

077718



SENTRUM VIR BIBLIOTEK- EN  
INLIGTINGSDIENSTE  
CENTRE FOR LIBRARY AND  
INFORMATION SERVICES

VERVALDATUM/DATE DUE

24 JUL 1999

00002067916



001.3072068 HSRC LEXI 6



\* 206791 \*

# Rekenaargesteunde vertaling

(TRANSNET-projek in die LEXINET-program)

Verslag LEXI—6

# Rekenaargesteunde vertaling (TRANSNET-projek in die LEXINET-program)

R. Morris  
H.M. Retief

Pretoria  
Raad vir Geesteswetenskaplike Navorsing  
1988

R. Morris  
H.M. Retief

Afdeling Leksikologie

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

ISBN 0 7969 0685 6

© Raad vir Geesteswetenskaplike Navorsing, 1988

Gedruk en uitgegee deur die RGN  
Pretoriusstraat 134  
Pretoria

## VOORWOORD

Die Afdeling Leksikologie in hierdie Instituut het teen die einde van 1986 'n ondersoek geloods na die gebruik van die rekenaar in die verwerking van taaldata. In die buiteland vind daar snelle ontwikkelings op die gebied plaas terwyl relatief min aandag in Suid-Afrika daaraan bestee word. Die doel van die ondersoek was om die belangrikheidswaarde 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 die 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 inligtingsdokumente en vertrekstukke vir die verdere uitbouing van 'n betreklik nuwe vakgebied in Suid-Afrika, naamlik die rekenaarhantering van taal.

Die ondersoek waaroor dit in die onderhawige verslag gaan (die TRANSNET-ondersoek) was 'n verkenning van die rekenaar se huidige en potensiële rol in die vertaalwese. Die ondersoek is uitgevoer deur dr. R. Morris en mej. H.M. Retief, albei navorsers in hierdie Instituut.

'n Groot aantal persone en instansies was behulpsaam met die verskaffing van gegewens. Daarbenewens het 25 taalkantore in die land deelgeneem aan 'n vraelysondersoek en onderhoudreeks, en 'n tiental universiteitsdepartemente het stof verskaf oor projekte in die buiteland. Aan al hierdie persone en instansies word opregte dank vir hul samewerking en aktiewe belangstelling in die projek betuig.

- Ten slotte spreek ek graag my waardering uit vir die wyse waarop dr. Morris as projekteier vir hierdie projek en as programleier vir die hele reeks projekte haar taak deurgevoer het.

K.P. PRINSLOO

UITVOERENDE DIREKTEUR: INSTITUUT VIR TAAL- EN KUNSTENAVORSING

<b>RGN BIBLIOTEK</b> 1989 2. 0 8.	
<b>HSRC LIBRARY</b>	
STANDKODE 001-3072068 HSRC LEXI 6	AANWINSNOMMER 077718

## INHOUDSOPGAWE

		Bladsy
1.	INLEIDING	1
1.1	RAAMWERK VAN DIE ONDERSOEK	1
1.2	TAAKOMSKRYWING	2
1.3	AGTERGROND	3
1.4	WERKSWYSE	4
2.	REKENAARGESTEUNDE VERTALING AS ONTWIKKELENDE TERREIN	6
2.1	KORT GESKIEDENIS VAN REKENAARGESTEUNDE VERTALING	6
2.2	DIE VERANDERENDE ROL VAN DIE VERTALER	7
2.2.1	Soorte redigering	8
2.2.2	Redigeringsprobleme	10
2.2.3	Faktore wat naredigering beïnvloed	11
2.2.4	Opleiding	13
2.3	VOORBEELDE VAN VERTALING EN REDIGERING	15
2.3.1	Voorbeeld 1: SYSTRAN, Frans-Engels	15
2.3.2	Voorbeeld 2: Preredigering	17
2.3.3	Voorbeeld 3: Vinnige naredigering	18
2.4	SAMEVATTEND	19
3.	OORSIG VAN BESTAANDE STELSELS	21
3.1	STELSELTIPIES	21
3.1.1	Die aard van die bronmateriaal	21
3.1.2	Werkverdeling tussen die vertaler en die vertaalprogram	22
3.1.3	Die tussentaalse raakvlakstrukture in die program	22
3.2	TAUM-METEO	26
3.2.1	Rekenaargesteunde vertaling in Kanada	26
3.2.2	Oorsig van die METEO-projek	26
3.2.3	Taalkundige beginsels van die METEO-program	27
3.2.4	Die METEO-stelsel in daaglikse gebruik	27
3.2.5	Vooruitsigte	28
3.3	TAUM-AVIATION	28
3.3.1	Oorsig	28
3.3.2	Kenmerke van die subtaal van die handleidings	29
3.3.3	Taalkundige beginsels in die vertaalprogram	29

3.3.4	Resultate	31
3.3.5	Vooruitsigte	31
3.4	ALPS	32
3.4.1	Oorsig	32
3.4.2	Die ALPS-program	33
3.4.3	Beskikbare programmatuur	34
3.5	WEIDNER COMMUNICATIONS CORPORATION (WCC)	36
3.5.1	Oorsig	36
3.5.2	Die Weidner-vertaalprogram	37
3.5.3	Weidner-woordeboeke	38
3.5.4	Vooruitsigte	40
3.6	TITUS	41
3.7	CULT	42
3.7.1	Oorsig	42
3.7.2	Die CULT-program	42
3.8	LOGOS	43
3.8.1	Oorsig	43
3.8.2	Die LOGOS-program	44
3.8.3	Vooruitsigte	46
3.9	DIE CARNEGIE-MELLON-PROJEK	47
3.9.1	Agtergrond	47
3.9.2	Kriteria vir die stelsel	47
3.9.3	Oorsig van die stelsel	48
3.9.4	Entiteit-georiënteerde grammatika-formalisme	50
3.9.5	Funksionele grammatika-formalisme	52
3.9.6	Vooruitsigte	52
3.10	SYSTRAN	53
3.10.1	Agtergrond	53
3.10.2	Oorsig van die EEG-stelsel	54
3.10.3	Die woordeboeke	57
3.10.4	Vooruitsigte	58
3.11	EUROTRA	59
3.11.1	Agtergrond	59
3.11.2	Die EUROTRA-navorsingsgroepe en -skakelmeganismes	61
3.11.3	Die EUROTRA-stelselontwerp	61
3.11.4	Die programmatuur en die taalkundige ontwerp	62
3.11.5	Vooruitsigte	64
3.12	SUSY	64
3.12.1	Agtergrond	64
3.12.2	Onderliggende SUSY-beginsels	65
3.12.3	Die SUSY-woordeboeke	66
3.12.4	Stappe in die SUSY-vertaalproses	67
3.12.5	Hoofverskille tussen SUSY 1 en SUSY 2	71
3.12.6	SUSY-verwante projekte	71

3.13	METAL	72
3.13.1	Agtergrond	72
3.13.2	Oorsig van die METAL-stelsel	73
3.13.3	Die METAL-grammatikas	74
3.13.4	Die METAL-woordeboeke	75
3.13.5	Resultate en vooruitsigte	75
3.14	SAMEVATTEND	77
4.	STANDBESKRYWING EN KOSTE-IMPLIKASIES	78
4.1	DIE VERTAALWESE IN SUID-AFRIKA	78
4.1.1	Die omvang van die vertaalmark	78
4.1.2	Die aard van die vertaalmark en die tale wat figureer	81
4.1.3	Die stand van rekenarisering in die Suid-Afrikaanse vertaalwese	82
4.1.4	Slotopmerkings	86
4.2	DIE KOSTE-EFFEKTIWITEIT VAN REKENAAR-GESTEUNDE VERTALING	86
4.2.1	Ontwikkelingskoste	87
4.2.2	Implementeringskoste	89
4.2.3	Bedryfskoste	90
5.	SAMEVATTING EN AANBEVELINGS	92
5.1	AGTERGROND	92
5.2	OORSIG VAN DIE GEBIED	92
5.3	OORSESE VERTAALSTELSELS	93
5.3.1	Rekenaargesteunde vertaling as ontwikkelingsgebied	93
5.3.2	Die aard en gehalte van bestaande vertaalstelsels	93
5.3.3	Bepanning en finansiering	94
5.4	STANDBESKRYWING VAN DIE SUID-AFRIKAANSE VERTAALWESE	94
5.4.1	Die omvang van die vertaalmark	94
5.4.2	Die aard van die mark	95
5.4.3	Die tale wat vertaal word	95
5.4.4	Die stand van rekenarisering in die vertaalwese	95
5.5	DIE KOSTE-EFFEKTIWITEIT VAN REKENAAR-GESTEUNDE VERTALING	95
5.5.1	Ontwikkelingskoste	95
5.5.2	Implementeringskoste	96
5.5.3	Bedryfskoste	96
5.6	AANBEVELINGS	96



5.7	SLOTOPMERKING	98
	BYLAE	100
	BIBLIOGRAFIE	103

## HOOFSTUK 1: INLEIDING

### 1.1 RAAMWERK VAN DIE ONDERSOEK

Die ondersoek waaroor daar hier verslag gedoen word (die TRANSNET ondersoek) is een van die projekte in die navorsingsprogram LEXINET en die rekenarisering van taal.

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

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

- Die rekenarisering van terminografiese prosesse (TERMNET)
- Die rekenarisering van leksikografiese prosesse (WORDNET)
- Rekenaargesteuende vertaling (TRANSNET)
- Rekenaartoepassings in die taalwetenskap (TEXTNET)
- Kunsmatige intelligensie en die prosessering van 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 rekenaargesteuende taalondersoek (die TEXTNET-verslag) enersyds sprake van die rekenaar in diens van taalnavorsing en andersyds bespreking van taalnavorsing in diens van gerekenariseerde taalverwerking.

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

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

## 1.2 TAAKOMSRYWING

Die doelstellings met die TRANSNET-ondersoek was die volgende:

- om die aard en omvang van die gebruik van die rekenaar in die vertaalwese in Suid-Afrika vas te stel;
- om ondersoek in te stel na gerekenariseerde vertaalstelsels in die buiteland, en na ontwikkelingsprojekte aldaar;
- om in die lig van behoeftes in die Suid-Afrikaanse vertaalwese voorstelle te doen oor moontlike rekenaartoepassings op die terrein;

- om die implikasies van die ontwikkeling van vertaalprogrammatuur vir Suid-Afrika uiteen te sit.

### 1.3 AGTERGROND

Bogenoemde navorsingsdoelstellings spruit uit:

- die besef dat vertaalprobleme in die Suid-Afrikaanse meertalige situasie eskaleer, en
- die feit dat die rekenaar in die buiteland toenemend ingespan word om vertaalprobleme die hoof te bied.

Vertaling speel in Suid-Afrika 'n sleutelrol in die bevordering van effektiewe kommunikasie tussen die verskillende taalgroepe. Reuse hoeveelhede vertaling word gedoen, in die eerste plek tussen die twee amptelike tale, maar toenemend na 'n verskeidenheid ander tale. Met die uitbreiding van die gebruik van die Afrikatale op ekonomiese, politieke, maatskaplike en sosiale vlak neem die verkeer tussen die twee amptelike tale en die Afrikatale snel toe. Die taalbeleid in regshowe dra ook hiertoe by aangesien getuienis in die betrokkenes se moedertaal gelewer mag word. Dit bring baie tolk- en vertaalwerk mee.

Ten einde buitelandse nie-Engelse materiaal in Suid-Afrika, en Suid-Afrikaanse werke oorsee (en in Afrika), beskikbaar te stel, is daar verder 'n toenemende behoefte aan vertalingskanale na en uit die tale van ander lande.

Die 1986 ondersoek na vertaleropleiding in Suid-Afrika (vgl. Pinchuck 1986) het die volgende hoofkenmerke van die Suid-Afrikaanse vertaalwese uitgewys:

- die vertaalmark brei snel uit;
- daar is 'n tekort aan opgeleide mannekrag, veral op die gebied van tegniese vertaling en ten opsigte van die Afrikatale;
- 'n beduidende deel van die aanvraag na vertaling behels dokumente van alledaagse administratiewe aard, waarby die taalgebruik in 'n bepaalde sin geyk is en die oordrag van inligting belangriker is as individuele styl; juis hierdie soort teks is baie geskik vir rekenaargesteunde vertaling;
- die koste verbonde aan vertaling styg na vlakke wat vir sommige doeleindes onekonomies hoog is.

Selfs uit 'n vlugtige beskouing van vertaalmarkte in die buiteland kan afgelei word

- dat hoofsaaklik dieselfde knelpunte daar ontstaan het, naamlik uitbreidende aanvraag, stygende koste, mannekragtekorte en al meer tale waaruit en -na vertaal moet word;
- dat bepaalde rekenaartoe toepassings in vertaling reeds gedeeltelike oplossings tot die probleme bied, byvoorbeeld in tegniese vertaling en in die verskaffing van onafgewerkte vertalings wat uitsluitlik vir inligtingsherwinning bedoel is;
- dat rekenaargesteuende vertaling geleidelik al meer koste-effektief word.

Uit bostaande kan afgelei word dat ontwikkelings op die gebied van rekenaargesteuende vertaling aandag verdien. Daarbenewens is dit duidelik dat indien stelsels vir Afrikaans en die Afrikatale in die vooruitsig gestel word, dit plaaslik onderneem sal moet word. Daar kan egter gesteun word op oorsese voorbeelde van stelselontwerp vir bepaalde toepassings. Hierdie verslag verskaf onder meer beskrywings van 'n aantal oorsese projekte.

#### 1.4 WERKWIJSE

Die ontleding van vertaalaktiwiteite in Suid-Afrika is gegrond op gestruktureerde telefoniese onderhoude met vertalers in 25 taalkantore, opgevolg deur besoeke aan 'n aantal van die grootste kantore, en aangevul deur die bevindinge van die Pinchuck ondersoek (1986).

Ten einde vas te stel of navorsing en ontwikkelingswerk op die gebied van rekenaargesteuende vertaling in die land onderneem word, is die vraelysinligting van twee ander ondersoeke in die LEXINET-reeks gebruik, naamlik die ondersoeke na gerekenariseerde taaldatabasisse in Suid-Afrika (die DOCNET-ondersoek) en na rekenaargesteuende taalnavorsing (die TEXTNET-ondersoek).

Inligting ten opsigte van vertaalstelsels in die buiteland is eerstens deur literatuurstudie en tweedens (en hoofsaaklik) deur persoonlike kontak tydens oorsese reise ingesamel. In die loop van studiereise en kongresbywoning in die buiteland het die volgende persone, afgesien van dr. R. Morris, besoeke afgelê en inligting versamel (meestal ongepubliseerde stof oor lopende projekte) en dit aan die projek beskikbaar gestel: professore W.T. Claassen (Univeriteit Stellenbosch), J.C. Roux (Univeriteit Stellenbosch), V.N. Webb (Univeriteit Pretoria), P.H. Swanepoel (UNISA) en A.D. de V. Cluver (UNISA). Hulle bydraes word met dank erken.

Tydens die opstel van die verslag het daar steeds nuwe inligting oor projekte in die buiteland ingevloei. Die terrein is tans so dinamies dat enige standbeskrywing noodwendig onvolledig sal wees. Terselfdertyd is daar vanweë kommersiële oorwegings 'n beduidende mate van geheimhouding oor projekbesonderhede, dus is die projekbeskrywings in hierdie verslag nie almal ewe volledig nie. Oor projekte in Japan (waar die grootste besteding aan ontwikkeling plaasvind) is die geheimhouding sodanig dat geen noemenswaardige inligting in die literatuur beskikbaar is nie.

Die verslag is opgestel deur dr. R. Morris, bygestaan deur mej. H.M. Retief. Mnr. A.J. Holtzhausen het redaksionele hulp verleen.

## HOOFSTUK 2: REKENAARGESTEUNDE VERTALING AS ONTWIKKELENDE TERREIN

Hierdie hoofstuk behandel die volgende aspekte van rekenaargesteunde vertaling:

- 'n Kort geskiedenis van rekenaargesteunde vertaalstelsels
- Die veranderende rol van die vertaler
- Voorbeelde van vertaling en redigering

### 2.1 KORT GESKIEDENIS VAN REKENAARGESTEUNDE VERTALING

Reeds in 1933 is daar in Moskou melding gemaak van die moontlikheid van rekenaargesteunde vertaling deur Smirnov-Troyanski (Van Schalkwyk 1985:11). In die vyftiger- en sestigerjare is daar veral in Amerika navorsing op die gebied gedoen, hoofsaaklik vir die vertaling van Russiese tekste na Engels, vir militêre en inligtingsdoeleindes. In daardie stadium was beperkings ten opsigte van apparatuur en die ontoereikendheid van programmeringstale bepalende faktore. Daarbenewens het ~~politieke- en ekonomiese druk befondsing beperk tot projekte vir enkele groot tale soos Engels en Russies~~ (Hutchins 1982:21).

Gedurende die sestigerjare het navorsers al hoe meer oor vertaalstelsels ontugter geraak omdat daar toe ingesien is dat vertaling veel meer taalkundige probleme inhou as wat aanvanklik beseft is. In 1966 het die Amerikaanse National Academy for Science vertaalstelsels laat ondersoek en in die verslag wat uitgebring is (die ALPAC-verslag), was die uitsprake oor rekenaartoe toepassings in vertaling so negatief dat bykans alle befondsing deur die regering gestaak is. Teen 1973 was daar slegs drie staatsgesteunde vertaalprojekte, en teen 1975 is ook hierdie befondsing gestaak. 'n Klein aantal projekte het met befondsing uit ander oorde voortgegaan, terwyl die reeds gevestigde stelsels, soos SYSTRAN wat deur die VSA se lugmag vir inligtingsherwinning aangewend is, wel in gebruik gebly het.

In die laat sewentiger- en in die tagtigerjare het daar 'n hernude belangstelling in rekenaargesteunde vertaling ontstaan. Deels was hierdie herlewing toe te skryf aan groter insig in die intrinsieke probleme in enige vertaalproses, wat meer realistiese verwagtinge ten opsigte van die rekenaar se potensiele rol in vertaling meegebring het. Die beseft dat volmaakte vertalings nie haalbaar is nie, het gepaard gegaan met die insig dat 'onvolmaakte' vertalings nogtans op verskillende wyses, en afhangend van die doel met die vertaling, bruikbaar is, of vinnig bruikbaar gemaak kan word. Terselfdertyd het ontwikkelings in die rekenaartegnologie nuwe moontlikhede vir stelselontwerp in die vooruitsig gestel en dit het besteding aan vertaalstelsels aansienlik gestimuleer (Slocum 1984:1).

Die belangrikste faktor in die herlewing van belangstelling in rekenaargesteunde vertaling was egter die besef dat vordering op die gebied van kunsmatige intelligensie (KI) ten nouste saamhang met die ontwikkeling van stelsels wat natuurlike taal kan ontleed en verwerk. Taalontleders (*parsers*) is 'n kernkomponent beide in vertaalprogramme en in deskundige stelsels (*expert systems*), dus trek hierdie terreine gesamentlik uit vordering op die gebied van taalprosessering voordeel. Dit is dan ook juis in lande waar kunsmatige intelligensie-navorsing as 'n hoë prioriteit gestel is dat vertaalprojekte grootskaalse befondsing geniet, byvoorbeeld Japan, Wes-Europa, die Verenigde Koninkryk en die Verenigde State van Amerika (vgl. die AILANG-verslag vir verdere besonderhede oor die samehang tussen kunsmatige intelligensie en rekenaargesteunde vertaling).

Vandag is daar reeds 'n aantal operasionele vertaalstelsels, waaronder die volgende in die Westerse wêreld algemeen bekend is: SYSTRAN, LOGOS, WEIDNER, ALPS, SPANAM, TITUS en TAUM-METEO. Naas die tientalle programme wat nog ontwikkel word, word die programme wat reeds operasioneel is, ook deurlopend uitgebrei en verfyn. Die kommersieel-beskikbare stelsels soos LOGOS, ALPS en WEIDNER verkoop wêreldwyd besonder goed, wat onder andere meebring dat nuwe tale tot die stelsels bygevoeg word.

Opsommend kan daar dus beweer word dat rekenaargesteunde vertaling sodanig gevestig geraak het dat daar nie meer gewonder word, soos in die sestigerjare, of die terrein finansiering vir ontwikkelingswerk regverdig nie. Inteendeel, tans word daar gewaarsku teen die verontagsaming van ontwikkelings op die gebied van rekenaarvertaling (en van taalprosessering in die breë) (Hutchins 1986:17).

In 3.2 tot 3.13 van hierdie verslag word verskillende vertaalstelsels afsonderlik bespreek, en daar word dan ook na verdere aspekte van die geskiedenis van rekenaargesteunde vertaling verwys. In die Bylae is daar 'n opsomming van die bekendste projekte, met besonderhede oor die tydperke waarin dit onderneem is, die plekke waar dit setel, die projekteleiers, die benaderings wat gevolg is, en die tale wat betrek is.

## 2.2 DIE VERANDERENDE ROL VAN DIE VERTALER

Wanneer tegnologiese vernuwing plaasvind, heers daar soms 'n tydlank die vrees dat masjiene die mens sal vervang. Dit was dan ook vertalers se eerste algemene reaksie op die moontlikheid van outomatiese vertaling, en tans is daar in 'n mate nog vrese en verset (Pigott 1987).

Oor die afgelope dekades het dit egter duidelik geword dat die vertaler altyd 'n sleutelrol in die vertaalproses sal speel. Hoë gehalte outomatiese vertalings is slegs in beperkte omstandighede haalbaar, byvoorbeeld met programme wat slegs 'n omskrewe woordeskat en 'n beperkte aantal sinskonstruksies kan hanteer, soos TAUM-METEO, wat weerberigte vertaal. Ander sub-tale is ook al geprogrammeer, byvoor-



beeld die taal van die tekstielbedryf in Wes-Europa, die TITUS-program. Dit lyk egter heel onwaarskynlik dat vertaalprogrammatuur vir algemene tekste binne die afsienbare toekoms sodanig verfyn sal word dat redigering van die vertalings nie meer nodig sal wees nie (Sereda 1982:120). Daarbenewens blyk dit dat die vertaler se spesialiskennis altyd vir sake soos die aanpassing van programme en van die woordeboeke in die programme nodig sal wees.

Hoewel die vertaler in die rekenaaromgewing dus steeds 'n sleutelrol sal speel, is die vertaler se take anders as voorheen. Die hoofzaak word nou redigering, wat voor, tydens of ná die vertaalproses gedoen word. Hierdie verskillende soorte redigering word vervolgens bespreek.

### 2.2.1 Soorte redigering

#### Voorredigering

Die eerste soort redigering is voorredigering (*pre-editing*). Afhangend van die aard van die teks en van die beskikbare vertaalprogram, is dit gewoonlik nodig om eers die taalgebruik na te gaan om meerduidighede uit die weg te ruim, om ingewikkelde konstruksies te vereenvoudig en om die ~~woordeskat in die vertaalprogram aan te vul~~ (lg. kan semi-outomaties gedoen word). Met sommige vertaalprogramme is dit daarna nodig om die teks te kodeer, byvoorbeeld deur rededele en klousgrense aan te dui.

Ondersoeke na die verskillende aspekte van teksvoorbereiding deur voorredigering verskaf interessante inligting oor die oorsaak van verskillende tipes foute wat in vertaling algemeen voorkom. Volgens Ladd (1986:217) kan 'n beduidende hoeveelheid foute in rekenaarvertaalde tekste ~~aan oorsake ekstern aan die vertaalprogram toegeskryf word~~. 'n Ontleding van die WEIDNER-stelsel het aangetoon dat 38% van die foute toe te skryf is aan verskille tussen die bron- en doeltaal wat nie deur die program ondervang word nie. 'n Verdere 4,6% is die gevolg van drukfoute, taalfoute en swak formulering in die bronteks, en 9,8% was die gevolg van onvoldoende of verkeerde inskrywings in die woordeboeke in die program. Deur middel van voorredigering kon die aantal foute in die rekenaarvertaalde tekste met 27,5% verminder word, en daar was ook 'n aansienlike verbetering in die leesbaarheid van die vertalings. Foute kan verder deur die uitskakeling van formateringsprobleme tussen die bron- en doeltteks verminder word. Vergelyk 2.3.2 vir 'n voorbeeld van voorredigering.

