiperteks 6-7, 16, 20-23, 28, 57-60, 69-100, 105, 116, 123
- Hiperteksdatabasisse 60
- Hiperteksfasiliteite 64
- Hiperteksdokument 20, 70, 76-83, 90
- Hiperteksprogram 8, 57, 73-81, 86-90, 98, 113, 122
- Hiperteksstelsel 29, 41, 45, 73-112
- Hipertekstoepassing 16, 74, 77, 88
- Homograwe 47
- HyperCard** 74-77, 82, 84, 86-87
- HyperCard "stacks"** 74, 76
- Hyperdoc** 98
- Hyperties** 75, 84
- IBIS** 98
- IBYCUS** 54, 112
- IDEX** 75, 80, 82, 98, 105
- Ikone 21
- IndexETC** 51-52
- Inligtingherwinning 51, 59-60, 63
- Inligtingsera 10-13, 20, 42, 95, 101, 117, 119, 121
- Inligtingsoekende gedrag 10, 15, 22
- Inligtingverkenning 60
- Inligtingverpakkingsbedryf 12
- Interlinear Text Processing (IT)** 6, 44, 56-57, 91
- Intermedia** 23, 73-76, 87
- Kantooroutomatisasie 100, 103
- Kantoorpublikasie 24-25, 40, 67, 100
- Kennisverkennende stelsel 72
- Kennisverkenning 87-88
- KI vgl. Kunsmatige intelligensie
- KnowledgeMaker** 83
- KnowledgePro** 75, 80, 83, 88 89
- Kognitiewe prosesse 5, 6, 22, 28, 71, 77, 90, 95, 96, 122
- Komentariëring van teks 7, 56-57, 85
- Konkordansie 9, 16, 20, 31, 45-55, 58, 61, 67, 69, 99, 116
- Konteks 21-22, 35-36, 39, 46-47, 50, 53, 59, 68, 93
- Kontrole van grammatika en styl vgl. Grammatika- en stylkontroleprogramme
- Kopieregwet 115
- Kopiëring 115
- Kunsmatige intelligensie 1-2, 7, 15, 24, 20-21, 39, 83, 86, 89, 91, 120
- Lbase** 44, 57
- Leksikografiese prosesse 45
- Linguist's Workbench** 44
- Linguistic Database** 44
- Literêre tekste 31, 85
- Logistiek 5-6, 13-14, 93
- MacIntosh** 29, 73-75, 81-84, 109-111, 113
- Mediese vakgebied 85

- Mens-masjienraakvlakstelsels 1
- Micro-OCP 49-51
- MS-DOS 16, 26, 30, 38, 43, 54, 74-75, 80-82, 107-111
- Muis 16, 21, 29, 52, 74, 78-82
- Multimedia 76
- Nasionale Vakterminologiediens 43, 118
- Navigasieprobleem 68
- Netwerk 75, 98, 110
- Netwerkprogrammatuur 117
- Netwerkvorming 98
- Nodi 29, 70-71, 78-80, 86-89
- Nota Bene 30
- NoteCards 73-74
- Novell 117
- Office Document Architecture 102
- Optiese lees van karakters 40
- Optiese skyf 14-15, 98, 113
- Oracle 43, 109
-
- Oriëntalistiek 85
- Othocheck 36
- Oxford Concordance Package (OCP) 49, 102
- Pascal 43
- Perseusprojek 84-85, 88
- PHI-CCAT 15
- Poortprogramme 109
- PostScript 13, 31, 42, 100
- "Precision" 62
- Psigodinamika 5, 32, 68
- Q & A 63
- QUEST 69
- Raamwerkfasiliteite 26-28, 40, 81
- "Recall" 62
- Rekenaar-gebruiker-raakvlak 17, 21-23, 63, 73, 113
- Rekenaargesteunde onderrig 68, 81, 87-88
- Rekenaaromgewing 26, 74, 94, 105, 107-110
- Rondblaaier ("browsing") 59-60, 63, 81
- Sentralisering 13-14
- SGML 80-82, 90-91, 103-106, 123
- Sigbladverwerker 69
- Skakelarea 78, 80, 87-90
- Skakels 29, 70-82, 87-90
- Sneeubal-effek 57
- "Soundex" 37
- Spelkontroleprogram 18, 21, 32, 35-37
- SQL 13, 42-43, 95
-
- "Stacks" vgl. HyperCard "stacks" 74, 76
- STAIRS 66
- Standaard 3, 10, 13, 42, 80, 96, 99-102, 105, 108, 112-113, 117, 121
- Stapel 74
- Statistiek 46, 50
- Stelselanalise 93-96, 114, 121
- Stelselontleding 17
- Storyspace 29
- Stylstelle 25, 101
- Taaldatabasisse 44
- Teks-aanbieding 60, 66-67, 96

Teksgenerering 19, 28, 31-32, 38, 64, 86, 96
 Teksherwinning 53, 61
 Tekskoderingstandaarde 6, 48, 80, 99,
 100-102, 105-106
 Tekskritiese apparaat 31
 Teksverwerking 7
 Terminologie 7, 11, 35, 37, 43, 51, 55, 58-
 61, 64-66, 69, 71, 76-78, 114, 118
 Tesaurus 18, 20, 32-34, 35, 37, 40, 51-52,
 55, 59-66, 86, 89, 122
 TEX 41, 98
 Textcode 102
 Thesaurus Linguae Graecae (TLG) 15, 44
 ThinkTank 26-28
 Thoth-II 73
 Tipografie 25, 27, 30-31, 40, 55, 67-68, 78,
 81-82, 91, 94, 99
 TOSCA-projek 44
 Turbo Lightning 33
 Turbo Pascal 43, 45
 TurboPROLOG 83
 TUSTEP 30-31
 UNIX 38, 74, 81, 107-111
 VAX 30
 Ventura 25, 100
 Vertalingsondersteuning 32
 Vertalingswoordeboek 33-35
 ViewETC 51-52
 Volkswriter 37
 Volteks-databasisse 58-63, 87, 100
 Volteksmateriaal 44
 Werkstasie-rekenaars 15, 74
 Wisselkarakters 51, 62, 64
 Woord-indeks 47
 Woordeboek 20, 114
 Woordelyste 45-46, 62
 Woordeskatstatistiek 45
 Woordverwerking 4-7, 18-19, 24-41, 49,
 52, 56, 67, 81, 86, 99-101, 104, 109, 121
 WORD (MS WORD) 24-30, 34-35, 101-
 104, 109
 Word Finder 33-34
 WordCruncher 16, 50-55, 66, 116
 WordPerfect 25, 30, 37, 52
 WordStar 24
 WordStar 2000 24
 Writer's Workbench 38-39
 Writing Environment 24, 28-29, 86
 Xanadu 75
 Zoom 26-27

Doc no 242899
Copy no 243245



R32,00