Voorredigering verbeter die gehalte van die vertaling maar dit bring nie noodwendig mee dat naredigering (*post-editing*) onnodig is nie (Lawson 1982:6).

Voorredigering is in wese die kontrolering van die invoer na die rekenaar. Invoer kan nog meer radikaal 'gesuiwer' word in gevalle waar die skrywers van die bronmateriaal bepaalde taalriglyne moet volg. As

voorbeeld: die Xerox-maatskappy laat alle dokumentasie oor sy produkte in 'n voorgeskrewe weergawe van Engels opstel, Multinational Customized English, wat uit 'n omskrewe woordeskat en stel sinskonstruksies bestaan. Die vertaalprogrammatuur is op hierdie Engels ingestel en kan tekste wat streng daarvolgens opgestel is, outomaties na 'n aantal ander tale vertaal (vgl. Morris 1986). Naredigering word sodoende feitlik heeltemal uitgeskakel.

### Naredigering

Naredigering behels 'n aantal take, wat egter nie noodwendig almal op elke vertaling uitgevoer hoef te word nie. Feitelike korrektheid moet altyd gekontroleer word, en so ook die juistheid van die terminologie in die stuk. Taalkorrektheid en stylsake is egter dikwels van minder belang, veral wanneer die vertaling hoofsaaklik vir inligtingsherwinning gebruik gaan word.

Die verbetering van taal- en stylfoute is tydrovende aktiwiteite wat die koste van naredigering aansienlik kan verhoog. Daarom word dit al hoe meer algemeen dat vertaalkantore aan kliënte keuses bied ten opsigte van die graad van afwerking wat hulle verlang. Die tariewe vir vinnige naredigering (*rapid post-editing*) word bereken volgens die tyd per bladsy wat daaraan bestee word. Akkuraatheid ten opsigte van die feitelike inhoud van die stuk word gewaarborg, en leesbaarheid word nagestreef, maar verder hang die gehalte van die vertaling grotendeels af van die tydsinset waarvoor die kliënt betaal. (Pigott 1984:3). Daar is 'n toenemende aanvraag na vinnig geproduseerde vertalings van laer kwaliteit, aangesien al hoe meer tekste vir inligtingsdoeleindes vertaal moet word. Vergelyk 2.3.3 vir 'n voorbeeld van vinnige naredigering.

Baie instansies gebruik beide voor- en naredigeringstegnieke. So byvoorbeeld gebruik General Motors 'n woordverwerking- en teksversorgingstelsel waarby die bron- en doelt tekste langs mekaar op die rekenaarskerm vertoon word. Voorredigering geskied op die skerm, dan vertaal die program die teks en daarna word dit, nog steeds op die skerm, afgewerk. Allerlei redigeringshulpmiddels kan opgeroep word, soos woordeskat hulp en teksmanipulering. Sodoende word die sleuraspekte van redigering grotendeels uitgeskakel en die redigeerder kan hom op die oplos van taalkundige probleme toespits. Die tempo waarteen vertaling gedoen word, is by General Motors aansienlik verhoog (Sereda 1982:6).

### Interaktiewe redigering

Interaktiewe redigering (*inter-active editing*) is 'n fasiliteit wat in sommige stelsels ingebou is, byvoorbeeld die ALPS-programme. Die program vertaal die teks een sin op 'n slag en die vertaler wysig die sin, indien nodig, voordat die volgende sin op die skerm verskyn. Die program verskaf hulpmiddels soos sinoniemlyste, en wys meerduidighede, drukfoute en ander probleme uit. Een van die groot voordele van interaktiewe stelsels is dat die vertaler vroegtydig terminologie- en

ander probleme wat by herhaling verder in die teks weer kan opduik, kan opspoor en oplos. 'n Groot nadeel aan interaktiewe redigering is dat die vertaler die vertaalproses intyds moet begelei, wat veral by eenvoudige vertalings onnodig tydrowend is. Verder is dit 'n nadeel dat daar nie, soos by die nie-interaktiewe stelsels, uit die gebruik van goedkoper rekenaartyd in die nag koste-voordele getrek kan word nie.

### 2.2.2 Redigeringsprobleme

In die bedryf van rekenaargesteuende vertaalstelsels is die duurste aspek tans die naredigeringsfunksie. Vanselfsprekend is dit die mees gespesialiseerde taak in die vertaalproses, dus kán dit nie goedkoop wees nie. Daar bestaan egter maniere om dit meer ekonomies te maak, waarvan die belangrikste die opleiding van redigeerders is (Sager 1982:14; vgl. 2.2.4 vir 'n bespreking van die opleiding wat vereis word).

Die foute wat in rekenaarvertaalde tekste voorkom, is van 'n ander aard as dié wat vertalers maak, dus moet die taak anders gestruktureer word. Daarby moet redigeerders opgelei word in die doeltreffende gebruik van die (soms verbysterende) verskeidenheid redigeringshulpmiddels wat nou beskikbaar is.

Die taalfoute wat tipies deur vertaalprogramme begaan word, word breedvoerig in Lavorel 1982 en Green 1982 bespreek. Hier word 'n aantal daarvan kortliks genoem:

- Voorsetsels: Dit is die reedeel wat in die Europese tale die ~~mins patroonmatig voorkom.~~ Afgesien van verskille in voorsetselgebruik tussen tale, kan voorsetselkeuse deur 'n aantal elemente in 'n sin bepaal word; gevolglik is dit moeilik om presiese riglyne in die program in te skryf.
- Werkwoordvorme: By gebrek aan 'n een-tot-een korrelasie van tydsvorme (*tenses*) tussen tale, kan die program slegs algemene riglyne volg, byvoorbeeld dat die (een) Franse teenwoordige tydsvorm gelykstaan aan die Engelse progressiewe vorm, behalwe in die geval van statiewe werkwoorde, dan word die Engelse enkelvoudige vorm gebruik, en so ook in die geval van generiese en gewoonte-aanduidende gebruike, en in verhalende kontekste.
- Idiome: Idiome word dikwels letterlik vertaal.
- Rededeelherkenning: Sommige woorde behoort aan verskeie reedeelklasse, byvoorbeeld *loop*, *meer*, *spook*, *lig*. Indien die verkeerde klas toegeken word, kan dit die res van die sinsontleding raak.

- Frase- en klousindeling: Verkeerde indelings kan veroorsaak dat die redigeerder die hele sin moet oorskryf.
- Stereotipering: Al, bevat die program se woordeboeke sinoniemlyste, gebeur dit dikwels dat bepaalde uitdrukkings óorgebruik word, of dat die program naas-sinonieme nie gepas kan gebruik nie. As voorbeeld: die uitdrukkings *ten opsigte van, met betrekking tot, insake* en wat *betref* is naas-sinonieme, maar hulle is nie orals ewe gepas nie.
- Ellipsis: Weglatings kan soms nie deur die program ondervang word nie, byvoorbeeld *The price of raw materials has risen steeply, so consumer prices will too.*
- Samestellings van naamwoorde: (*noun stacking*) wanneer 'n naamwoordstuk 'n reeks naamwoorde bevat, onderskei die meeste programme die verhoudings tussen die naamwoorde betreklik swak. As voorbeeld: *mainframe memory organization systems user manual*, wat ontleed moet word as *user's manual for memory organization in mainframe systems.*
- Lidwoorde: Dit is uiters moeilik om die verskille in lidwoordstelsels tussen tale volledig in vertaalprogramme te dek. Selfs Engels en Afrikaans, met op die oog af dieselfde lidwoordstelsels (*die-the, 'n-a*) verskil, byvoorbeeld *South African society is changing - Die Suid-Afrikaanse gemeenskap verander.*

Sommige programme begaan ook met verbuigings en vervoegings heelwat foute. Dié soort fout, asook foute met voornaamwoorde en lidwoorde, is voorbeelde van foute wat vertalers betreklik min maak. Redigeerders beskou dit as 'dom' foute, en ondervind blykbaar aanvanklik groot irritasie. Dit is egter foute wat met die rekenaar redelik maklik opgespoor en gekorrigeer kan word (Green 1982:10).

Redigeerders wat met die rekenaar se tipiese 'dommighede' vertrou is, kan dit doeltreffend hanteer. 'n Ervare vertaler van tegniese stukke kan op die tradisionele manier tussen 800 en 1500 woorde per dag vertaal, afhangende van die moeilikheidsgraad van die teks. Dieselfde hoeveelheid stof kan in minder as twee minute deur die rekenaar vertaal word en dan kan 'n redigeerder, om General Motors se situasie as voorbeeld te neem, die teks so vinnig afwerk dat die hele vertaalproses drie- tot viermaal vinniger as vertaling op die gewone manier is (Sereda 1982:120).

### 2.2.3 Faktore wat naredigering beïnvloed

Die volgende faktore is vir die minimalisering van naredigering bepalend:

- Die taalverwerkingsgehalte van die program
- Die aard en gehalte van die brontekste
- Die toeganklikheid van geskikte terminologie
- Die vaardigheid van die redigeerders
- Die integrering van die program in die vertaaldiensgeheel

#### Die taalverwerkingsgehalte van die program

Die taalverwerkingsgehalte van die program is die deurslaggewende faktor. Indien foutiewe analyses en genererings gereeld voorkom, kan redigering meer tyd in beslag neem as gewone vertaling. Programme wat basies niks meer as woord-vir-woord vervanging verrig nie, kan slegs in heel beperkte toepassings doeltreffend funksioneer. 'n 'Intelligente' ontleder (*parser*) is dus vir koste-effektiewe vertaling 'n voorvereiste.

#### Die aard en gehalte van die brontekste

Die aard en gehalte van die brontekste is ook belangrike bepalers van die sukses wat met rekenaargesteunde vertaling behaal kan word. Tekste wat inhoudelik en struktureel besonder ingewikkeld is, kan in baie gevalle vinniger deur 'n vertaler as deur 'n vertaalprogram en nareediging afgehandel word. Dit verg egter 'n deeglike begrip van die vermoëns van die beskikbare vertaalprogram, sowel as kennis van taalkundige kriteria van ingewikkeldheid, om te kan oordeel watter weg die mees koste-effektief sal wees.

Brontekste van swak gehalte, byvoorbeeld ten opsigte van drukfoute, onvolledige sinne, swak punktuasie, inkonsekwente gebruik van afkortings, meerduidighede, en so meer, kan nie bevredigend deur vertaalprogramme hanteer word nie. ~~Die materiaal moet of voorredigeer word, of dit moet deur vertalers vertaal word. Dit verg ook gespesialiseerde kennis om die gehalte van die bronmateriaal te beoordeel.~~

#### Die toeganklikheid van geskikte terminologie

Doeltreffende terminologievoorsiening is in 'n vertaalstelsel kernbelangrik. Die rekenaar se vermoë om woorde vinnig op te soek, en om termkonsekwentheid te handhaaf, is een van sy nuttigste funksies as hulpmiddel by vertaling. Die opsporing van geskikte terme is normaalweg 'n tegniese vertaler se tydrowendste taak. Dit is ook moeilik om konsekwentheid noukeurig te handhaaf. Die rekenaar kan hierdie take uiters doeltreffend verrig, maar slegs op voorwaarde dat die program se vakwoordeboeke van hoë gehalte is. Enige onnoukeurighede in die woordeboekinskrywings sal foute in die vertaling tot gevolg hê.

'n Verdere voorwaarde vir die effektiewe funksionering van 'n program is dat die meeste van die terme in die bronteks reeds in die program se woordeboeke moet wees vóór die teks ingevoer word. Indien 'n groot

aantal terme in die naredigeringfase ingevoeg moet word, styg die koste van vertaling na onekonomiese vlakke.

#### Die redigeerders se vaardighede

Soos hierbo genoem, is die rol van die vertaler in rekenaaromgewings hoofsaaklik dié van redigeerder. Tensy die redigeerders vir die verwerking van die tekortkominge eie aan rekenaarvertaalde stukke opgelei word, bestaan daar die neiging om vertalings te herskryf. Sodoende word die koste- en tydbesparende voordele van rekenaarsteun verbeur (vgl. Rossi 1982).

#### Die integrering van die program in die vertaaldiens

Ten einde optimaal voordeel uit die gebruik van rekenaargesteunde vertaling te trek, behoort die vertaalkantoor se dienste in die geheel so ver moontlik gerekenariseerd te wees. Ideaalgesproke kan die diensvloei soos volg wees:

- kliënte lewer dokumente per elektroniese pos;
- die tekste word op skerm voorredigeer;
- die program vertaal die tekste;
- die naredigeerders versorg op skerm die tekste;
- die dokumente bereik die kliënte per elektroniese pos;
- koste-faktore (omvang van die tekste, moeilikheidsgraad, redigeringstyd, ens.) word in die loop van die vertaalproses deur die rekenaar geregistreer en die gegewens voer dan in na die kantoor se administratiewe rekenaardiens, sodat koste bereken en rekenings uitgestuur kan word.

Bostaande tipe diensvloei-stelsel bestaan reeds in sommige groot vertaalkantore in die buiteland, byvoorbeeld ITT in Engeland. Die integrering van vertaalprogramme in die vertaaldiens se werksvloei-stelsel is 'n saak wat weer in Hoofstuk 4 ter sprake kom. Dit spreek vanself dat effektiewe metodes vir die ontvangs, verwerking en versending van werkstukke die koste van vertaling kan verlaag, al gebruik 'n vertaalkantoor glad nie rekenaarsteun nie.

#### 2.2.4 Opleiding

Uit die literatuur blyk (vgl. veral Lavorel 1982 en Pigott 1987) dat die eise wat aan vertalers in rekenaaromgewings gestel word, groter aanpassings vra as wat aanvanklik verwag is. Opleiding, of heropleiding, is nodig veral ten opsigte van:

- Rekenaargeletterdheid
- Die rekenarisering van leksikons

- Redigeringstegnieke
- Outomatisering in kantooromgewings

### Rekenaargeletterdheid

Onkunde oor die rekenaar is een van die redes vir verset teen rekenaar-gesteunde vertaling. Aan die een kant bestaan die vrees dat die rekenaar werkloosheid onder vertalers gaan veroorsaak. Aan die ander kant is daar irritasie en frustrasie omdat die rekenaar 'dom' is. Dit het nie die oordeel, en in vele opsigte ook nie die kennis, waaroor selfs swak vertalers beskik nie. Blykbaar ervaar sommige vertalers die rekenaar as 'n voorskriftelike mag, en hulle voel dat hulle 'n ondergeskikte rol speel: die rekenaar voer die take uit waartoe dit in staat is, en wat oorbly moet die vertaler doen.

Verder is dit vir party moeilik om aan te pas by die idee dat ruwe vertalings, of vinnig geredigeerde vertalings, vir sommige doeleindes deur kliënte verkies word. Dit druis blykbaar teen beroepseergevoel in om enigiets behalwe behoorlik afgewerkte produkte aan kliënte te lewer. Laastens verg dit aanpassing om in vennootskap met die rekenaar vertalings te doen. Tradisioneel is vertaalwerk solowerk, en die eindprodukt is dan die vertaler se eie prestasie.

Vertroudheid met die rekenaar kan meebring dat vertalers die saak in 'n ander perspektief stel. Die rekenaar is 'n hulpmiddel vir die mens, nie andersom nie. Dit is 'n hulpmiddel wat besonder nuttig is, veral vir die hantering van die sleuraspekte van vertaling. Die werklik uitdagende aspekte van die werk, soos fyn keuses tussen uitdrukkings en konstruksies om betekenisnuanses getrou weer te gee, bly die vertaler se domein.

### Die rekenarisering van leksikons

Vir die vertaler wat nie oor 'n vertaalprogram beskik nie, is die nuttigste steun deur die rekenaar seker die woordeboek-fasiliteite wat in sommige woordverwerkingspakkette ingebou is. Vir rekenaargeletterdes verg dit geringe opleiding om die tegnieke aan te leer vir die uitbou van 'n woordeboek na gelang van eie behoeftes.

### Redigeringstegnieke

Die bespreking in 2.2.1 toon dat voor- en naredigering in die rekenaar-omgewing verskil van die soort redigering wat tradisioneel gedoen is. Voorredigering het vroeër gewoonlik hoogstens beteken dat die vertaler terminologie-naslaanwerk vooraf doen. Dit was onnodig om tekortkominge in die bronteks reg te stel, soos spelfoute, swak formulerings en meerduidighede. Dit moes net in die doelteks vermy word. Vir invoer na 'n vertaalprogram moet die doelteks egter deeglik gesuiwer word. Daarvoor is spesiale vaardighede nodig, byvoorbeeld kennis en oordeel

ten opsigte van moontlike probleme wat die bepaalde vertaalprogram met die bronteks sal ondervind. Blykbaar kan dit ook opleiding vereis om te keer dat vertalers te veel voorredigering, en selfs herskrywing van die bronteks, uitvoer.

Opleiding in naredigeringstegnieke behels hoofsaaklik sensitering vir die tipe foute wat vertaalprogramme begaan en die aanwending van die teksversorgingsfasiliteite wat die rekenaar bied. Dit is byvoorbeeld onnodig om die teks te kontroleer vir spelfoute, want dit kan outomaties gedoen word, tog kom redigeerders blykbaar moeilik uit die gewoonte om spelling woord-vir-woord na te gaan.

#### Outomatisering in kantooromgewings

Soos reeds vroeër genoem, kan die koste- en tydsbesparende voordele van rekenaargesteuende vertaling genegeer word indien die kantooromgewing nie vir effektiewe werksvloei ingerig is nie. Rekenarisering speel vanselfsprekend deesdae 'n groot rol in doeltreffende kantoorstelsels, maar sonder personeelopleiding kan die voordele van die stelsels nie ten volle benut word nie. Vir vertalers wat tradisioneel by boekomringde lessenaars in betreklike afsondering gewerk het, kan aanpassing by die nuwe omgewing veeleisend wees.

### 2.3 VOORBEELDE VAN VERTALING EN REDIGERING

Die voorbeeld van rekenaargesteuende vertaling wat hier bespreek word is 'n ongeredigeerde vertaling deur die EEG se SYSTRAN-program (vgl. 3.10). Terwille van die bespreking wat volg is die teks in genommerde groepe frases ingedeel, en die vertaling daarvan verskyn direk daaronder.

#### 2.3.1 Voorbeeld 1: SYSTRAN, Frans-Engels (vgl. 3.10)

1. FSSRS (Farm Structure Survey Retrieval System) est un logiciel  
FSSRS (Farm Structure Survey Retrieval System) is a software
2. qui permet: la consultation, l'extraction et  
which allows: the consultation, the retrieval and
3. la manipulation des données des enquêtes sur la structure  
the manipulation of the data investigations about the structure
4. des exploitations agricoles.  
of the farms.
5. Ce logiciel, suite à une demande de la division E4  
This software, further to a request of the division E4
6. 'Comptes et Structures Agricoles' de l'OSCE,  
'agricultural accounts and structures' of the SOEC,
7. a été conçu par Monsieur Guido Vervaet de l'unité A1  
was conceived by Mr Guido Vervaet of the unit A1
8. de l'OSCE et réalisé par Messieurs Neal Garlick et  
of the SOEC and carried out by Messrs Neal Garlick and



9. Steve Ditchburn de la societ  ARONWORTH.  
Steve Ditchburn of 'ARONWORTH' company.
10. FSSRS est operationnel depuis le d but de 1984 et  
FSSRS is operational since the beginning of 1984 and
11. tourne actuellement sur un ICL 39/80.  
turns currently on ICL 39/80.
12. La gestion de la base et l'aide aux utilisateurs  
The management of the base and the aid for users
13. sont assur es par M. Jean Marie LENTSCHAT.  
is ensured by Mr Jean Marie LENTSCHAT.
14. L'utilisation de FSSRS est bas e sur un syst me  
The use of FSSRS is based on a system
15. de menus hi rarchis s d'un usage assez ais .  
of hierarchised fragments rather easy use.
16. L'utilisateur a la possibilit  de travailler en  
The user has the possibility of working as
17. Fran ais, Anglais ou Allemand. Ce manuel de consultation  
Frenchmen, Englishmen or German. This consultation handbook
18. et le manuel 'Contenu de la base' ne sont disponibles qu'en  
and the handbook 'contents of the base' is available only as
19. Fran ais et en Anglais; cependant si la demande existe  
Frenchmen and as Englishmen; meanwhile if the request exists
20. ils pourraient  tre traduits en Allemand.  
~~they-could-be-translated-into-a-German.~~

Kommentaar oor bostaande vertaling:

Die foute en swak formulerings kan in drie kategorie  ingedeel word: foute in die bronteks, funksieverskille tussen Frans en Engels, en defektiewe woordeboekiniigting.

(a) Foute in die bronteks

Die eienaardige voorkoms in (17), (19) en (20) van *Frenchman* en so meer, is gewoon daaraan toe te skryf dat die betrokke woorde in Frans nie met 'n hoofletter moet begin nie. Met 'n hoofletter verwys die betrokke naamwoorde na mense, en met 'n kleinletter na tale. Dit is die soort fout wat vir redigeerders irriterend is, want 'n vertaler sou dit nie maak nie, maar die vertaalprogram is nie die sondebok nie. Die skrywer van die bronteks het di  foute begaan.

(b) Funksieverskille tussen Frans en Engels

Verskille tussen die twee tale in lidwoordgebruik veroorsaak die lidwoordfoute in (1), (4) en (9), en die oorbodige lidwoorde in (2), (3) en (12). In (3) is die foutiewe *data investigations* (die bedoeling is: *data from investigations*) ook toe te skryf aan 'n verskil in lidwoordgebruik tussen die tale.

Die fout in (10) spruit voort uit die feit dat die werkwoordtyd wat met *depuis/since* voorkom, die teenwoordige tyd in Frans en die perfektum in Engels is. Hierdie verskil tussen die tale kan deur annotering van die items in die woordeboek ondervang word.

(c) Defektiewe woordeboekinligting

Die volgende verbeterings sou deur die verfyning van die woordeboek-  
inskrywings bewerkstellig kon word:

- (4) *farms* moet eerder *agricultural undertakings* wees;
- (5) *further to*, eerder: *following on*;
- (11) *turns*, eerder: *operates*;
- (12) *aid for users*, eerder: *user support system*;
- (13) *is ensured*, eerder: *are the responsibility of*
- (15) *hierarchised fragments* = *hierarchical menus*;
- (15) *rather easy use* = *which are quite easy to use*;
- (17) *consultation handbook* = *reference manual*.

Punt (17) is ooglopend die gevolg daarvan dat die woordeboekinskrywing deur 'n nie-moedertaalspreker verskaf is, en van die ander voorbeelde kan waarskynlik ook daaraan toegeskryf word.

Die vertaling, en bostaande ontleding, illustreer onder andere dat rekenaarvertalings sonder enige redigering heel verstaanbaar en bruikbaar kan wees ten spyte van sekere gebreke. Daarby illustreer die voorbeeld dat vertalers deur foutontleding die gehalte van die vertaalprogram, en veral die gehalte van die woordeboeke, kan verbeter.

2.3.2 Voorbeeld 2: Voorredigering

Die volgende is 'n voorbeeld van die voorredigering van 'n Engelse teks vir vertaling na Frans.

Oorspronklike teks:

This chapter describes the use of recorders for recording and loading programs. The 1770-SA recorder is also described in the Digital Cassette Recorder User Manual (document 1770-86).

This chapter describes the use of printers for printing memory data values, ladder diagram rungs, the text of message files in the message area, and memory map reports. To use a printer or CRT terminal for programmable report generation, refer to the I/O Scanner-Message Handling Module User Manual (publication 1770-806).

When operating peripheral devices, it may be necessary to enter or interpret an extended address. For detailed information on memory organization and extended addressing, refer to the PLC-3 Programming Manual (publication 1775-801).

### Voorredigeerde teks:

This chapter describes the use of recorders to record and to load programs. The recorder 1770-SA is also described in the user's Manual of the digital cassette recorder (document 1770-86).

This chapter describes the use of printers to print values in the memory, ladder diagram rungs, the text of message files in the message area and memory states. To use a printer or a CRT terminal for the generation of programmable reports, please see the user's Manual of the message handling module of the I/O (document 1770-806).

During the use of peripheral devices, it may prove necessary to enter or to interpret an extended address. For detailed information on memory organization and on extended addressing, please see the programming Manual of the PIC-3 (document 1775-801).

### 2.3.3 Voorbeeld 3: Vinnige naredigering

Die volgende is 'n voorbeeld van vinnige naredigering. Die redigeerder het 20 minute lank aan die stuk gewerk.

#### Ongeredigeerde vertaling:

~~The borders between countries industrial advanced and NPI, or NPI and other developing countries, evolve permanently. The countries form according to the expression by OECD, a "dynamic continuum" in the process of development. They reach for certain products, at the competitive country row on the world market, but can lose or strengthen this position within the variable time according to the evolution of their national productive structures, of the forms of competition on the world market, of the technological transformation, etc. ...~~

~~In the Sixties, a first group of developing countries emerges on the basis of the growth rates of their internal production and their very high exports. This group comprises four states from Asia of the south - is: Hong Kong, South Korea, Taiwan, Singapore; four countries from Europe of the south: Greece, Yugoslavia, Spain, Portugal; and two from Latin America: Brazil and Mexico.~~

A second wave of exporting developing countries of manufactured products develop in the decade seventy: the Philippines, Malaysia, Thailand, Macao, Indonesia, India, Argentina, Tunisia, Morocco, Mauritius, etc. ... Some among to seem revolving plates between first NPI and the developed economy markets. First NPI has there, indeed, delocalized or sub-contracted a number of activities under multinationals of the Third World. This strategy facilitated the development of the south-south trade, but it also strengthened the capacity of export of a growing number of developing countries which became competitive countries on the world market.

### Geredigeerde teks:

The borderline between advanced industrial countries and NICs, or NICs and other developing countries, changes constantly. The countries form, to use the expression of the OECD, a 'dynamic continuum' in the development process. For certain products they attain competitive country status on the world market, but can lose or strengthen this position in time, depending on trends in their national production structures, the forms of competition on the world market, technological change, etc.

In the Sixties, a first group of developing countries emerged on the basis of the growth rates of their domestic production and their very high exports. This group comprises four states of South-East Asia: Hong Kong, South Korea, Taiwan and Singapore; four countries from southern Europe: Greece, Yugoslavia, Spain and Portugal; and two from Latin America: Brazil and Mexico.

A second wave of developing countries exporting manufactured goods arose in the seventies: the Philippines, Malaysia, Thailand, Macao, Indonesia, India, Tunisia, Argentina, Morocco, Mauritius, etc. Some appeared as centres trading between the first NICs and the developed-economy markets. Indeed, the first NICs set up or sub-contracted locally a number of activities through Third World multinationals. This strategy facilitated the development of south-south trade, but it also strengthened the export capacity of a growing number of developing countries which have become competitive on the world market.

Die bostaande voorbeelde van rekenaarvertaling en van redigeringsstegnieke toon dat 'n aanvaarbare teksgehalte met betreklik min tydsinset behaal kan word mits die redigeerder oor 'n deeglike kennis van die vertaalprogram se tekortkominge beskik. In hierdie verband illustreer Voorbeeld 1 dat selfs net 'n paar minute van voorredigeringsaandag voldoende sou gewees het om 'n aantal van die foute te voorkom.

In die literatuur oor die rolverdeling tussen die vertaler en die vertaalprogram (vgl. Ladd 1986, Evans 1986, Habermann 1986, Pigott 1984) word telkens genoem dat vertalers sleutelbydraes tot die verbetering van vertaalstelsels kan maak deur foute-analises uit te voer en die taalkundige redes vir die foute te identifiseer. 'n Verdere belangrike bydrae van die kant van vertalers is die opgradering van die woordeboeke in die stelsels - wat dikwels inhou dat implisiete kennis wat die mens onbewustelik aanwend wanneer 'n woordeboekinskrywing geraadpleeg word, vir die rekenaar uitgestippel moet word.

### 2.4 SAMEVATTEND

Die hoofpunte in hierdie hoofstuk kan soos volg saamgevat word:

- Rekenaargesteeunde vertaling het 'n wankelende vroeë geskiedenis deurgemaak maar dit het nou sterk gevestig geraak,

beide in die vertaalwese en as navorsings- en ontwikkelingsgebied.

- Die vertaler speel in rekenaaromgewings al hoe meer die rol van 'n hoogs gespesialiseerde redigeerder, wat meebring dat meer gevorderde opleiding in die linguistiek, in rekenaargeletterdheid en in redigeringsstegnieke nodig is.
- Die voorbeeldmateriaal wat hierbo aangebied is, toon dat bestaande stelsels betreklik goed vertaal maar dat dit die vertaler hoegenaamd nie vervang nie.

In die hoofstuk wat volg, word daar breedvoerig ingegaan op verskillende soorte vertaalstelsels. 'n Belangrike punt wat dan ook verder beklemtoon word, is dat rekenaarsteun by vertaling 'n breë spektrum gebruikdek, vanaf die gebruik van niks meer as gewone woordverwerkings- en woordeboekfasiliteite nie tot ten volle outomatiese vertaling, met, tussenin, verskeie modelle vir interaktiewe werkswerrigting deur die vertaler en die rekenaar saam.

## HOOFSTUK 3: OORSIG VAN BESTAANDE STELSELS

As inleiding tot die beskrywing in hierdie hoofstuk van 12 vertaalstelsels wat in die buiteland ontwikkel is, word daar eers 'n kort uiteensetting van die hoofkenmerke van verskillende stelseltipes gegee. Die 12 stelsels wat dan beskryf word, is van die bekendstes in die Westerse wêreld en hulle illustreer verteenwoordigend 'n aantal benaderings tot die problematiek van rekenaargesteunde vertaling. Soos reeds genoem, is Japan tans die grootste besteder op die gebied van die rekenaarhantering van taal, maar weinig inligting oor die Japannese projekte word beskikbaar gestel.

### 3.1 STELSELTIPES

Die hoofverskille tussen vertaalprogramme raak die volgende drie aspekte van die vertaalproses:

- Die aard van die stof wat deur die program vertaal moet kan word.
- Die werksverdeling tussen die vertaler en die vertaalprogram.
- Die tussentaalse raakvlakstrukture in die program.

#### 3.1.1 Die aard van die bronmateriaal

Sommige stelsels is op die vertaling van slegs een spesifieke tipe teks toegespits, bv. TAUM METEO, vir weerberigte (vgl. 3.2), TAUM-AVIATION, vir handleidings vir die instandhouding van lugvaartuie (vgl. 3.3), TITUS, vir opsommings vir die tekstielbedryf in Europa (vgl. 3.6) en CULT, vir wiskundige tekste (vgl. 3.7).

Met sodanige programme word vertaalprobleme op twee wyses vereenvoudig. Eerstens, die beperking tot een vakgebied skakel heelwat terminologieprobleme uit, asook probleme van meerduidigheid. Tweedens, die beperking tot een toepassing binne die gekose vakgebied (bv. handleidings, of opsommings, of berigte) vereenvoudig die grammatika-vereistes. As voorbeeld: die sintaksis in handleidings kan betreklik 'telegrafies' wees. Daarby bestaan die teks gewoonlik uit genommerde sinne of kort paragrawe, wat die vele probleme met die vertaling van kohesie-strukture in aaneenlopende tekste grotendeels uitskakel.

'n Voortsetting van die idee dat 'n vertaalprogram beter kan funksioneer indien dit beperk word tot een vakgebied en een toepassing, is die ontwikkeling van gekontroleerde invoertale, byvoorbeeld Caterpillar English en Customized Multinational English. Soos genoem in 2.2.1, is dit weergawes van Engels wat tot 'n omskrewe stel sinskonstruksies en 'n afgebakende algemene woordeskat beperk is (waartoe terminologie vir bepaalde doeleindes dan gevoeg kan word).

Vertaalstelsels wat nie tot een vakgebied, subtaal of toepassing beperk is nie, is vanselfsprekend aansienlik ingewikkelder en terselfdertyd in die lewering van vertalings van hoë gehalte oor die algemeen minder suksesvol. Die EEG se SYSTRAN-stelsel (vgl. 3.10) en EUROTRA (vgl. 3.11) is voorbeelde van sodanige groot stelsels.

### 3.1.2 Werksverdeling tussen die vertaler en die vertaalprogram

Die hoofonderskeid wat onder hierdie rubriek getref word, is tussen interaktiewe programme en outomatiese stelsels, soos bespreek in 2.2 hierbo. Soms word die term *rekenaargesteunde vertaling* met verwysing na interaktiewe stelsels gebruik, terwyl die term *masjienvertaling* na outomatiese vertaling verwys. Die interaktiewe stelsels wat in hierdie hoofstuk bespreek word, is ALPS (3.4), CULT (3.7) en Carnegie-Mellon (vgl. 3.9). Die ander stelsels onder bespreking is outomaties, dit wil sê die teks word, sonder vertalerstussenkoms tydens die vertaalproses self, deur die rekenaar vertaal. Soos geblyk het uit die bespreking in 2.2, lewer die vertaler sy bydrae in die voor- en naredigeringsfases.

### 3.1.3 Die tussentaalse raakvlakstrukture in die program

Daar is basies drie benaderings tot die kwessie van watter struktuur gebruik moet word om 'n teks van een taal na 'n ander om te sit:

- Direkte vertaling
- Vertaling via 'n intertaal
- Vertaling via 'n oordragstruktuur

#### Direkte vertaling

Die onderskeidende kenmerk van direkte stelsels is dat 'n afsonderlike program vir elke taalpaar ontwerp word. Die grondbeginsel is dat die program volgens die spesifieke probleme wat vertaling uit brontaal X in doeltaal Y meebring, ontwerp word. As klein voorbeeld: vir vertaling uit Engels na Afrikaans moet die program rekening hou met die feit dat die werkwoord *know* in Afrikaans deur *weet* of *ken* vertaal kan word, maar dat die twee werkwoorde nie uitruilbaar is nie, byvoorbeeld *I know that man/Ek weet daardie man* is nie korrek nie. 'n Program vir vertaling uit Afrikaans na Engels sou daarenteen die inligting bevat dat die twee Afrikaanse werkwoorde deur die een Engelse werkwoord vertaal kan word.

Die voordeel aan die direkte benadering is dat volledige ontleding van die brontaal by taalpare wat verwant is onnodig is. Ooreenkomste tussen die twee tale, bv. woordorde-patrone, hoef nie reëlsgewys in die program ingebou te word nie want dit word by wyse van verstek (*default*) gehandhaaf. Direkte stelsels steun besonder swaar op vertaalekwivalente wat in die woordeboek gestoor is en gewoon opgesoek en in die doeltaaltekste ingevoeg word, sonder 'n voorafgaande volledige ontleding van die taalstrukture. Die volgende sinne is byvoorbeeld woord-vir-woord

weergawes van mekaar: *The social history of this tribe is interesting/Die sosiale geskiedenis van hierdie stam is interessant.*

Die twee groot nadele aan hierdie benadering is dat woord-vir-woord vertaling slegs in 'n heel klein persentasie van sinne goeie vertalings oplewer, al is die betrokke tale baie eenders; tweedens, dit is verkwistend om vir elke taalpaar, en vir albei rigtings vir elke paar, afsonderlike programme (dus ook afsonderlike woordeboeke) te bou. In 'n multitalige situasie soos die EEG, waar 9 tale nou geakkommodeer moet word, sou 72 programme nodig wees. Dit is die enkel belangrikste rede waarom SYSTRAN, wat in wese 'n direkte stelsel is (alhoewel dit oor die jare heen aanpassings ondergaan het, vgl. 3.10), deur 'n nie-direkte stelsel vervang gaan word.

Vertaling via 'n intertaal

Die grondliggende gedagte in hierdie benadering is dat die brontaaltekste in 'n intertaal wat as brug kan dien na verskeie ander tale omgesit word. Die veronderstelling is dat 'n kunsmatige taal ontwerp kan word (of 'n bestaande kunsmatige taal soos Esperanto aangepas kan word) om 'n 'skarnier-funksie' tussen tale te vervul. Die intertaal sou 'n taalgemeenskaplike metasintaksis en universele stel semantiese primitiewe bevat, asook 'n meta- en intertalige leksikon.

Die teoretiese voordeel aan hierdie benadering is dat 'n multitalige vertaalstelsel daargestel, en na bykomende tale uitgebrei, sou kon word met 'n minimum van taalspesifieke programme (in teenstelling met die direkte benadering hierbo): vir elke taal is net twee modules nodig, dus 'n totaal van 18 vir die EEG se 9 tale.

Die eerste module is die omskakelingsprogram van die taalspesifieke na die intertaal, en die tweede is die generering van die taalspesifieke op grond van die intertaal. As voorbeeld: Die Engelse werkwoordvorme *ate, have eaten* en *were eating* sou in die intertaal al drie as een vorm verskyn: die intertaal se verlede-tydsvorm vir *eat*. Die grammatika van die intertaal sal 'n stel opsies met betrekking tot aspek (en ander werkwoordverbonde verskynsels) bevat. Vir die generering, vanuit die intertaal, van vertalings van sinne met bogenoemde Engelse vorme, sal die keuse van die opsies in die intertaal afhang van die doeltaal. Sodoende sal Afrikaans, met sy een verlede-tydsvorm, die intertaalvorm van die verlede tyd van *eat* omsit na *het geëet*. Daarenteen sal Frans, met sy verskeidenheid verledetydsvorme, 'n aantal (baie ingewikkelde) opsies vereis.

Voorgaande voorbeeld illustreer op 'n eenvoudige wyse juis waarom die intertalige benadering (nog) nie geslaagde stelsels oplewer het nie. Navorsing is deur verskeie groepe onderneem (Cambridge, Leningrad, Grenoble, Texas, vgl. Hutchinson 1986:55), maar geeneen het 'n volledige intertaalmodel kon voortbring nie. Afgesien van die omvang en



ingewikkeldheid van die taak, wat eintlik neerkom op die ondervanging van alle taalverskynsels in een kunsmatige taal, is dit te betwyfel of dit 'n funksioneel-doeltreffende benadering is. Om terug te kom na die voorbeeld hierbo: Vir die korrekte vertaling van die genoemde Engelse werkwoordtye na Frans is die teksaard en struktuur van groter belang as die Engelse vorms as sodanig. Met die vertaling eers na 'n intertaal, en vandaar na die doeltaal, val noodsaaklike inligting uit.

Opsommend kan daar gesê word dat die droom van 'n universele taal wat alle taalverskynsels ondervang, en sonder verskraling 'n brug tussen tale kan slaan, (nog) nie verwesenlik is nie.

#### Vertaling via oordragstelsels

Hierdie derde benadering skyn tans die mees belowende te wees (vgl. die bespreking van EUROTRA in 3.11). Dit is gegrond op ervaring van die voor- en nadele van die twee voorgaande benaderings:

- Die direkte metode het die voordeel gehad dat die raakvlakstruktuur tussen 'n taalpaar taalspesifiek was, dus kon daar op bepaalde ooreenkomste en verskille tussen die betrokke twee tale ingespeel word; by die intertaalbenadering is die raakvlakstruktuur universeel, dus verbeur die vertaalprogram die voordeel van bestaande brûe tussen bepaalde tale.
- Die grootste nadeel van die direkte benadering is dat dit die multitaligheid van stelsels beperk. Toe twee tale 'n paar jaar gelede by die bestaande sewe tale van die EEG bygekom ~~het, het dit meegebring dat die aantal programme wat vir die~~ Gemeenskap se vertaalstelsel nodig sou wees van 42 na 72 vermeerder het. Daar is tans sprake dat nog drie tale sal bykom, wat die totaal programme wat ontwikkel moet word - indien die direkte metode gevolg word - tot 132 vermeerder. Die intertaalbenadering, daarenteen, bring met die byvoeging van tale slegs twee programme per taal by, dus vir 'n twaalfstalige stelsel 'n totaal van 24 programme in plaas van 132.

Daar moet, tussen hakies, ook in gedagte gehou word dat benewens die groot taak van die ontwikkeling van 'n program, deurlopende instandhouding, hersiening, verfyning en uitbreiding (ook van die woordeboeke, wat in vertaalstelsels baie lywig is) moet geskied. Dit is dus noodsaaklik om die aantal programme sover moontlik te beperk.

Die oordragbenadering bou op die voor- en nadele van die ander twee soos volg. Ten einde die aantal programme te beperk, word daar vir elke taal slegs een ontledingsprogram en een genereringsprogram geskryf. Die

ontledingsprogramme is brontaalspesifiek en doeltaalonafhanklik, terwyl die genereringsprogramme brontaalonafhanklik en doeltaalspesifiek is.

Ten einde, aan die ander kant, 'n direkte en eenvoudige brug tussen taalpare te slaan, word daar dan vir elke taalpaar 'n paarspesifieke raakvlakstruktuur ontwikkel. Die invoer na die oordragstruktuur (d.w.s. die raakvlakstruktuur) is die analiesestruktuur van die brontaal, en die afvoer is die sintesestruktuur van die doeltaal.

Die stappe in die vertaalproses is dus soos volg:

- Ontleding van die bronteks, sonder inagneming van watter taal die doeltaal gaan wees.
- Inskakeling van die ontleding in 'n oordragstruktuur wat spesifiek vir die paring van die brontaal en 'n bepaalde doeltaal ontwerp is.
- Oorplasing van die afvoer van die oordragstruktuur na die genereringsmodule van die doeltaal.
- Generering van die vertaling.

Uit bostaande kan daar afgelei word dat die byvoeging van tale na die stelsel per nuwe taal die volgende sal behels:

- Die ontwerp van 'n analiese-module vir die taal.
- Die ontwerp van 'n sintese-module (wat in groot mate die omkering van die analiese-module beteken).
- Die ontwerp van 'n oordrag-module vir elke paring van die taal met ander tale.

Die oordrag-module word so klein en eenvoudig moontlik gehou, dus kan tale met 'n minimum uitwerking op die stelsel as geheel bygevoeg word. Hierdie oorweging word orals in die literatuur beklemtoon, want daar is aanduidings dat die Europese talenewerk tot 'n groter internasionale uitkringing van die netwerk, moontlik uiteindelik na 'n wêreldnetwerk, kan lei. Vir kleiner lande, wat veronagsaming van hul tale wil voorkom, is dit belangrik om in beplanning met hierdie aspek van ontwikkelings op die gebied rekening te hou.

In die afdelings wat volg, word afsonderlike stelsels bespreek. Soos reeds genoem in die inleiding tot hierdie verslag, is dit moeilik om die jongste inligting te bekom, behalwe deur besoeke en persoonlike gesprekke - wat wel sover moontlik gedoen is. Oor twee belangrike

aspekte van projekte is daar ongelukkig, maar begryplik, groot geheimhouding: fyn besonderhede oor die 'binnewerkinge' van die stelsels, en presiese syfers ten opsigte van navorsings- en ontwikkelingsbesteding.

### 3.2 TAUM-METEO

#### 3.2.1 Rekenaargesteunde vertaling in Kanada

Navorsing oor rekenaargesteunde vertaling is in 1965 in Kanada van stapel gestuur deurdat die Kanadese navorsingsraad projekte by die Universiteit van Saskatchewan, die taalnavorsingseenheid by Cambridge, en die Universiteit van Montreal geborg het. In November 1965 het die Kanadese navorsingsraad CETADOL (Centre de Traitement Automisé des Données Linguistiques) onder leiding van Guy Rondeau by Montreal gestig. Aanvanklik het die groep hoofsaaklik aan Engelse morfologie, sintaksis en grammatikale klassifisering, en aan algemene probleme met betrekking tot die verwerking van natuurlike taal, aandag geskenk. Sedert 1967, toe die Kuno-Woods-ontleder beskikbaar geword het, word navorsing oor vertaling gedoen. Die groep se naam is in 1970 verander na TAUM (Traduction Automatique de l'Université de Montréal), en dit het die ontwikkeling van 'n operasionele vertaalstelsel (Engels-Frans) ten doel gehad (Hutchins 1982:29).

#### 3.2.2 Oorsig van die METEO-projek

Teen 1973 het die TAUM-navorsingsgroep by die Universiteit van Montreal 'n eksperimentele masjienvertalingstelsel (Engels-Frans) wat die meeste van die kompleksiteite van Engelse invoersinne kon hanteer, gedemonstreer. Met behulp van hierdie stelsel kon die korrekte vertaling van individuele woorde en strukture om aanvaarbare Franse sinne te genereer, verkry word. Een van die grootste probleme was egter die omvang en kompleksiteit van die sintaktiese reëls wat nodig was. (Kittredge 1981:100).

In 1974 het die borge op TAUM druk uitgeoefen om die stelsel prakties toe te pas. TAUM het na 'n toepassing gesoek binne 'n beperkte domein waarin die grootte van die woordeskate, moontlike woordbetekenisse, en grammatikale reëls baie meer beperk as in natuurlike taal sou wees. As gevolg van die tweetaligheidsbeleid wat die Kanadese regering handhaaf, word weervoorspellings dwarsdeur Kanada in sowel Engels as Frans uitgesaai. Daarom het die Kanadese regering besluit om aan die TAUM-groep die taak op te dra om 'n outomatiese vertaalstelsel vir weerberigte te ontwikkel. Hierdie kontrak vir die ontwikkeling van die stelsel het gevolg op tien jaar se navorsing wat gefinansier is deur die nasionale navorsingsraad en die regering (Thouin 1982:39).

Na twee jaar van uitvoerbaarheidstudies en ontwerpnavorsing het die TAUM-groep in 1976 'n eerste weergawe van die stelsel aan die Kanadese regering gelewer. Dit het nog 'n jaar geduur voordat die stelsel by die

vertaalburo van die Kanadese meteorologiese sentrum geïmplementeer is. Sedert 1977 is die METEO-stelsel in daaglikse gebruik, en dit word slegs vir instandhoudingswerk aan die rekenaartoerusting onderbreek. Die stelsel is ten volle in die Kanadese netwerk vir die transmissie van weerberigte geïntegreer. Die tekste word dus elektronies in- en afgevoer.

### 3.2.3 Taalkundige beginsels van die METEO-program

Die beperkte woordeskat en stereotipe sintaksis van meteorologiese berigte het die ontwerpers van TAUM in staat gestel om die basiese TAUM-stelsel in 'n groot mate te vereenvoudig. Die belangrikste opsigte waarin dit verskil van die TAUM-prototipe is die uitskakeling van 'n oordragkomponent. Die meeste van die verwerking wat gewoonlik tydens oordrag plaasvind, is in die analisekomponent geïnkorporeer. Dit kom daarop neer dat METEO 'n direkte stelsel is, alhoewel dit 'n onafhankliker brontaalanalise as die meeste ander direkte stelsels het.

Die stelsel is verder vereenvoudig deurdat morfologiese analise nie meer voor die woordeboeksoektog gedoen word nie, aangesien daar so min variante vorms van Engelse woorde in hierdie subtaal voorkom. Die indeling van die woordeboeke is ook verander. In plaas van drie woordeboeke, was een voldoende om die Franse ekwivalente van Engelse uitdrukkings en Franse morfologiese data weer te gee. Hierdie woordeboeke is slegs tydens die analisefases geraadpleeg. Die beperkte morfologiese module wat vir weerberigte taal nodig is, word in METEO met 'n 'semantiese grammatika' gegrond op begripkategorieë soos *atmosferiese toestand, toestandsverandering, plek, tyd*, aangevul. Die woordeboekinskrywings bevat beide sintaktiese en semantiese kategorie-inligting, terwyl boomstruktuurreëls die moontlike woordkombinasies binne elke konstituent bepaal (Kittredge 1981:101).

### 3.2.4 Die METEO-stelsel in daaglikse gebruik

Gewoonlik verwerk die METEO-vertaalstelsel sowat 30 000 woorde per dag. Die grootste hoeveelheid invoersinne word deur die stelsel vertaal, en die menslike vertaler doen net af en toe 'n steekproef. Een van die grootste voordele van hierdie stelsel is die betroubaarheid van die vertalings. Dit is die gevolg van 'n baie konserwatiewe benadering in die skryf van taalkundige reëls vir die analiseprogram. Om die analisefase te slaag, moet al die woorde van die sin in die woordeboek gevind word en elke grammatikale struktuur moet herkenbaar wees. Anders word geen boomstruktuur aan die sin toegeken nie. Geen metaforiese uitbreidings word toegelaat nie, en die program 'raai' nie wanneer iets nie 'verstaan' word nie. Hierdie 'probleemsinne' word aan die vertaler oorgelaat (Kittredge 1981:101).

### 3.2.5 Vooruitsigte

Die operasionele sukses wat oor die afgelope jare met die METEO-stelsel behaal is, het bewys dat rekenaargesteunde vertaling wel binne beperkte subtale moontlik is. Alhoewel die stelsel in omvang beperk is, hanteer dit meer as tien miljoen woorde jaarliks. Dit is sover bekend die eerste vertaalprogram wat vertalings wat nie geredigeer hoef te word nie, voortbring. Die feit dat die stelsel in 'n uitgebreide telekommunikasienetwerk geïntegreer is, maak dit nog merkwaardiger. TAUM-METEO is ook waarskynlik die eerste masjienvertalingstelsel waarin vertalers by al die fases betrokke was.

METEO is egter nie uitbreibaar nie, en teen 1981 het dit geblyk dat die interne beperkings in METEO se teoretiese basis verdere verbeterings onmoontlik maak (Slocum 1984:12). Dit sal nie ekonomies wees om die stelsel uit te brei sodat dit minder algemene sinne kan hanteer nie (Thouin 1982:44).

---

### 3.3-TAUM-AVIATION

#### 3.3.1 Oorsig

Die TAUM-AVIATION-projek is 'n voorbeeld van 'n vertaalprogram waarin die teks nie beperk word met betrekking tot sintaksis nie, maar die soort teks wat vertaal word, is wel vooraf bepaal.

~~Die sukses wat met die TAUM-METEO-stelsel behaal is, het in 1977 daartoe gelei dat 'n meer ambisieuse projek, vir die vertaling van handleidings vir die instandhouding van lugvaartuie uit Engels in Frans, aangepak is. Die vertaalburo van die Kanadese Staat het aan TAUM die opdrag gegee om binne drie jaar 'n stelsel vir die vertaling van die handleidings vir die CP-140 kuspatrollie-lugvaartuie te ontwikkel, wat die vertaling van ongeveer 90 miljoen woorde sou behels (Hutchins 1986:225,231).~~

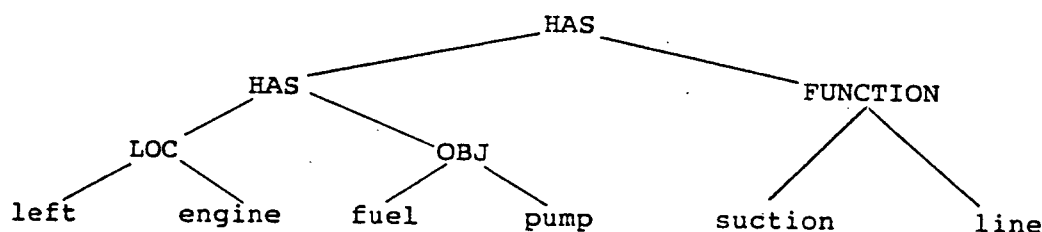
Die lugvaarthandleidingssubtaal is een van die ingewikkeldste tale waarin outomatiese verwerking nog aangepak is (Kittredge 1981:102). Toe die AVIATION-projek aangepak is, is die navorsing vir byna twee jaar vertraag as gevolg van organisatoriese probleme. Nogtans is daar in 1979 'n suksesvolle openbare demonstrasie van die stelsel gehou, en toe die sperdatum in 1980 bereik is, is 'n operasionele prototipe aan die borg gelewer. Hierdie prototipe is geïmplementeer op 'n Cyber 173. 'n Woordeboek bestaande uit 7 000 inskrywings is in die program ingebou.

### 3.3.2 Kenmerke van die subtaal van die handleidings

Die onderskeidende kenmerke van die subtaal is die volgende:

- Dit bevat geen vraag- en uitroepsinne nie,
- bevele bevat geen lidwoorde, koppelwerkwoorde of anafore nie (byvoorbeeld *Remove used filter and discard; Check reservoir full*),
- dit bevat baie komplekse naamwoordstukke (byvoorbeeld *hydraulic ground test stand pressure and return line filters*).

Om die komplekse naamwoordstukke te kan hanteer, is daar voorgestel dat AVIATION semantiese kategorisering van naamwoordkonstituente insluit, sodat byvoorbeeld die volgende analise verkry sal word vir *left engine fuel pump suction line*:



Hierin word die funksionele (FUNCTION), lokatiewe (LOC), possessiewe (HAS), en objek-(OBJ)relasies gespesifiseer. Dié idee kon egter nie binne die sperdatum vir die projek geïmplementeer word nie.

### 3.3.3 Taalkundige beginsels in die vertaalprogram

Die AVIATION-stelsel verdeel elke Engelse teks in segmente wat verwerkingseenhede genoem word. Die verwerkingseenheid kan 'n volledige sin, die titel van 'n afdeling, of die etiket in 'n diagram wees. Elke verwerkingseenheid word apart hanteer, en dit volg outomaties die volgende stappe:

#### (a) Voorredigering en morfologiese analise

Ná die aanvanklike formatering, word die verboë vorme van woorde (bv. *locks*) vervang deur hul moontlike woordeboekrepresentasies (bv. die meervoud van die naamwoord *lock*, of derdepersoonsenkelvoud van die werkwoord *lock*). Die woordeboekrepresentasie bevat alle inligting oor die sintaktiese en semantiese potensiaal van die woorde.

(b) Sintaktiese analise

'n Sintaktiese verwerker ontleed die reeks woorde en hul woordeboekinligting van links na regs om 'n enkele sinstruktuur wat met al die moontlike gebruike van elke woord in die sin, versoenbaar sal wees te vind. 'n Boomstruktuur word dan saamgestel.

(c) Leksikale oordrag

Elke Engelse woord wat in die boomstruktuur voorkom, word deur die toepaslike Franse woord vervang. Hierdie vervanging geskied deur komplekse reëls wat die woord se onmiddellike omgewing in die boomstruktuur nagaan. Wanneer die korrekte vertaling vir die Engelse werkwoord *replace* byvoorbeeld gekies moet word, word daar in die leksikale oordrag nagegaan of die direkte objek van die werkwoord 'n naamwoord is wat deur die semantiese kenmerk *easily replaceable part* gemerk is. Indien dit só is, word die woord met die Franse *remplacer* (om te vervang), vertaal, maar indien nie, word dit met *remettre en place* ('om terug te plaas'). vertaal.

(d) Sintaktiese oordrag

In sommige gevalle moet die sintaktiese struktuur van die Engelse sin groot veranderings ondergaan voordat 'n aanvaarbare Franse sin geproduseer kan word. Só kan 'n Engelse passiefsin (*This experiment was performed several times*) soms ooreenstem met 'n Franse aktiefsin (*On a fait cette expérience plusieurs fois*). Nadat die nodige strukturele veranderings deur die oordragreëls aangebring is, kom die nuwe boomstruktuur (met Franse woorde) met 'n aanvaarbare Franse sin ooreen.

(e) Sintaktiese sintese

Tydens hierdie stap word sekere transformasies in 'n lineêre reeks Franse woordvorme omskep. Die woordvorme is nog nie verbuig nie, maar bevat gekodeerde morfologiese inligting om die korrekte verboë vorms te bepaal.

(f) Morfologiese sintese

Die string woordvorme word omskep in 'n leesbare Franse sin deur middel van reëls wat die korrekte verboë vorm van die gekodeerde inligting en basisvorm aflei. (Kittredge 1981:102.)

### 3.3.4 Resultate

Voorlopige resultate wat met AVIATION verkry is, het baie belowend gelyk. Die volgende is 'n voorbeeld van die program se werkverrigting:

Les clapets de décharge incorporés sont champignon, sont rappelés par ressort à la position fermée. Une pression de 3450 psi s'exercant sur le clapet-champignon est suffisante pour vaincre la force de rappel du ressort et le clapet-champignon se déplacera de son siège en couteau.

Oorspronklike Engelse teks:

The in-line relief of valves are poppet-type, springloaded to the closed position. A pressure of 3450 psi impinging on the poppet is sufficient to overcome the opposing force, and the poppet will move from its knife-edge seat.

Die enigste verandering wat met die hand aangebring moes word, is die vervanging van *sont* met *du type* wat voorkom tussen *incorporés* en *champignon*, en die vervanging van *s'écartera* met *déplacera* in die laaste klous.

### 3.3.5 Vooruitsigte

Alhoewel die kwaliteit van die vertalings wat verkry is, redelik goed was, kon die prototipestelsel nie daarin slaag om enige afvoer vir 20% tot 40% van die invoersinne te verkry nie. In 1980 is daar in 'n onafhanklike evaluering van die projek bevind dat daar weinig vooruitsigte vir die daarstel van 'n koste-effektiewe produksiestelsel bestaan het.

Die finale afkeuring van die stelsel was nie soseer op grond van die kwaliteit van die afvoer wat verkry is nie, maar eerder op grond van die hoeveelheid tyd en geld wat nog benodig sou word om die stelsel uit te brei sodat dit operasioneel kon word (Hutchins 1986:232). Daarby was die koste verbonde aan die uitbreiding van die program se woordeboeke ook uitermate hoog. Daar is byvoorbeeld gevind dat een inskrywing in die AVIATION-stelsel 40 dollar bedra het teenoor die minder as 5 dollar wat nodig was vir baie ander stelsels (Pigott 1983:3).

Die TAUM-navorsers was van mening dat die evaluering te gou gedoen is, aangesien die stelsel nog in 'n vroeë stadium van ontwikkeling was en snelle vordering binnekort gemaak sou word. In 'n interne ondersoek is bevind dat daar vir 70% van die mislukkings bekende oplossings bestaan het, en hierdie probleme sou binne 'n redelike tyd opgelos kon word. TAUM-AVIATION het egter nie 'n kans gehad om tot sy volle potensiaal te



ontwikkel nie. In Language Monthly January 1985:13-16 word daarop gewys dat die TAUM-AVIATION-projek 'n navorsingsprojek was, maar dit is 'n ontwikkelingsprojek geëvalueer.

In 1981 is 'n verdere uitvoerbaarheidstudie onderneem waarin TAUM-AVIATION met die Engels-Franse Systran-stelsel wat ontwikkel is vir die Kommissie van die Europese Gemeenskap, (vgl. 3.10) en met die interaktiewe stelsels ALPS (vgl. 3.4) en WIEDNER (vgl. 3.5) vergelyk is. Alhoewel die bedryfskoste en die koste verbonde aan woordeboekbywerking (*updating*) minder vir hierdie drie stelsels as vir TAUM was, is daar bevind dat die kwaliteit van dié drie stelsels baie laer was. Geeneen van hierdie stelsels is dus vir aankoop en gereelde gebruik in die Kanadese vertaaldienste aanbeveel nie (Hutchins 1986:232-233).

Die belangrikste en deurslaggewendste beswaar teen die TAUM-AVIATION-stelsel was dus die hoë koste verbonde aan die saamstel van die woordeboek. Die subtaalbenadering wat hoë kwaliteit vertaling binne 'n beperkte domein ten doel het, vereis sorgvuldige semantiese klassifisering van die woordeskat, en tydrawende analise van leksikale items in semantiese en sintaktiese kontekste. Sonder so 'n tydsame en moeisame analise kan dubbelsinnighede nie opgelos word nie. Daar kan natuurlik beweer word dat daar nie van enige outomatiese stelsel verwag kan word om alle dubbelsinnighede op te los nie, maar tog was een van die oogmerke van TAUM-AVIATION juis om hoë kwaliteit resultate te lewer. As gevolg van die ondersoek wat gedoen is, is die finansiële steun aan die projek gestaak, wat toe gelei het tot die finale ontbinding van die TAUM-AVIATION-projek in September 1981 (Hutchins 1986:223,225).

### 3.4 ALPS

#### 3.4.1 Oorsig

In 1971 is 'n projek aan die Brigham Young-Universiteit om Mormoonse kerktekste uit Engels in verskillende tale te vertaal, van stapel gestuur. Die uiteindelijke oogmerk was om 'n ten volle outomatiese masjienvertalingstelsel daar te stel, maar aanvanklik sou dit 'n rekenaargesteuende stelsel wees, met wisselwerking tussen die vertaler en die program.

Die BYU (Brigham Young-Universiteit)-projek het nooit 'n operasionele sisteem opgelewer nie. Omstreeks 1977 het 'n groep programmeerders BYU verlaat en by Weidner Communications, Inc. aangesluit om die ten volle outomatiese, direkte Weidner-vertaalstelsel te ontwikkel. Kort daarna, in 1980, het die meeste van die oorblywende BYU-projeklede ook die projek verlaat en Automated Language Processing Systems (ALPS) gestig, waar hulle voortgegaan het met die kommersiële ontwikkeling van die BYU-rekenaargesteuende stelsel (Slocum 1984:8-9).

Die maatskappy het geen finansiële verbintenisse met die Brigham Young-Universiteit of the Mormoonse kerk nie, en ALPS se hoofkantoor is in Provo, Utah. In 1983 is die ALPS-stelsel bemark, en in 1984 het die maatskappy 'n Europese kantoor in Neuchâtel, Switserland, geopen (Hutchins 1986:302).

Die ALPS-stelsel is ontwikkel met die doel om rekenaarhulp (hoofsaaklik teksredigering en woordeboekfasiliteite) aan vertalers te verskaf. Die stelsel is in alle opsigte interaktief en poog glad nie om onafhanklik te vertaal nie. Dit verskaf eerder aan die vertaler 'n stel programmeringshulpmiddele om baie van die take verbonde aan vertaling te outomatiseer. Aangesien ALPS die hulpmiddele wat by BYU ontwikkel is, oorgeneem het, was dieselfde taalpare aanvanklik beskikbaar, naamlik Engels na Frans, Duits, Portugees, en Spaans. Sedertdien is ander taalpare soos Engels-Italiaans en Engels-Arabies ontwikkel (Rouvray en Wilkinson 1984:20).

### 3.4.2 Die ALPS-program

ALPS bestaan uit 'n multitalige woordverwerker waarvan die werking bepaal word deur 'n spyskaartverwerker (*menu-driven processor*) wat aan 'n drukker verbind is. Die gebruiker kan hierdie fasiliteit met behulp van die volgende uitbrei:

#### 'n Selektiewe woordeboek-naslaanfasiliteit

Hierdeur word die gebruiker se eie woordeboeke (wat eentalig, tweetalig of multitalig kan wees) toeganklik.

#### 'n Outomatiese woordeboek-naslaanfasiliteit

Hiermee soek die rekenaar na brontaalwoorde, en vertoon doeltaal-ekwivalente. Die woorde wat gevind word, word met hul kontekste, gelys terwyl die woorde wat nie gevind word nie, vir woordeboekbywerking gelys word. Geen sintaktiese verwerking vind plaas nie, maar wel outomatiese morfologiese analise om die stamme te bepaal.

#### 'n Rekenaarvertaalstelsel

Hierdie stelsel skep 'n woordeboek van brontaaltekswoorde, stel vrae oor dubbelsinnighede in spelling, oor leksikale keuses en oor sintaktiese probleme. Dit sal, byvoorbeeld, die moontlike doeltaalekwivalente vir 'n brontaalwoord vertoon, en die vertaler vra om 'n keuse daaruit te maak, of as die woord nie in die woordeboek voorkom nie, die vertaler vra om een te verskaf.

Die voordeel verbonde aan die interaktiewe vertaalstelsels lê juis in die feit dat die samewerking tussen die menslike vertaler en die rekenaar ten beste benut word. Die rekenaar word slegs gebruik vir dit wat hy goed kan doen, naamlik die beskikbaarstelling van woordeskat, die verskaffing van morfologiese ontledings en die lewering van eenvoudige ruwe ontledings. Menslike vaardighede word weer gebruik in die

ingewikkelder prosesse soos semantiese ontledings, die oplos van dubbelsinnighede en die selektering van die toepaslike uitdrukking wanneer daar 'n keuse in die aantal moontlike vertalings is. In die ALPS-stelsel val die klem op die noue samewerking tussen vertalers en 'n gebruikersvriendelike interaktiewe stelsel wat die gebruiker kan aanpas na gelang van spesifieke behoeftes en organisatoriese vereistes (byvoorbeeld deur die ontwikkeling van persoonlike programme en die gebruik van ander rekenaartoerusting soos optiese karakterlesers). Die vertalings verskyn sin-vir-sin op die skerm sodat veranderings deurgaans aangebring kan word. Hierdie veranderings word deur die rekenaar gestoor en hergebruik wanneer dit benodig word.

Die kompleksiteit van ALPS se leksikon-inligting word tot die minimum beperk. Volgens die ontwerpers van interaktiewe stelsels soos ALPS is daar geen nut daarin om kunsmatige intelligensie-tipe representasies en gesofistikeerde semantiese representasies in die woordeboeke te inkorporeer nie, aangesien die vertaler tog self die beste bron van hierdie soort kennis is. Dit het tot gevolg dat die inligting wat in die woordeboek vervat is weinig meer is as data oor morfologiese variante, grammatikale kategorieë, en potensiële doeltaalekwivalente.

---

Besonderhede van ALPS se sintaktiese verwerker is om kommersiële redes nie beskikbaar nie, maar dit sluit natuurlik verbuiging, vervoeging, ooreenstemming en woordordening in. Dit blyk egter dat die BYU-verbindingsgrammatika (*junction grammar*) nie meer 'n belangrike rol speel in die stelsel nie.

### 3.4.3 Beskikbare programmatuur

---

**Vertaalprogrammatuur: CAT**

CAT ('Computer Assisted Translation') is 'n pakket met drie programme wat op hoofraam of mikrorekenaar werk: TSS ('Translation Support System'), Autoterm en Transactive. TSS is die program wat hierbo beskryf is.

AutoTerm en TransActive is die twee soorte geoutomatiseerde vertaalhulpmiddels wat saam met TSS beskikbaar is. Nie-tegniese dokumente soos briewe, brosjures, bemarkingsstukke en uitvoerdokumente is vir vertaling deur AutoTerm ideaal aangesien dit terminologie verskaf en algehele herstrukturering van sinne toelaat. Met behulp van AutoTerm kan woordeboeksoektogte outomaties uitgevoer word, en die vertaalekwivalente vir spesifieke terminologie word op die skerm vertoon. Die vertaler kan dan sy aandag by die styl en bewoording van die vertaling bepaal. Die terminologie wat gebruik word, sal ook konsekwent wees. AutoTerm is geskik vir vertaling uit die volgende brontale in enige taal wat gebruik maak van die Romeinse alfabet: Frans, Duits, Spaans, Italiaans en Engels.

TransActive is weer geskik vir die vertaling van dokumente van tegniese en herhalende aard, byvoorbeeld dienshandleidings, tegniese handleidings en gebruikersgidse as gevolg van hul eenvoudige grammatikas, herhaling van woordeskat en eenvoudige sinstruktuur. TransActive identifiseer outomaties sintaktiese, grammatikale en leksikale dubbelsinnighede, en skets 'n vertaling wat die vertaler kan redigeer. Taalpare wat beskikbaar is vir vertaling met TransActive is Engels na Frans, Duits, Spaans, Italiaans, asook Frans na Engels.

TSS is 'n moderne en gebruikersvriendelike stelsel. Aangesien terminologie gestandaardiseer word, het dit hoë kwaliteit vertalings tot gevolg.

Skryfprogrammatuur: CAW

CAW ('Computer Assisted Writing Software') bied 'n keuse tussen MacProof (wat op MacIntosh uitgevoer word) en DRF ('Document Revision Facility') wat op hoofraam uitgevoer word. MacProof kan ook ingeskakel word in 'n netwerk. MacProof is 'n teksredigerings- en hersieningsprogrammeringspakket. Spelling, leestekens en hoofletters kan hiermee nagegaan word. Dit sluit in 'n woordeboek van 93 000 woorde (wat deur die gebruiker aangepas/verander kan word), en hulpmiddels vir die ontleding van sins- en paragraafstruktuur. MacProof herken probleme soos die volgende: verwarrende woorde, seksistiese woorde, rassistiese woorde, nominaliserings, vae woorde en clichés. Aangesien dit terugvoer verskaf oor taalgebruik, styl en struktuur, kan MacProof gebruik word om skryfstyl en ander skryftegnieke te verbeter.

Programmeringshulpmiddels

Hierdie hulpmiddels is spesifiek ontwerp vir gebruik in natuurliketaalverwerking ('Natural Language Processing' (NLP)), en staan bekend as PeriPhrase. PeriPhrase verskaf 'n reëlgebaseerde omgewing vir die ontwikkeling van taalkundige programmatuur. Hierdie programmatuur is geskik vir verskeie take wat gesofistikeerde taalkundige verwerking vereis, byvoorbeeld vir korpus-analise en terminologie-uitligting. PeriPhrase is ook ideaal vir die ontwikkeling van natuurliketaal-koppelvlakke vir bedryfstelsels. Die databasis-subtale kan beperk of uitgebrei word na gelang van die toepassing. PeriPhrase kan ook vir onderrigtoepassings gebruik word. Dit kan gesofistikeerde taalkundige ontleders vir byvoorbeeld die ontleding van studente se werkstukke genereer. Tans word PeriPhrase vir die volgende tale ontwikkel: Engels, Frans, Duits, Italiaans, Nederlands, Deens, Sweeds, Noors, Russies en Spaans.

Elektroniese woordeboeke

Tweetalige, eentalige, tegniese en algemene woordeboeke is op disket beskikbaar. Hulle kan binne die CAT- en CAW-programmaturomgewing gebruik word.

Met die klem op interaksie, vereenvoudigde grammatikale data in die leksikons, en elementêre sintaktiese analise, kom dit voor asof ALPS nie veel meer as 'n gesofistikeerde teksverwerker met onmiddellike toegang tot tweetalige woordeboeke is nie. Ongetwyfeld lewer die stelsel egter goeie resultate en dit werk hoër produktiwiteit in die hand (Hutchins 1986:303).

### 3.5 WEIDNER COMMUNICATIONS CORPORATION (WCC)

#### 3.5.1 Oorsig

Die Weidner Communications Corporation is in 1977 in Provo, Utah, deur 'n groep, voorheen van die Brigham Young-Universiteit, gestig. Die wedywering wat tussen Weidner en die vertaalinstituut van die BYU ontstaan het, het oorgegaan in 'n wedywering tussen Weidner en ALPS, wat ook 'n uitvloeisel van die projek aan die BYU was (Slocum 1984:12). Alhoewel die projek van die BYU deur lede van die Mormoonse Kerk gestig is, is Weidner vandag 'n privaatonderneming wat geen finansiële bande met òf die Universiteit òf die Kerk het nie. In 1982 is Weidner se hoofkwartier na Northbrook, Illinois, naby Chicago verskuif terwyl die navorsings- en ontwikkelingsentrum steeds in Provo gesetel is.

~~Weidner is vandag kommersieel een van die mees suksesvolle vertaalstelsels ter wêreld. Die volgende taalpare is beskikbaar: Engels na Frans, Spaans, Duits, Portugees, Italiaans en Arabies; Frans na Engels en Spaans; Spaans na Engels; Duits na Frans en Engels; en Japanees na Engels (Rouvray en Wilkinson 1984:20).~~

In 1980 het Weidner 'n Engels-Franse stelsel aan Mitel in Kanada gelewer, en in dieselfde jaar 'n Engels-Spaanse stelsel aan die Siemens-korporasie (VSA). Mitel het in 1981 Weidner se Engels-Spaanse en Engels-Duitse stelsels aangekoop, en Bravis ('n vertaaldienstburo in Japan) het die Weidner Engels-Spaanse en Spaans-Engelse stelsels aangekoop (Slocum 1984:12). Gedurende 1981 het Weidner regte tot die Engels-Arabiese stelsel aan OMNITRANS verkoop (Slocum 1984:13).

Alhoewel die Weidnerstelsel gedurende vertaling ten volle outomaties is, geskied ander funksies soos die leksikale pre-analise van tekste, en die saamstel van woordeboeke, interaktief. Die woordverwerkingsprogrammatuur word met eksterne dienste geïntegreer, byvoorbeeld 'n laserdrukker. Die geformateerde dokument wat in die stelsel ingevoer word, word met formateringskodes gekodeer sodat 'n geformateerde vertaling verkry word.

Weidner bemark die stelsel op 'n draaisleutel-('turnkey') basis teen 'n vasgestelde prys vir 'n taalpaar. Daar is nou twee weergawes beskikbaar: MacroCAT wat op 'n DEC PDP-11 of DEC Vaxrekenaar loop, en MicroCAT wat op 'n IBM PC mikrorekenaar loop. Albei pakkette sluit rekenaars, terminale, die vertaalprogrammatuur, die teksverwerkingsprogrammatuur, 'n drukker, en 'n woordeboek met 'n woordeskat van 8 000 stamme en 3 000 uitdrukkings in. MacroCAT vertaal teen 'n tempo van ongeveer 1 400 woorde per uur, en MicroCAT teen ongeveer 1 600 per uur. Vir albei weergawes is daar 'n instandhoudingsooreenkoms. Weidner het ook sy stelsel op die IBM persoonlike rekenaar geïmplementeer om sodoende sy afhanklikheid van die PDP-11 te verminder (Slocum 1984:13).

### 3.5.2 Die Weidner-vertaalprogram

Presiese besonderhede oor die werking van die Weidner-stelsel is om kommersiële redes nie beskikbaar nie, maar dit is bekend dat die stelsel die volgende stappe uitvoer:

(a) Die invoerhanteerder

Die invoerhanteerder kontroleer die invoer van die brontaaltekst en verdeel die teks in woorde en sinne.

(b) Die woordeboeknaslaan-fasiliteit

Die woordeboeknaslaan-fasiliteit voer morfologiese analises uit (deur die verwydering van uitgange) en voer 'n woordeboeksoektog uit.

(c) Die idioomomsetter

Die idioomomsetter herken idiomatiese uitdrukkings wat deur die vertalers in die woordeboek ingevoer is.

(d) Homograafoplossing

Hier word die konteks geraadpleeg soos aangedui in die woordeboek. Voorbeelde van homograwe is *pers* (kleur) en *pers* (druk), en *kopie* (afskrif) en *kopie* (winkoop).

(e) Sintaktiese analise

Tydens hierdie stap word naamwoord- en werkwoordfrases sowel as die verhoudings tussen hulle geïdentifiseer om sodoende 'n soort frase-struktuur-analise te vorm.

(f) Herordening van die teks

Die finale stappe in oordrag en sintese sluit die herordening van teks volgens doeltaalstrukture, werkwoordvervoeging en ooreenstemming, dit wil sê die morfologiese sintese van doeltaal werkwoord- en naamwoordvorme en roetines vir die invoeging van doeltaallidwoorde en -voorsetsels.

(g) Die afvoerhanteerder

Die afvoerhanteerder herorganiseer die teks in 'n formaat wat ooreenstem met die oorspronklike brontaaltekst.

(h) Die teksbywerker

Die doeltaaltekst word dan deur die vertaler met behulp van die teksbywerker geredigeer. Die teksbywerker is 'n woordverwerkingsprogram met verskeie redigeringsfasiliteite, insluitende 'n sinoniemwoordeboek (Hutchins 1986:304).

### 3.5.3 Weidner-woordeboeke

In hierdie afdeling word die Mitel-weergawe van Weidner se woordeboekroetines beskryf. Hundt (1982) bied verdere besonderhede.

By Mitel ontvang die vertaler die brontekst in 'n pre-publikasieformaat. Die teks word eers deur 'n woordeskatsoektogprosedure gevoer om woordeskatleemtes in die woordeboek te identifiseer en aan te vul. Die aanvullingsproses is redelik ingewikkeld, soos blyk uit die volgende beskrywing van naamwoordinskrywings.

Eerstens vra die program aan die vertaler of die item 'n homograaf is. Byvoorbeeld, die woord *house* word in Spaans vertaal as *casa* of *albergar*, afhangend van die konteks (*a house/to house*). Die vertaler sleutel die korrekte item in en annoteer dit dan na aanleiding van die volgende vrae wat op die skerm verskyn:

- ~~- Rededeel? (Werkwoord, naamwoord, voegwoord, ens.)~~
- Geslag? (Manlik, vroulik, onsydig)
- Getal? (Meervoud, enkelvoud, taalverskille: sommige naamwoorde is in een taal altyd meervoudig en in ander altyd enkelvoudig, en omgekeerd)
- Semantiese merkers? (Menslik, groep, liggaamsdeel, abstrak, dierlik, konkreet, ens.)
- Eienaam? Pleknaam?

Interessantheidshalwe kan daar hier genoem word dat een spesifieke kenmerk van Engels deurlopend vir die rekenaar vertaalprobleme skep, dus moet dit sover moontlik in die voorbereidingstadium onderskep word, naamlik woorde wat op *-ing* uitgaan:

- *Hunting lions are dangerous*
- *Hunting lions is dangerous*
- *Hunting is dangerous*
- *Hunting they go every weekend*
- *Hunting they adore*
- *Hunting they were when the storm broke*

- *Hunting around they found some firewood*

Bostaande voorbeelde toon dat *-ing-*vorms 'n verskeidenheid funksies wat moeilik te onderskei is, vervul. Die Afrikaanse woord *jagtend(e)* is byvoorbeeld hoogstens in een van bostaande sinne gepas.

Die woordeboekinskrywings van werkwoorde word op min of meer dieselfde manier as die vir naamwoorde gedoen, na aanleiding van vroe soos die volgende:

- Agentskap? (Geen, rigting, plek, ens.)
- Transitief?
- Naamval van direkte voorwerp? (Akkusatief, datief, genitief)
- Indirekte voorwerp?
- Wederkerige vorm?
- Hulpwerkwoord? (Bv. *have* of *be*)
- Karakterlengte van skeibare prefiks?
- Klas? (Bv. sterk/swak, reëlmatig/onreëlmatig)

Op soortgelyke wyse word byvoeglike naamwoorde en bywoorde ontleed. Die hantering van idiome bied bepaalde probleme. Die idioom *to rain cats and dogs*, wat in Duits vertaal word as *in Strömen regnen*, word as voorbeeld gebruik (Slocum 1984). Eerstens word al die Engelse woorde afsonderlik in die woordeboek ingesleutel. Dan word die idioom as geheel ingesleutel, as een woordeboekitem waaraan 'n rededeelklas toegeken word. In hierdie geval word die idioom as werkwoord geklassifiseer en, na aanleiding van die program se vroe oor werkwoorde, geannoteer. Vervolgens vra die program na 'n woord in die doeltaal wat ooreenkom met 'n hoofwoord in die idioom, in hierdie geval *regnen*. Hierdie woord sal as die sleutelwoord vir kruisverwysing tussen die betrokke idioom in die twee tale dien. Die vertaler sleutel dan die doeltaalidioom in, en voeg inligting soos verbuigingsreëls by.

Soos in die meeste vertaalprogramme, is die omvang van die idioomwoordeboek groot. Die rede daarvoor is dat tale vele uitdrukkings bevat wat moontlik nie dadelik opval as idiome nie, maar wat wel as idiome hanteer moet word omdat dit gaan om figuurlike gebruik. As voorbeeld: *to ring a bell* het beide die letterlike betekenis van *to make a bell sound* en die figuurlike betekenis van *to rouse a response*. Die uitdrukking word in die idioomwoordeboek met slegs die figuurlike betekenis opgeneem.

Oorsigtelik kan die Weidner-stelsel as 'n interessante model van werksverdeling tussen die vertaler en die rekenaar gekenmerk word. Die interaktiewe benadering word in die voorbereidingsfase gevolg, waar die vertaler se gespesialiseerde taalkennis vir invoer na die woordeboek benut word. Daarna voer die program sonder hulp die vertaling uit. In die afrondingsfase is dit weer die vertaler wat die hoofrol speel, terwyl die rekenaar net as (voortreflik toegeruste) woordverwerker dien.



### 3.5.4 Vooruitsigte

World Communications Corporation, die Weidner-ouermaatskappy, brei steeds die stelsel na nuwe tale uit, soms in samewerking met ander instansies. So, byvoorbeeld, sal daar binnekort 'n Engels-Noorse program, wat in samewerking met die Universiteit van Bergen ontwikkel is, op die mark verskyn. Die program, ENTRA (English Norwegian TRANslation), is ontwerp vir die vertaling van tegniese en wetenskaplike tekste. Die Universiteit het die projek op grond van sy vorige ervaring met die vertaling van tekste in verband met die olie-nywerheid (die projek PETTRA) onderneem. ENTRA is op Weidner se Engels-Duitse program gebaseer (*Language Monthly* April 1987:6).

Volgens Hundt (1982) kan daar nie in die vooruitsig gestel word dat die Weidner-stelsel ooit meer as 'n hulpprogram vir vertaling sal wees nie. Die vertaalkomponent, wat op die direkte benadering gegrond is, is nie baie gesofistikeerd nie. Analise word byvoorbeeld tot die naamwoord- en werkwoordstukvlak beperk, met die gevolg dat kontekste wat op ander vlakke beskikbaar is, nie in ag geneem word nie.

~~Nietemin toon die klaarblyklike kommersiële sukses van die stelsel dat dit 'n nuttige hulpmiddel is. Daar word beweer dat redigeerders teen 'n tempo van 600 to 1000 woorde per uur vertalings kan afhandel (Hutchins 1986:305). Namate 'n gebruiker sy woordeboeke na eie behoefte uitbrei, word die stelsel al hoe meer koste-effektief. Soortgelyk neem doeltreffendheid toe namate vertalers ervaring opdoen in die gebruik van die hulpfasiliteite rondom die vertaalkomponent. Soos daar by herhaling elders in hierdie verslag genoem word, is die rekenaarbehandigtheid van vertalers en redigeerders 'n belangrike faktor in die effektiewe implementering van rekenaargesteuende vertaling.~~

Die volgende is voorbeelde van ongeredigeerde Weidner-vertalings:

#### Frans-Engels:

This function assigns ports to the exchange and to the workstations by the parameters introduction. She allows also to put these circuits into service or out of service and authorises the ports of the workstations to carry out some functions or all functions. At last she allows to assign periodically recording equipment for the output of the message.

#### Spaans-Engels:

Because of it these functions were considered as of prime importance during the development of the digital exchange ITT 1240. Taking advantage of to the maximum the inherent advantages of the digital technology and of the distributed control, the majority of the facilities of operation and maintenance has been incorporated directly to the equipment.

## Duits-Engels:

The automatic switching on of the operating voltage takes place after connection the supply voltage of the exchange after short voltage failures, after voltage fades and after the failure of a supply voltage branch. The manual switching on of a supply voltage branch is caused by pressing of the key 'voltage converter' at the signal field (MAP).

### 3.6 TITUS

Soos TAUM-METEO is TITUS 'n vertaalstelsel wat spesifiek vir die vertaling van tekste met 'n beperkte woordeskat en sintaksis ontwerp is. In die vroeë sewentigerjare is die ontwikkeling van die TITUS-stelsel deur die Institut Textile de France geborg, vir die vertaling van opsommings vanuit en in vier tale, naamlik Frans, Engels, Duits en Spaans. Die tekstiedokumente word in 'n standaard formaat wat die 'kanonieke dokumentasietaal' genoem word, opgestel. Hierdie 'taal' is spesiaal vir tekstiel-ekserpte ontwerp, en bestaan in wese uit 'n vasgestelde woordeskat en sestien sinspatrone. Die brontaaldokument word dan via 'n intertaal wat uit eenvoudige kodes bestaan, vertaal.

Opsommings vir TITUS word as reekse klouse geformuleer, wat uit toegelate sleutelwoorde (indeksterme) en 'n vaste stel funksiewoorde (voorsetsels, voegwoorde negatiewe, hulpwerkwoorde en modale woorde) bestaan. Die naamwoorde kan samestellings wees, die werkwoorde kan verbuig word, en die adjektiewe kan eenvoudig of saamgestel wees of hulle kan verbale afleidings wees.

Die dokumente word in-baan ingesleutel en nagegaan. Die stelsel is met ander woorde interaktief. Die menslike vertaler moet dele van die teks herskryf as dit nie volgens die TITUS-woordeskat of -sintaksis geformuleer is nie. Soos genoem, word opsommings in 'n intertaal gestoor, waaruit hulle in een of meer van die vier tale vertaal word. Die TITUS-sintaksis laat 'n reeks van tot vier klouse toe, naamlik 'n hoofklous gevolg deur ondergeskikte of relatiewe klouse. Elke klous bestaan uit die volgende reeks: subjek, voorsetselstuk, verbale groep, werkwoordkomplement en voorsetselstuk. Diakritiese tekens word gebruik om die begin van die ondergeskikte klouse aan te dui sodat daar nie verwarring tussen die voeg- en bywoordelike gebruike van woorde soos *before*, *after*, *since*, kan ontstaan nie.

In 1973 is TITUS 2 vrygestel, en die nuutste weergawe van TITUS, TITUS 4, is in 1980 geïmplementeer (Slocum 1984:11).

Die sukses van die beperkte stelsels soos METEO en TITUS kan grotendeels toegeskryf word aan die feit dat daar vroeg al beseft is dat vertaalstelsels meer suksesvol aangewend kan word indien die semantiese veld en die sintaktiese kompleksiteit van die tekste wat vertaal moet word, presies gespesifiseer word.

### 3.7 CULT

#### 3.7.1 Oorsig

In 1969 is daar by die Chinese University of Hong Kong met navorsing vir 'n Chinees-Engelse vertaalprogram begin. 'n Prototipe stelsel vir die vertaling van Chinese wiskundige tekste in Engels is in 1972 op 'n ICL 1904A-rekenaar geïmplementeer. In die daaropvolgende jaar het dit die CULT-stelsel geword, en met die borgskap van die Asia Foundation (later die Rockefeller Brothers Fund) is die stelsel vanaf 1975 vir die gereelde produksie van die vertaling van die wiskundige tydskrif Acta Mathematica Sinica in Engels gebruik. Vanaf 1976 is CULT ook vir die vertaling van die wetenskaplike tydskrif Acta Physica Sinica in Engels gebruik, maar dit is later gestaak (Slocum 1984:14).

#### 3.7.2 Die CULT-program

CULT is 'n interaktiewe stelsel wat op die direkte vertaalstrategie gebaseer is. Die sinne word een vir een in 'n reeks deurgange (*passes*), ontleed en vertaal. Die ontleder doen weinig meer as oppervlaktgroepeerings van grammatikale kategorieë soos naamwoord- en werkwoordgroepe en modifiseerders. ~~Ná elke deurgang word 'n doel van die vertaalproses uitgevoer.~~ Gedurende analise vind menslike tussenkoms vir die oplossing van homografe en sintaktiese dubbelsinnighede plaas, en gedurende sintese word Engelse lidwoorde, en die bepaling van werkwoordtipe en wyses (*modes*) ingevoeg. Indien die ontleding misluk, voeg die redigeerder inskrywings tot die leksikon by. Alhoewel analise en sintese geïntegreer is, kan fases in die vertalingsproses tog onderskei word:

- ~~(a) Voorbereiding van die bronteks.~~
- (b) Invoer (via Chinese sleutelbord).
- (c) Leksikale analise (d.w.s. brontaalwoordeboek-naslaan, insluitende in-baan bywerking van die woordeboek).
- (d) Sintaktiese en semantiese analise (bepaling van grammatikale kategorieë; identifisering van naamwoordgroepe, werkwoordgroepe, modifiseringsgroepe, en hulpwerkwoorde).
- (e) Relatiewe ordeningsanalise (herkenning van relasies tussen naamwoordgroepe, werkwoordgroepe en modifiseerders).
- (f) Analise van doeltaalekwivalente (oplos van veelvoudige betekenis).
- (g) Afvoer (doeltaalwoordeboek-naslaan).
- (h) Die verfyning van afvoer (herformatering; invoeging van wiskundige formules).  
(Hutchins 1986:298).

CULT is oorspronklik in bondelmodus (*batch mode*), geïmplementeer, en daar is probleme met die invoer van Chinese karakters ondervind. Dit is deur die gebruik van standaard telegraafkodes opgelos, en die invoerdata is op kaarte gepons. Dit was 'n moeisame proses en ook onderhewig aan baie foute. Om hierdie rede is baie aandag gegee aan die ontwikkeling van 'n Chinese sleutelbord wat in-baan invoer en 'n werklik interaktiewe stelsel tot gevolg sou hê. In 1978 is CULT op woordverwerkingstoerusting (PDP 11/34) vir in-baan data-invoering en redigering (voorredigering sowel as naredigering) geïmplementeer (Slocum 1984:14).

In die verdere ontwikkeling van CULT is daar hoofsaaklik op die ontwikkeling van die woordeboeke gekonsentreer. Later is daar ook eksperimente gedoen om CULT uit te brei om ook voorsiening te maak vir Engels-Chinese vertaling, en dit het die herorganisasie van woordeboekstrukture tot gevolg gehad (Hutchins 1986:298).

Die aanvanklike doel met CULT was om naredigering grotendeels deur middel van voorredigering van die invoer uit te skakel. Dit wil voorkom asof daar wel in hierdie doel geslaag is, aangesien die hoeveelheid naredigering wat benodig word tot die minimum beperk is. Die stelsel kan, byvoorbeeld, nie sekere formules en figure vertaal nie, en dit moet dan deur die vertaler vertaal word. As interaktiewe stelsel kan CULT as 'n baanbreker gesien word, want met die inwerkingtreding daarvan was in-baanstelsels nog baie ongewoon. CULT is ook ontwerp om spesifieke soorte tekste te vertaal. Dit plaas dus 'n beperking op die gebruik van die stelsel, met die gevolg dat CULT maar baie primitief voorkom in vergelyking met latere interaktiewe stelsels soos ALPS. Daar moet egter in gedagte gehou word dat die ontwerpers van CULT slegs in die lewering van vertalings vir 'n spesifieke doel belanggestel het, en nie in die lewering van hoë-kwaliteit rekenaargesteunde vertaling nie. In hierdie opsig kan dus gesê word dat CULT suksesvol is: die stelsel werk goed en die tydskrif verkoop goed.

### 3.8 LOGOS

#### 3.8.1 Oorsig

Die LOGOS-ontwikkelingskorporasie is in 1969 in die lewe geroep. Dit is deur die Amerikaanse lugmag geborg en moes navorsing vir die ontwikkeling van 'n Engels-Viëtnameese stelsel voortsit. In 1970 is die stelsel LOGOS I, gebaseer op 'n korpus van slegs 1300 woorde, gedemonstreer, en daar is bevind dat dit vir die Amerikaanse lugmag vir die vertaling van opleidingshandleidings bruikbaar sou kon wees. Verdere ontwikkeling daarvan is aanbeveel, maar daar is gewys op die feit dat die stelsel nog nie goed genoeg was vir grootskaalse vertaling van tegniese materiaal nie. Dit is ook benadruk dat LOGOS-vertaalde stukke aansienlike naredigering vereis.

### 3.8.2 Die LOGOS-program

In LOGOS I is die brontaalanalise en doeltaalsintese heeltemal van mekaar geskei. Ook die rekenaarprogrammeringsprosesse en taalkundige data en reëls het onafhanklik gefunksioneer. Volgens die ontwerpers, was die programme heeltemal taal-onafhanklik alhoewel die prosedures self en die woordeboeke spesifiek vir een bepaalde brontaal-doeltaal-paar, ontwerp is, naamlik Engels en Viëtnameses. Daar kan gesê word dat LOGOS I 'n hibriede direkte oordrag-stelsel is waarin tweetalige woordeboeke aan apart gedefinieerde fases van analise, oordrag en sintese gekoppel is.

Dit was glad nie nodig om die teks vooraf te redigeer nie. Die program het outomaties woord- en sinsgrense opgespoor en ook parentetiese reekse geïdentifiseer. Twee woordeboeke is gebruik, naamlik een met hoë frekwensie woorde wat in die rekenaargeheue gestoor is, en 'n groot woordeboek (wat aanvanklik sowat 100 000 inskrywings bevat het) op band. Basisvorms met aparte uitgange is in die inskrywings gebruik.

Ná die woordeboeksoektog voer die sintaktiese ontleder die volgende funksies uit (eers van regs na links in die sin en daarna van links na regs): die oplos van woordklas-dubbelsinnighede, die uitken van naamwoordstukke, onderwerpe, predikate, kongruensie, en koördinasie, en die oplos van sekere semantiese dubbelsinnighede. Geen poging om volledige stukbou-analises uit te voer, is aangewend nie. Die strukture is net so ver geanaliseer as wat nodig was vir die verkryging van genoeg inligting vir die omskakeling van Engelse strukture na aanvaarbare Viëtnamese sintaktiese vorme. In die analise-roetines is veral op die sintese van sintaktiese en semantiese inligting klem gelê.

~~In die oordragfase is die Engelse strukture na tipiese Viëtnamese strukture verander. Die volgende dien as voorbeelde hiervan: die adjektief-naamwoordgroepe (in Engels) is omgeskakel na Viëtnamese naamwoord-adjektiefgroepe, komplekse werkwoorde is vereenvoudig, genitiewe naamwoorde het of-vorms geword (bv. *pilot's compartment: compartment of the pilot*), en passiewe konstruksies is verander in aktiewe konstruksies, aangesien Viëtnameses geen passiewe vorm het nie (bv. *wires can be disconnected upon removal of clip/ (you) can disconnect wires when you have removed clip*). In die finale fase word die Engelse leksikale vorms deur Viëtnamese vorms vervang.~~

Die resultaat word sin-vir-sin gelewer. Die ontwerpers van die sintaksis-georiënteerde stelsel was van die beperkings van hierdie stelsel bewus. Dit is juis hierdie beperkings wat naredigering noodsaaklik gemaak het.

In 1972 is LOGOS I deeglik op basis van 'n vertaalde lugmaghandleiding geëvalueer. Die resultate wat hiermee verkry is, was nie baie bemoedigend nie. Nie lank daarna was daar nie meer 'n behoefte aan Viëtnamese lugmaghandleidings nie. Intussen is daar gewerk aan LOGOS II en teen 1973 is 'n Engels-Russiese weergawe, LOGOS III, ontwikkel.

Later is daar aan stelsels vir die vertaling van Engels in Frans, Spaans en Duits gewerk. Die LOGOS-korporasie het in die middel sewentigerjare 'n kontrak ontvang van die Iranese regering vir die ontwikkeling van 'n multitalige stelsel vir vertaling in Persies. Aanvanklik is slegs die taalpaar Engels-Persies ontwikkel, maar later is dit uitgebrei om ook Russies, Duits, Frans, Spaans en Arabies as brontale in te sluit.

Hierna is daar egter nie verder van LOGOS gehoor nie totdat die 'LOGOS Intelligent Translation System' in 1982 van stapel gestuur is. Hierdie stelsel was 'n kommersiële produk wat spesifiek vir vertaalagentskappe bedoel was. Voorheen was die LOGOS-stelsel slegs vir regerings- en militêre doeleindes bedoel. Die eerste kommersiële program was 'n Duits-Engelse stelsel. Later is 'n Engels-Duitse weergawe ontwikkel, en daar is ook begin met werk aan 'n Franse weergawe.

LOGOS ontwikkel hul stelsels op hoofraam en ontwerp hul programme om rekenaar-onafhanklik te wees. Tans is die Duits-Engelse stelsel egter net op die WANG OIS ('Office Information System')-woordverwerker en die WANG VS-mikrorekenaar beskikbaar.

LOGOS stel vertalers in staat om in 'n woordverwerkingsomgewing woordeboeke aan te pas by hul eie behoeftes, en die vertaalprogram te gebruik en deurlopend te hersien. Hierdie take kan almal op dieselfde rekenaar- (of woordverwerker-)terminaal uitgevoer word. Daar word beweer dat meer as 20 000 woorde binne 24 uur deur die LOGOS-stelsel vertaal kan word.

Die Duits-Engelse stelsel bevat 'n basiese woordeboek van meer as 100 000 inskrywings. Spesifieke terminologie kan met behulp van 'n interaktiewe stelsel vir die saamstel van woordeboeke (ALEX) hierby gevoeg word. Hierdie stelsel stel vrae oor die sintaktiese en semantiese eienskappe van die woorde wat ingesleutel word, en verseker ook dat die kodering van hierdie woorde versoenbaar is met inskrywings wat reeds in die woordeboek is. Gebruikers kan ook spesifieke vakgebiede vir spesifieke inskrywings spesifiseer. Sodoende word verseker dat poliseme so ver moontlik binne die konteks van 'n spesifieke vakgebied vertaal word.

Basies is die eienskappe van die vroeëre LOGOS-weergawe in die Duits-Engelse stelsel behou, alhoewel dit nou 'n ware oordragstelsel geword het met taal-onafhanklike programmering en skeiding van bron- en doeltaal-datalêers. Daar is baie aandag aan die integrering van semantiese inligting in die vertaalalgoritme geskenk. Gedurende die sewentigerjare is 'n semantiese abstrakte taal ('Semantic Abstraction Language') deur LOGOS ontwikkel en verfyn. Dit hou in dat semantiese inligting tydens elke deel van die analise teenwoordig is, en dat hierdie inligting beskikbaar is (soos wat dit benodig word) vir die oplos van dubbelsinnighede op leksikale, sintaktiese of semantiese vlak.

Uit die praktiese ervaring met LOGOS, blyk dat dit 'n heel bruikbare stelsel is. Een agentskap wat LOGOS in Desember 1982 geïnstalleer het, het in 1983 bereken dat die kwaliteit van die LOGOS-produkte sonder enige naredigering vir 60% tot 80% van die teks voldoende vir voorlopige inligtingsdoeleindes is. Naredigering word vir die lewering van hoë-kwaliteit tekste vereis. Vertaalagentskappe in Duitsland vind dat die Duits-Engelse stelsel hul produktiwiteit met tot 100% (en ten minste 30%) kan verhoog. Terselfdertyd ondervind hulle ook verbeterings in die kwaliteit en konsekwentheid van hul vertalings. Daar is gevind dat LOGOS die geskikste vir hoogs gespesialiseerde tekste is, waarvoor die agentskap self spesifieke terminologie insleutel. Dit is die mins bevredigend vir algemene korrespondensie en tekste met redelik min tegniese terme, en vir interdisiplinêre tekste.

### 3.8.3 Vooruitsigte

Die LOGOS-stelsel word deur die navorsing wat die maatskappy uitvoer, asook op grond van terugvoer van die gebruikers van die stelsel, voortdurend verbeter. Soos die ander verskaffers van kommersieel beskikbare stelsels, ALPS en Weidner, volg die LOGOS-maatskappy die filosofie dat vertaalprogramme bloot hulpmiddels is wat verbeter moet word, maar intussen in die bedryfswêreld bruikbaar is. Die filosofie wat hierteenoor staan - en wat kenmerkend van universiteitsondernemings op die gebied is, vergelyk byvoorbeeld SUSY 3.12 - is dat daar ten opsigte van die taalkundige 'intelligensie' van stelsels hoër gemik moet word. Die basiese struktuur van die kommersiële vertaalprogramme is sodanig oppervlakkig (ten opsigte van taalkundige ontleding) dat verbeterings aan die programme nie ingrypend aan die doeltreffendheid van die stelsel kan verander nie.

~~Die volgende is 'n ongeredigeerde LOGOS-vertaling van Duits na Engels.~~

#### Oorspronklike teks:

Eine feldweise Steuerung des Auftretens der verschiedenen Felber ist glücklicherweise nicht nötig. Es gibt Gruppen von Feldern, die stets gemeinsam auftreten oder nicht auftreten und die wir deshalb zu sog. 'Feldgruppen' zusammenfassen können. Tatsächlich genügt es, die oben erwähnte Steuerung (in Abhängigkeit von Konto und Buchschlüssel) auf Feldgruppen anstatt auf Felder zu beziehen, was eine erhebliche Vereinfachung bedeutet.

#### Ongeredigeerde vertaling:

A field-by-field control of the occurrence of the different fields is luckily not necessary. There are groups of fields which occur always commonly, or occur not and which we sucked therefore too. Field-groups are able to combine. In fact, it suffices to relate to the control mentioned above (depending on account and posting key) to field-groups instead of on fields what a considerable simplification means.

### 3.9 DIE CARNEGIE-MELLON-PROJEK

#### 3.9.1 Aqtergrond

Jaime Carbonell het onlangs sy kennis-gebaseerde masjienvertalingsprojek (wat hy by Yale begin het) verder begin ontwikkel. Sy nuwe projek word by die Carnegie-Mellon-Universiteit by Pittsburgh aangepak. Die 'ou' Yale-projek word egter nou as grondslag vir die uitvoer van 'n bewerking in 'n operatiewe modus gebruik. Daar word beplan om hierdie benadering met die interaktiewe stelsel van Masuru Tomita wat deur Nishida en Doshita by die Kyota-Universiteit ontwikkel is, te integreer. Die uiteindelige doel is 'n Engels-Japannese stelsel wat vir die saamstel en gelyktydige vertaling van kort persoonlike tekste in interaktiewe modus gebruik sal word.

#### 3.9.2 Kriteria vir die stelsel

Met die Carnegie-Mellon-projek word 'n vertaalstelsel in die vooruitsig gestel wat in staat sal wees om veel meer te verrig as wat moontlik is met die huidige tegnologie. Die ontwerp van die nuwe stelsel sal aan die volgende kriteria moet voldoen:

- (a) Akkuraatheid  
Vertalings in goed gedefinieerde gebiede moet foutvry wees. Met akkuraatheid word behoud van betekenis en styl in die vertaalproses bedoel.
- (b) Intydse vermoë  
Dit behoort nie langer te neem om 'n brief of teleksboodskap te vertaal as wat dit neem om dit in die stelsel in te sleutel nie.
- (c) Veeltaligheid  
Vertalings in baie verskillende tale uit een brontaaltekst moet moontlik wees.
- (d) Interaktiwiteit  
Die masjienvertalingstelsel moet interaktief en gebruikersvriendelik wees. Dit moet vrae aan die gebruiker kan stel wanneer 'n woord nie verstaan word of 'n paragraaf in die teks nie goed gevorm is nie.
- (e) Onafhanklikheid  
Dit moet nie nodig wees vir 'n vertaler om die vertaalde teks na te gaan en te korrigeer nie.
- (f) Lae-koste apparatuur  
Die stelsel moet op lae-koste apparatuur bedryf kan word. In beginsel moet dit moontlik wees om die stelsel op rekenars soos die kragtige 32-bis-mikrorekenaars, wat eers oor twee of drie jaar algemeen in 'n kantooromgewing beskikbaar sal wees, te gebruik.



(g) Uitbreibaarheid

Dit moet moontlik wees om die leksikon, grammatika en kennisdatabasis van elke stelsel te kan uitbrei sonder dat dit enige komplikasies in die stelsel veroorsaak. Die stelsel moet só opgestel word dat die gebruiker self die stelsel volgens sy spesifieke behoeftes kan uitbrei.

(h) Integreerbaarheid

Die stelsel moet uiteindelik op dieselfde rekenaar geïntegreer kan word, met woordverwerkingsprogrammatuur, elektroniese pos, en toepassingsprogramme.

(i) Veeldoeligheid

Die natuurlike taal interpreteerder en die taalgenereerder moet ook vir ander doeleindes gebruik kan word, byvoorbeeld vir veeltalige databasisnavrae, of vir geoutomatiseerde verskaffing (in een of meer doeltale) van opsommings van lang tekste.

Om bogenoemde oogmerke te bereik, moet verskeie nuwe tegnologiese rigtings gevolg word. Vir Carbonell en Tomita (1987) is veral die volgende van belang:

(a) Modulariteit

Sintaktiese en semantiese kennis wat vir akkurate vertaling benodig word, word in aparte goed gedefinieerde kennisbronne geplaas. Modulariteit hou veral twee voordele in: namate nuwe semantiese domeine tot die stelsel gevoeg word (soos banktelemekse, reisnavrae, aankope, bestellings, ens.), bly alle vorige taalkundige kennis toepaslik, en hoef dit nie herhaal te word nie. Soos wat nuwe tale tot die stelsel gevoeg word, kan alle vorige kennis oor die semantiese domein in die vertaling van tekste in enige bekende domein in die nuwe taal gebruik word.

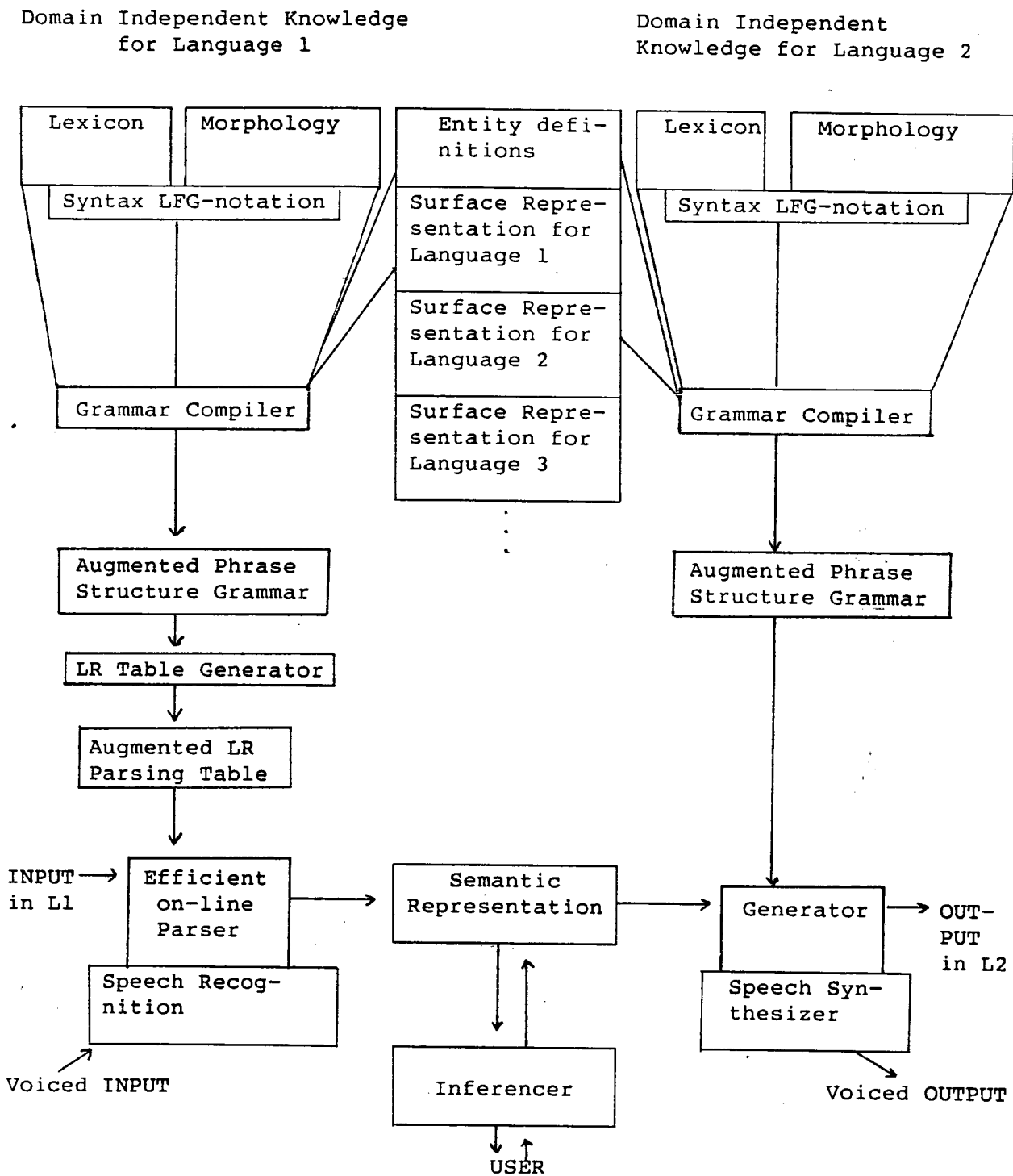
(b) Effektiewe ontleding

Om intydse vertaling op lae-koste apparatuur te bereik, is besonder effektiewe metodes vir ontleding en generering nodig.

### 3.9.3 Oorsig van die stelsel

Figuur 3.1 is 'n voorstelling van die stelsel. Modulariteit van domein-spesifieke semantiese kennis en domein-onafhanklike (maar taal-spesifieke) sintaktiese kennis word in die stelsel beoog. Semantiese kennis word as 'n entiteit-georiënteerde grammatika voorgestel, en sintaktiese kennis as 'n funksionele grammatika. Albei hierdie grammatikas verteenwoordig onlangse vooruitgang op die gebied van die rekenaarlinguistiek. Hierdie twee grammatikas word dan in 'n enkele groot grammatika vir effektiwiteit saamgevoeg. Dié saamgestelde grammatika word vooraf in 'n ontledingstabel saamgestel. Met behulp van hierdie ontledingstabel word invoersinne op 'n baie effektiewe manier ontled.

Figuur 3.1 Stelselstructuur



(Carbonell en Tomita 1987:87)

### 3.9.4 Entiteit-georiënteerde grammatika-formalisme

Die entiteit-georiënteerde benadering tot beperkte-domein-ontleding is in 1984 deur Hayes voorgestel as 'n metode vir die organiseer van semantiese en sintaktiese inligting oor alle domeinkonsepte rondom 'n versameling entiteite (bv. objekte, gebeure, bevele en toestande) wat deur 'n spesifieke stelsel herken moet kan word. 'n Entiteit-definisie bevat inligting oor die interne struktuur van die entiteite, oor relasies met ander entiteite, oor die manier waarop die entiteite in die invoer van die natuurlike taal gemanifesteer word, en oor die ooreenkomste tussen die interne struktuur en die veelvoudige oppervlakovorme van elke entiteit.

Om die entiteit-georiënteerde grammatika te illustreer, kan die gesprek tussen 'n dokter en pasiënt (en in besonder die pasiënt se aanvanklike klagte oor 'n pyn) as voorbeeld geneem word. Entiteite wat in hierdie domein ingesluit word, is 'n gebeure-entiteit, pasiënt-klagte-handeling, en objek-entiteite soos pyn, menslik, en so meer. Sinne met verskillende oppervlakovorme kan herken word as voorbeelde van 'n entiteit:

*I have a head ache.*  
*I have a burning pain in the chest.*  
*I don't feel any pain.*  
*Did you have a dull ache in your head?*

Die definisie van die entiteit lyk soos volg:

[Entity Name: PATIENT-COMPLAINT-ACT

Type: STRUCTURED

Agent: -HUMAN ; Semantic restriction on the agent.

Pain: PAIN

SurfaceRepresentation:

[Syntax type: SENTENTIAL

Head: (have | feel)

Subj: (+ Agent)

;+ Agent refers to the semantic case 'Agent'

DObj: (+ Pain ) ]]

;+ Pain refers to the semantic case 'Pain'

```

[Entity Name: PAIN
Type: STRUCTURED
Location: BODY-PART           ;Semantic restriction on the location.
PainKind: PAIN-FEEL
SurfaceRepresentation:
[Syntax Type: NOUNPHRASE
Head: (pain | ache)
PP: ({Prep: in
      Comp: (+Location)})
Adj: (]AdjPhrase: (sharp | stabbing | acute)
      Component: PainKind
      Value: SHARP ]
     [AdjPhrase: burning
      Component: PainKind
      Value: BURN ])

```

Die finale semantiese representasie van die sin *I have a dull ache in my chest* word geïllustreer deur die volgende:

```

[cfname: MEDICAL-COMPLAINT-ACT
type: SENTENTIAL
agent: ]cfname: PERSON
      name: *speaker*} ;Whoever is the 'I' who has the chest ache.
pain: [cfname: PAIN
      location: [cfname: BODY-PART
                name: CHEST ]
      pain-kind: DULL]

```

Die oppervlakrepresentasie-dele van 'n entiteit lei die ontleding deur die sintaktiese strukture te verbind met die semantiese dele van die entiteit. As gevolg van die ontleding van 'n sin, word 'n samestelling van die semantiese dele van die entiteite gelewer.

Die grootste voordeel van die entiteit-georiënteerde grammatika formalisme is die feit dat die subtaal baie duidelik gedefinieer kan word.

Aangesien alle inligting wat verband hou met 'n entiteit, saamgegroeper word, kan duidelik gesien word wanneer 'n definisie volledig is. Die gevolg van enige toevoeging tot die definisie of weglating uit die definisie is ook baie duidelik. Op dieselfde manier word die oppervlak sintaktiese (en strukturele) inligting saamgegroeper, en daarom kan die oppervlakinligting op 'n duidelike en samehangende manier na die struktuur verwys.

Omdat die semantiese dele van die entiteite heeltemal taalonafhanklik is, kan een stel entiteit-definisies gebruik word vir alle tale. Daar word slegs vereis dat elke entiteit verskillende oppervlakvorme kan hê, een of meer vir elke taal. Op hierdie manier kan verseker word dat die semantiese domeine van elke taal op dieselfde manier hanteer word.

Nog 'n voordeel van die entiteit-georiënteerde benadering is kragtigheid in die hantering van buite-grammatikale invoer.

### 3.9.5 Funksionele grammatika formalisme

Behalwe die entiteit-georiënteerde grammatika vir semantiese kennis, word daar in die Carnegie-Mellon-projek ook gebruik gemaak van die funksionele grammatika vir die definiëring van die taal-spesifieke (maar domein-onafhanklike) sintaksis.

Daar is twee hoofvoordele verbonde aan die gebruik van die funksionele grammatika formalisme in praktiese masjienvertalingstelsels:

#### (a) Omkeerbaarheid

'n Stelsel wat streng binne die funksionele grammatika formalisme geïmplementeer word, sal omkeerbaar wees. Daar hoef dus nie twee aparte grammatikas opgestel te word vir ontleding en generering nie. Die grammatika (in ontledingsmodus) word dus toegepas vir die brontaal, en 'n ander grammatika (in die genereringsmodus) vir die doeltaal. Dieselfde twee grammatikas kan in die omgekeerde volgorde gebruik word om in die ander rigting te vertaal.

#### (b) Bekendheid

Funksionele grammatika formalismes soos die 'Functional Unifications Grammar' en die 'Lexical Functional Grammar' is bekend aan rekenaarlinguïste. Hulle sal bestaande formalismes bo stelsel-spesifieke grammatika formalismes verkies.

~~Daar is ook 'n nadeel verbonde aan die implementering van funksionele grammatika formalismes. Dit blyk dat funksionele grammatika formalismes glad nie so doeltreffend is soos 'Augmented Transition Network'- of stukbou-grammatikas nie. In die Carnegie-Mellon-projek word hierdie probleem opgelos deur 'n saamgestelde grammatika binne die funksionele grammatika formalisme in 'n uitgebreide konteksrye grammatika.~~

### 3.9.6 Vooruitsigte

Afgesien van die vertaalprogram en die vraag-antwoordstelsel waarna hierbo verwys is, onderneem die Carnegie-Mellon-span 'n groot projek waarin 'n kennisgebaseerde stelsel wat die domein van handleidings vir mikrorekenaars dek, opgebou word (die XCALIBUR-projek). Daarby word 'n beheersisteen vir kennisgebaseerde stelsels ontwikkel. Hierdie projekte is boustene vir 'n beoogde kennisgebaseerde vertaalstelsel wat moontlik in 1989 die lig sal sien.

### 3.10 SYSTRAN

#### 3.10.1 Agtergrond

Die eerste grootskaalse vertaalprogram is in Amerika aan die Universiteit van Georgetown in die vyftiger- en sestigerjare onderneem. Die doel daarmee was 'n stelsel vir die vertaling van Russiese wetenskaplike en tegniese dokumente in Engels, hoofsaaklik vir militêre doeleindes. Een van die Georgetownspan, Peter Toma, het in 1968 sy eie maatskappy gestig en 'n eie stelsel ontwikkel, SYSTRAN. Oor die volgende dekade het Toma ooreenkomste gesluit met verskillende groot gebruikers vir en deur wie die stelsel toe na eie behoeftes aangepas en ontwikkel is. Gevolglik bestaan daar 'n groot aantal verskillende weergawes van SYSTRAN, wat nietemin basies dieselfde struktuur het, soos blyk uit die verrigtinge van die World Systran Conference wat in 1986 deur SYSTRAN-gebruikers belê is (Pigott 1987:7).

Die hoofgebruikers van die SYSTRAN-stelsel is die Europese Gemeenskapskommissie (EEG), die Amerikaanse lugmag (USAF), General Motors, Xerox, Wang Laboratories, die Amerikaanse atoomkragskommissie, die National Aeronautic and Space Administration (NASA), die European Atomic Energy Authority (EURATOM), 'n aantal Japannese maatskappye, die Centre National de Recherches Scientifiques (CNRS), NAVO, die United Nations Industrial Development Organization (UNIDO), die internasionale atoomenergiesentrum in Wenen, en die Duitse kernnavorsingsentrum in Karlsruhe. Burodiensies soos CSATA (Italië), ECAT (Luxemburg), Gachot (Frankryke) en ORDA-B (België) besit ook regte vir die gebruik van SYSTRAN. Verder is daar talle kleiner instansies wat die stelsel gebruik.

Die taalpare wat ontwikkel is, of word, is

- Engels na Frans, Italiaans, Spaans, Duits, Nederlands, Portugees, Japanees, Arabies en Russies;
- Frans na Engels, Duits, Nederlands;
- Duits na Engels, Frans, Spaans, Italiaans;
- Portugees-Engels;
- Italiaans-Engels;
- Chinees-Engels;
- Russies-Engels (Pigott 1987).

Aangesien die groot gebruikers SYSTRAN na eie behoefte aangepas het, bestaan daar vir sekere taalpare, soos Engels-Frans, meer as een program. Byvoorbeeld, die EEG gebruik die stelsel om brontekste van allerlei aard, en met betreklik min beperkings ten opsigte van vakgebiede, te vertaal. Derhalwe is sy programme, en die woordeboeke wat die programme ondersteun, op 'n breë spektrum van taalgebruik geskoei, terwyl gebruikers wat op een vakgebied toespits, soos die atoomkragsentra, hul programme grotendeels beperk tot die betrokke vaktaal.

Hieronder word die SYSTRAN-stelsel aan die hand van die EEG se weergawe daarvan verder beskryf.

### 3.10.2 Oorsig van die EEG-stelsel

Die stelsel bevat drie hoofkomponente: die beheerprogram van ongeveer 100 000 saamstellingsreëls, die taalkundige program ook met ongeveer 100 000 makroreëls per taalpaar, en die woordeboeke wat vir sommige tale tot 150 000 inskrywings bevat (Pigott 1987).

Hier word daar nie ingegaan op besonderhede van die beheerprogram nie, behalwe om te noem dat dit IBM-gebaseer is en dat die program op 'n verskeidenheid van apparatuur gebruik word, onder andere op 'n Amdahl-hoofraam en op mikrorekenaars van verskillende herkoms.

Die taalkundige program word in Figuur 3.2 skematies voorgestel. Die opeenvolgende stappe behels die volgende:

(a) Woordidentifisering

Aan die hand eers-van-die-frekwensiewoordeboek en dan die hoofstamwoordeboek word elke woord nageslaan ten einde brontaalinligting (sintakties en semanties) aan elke item te heg en moontlike ekwivalente in die doeltaal te vind.

(b) Homograwe

Woorde wat deur die woordeboeke as homograwe aangedui is, word op grond van rededeel- en kontekstuele inligting ondersoek en geklassifiseer (bv. *light* kan as naamwoord, werkwoord of adjektief vertolk word afhangend van die sintaktiese plasing daarvan).

(c) Saamgestelde naamwoorde

'n Bepaalde semantiekwoordeboek verskaf riglyne met betrekking tot moontlike samestellings (bv. *blast furnace* bevat items wat afgebakende samestellingsmoontlikhede toon).

(d) Frase-identifisering

Frasegroepe word deur die uitkenning van puntuasietekens, voegwoorde, relatiewe voornaamwoorde en ander klousgrensaanduiders geïdentifiseer.

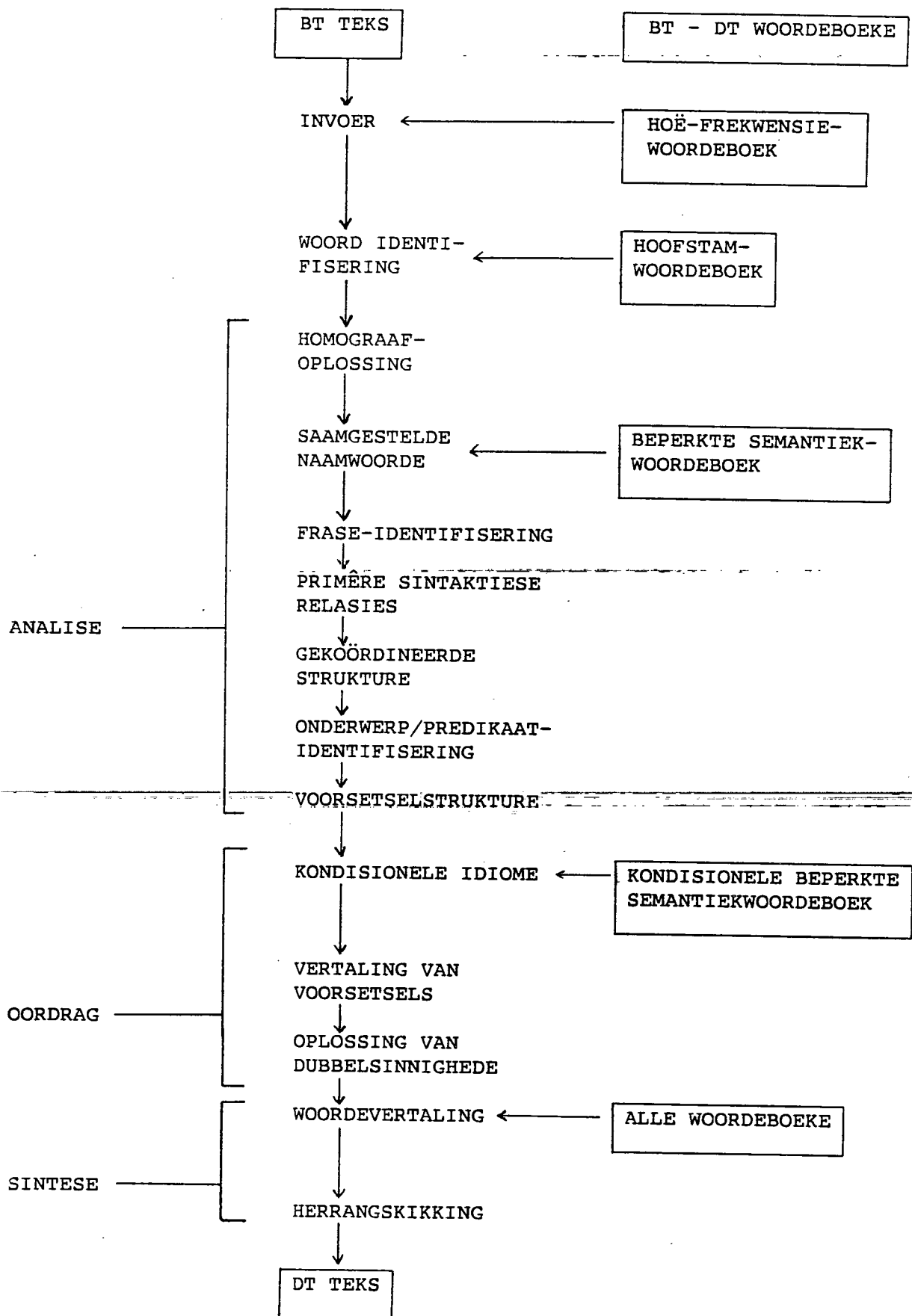
(e) Primêre sintaktiese verhoudings

Afhanklikheidsverhoudings word deur die merkers van kongruensie, regering en apposisie uitgeken.

- (f) Gekoördineerde strukture  
Binne die geïdentifiseerde frasegroepe word daar tussen koördinering en neweskikking (bv. die onderskeid tussen *The new and the subsidized industries/The new subsidized industries*) onderskei.
- (g) Onderwerp- en predikaatidentifisering  
Op grond van kongruensiemerkers, kasusrolle en verspreiding word die onderwerp van die predikaat geskei.
- (h) Voorsetselstrukture  
Die bestek van voorsetselgroepe word so ver moontlik geïdentifiseer (bv. die onderskeid tussen *The man with the telescope saw us/The man saw us with the telescope*).
- (i) Kondisionele idiome  
Op hierdie stadium vind die eerste omsetting na doeltaalwoorde plaas. Alvorens die doeltaalekwivalente wat in stap (a) nageslaan is, ingevoer word, moet daar vasgestel word of die inligting wat in die daaropvolgende stappe ingewin is, enige uitwerking op die keuse van doeltaalitems het. As voorbeeld: die Franse uitdrukking *être d'accord*, waarskynlik die mees frekwente vertaling vir *agree*, kan nie in onpersoonlike passiefsinne voorkom nie.
- (k) Oplossing van dubbelsinnighede  
Die oplos van die oorblywende dubbelsinnighede word so ver as moontlik bewerkstellig deur (soms redelik *ad-hoc*) inligting wat in die woordeboeke aangebring is om probleme tussen 'n spesifieke taalpaar te ondervang (bv. negeringsbestek in 'n taalpaar soos Frans en Engels, waar die een taal 'n dubbel-negatiefstruktuur het).
- (l) Sintese  
In hierdie laaste fase word doeltaalsinne gegenereer deur die invoeging van ekwivalente vir die res van die leksikale items, die aanbring van verbuigings- en vervoegingsmerkers, en die herrangskikking van woordorde, bv. die Engelse adjektief-naamwoord volgorde verander na die Franse naamwoord-adjektief volgorde.



Figuur 3.2



(Hutchins 1982:25)

Die stelselstruktuur van SYSTRAN het gaandeweg al hoe meer modulêr geword. Daarmee word bedoel dat die subkomponente en die subprogramme in die stelsel al hoe meer geskei is sodat veranderings in 'n bepaalde subdeel van die stelsel nie 'n 'domino-effek' oor die hele stelsel tot gevolg sal hê nie. Dit is veral die toenemende onderlinge onafhanklikheid van die drie hoofmomente, analise, oordrag en sintese, wat die neiging tot modulariteit illustreer. Die volle voordele van modulariteit kan egter nie benut word nie omdat die stelsel met *ad-hoc* reëls deurspek is. Dit geld veral vir die semantiese komponent en die semantiese inligting in die woordeboeke (Hutchins 1982:26).

### 3.10.3 Die woordeboeke

In die SYSTRAN-stelsel word woordeboeke wat afsonderlike woorde behandel en dié wat uitdrukkings hanteer, geskei (Siebenaler 1986:57). Vir die opstel van die sogenaamde idioomwoordeboeke word daar nie onderskei tussen werklike idiome en ander uitdrukkings wat, hoewel nie streng gesproke vaste idiomatiese uitdrukkings nie, tog nie ten beste deur die woord-vir-woord benadering vertaal kan word nie. In sommige kontekste neem bepaalde woorde 'n spesiale, dikwels figuurlike, betekenis aan. Deur sodanige gebruike in die idioomwoordeboeke op te neem, kan die stelsel heelwat konteks-afhanklike gebruike doeltreffend hanteer (Hartmann 1983:191).

Die hoofstamwoordeboek, wat woorde afsonderlik behandel, bevat vir elke inskrywing morfologiese, sintaktiese en semantiese inligting, addisionele vorms, homografiese kenmerke, en taalpaarspesifieke annoterings. 'n Beperkte stel semantiese merkers verskaf op binêre wyse beknopte inligting oor elke item. Verder is daar 'n stelsel van ongeveer 500 semantiese primitiewe wat aanvanklik hoofsaaklik gedien het om vakgebiede en onderwerpe aan te dui, maar dit is ook bruikbaar vir semantiese ontleding. Daar bestaan ook 'n aantal sub-roetines, byvoorbeeld een vir die uitkenning van reëlmatige uitgange soos *-ologie* (tak van 'n wetenskap), *-meter* (meganiese teller), met ekwivalente in ander tale.

In die verlede is 'n hoofstamwoordeboek vir elke taalpaar wat deur SYSTRAN gedek word saamgestel en in stand gehou. Dit het meegebring dat met die byvoeging van 'n taal tot die stelsel nuwe woordeboeke gebou moes word, nie net vir die bykomende taal nie, maar ook vir die bestaandes. Afgesien van die koste en die duplisering van werk, is daar gevind dat vele foute en inkonsekwentheid wat in die bestaande woordeboeke opgespoor en uitgeskakel is, in die nuwes weer voorkom. Vandag is daar slegs een woordeboek vir elke brontaal, gevolglik funksioneer die analisefase grotendeels doeltaalonafhanklik. Die doeltaalwoordeboeke is egter steeds taalspesifiek. SYSTRAN is dus nou 'n gemengde stelsel, terwyl dit oorspronklik 'n tweetalige direkte stelsel was (Homer 1986:120).

Die SYSTRAN-woordeboeke het met verloop van tyd 'n al hoe meer problematiese aspek van die stelsel geword. Afgesien van die groot aantal inskrywings (tot 150 000 in een woordeboek), het die inskrywings self al hoe lywiger en onhanteerbaar geword. Die oorsaak en implikasies van hierdie probleem word hieronder in die konteks van die EEG se besluit om SYSTRAN te vervang, verder bespreek.

#### 3.10.4 Vooruitsigte

Soos blyk uit die voorgaande bespreking van SYSTRAN, het die stelsel 'n lang geskiedenis en dit is wêreldwyd by groot gebruikers gevestig. Die gehalte van die vertalings wat gelewer word is nie hoog nie (vgl. die voorbeeldmateriaal in 2.3) dus moet naredigering gedoen word. Vir sommige doeleindes is ruwe vertalings egter voldoende, byvoorbeeld die Amerikaanse lugmag se gebruik van SYSTRAN vir inligtingherwinning. Vir 'n gebruiker soos XEROX, wat slegs Customized Multinational English vir invoer gebruik, is die stelsel ook doeltreffend. Verder is die stelsel, soos alle rekenaargesteunde vertaalstelsels, veral effektief in 'n beperkte tegniese omgewing, soos blyk uit die feit dat 'n toenemende aantal atoomkraginstansies SYSTRAN gebruik.

SYSTRAN was by die EEG skaars 'n jaar in gebruik toe daar in 1977 tot die gevolgtrekking gekom is dat die stelsel om diepliggende redes nooit 'reggeskaaf' kan word om in die vertaalbehoefte van die Gemeenskap te voorsien nie. Daar is toe begin met die beplanning van 'n geheel nuwe stelsel. Dit sou 'n stelsel wees

- wat vir multitaligheid ontwerp is (anders as SYSTRAN);
- ~~- wat 'n groot spektrum vakgebiede kan hanteer;~~
- waartoe tale bygevoeg kan word;
- wat taalkundig voortdurend uitgebrei kan word;
- wat vorderings, en selfs omwentelings, in die rekenaartegnologie kan akkommodeer en benut;
- waarvan komponente, soos die woordeboeke en teksverwerkingsprogramme, multidoelig aangewend kan word (bv. vir kunsmatige intelligensie toepassings van natuurliketaalverwerking).

SYSTRAN beantwoord nie aan hierdie vereistes nie. Dit is in wese 'n tweetalige stelsel, wat meebring dat die toevoeging van tale nie koste-effektief is nie en tot onnodige ingewikkeldheid en lompheid lei. Dit is nie op eksplisiete taalteoretiese en -metodologiese beginsels gegrond nie, gevolglik geskied uitbreidings op 'n ongerigte wyse. Die benadering in SYSTRAN is dat vertaling basies 'n woord- en frasevervangingsproses is. Derhalwe word daar swaar op woordeboekinligting gesteun terwyl sintaktiese en semantiese ontleding 'n heel beperkte rol speel.

Om woordeboeke in so 'n groot mate die spil van die stelsel te maak het verskeie nadele, afgesien van taalteoretiese oorwegings. Woordeboeke is itemgerig, nie patroongerig nie. Dit bring eerstens mee dat taalkundige inligting aan afsonderlike woorde gekoppel word, terwyl dit meer sinvol is om heelwat daarvan in samehang met die taal se struktuur uiteen te sit. Tweedens, dieselfde tipe inligting word herhaaldelik ingeskryf, wat onekonomies is. Derdens, namate die woordeboek opbou, bestaan daar al hoe meer die gevaar dat inskrywings idiosinkraties word. Dit het by EEG-SYSTRAN gebeur as gevolg van die gebrek aan 'n omlynde taalkundige model (bv. opeenvolgende spanne het die stelsel van semantiese merkers grondig gewysig). Vierdens, hoe meer die woordeboeke uitbrei hoe moeiliker word dit om 'n effektiewe kruisverwysingstelsel te handhaaf. Vyfdens, hoe groter die woordeboeke, hoe duurder elke inskrywing wat toegevoeg word.

Die stelsel wat ontwerp word om SYSTRAN by die EEG te vervang, EUROTRA, word in die volgende afdeling bespreek. Ten spyte van die tekortkominge van SYSTRAN, veral vir die EEG se doeleindes, kan daar enersyds beweer word dat SYSTRAN 'n belangrike mylpaal en aansporing op die gebied van rekenaargesteunde vertaling is, en, andersyds, dat die stelsel waarskynlik lank nog gebruik en verfyn gaan word deur sy huidige ander groot gebruikers, soos die Amerikaanse lugmag, die atoomkrag-organisasies, Xerox, NASA, Wang Laboratories en General Motors.

### 3.11 EUROTRA

#### 3.11.1 Aqtergrond

Die EEG se EUROTRA-program is die grootste onderneming op die gebied van rekenaargesteunde vertaling in die Westerse wêreld. Al die EEG-lidlande neem daaraan deel en altesaam is daar meer as 'n honderd navorsers en talle administratiewe en tegniese ondersteunende kragte by die program betrokke. Inderdaad, selfs net die beplannings- en logistiese aspekte van die onderneming is insiggewend, afgesien van die navorsing self.

Beplanning het in 1978 'n aanvang geneem, die eerste voorlegging is in 1980 aan die Europese Kommissie gedoen, en twee jaar later het die Raad die beplanningsdokumente goedgekeur. Die Raadsbesluit kan soos volg opgesom word:

- (a) EUROTRA is 'n navorsings- en ontwikkelingsprogram wat 'n gevorderde vertaalstelsel ten doel het. Die stelsel moet in staat wees om al die amptelike tale van die EEG te kan hanteer.

- (b) Aan die einde van die program moet 'n operasionele prototipe beskikbaar wees. Dit moet 'n klein aantal vakgebiede dek, bepaalde tekstipes hanteer en 'n woordeskat van ongeveer 20 000 inskrywings bevat. Die verdere ontwikkeling van die stelsel sal dan op industriële skaal onderneem word.
- (c) Die program vir die ontwikkeling van die prototipe sal vyf en 'n half jaar duur en 16 miljoen ECU (ongeveer 25 miljoen rand) word daarvoor begroot. Dit sal aangevul word met 11 miljoen ECU (ongeveer 18 miljoen rand) wat deur die lidlandse regerings bygedra sal word.
- (d) Die program word in drie fases verdeel:
- (i) 'n Voorbereidingsfase van 2 jaar (2 miljoen ECU (ongeveer 3 miljoen rand)) vir die opstel van die stelselspesifikasies, die juridiese ooreenkomste en die organisatoriese raamwerk.
  - (ii) 'n Basiese en toegepaste taalnavorsingsprogram van twee jaar (8,5 miljoen ECU (ongeveer 14 miljoen rand)) om 'n klein prototipe-stelsel, gebaseer op 'n korpus van ongeveer 2 500 inskrywings, te bou.
  - (iii) 'n Evaluerings- en uitbreidingsfase van 18 maande (5,5 ECU (ongeveer 9 miljoen rand)) waartydens die stelsel uitgebrei sal word om meer vakgebiede te dek en 'n groter woordeskat te hanteer (ongeveer 20 000 inskrywings). Aan die einde van hierdie fase moet voorstelle vir die ontwikkeling van die stelsel op ~~industriële skaal godoen word.~~
- (e) Die organisatoriese reëlings is soos volg: die algemene spesifikasies vir die taalkundige en programmatuurmodelle word deur die Kommissie se projekspan ontwerp. In elke lidland sal daar 'n navorsingspan in die lewe geroep word om vir die land se taal analise- en sinteseprogramme te ontwikkel, asook oordragkomponente na die ander tale.

Die program is tans in die tweede fase. Toe Spanje en Portugal in 1986 by die EEG aangesluit het, het die aantal EEG-tale van sewe na nege vermeerder (en die taalpare van 42 na 72). Dit het vanselfsprekend finansiële gevolge ingehou en die skedulering van die program beïnvloed. Die Kommissie het besluit om die program met 18 maande te verleng (na sewe jaar) en om die begroting te vergroot.

### 3.11.2 Die EUROTRA-navorsingsgroepe en -skakelmeganismes

Soos reeds genoem, is daar in elke lidland 'n EUROTRA-span wat vir die land se taal stelselkomponente ontwikkel. België, met sy twee amptelike tale, het twee spanne, wat 'n inset lewer by die Franse en Nederlandse projekte. Luxemburg en Ierland hanteer bepaalde algemene sake omdat hul amptelike tale deur ander spanne behartig word (Iers is nie 'n EEG-taal nie).

In elke land is daar 'n projekondersteunende instansie (in die VK is dit byvoorbeeld die departement van handel en nywerheid), 'n projekleier en 'n navorsingspan van ongespesifiseerde grootte, waarvan die lede meestal akademiëci is wat deelyds of voltyds na die projek gesekondeer is.

EUROTRA se sentrale span bestaan uit verteenwoordigers van die nasionale spanne, EEG-amptenare en 'n aantal ander kundiges uit 'n verskeidenheid spesialiteite. Dit is die sentrale span se verantwoordelikheid om gemeenskaplike sake te hanteer en om die nasionale projekte te koördineer. Die sentrale span is ook gemoeid met die implementering van die programmatuur wat deur die verskeie sentra verskaf word, en met die ontwikkeling van 'n gemeenskaplike gebruikerstaal en 'n gebruikersvriendelike omgewing (Somers 1986:131-134).

### 3.11.3 Die EUROTRA-stelselontwerp

Die drie grondbeginsels in die stelselontwerp is multitaligheid, modulariteit en uitbreibaarheid.

#### (a) Multitaligheid

Multitaligheid hou in dat die ontledings- en genereringskomponent vir elke taal onafhanklik van die ontledings- en genereringskomponente van die ander tale moet wees. Slegs in die oordragkomponent word vir taalparing voorsiening gemaak. Met hierdie benadering kan tale met 'n minimum van koste vir die bestaande lede tot die stelsel toegevoeg word. Daar hoef slegs 'n oordragprogram na elke nuwe taal geskryf te word. Ten einde koste nog verder te verlaag, is daar ook besluit om die oordragstrukture so eenvoudig moontlik te hou. Dit bring mee dat die ontledings- en genereringskomponente besonder gedetailleerd moet wees, maar vanweë die feit dat hierdie modules vir elke taal eenmalig ontwikkel word, is hierdie benadering nogtans die mees koste-effektief.

(b) Modulariteit

Modulariteit hou in dat die verskillende stelseltake deur modules wat so ver moontlik van mekaar onafhanklik is, verrig word. In 'n ten volle modulêre stelsel kan 'n module gewysig word, of deur 'n verbeterde weergawe vervang word, sonder dat ander modules daardeur geraak word. Modules kan ook verder uit kleiner modules bestaan, wat dus ook afsonderlik gewysig kan word, maar daar is vanselfsprekend perke aan hierdie indelingsproses.

(c) Uitbreibaarheid

Uitbreibaarheid is 'n uitvloeisel van die multitalige en modulêre ontwerp van die stelsel. Dit hou in dat nuwe tale, taalkundige teorieë, tekssoorte en vakgebiede bygevoeg kan word sonder dat enige deel van die bestaande stelsel daardeur geraak word.

Indien die basiese oogmerke van veeltaligheid en modulariteit bereik word, sal die prototipe uitbreibaar wees, en die basis kan vorm vir deurlopende verbeterings en selfs ingrypende veranderings (Lau 1983:31-32). ~~Op grond van vorderings op die gebied van kognitiewe in-~~  
~~terrensiesie word daar byvoorbeeld reeds besef dat die toevoeging van~~  
kennisdatabasisse en inferensiemeganismes oorweeg moet word.

Hoewel die EUROTRA-stelselontwerp uitdruklik vir toekomstige ontwikkelings voorsiening maak, is die huidige aktiwiteite van die navorsingspanne hoofsaaklik daarop toegespits om die reeds verworwe kennis en kundighede op die gebied van vertaalstelsels saam te snoer, nie om van vooraf nuwe teorieë te ontwikkel en op die proef te stel nie. ~~In dié sin is EUROTRA nie 'n eksperimentele program nie. Dit is egter~~  
eksperimenteel in die sin dat die stelsel op die konsep van 'n ontwikkelingsomgewing gegrond is.

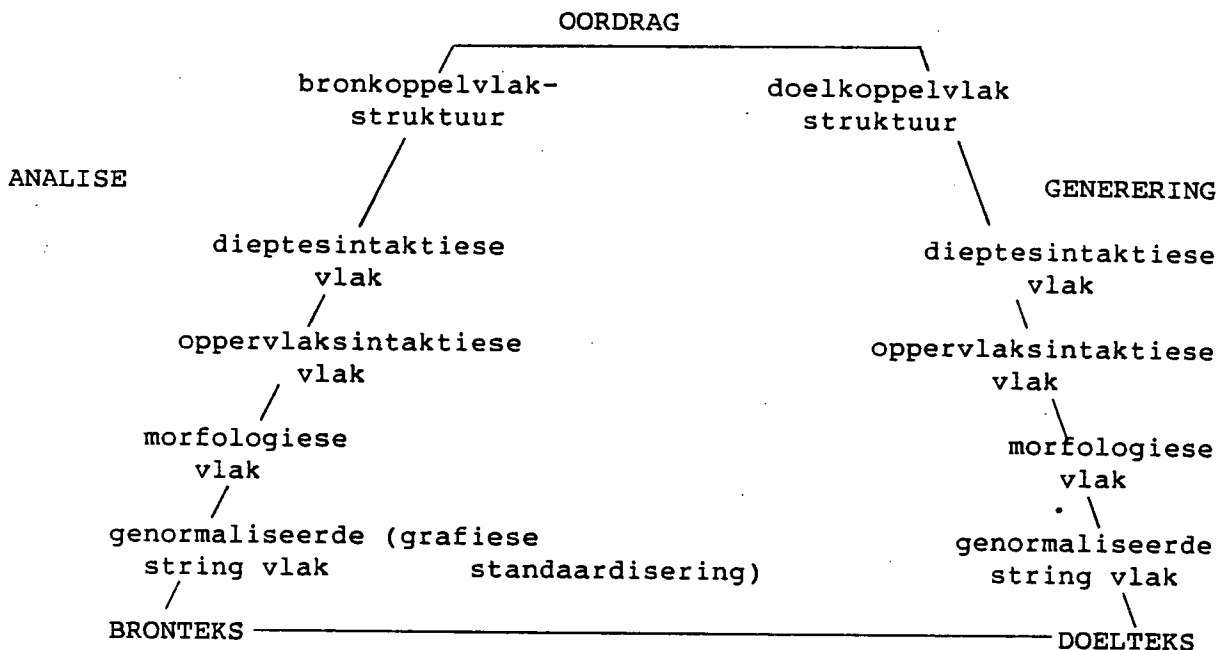
3.11.4 Die programmatuur en die taalkundige ontwerp

Die ontwikkeling van die programmatuur word heeltemal van die ontwikkeling van die taalkundige reëls geskei. Die basiese reëlvertolker is 'n algemene herskryfstelsel met 'n kontroletaal. Die taalkundige reëls word in 'n bondel subgrammatikas saamgevoeg, en die taalkundiges word van metodes vir die kontrolering van toepassingskeuses voorsien. Die afsonderlike reëls is nie-destruktiewe herskryfreëls ten einde te voorkom dat inligting geskrap word wanneer nuwe strukture geskep word. (Slocum 1984:20).

Wat die taalkundige komponent betref, is daar ook binne die algemene raamwerk 'n groot mate van vryheid. Die verpligte basiese strukture vir betekenisrepresentasie is afhanklikheidsstrukture. Die strukture word geannoteer met kenmerkwaardepare wat deels deur die afsonderlike spanne en deels deur onderlinge ooreenkoms tussen die spanne bepaal is. Katz en Fodor se merkerteorie en Fillmore se kasushipotese vorm die grondslag van die semantiese sisteem, terwyl die sintaktiese sisteem op die herskryfreëlkonvensies van die generatiewe grammatika gegrond is.

Die verskillende ontledingsvlakke in die taalkundige komponent word in Figuur 3.3 voorgestel. Een van die interessante aspekte van die sisteem is dat daar van vlak na vlak heen en weer beweeg kan word. Met ander woorde, analise op die eerste vlak hoef nie volledige deurgevoer te word voordat daar na die tweede vlak beweeg word nie. Hierdie soepelheid bring teweeg dat die program inligting vanaf ander vlakke kan gebruik om probleme op te los.

Figuur 3.3



Tegniese besonderhede oor die stelsel is in Jaspaert 1986, Van Eynde 1986, en Lau 1985 beskikbaar. Volgens hierdie skrywers kom die navorsingspanne veel ingewikkelder probleme teë as wat hulle te wagte was, veral op die semantiese vlak en ten opsigte van die *Euro-versale* waarop die koppeling tussen taalpare gebaseer moet wees. Daar is ook nog 'n aantal belangrike taalkundige besluite wat in hierdie laat stadium nog nie geneem is nie (persoonlike inligting, 1988).



Soos vir die taalkundige komponent, geld die beginsels van modulariteit en uitbreikbaarheid ook vir die ontwerp van die programmatuur. Die EUROTRA-programmatuur kan enige programmeertaal insluit mits die taal met die definisie van die stelsel se skynmasjien versoenbaar is. Die skynmasjien ('virtual machine') is 'n abstrakte definisie van 'n masjien wat die funksionele vereistes waaraan apparatuur en programmatuur moet voldoen om met die stelsel versoenbaar te wees, stel. Hierdeur word 'n aansienlike vryheid geskep (Lau 1985:31-32).

### 3.11.5 Vooruitsigte

Die ontwikkelingsplan wat in 1982 deur die Kommissie goedgekeur is, is veronderstel om in 1989 met die lewering van die prototipe-stelsel af te sluit. Daar bestaan in sommige EUROTRA-kringe twyfel oor die haalbaarheid van die mikpunt, maar selfs al sou die prototipe betyds voltooi word, sal dit nog 'n aantal jare duur voordat die program en die leksikons al die vakgebiede wat vir die EEG tersaaklik is, sal kan hanteer. Gevolglik sal SYSTRAN tot diep in die negentigerjare die hoofprogram bly. Daarna sal dit ook nog gebruik word vir bepaalde doeleindes (Pigott 1984).

In ag genome die grootte van die EUROTRA-program, die omvang van die koste daarvan, en die feit dat dit uiteindelik 'n vertaalstelsel sal oplewer waarvan die produkte nog steeds geredigeer sal moet word (dit was altans nog altyd die verwagting), kan daar met reg gewonder word of dit 'n regverdigbare onderneming is. Dit is egter 'n ingewikkelde saak om te evalueer. Vir die EEG is dit 'n belangrike oorweging dat vertaalkoste reeds in die tyd toe daar slegs sewe lidlande was meer as 'n derde van die organisasie se bedryfsbegroting verteenwoordig het. ~~Toe twee lande in 1986 by die Gemeenskap aangesluit het, het die omvang van die vertaalaktiwiteite eksponensieel gegroei. Nou is daar sprake van nog drie nuwe lidlande.~~ SYSTRAN (vgl. 3.10), met al sy beperkthede, het getoon dat rekenaargesteunde vertaling groot potensiaal het.

Vanselfsprekend word daar dikwels gevra waarom die EEG nie een, of twee-drie, tale as amptelike tale benoem nie. Die antwoord is kort en kragtig: dit is 'n EEG-geloofsartikel dat die hoogs gesonde same-werkingsklimaat wat in die Gemeenskap heers, in die eerste plek aan die gelykberegtiging van die lidlande se tale toe te skryf is.

### 3.12 SUSY

#### 3.12.1 Aqtergrond

Navorsing oor rekenaargesteunde taalverwerking het in Saarbrücken (Wes-Duitsland) in die middel sestigerjare 'n aanvang geneem. Die eerste doelwitte was 'n eksperimentele ontleder vir Duits en 'n eksperimentele Latyn-Duitse stelsel (SALADIN). Tussen 1967 en 1970 is die moontlike

verdere ontwikkeling van 'n vroeë weergawe van SYSTRAN se Russies-Engelse stelsel ook ondersoek. Hierdie ondersoek het op die besluit uitgeloop om van vooraf 'n Russies-Duitse prototipe-stelsel te ontwerp. Dit was die beginpunt van die navorsing wat gelei het tot die ontwikkeling van die multitalige vertaalprogram wat bekend sou staan as SUSY (Saarbrücker ÜberSetzungssYstem). Anders as by die meeste ander rekenaargesteurde vertalingsprojekte, het dit die Saarbrücken-groep vrygestaan om op navorsing eerder as op praktiese toepassings te konsentreer.

SUSY is ontwerp as 'n multitalige stelsel, en programme vir die volgende tale is ontwikkel: Engels, Russies, Duits, Frans en Esperanto. In 'n stadium is daar ook aan Deens en Nederlands aandag gegee (Back en Wilms 1985:7).

Oor die jare heen het die SUSY-projek verskeie ingrypende veranderings ondergaan. Aan die einde van 1980 is die navorsing selfs gestaak, maar na herbeplanning is dit in 1981 voortgesit. Voortaan sou dit in twee rigtings stuur: ener syds die ontwikkeling van 'n nuwe weergawe van SUSY, SUSY 2, en andersyds 'n ondersoek na die moontlike toepassings van die basiese SUSY-stelsel, SUSY 1. Laasgenoemde ondersoek het in 1985 uitgeloop op die MARIS-projek, 'n prakties-georiënteerde vertaalprogram wat deur die Wes-Duitse federale ministerie vir navorsing en ontwikkeling ook te Saarbrücken gevestig is.

Volgens Luckhardt (1987) word die Saarbrücken-navorsing van tyd tot tyd deur befondsings- en herbeplanningsprobleme geteister, maar die sentrum is as 'n leier op die gebied van taalprosessering sterk gevestig.

### 3.12.2 Onderliggende SUSY-beginsels

Die SUSY-stelsel is op grond van die volgende beginsels ontwerp:

#### (a) Multitaligheid

Die stelsel is uit die staanspoor vir multitaligheid ontwerp. Dit hou in dat die ontledingsmodule doeltaalonafhanklik moet wees, en die genereringsmodule brontaalonafhanklik. Gevolglik moet taalparing deur 'n taalpaarspesifieke oordragmodule bewerkstellig word (vgl. 3.11.3, die EUROTRA-stelselontwerp).

(b) Uitbreikbaarheid

Afgesien van uitbreikbaarheid in die sin van die toevoeging van nuwe tale, hou uitbreikbaarheid ook in dat die subdele van die stelsel sodanig van mekaar onafhanklik moet wees dat verbeterings en vervangings sonder 'domino-effekte' ingevoer kan word. Dit geld ook vir die beskerming van die taalkundige komponent teen programmatuur- en apparatuurveranderinge.

(c) Heterogeniteit

In hierdie konteks dui heterogeniteit op die beginsel dat verskillende reëlformalismes in die verskillende dele van die stelsel gebruik kan word. Hoewel dit in beginsel elegant lyk om 'n ontleder op een reëlformaat te grond, is daar in die praktyk gevind dat die funksionele vereistes van die verskillende onderdele van vertaalprogramme verskillende benaderings noodsaak. As ooglopende voorbeeld: intra- en intertalige reëls moet aan uiteenlopende vereistes voldoen.

In die SUSY-stelselontwerp word daar besondere aandag gegee aan die wyse waarop data gestruktureer en gekoppel word, asook aan die beheermechanismes in die stelsel. Dit bemoontlik onder andere maklike toegang na intermediêre fases van die vertaalproses. Vir verdere besonderhede hieroor, vergelyk Luckhardt (1987), Hutchins (1986) en Back en Wilms (1985).

3.12.3 Die SUSY-woordeboeke

Die modulariteit van SUSY word in die leksikonstelsel weerspieël. Die stelsel sluit die volgende woordeboeke in (vgl. Luckhardt 1987):

- Brontaal morfo-sintaktiese woordeboeke
- Brontaal leksikaal-semantiese woordeboeke
- Brontaal-doeltaalwoordeboeke
- Doeltaal leksikaal-semantiese woordeboeke
- Doeltaal morfo-sintaktiese woordeboeke

Hutchins (1986) wys daarop dat die SUSY-woordeboeke, met uitsondering van die Duitse woordeboeke, nog betreklik beskeie van omvang is. Vir die ander tale is die grootste woordeboek die morfo-sintaktiese woordeboek vir Russies as brontaal, met 15 000 inskrywings.

### 3.12.4 Stappe in die SUSY-vertaalproses

In hierdie afdeling word analise, oordrag en generering in die SUSY-stelsel stapsgewys kortliks beskryf (vir verdere besonderhede vgl. Back en Wilms 1985 en Hutchins 1986).

#### Analise

- (a) LESEN  
Dit is die invoerfase, waartydens woord- en sinsgrense ook geïdentifiseer word.
- (b) WOBUSO  
Die program raadpleeg die morfo-sintaktiese woordeboek vir die brontaal, identifiseer stamme en verbuigings, en verskaf voorlopige inligting oor woorde wat nie in die woordeboek voorkom nie.
- (c) DIHOM  
Homograwe word op verskillende wyses geïdentifiseer en ontleed: op grond van die woordklasse van die aanliggende items (die LEMMAT-roetine), toetse vir onreëlmatighede en teenstrydighede (die TABHOM-roetine), en 'n statistiese evaluering van waarskynlikhede, voorkeure en die 'gewig' van woordklas-afparings (die GEWICHTE-roetine).
- (d) SEGMENT  
Hier word die sin in klouse en subklouse ingedeel (op basis van interpunktuasie, voegwoorde, en so meer), en 'n voorlopige afhanklikheidsstruktuur word verskaf (op basis van voegwoorde, relatiewe voornaamwoorde, en so meer).
- (e) NOMA  
Die naamwoordgroepe word deur middel van reëls vir naamwoordstukbou in die brontaal geïdentifiseer.
- (f) VERA  
Die werkwoordgroepe word op soortgelyke wyse as in NOMA ontleed.
- (g) KOMA (voorheen SYSNAN genoem)  
In hierdie fase word die frasestrukture vir oordrag gebou. Naamwoordgroepe word met regerende werkwoorde op grond van valensie-inligting verbind, relatiewe voornaamwoorde en refleksiewe word deur hul antesedente vervang, en ondergeskikte konstruksies (soos adjektief- en deelwoordklouse) word aangedui.

(h) SEDAM

Hierdie finale stap in die ontledingsfase los eerstens oorblywende meerduidigheidsprobleme op leksikale vlak op (aan die hand van die leksikaal-semantiese woordeboek). Dit verskaf ook intertalige kasusmerkers vir sommige voorsetsels en verbuigings. Laastens identifiseer dit die valensierelasies van verbale naamwoorde wat as genominaliseerde regeerders funksioneer. Die afvoer is 'n voorstelling van die sintaktiese dieptestruktuur aangevul deur leksikale inligting wat reeds intertalige semantiese elemente bevat.

Oordrag

Oordrag word in een stap uitgevoer. Deur middel van die tweetalige woordeboek word die leksikale items van die bronteks met doeltaalitems vervang. Woorde wat nie in die woordeboek voorkom nie, word morfologies (deur die WOBUSO-roetine) ontleed ten einde moontlike afleidings te bepaal. Sintaktiese veranderings wat byvoorbeeld deur verskille in valensierame veroorsaak word, word sover moontlik in die woordeboekinskrywings aangedui. Indien dit nie itemspesifiek verskaf is nie, word standaard brontaal-doeltaal sintaktiese patrone toegepas.

Sintese

In vertaalprogramme kan sintese (die genereringskomponent) heelwat eenvoudiger as analise (die ontledingskomponent) wees, mits laasgenoemde volledig genoeg is. Sintese in die SUSY-stelsel bevat net drie stappe:

(a) SEMSYN

Deur middel van die leksikaal-semantiese woordeboek vir die doeltaal, word idiomatiese en ander vaste uitdrukkings eers vertaal. Dan word werkwoordmodaliteite in die doeltaal se sisteem vertolk, en laastens word kasusmerkers in doeltaalterme uitgedruk.

(b) SYNSYN

Die frasestrukture word deur middel van die morfologiese sintaktiese woordeboek vir die doeltaal na doeltaalkonstruksies verander. Dan verskaf die program reekse doeltaalwoorde of -stamme, aangevul deur itemspesifieke inligting.

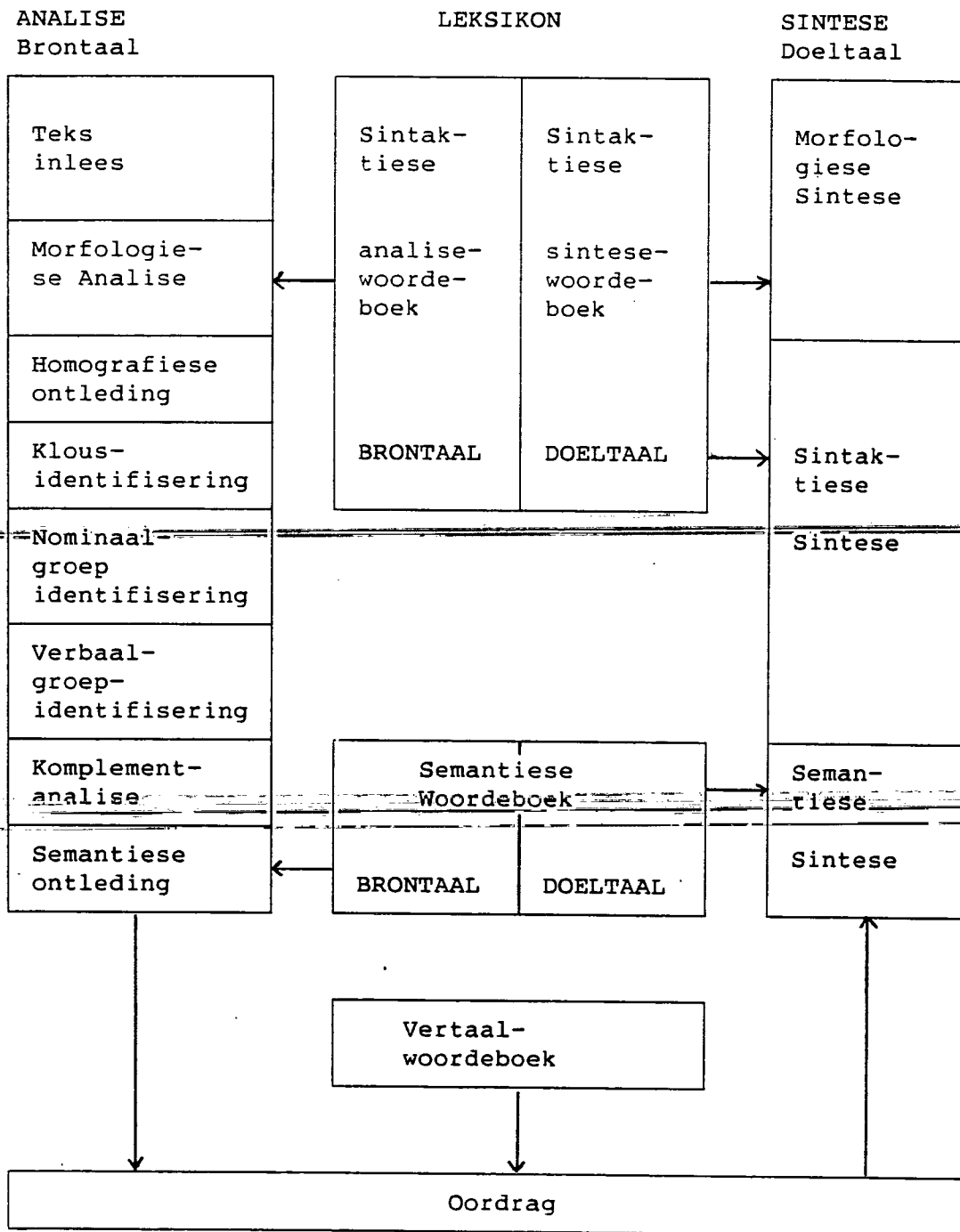
(c) MORSYN

In hierdie finale stap word die afvoer van SYNSYN gebruik vir die afleiding van korrekte morfologiese vorme.

In SUSY 1 is daar 'n meganisme om te voorkom dat die program vashaak. Die RESCUE-roetine kontroleer die afvoer van SEGMENT, NOMA, VERA en KOMA. Indien inkonsekwentheid gevind word, word minder streng reëls geaktiveer en die ontleding word herhaal. In SUSY 2 het RESCUE oorbodig geword omdat sekere voorheen verpligte patrone, byvoorbeeld valensierame, vryer vertolk word.

Figuur 3.4 bied 'n skematiese voorstelling van die stappe in die SUSY-program.

Figuur 3.4 Die volgende is 'n skematiese voorstelling van die werking van SUSY:



(Back en Wilms 1985:8)

### 3.12.5 Hoofverskille tussen SUSY 1 EN SUSY 2

SUSY 2 verskil in sommige van die ontledings- en genereringstappe van SUSY 1, terwyl die oordragproses wesenlik dieselfde is. In die ontledingskomponent word homograafoplossing in SUSY 2 tydens die sintaktiese ontleding saam met die ander prosesse vir die oplos van meerduidighede hanteer. Na invoer (LESEN) en morfologiese analise (WOBUSO) is daar in SUSY 2 net drie ontledingstappe. In die eerste word onproblematiese naamwoord- en werkwoordgroepe ontleed. In die tweede stap word die problematiese gevalle deur middel van toetse vir koördinering, attribusie, modifisering, en so meer, toegepas. In hierdie stap word 'n punt vir semantiese versoenbaarheid ook toegeken, en daarvolgens word die waarskynlik mees gepaste vertolking geïdentifiseer. Die derde stap hanteer hoofsaaklik klousstrukture en -relasies.

In die genereringskomponent is die hoofverskil tussen SUSY 1 en 2 die gebruik, in laasgenoemde, van deterministiese boomsamestellers ('tree transducers') wat meer doeltreffend as SUSY 1 se frasestrukture doeltaalstrukture voorspel. Verder verloop die genereringsproses basies dieselfde.

In die geheel gesien is SUSY 2 'n ryker en buigsamer program as SUSY 1. Dit is ryker omdat daar van 'n nog groter verskeidenheid van reël-formalismes gebruik gemaak word, en dit is buigsamer omdat waarskynlikheidstoetsing en voorkeurprosedures op verskillende vlakke ingebou is (vgl. Maas 1985 en Back en Wilms 1985 vir verdere besonderhede).

### 3.12.6 SUSY-verwante projekte

Uit die oogpunt gesien van die optimale benutting van navorsingsresultate is die Saarbrücken-groep se benadering besonder interessant. Met SUSY as die middelpunt van die navorsing, is daar ook voorsiening gemaak vir voortvloeiende projekte, waarin komponente van SUSY vir ander aanwendings ontwikkel kon word, byvoorbeeld vir die saamstel van woordeboeke en glossariums, die vertaling van titels en opsommings, inligtingsherwinning, en vraag-antwoord stelsels.

Een van die eerste praktiese toepassings van SUSY was SALEM (Saarbrücker Lemmatisierungssystem). Hierdie stelsel het die ontledingsprogram van SUSY vir die outomatiese saamstel van konkordansies van Duitse tekste gebruik, en vir navorsing oor die sintaktiese en semantiese aspekte van die Duitse woordeskat. Die SUSY-ontleider is ook in 'n inligtingsherwinningstelsel (TRANSIT) vir die outomatiese indeksering van regsdokumente gebruik. Tot dusver het die stelsel sowat 7 000 000 woorde ontleed en outomaties 'n leksikon van 110 000 Duitse samestellings opgestel.



SUSANNAH (SUSy ANwenderNAH) is 'n hulpprogram vir vertalers. Die program bied op elke stadium, vanaf invoer tot finale druk, groot steun. Afgesien van die vertaalkomponent self (wat nie hoë gehalte vertalings lewer nie, maar met behulp van die ingeboude naredigingsfasiliteit kan tekste vinnig verbeter word), verskaf die program woordverwerkings-, formaterings-, en terminologiese (en ander dataverskaffende) fasiliteite.

SUSY word sedert 1983 in samewerking met die Universiteit van Kyota ook in 'n projek vir die outomatiese vertaling van Duitse dokumenttitels in Japanees, en omgekeerd gebruik. 'n Buitengewone aspek van die projek is die gebruik van Engels as 'n intermediêre taal. Vertaling uit en in Duits en Engels word deur SUSY gedoen, terwyl dit uit en in Engels en Japanees deur Kyoto se TITRAN-program hanteer word. Die vertalings is stilisties nie suiwer nie maar dit voldoen aan die vereistes van akkuraatheid en verstaanbaarheid.

As gevolg van die SUSY-navorsing was Saarbrücken die natuurlike keuse vir die plasing van die Duitse EUROTRA-span.

---

Volgens Luckhardt (1987) is een van die grootste bydraes van die SUSY-navorsing die feit dat dit resultate wat in ander projekte ook ontplooi kan word, voortbring het.

### 3.13 METAL

#### 3.13.1 Agtergrond

---

Die Universiteit van Texas het vanaf 1961 van die Amerikaanse regering steun ontvang vir die vestiging van sy Linguistics Research Center, om navorsing oor rekenaargesteunde vertaling te doen. So het die METAL-projek (MEchanical Translation and Analysis of Language) vir Duits-Engelse vertaling sy ontstaan gehad.

Aanvanklik was die navorsing suiwer teoriegerig en in die eerste fase is die geskiktheid van Chomsky se TGG-model deeglik ondersoek. Nadat daar tot die slotsom gekom is dat die TGG-model nie geskik is nie, en terwyl 'n ander model gestalte begin aanneem het, het die regering besluit om alle befondsing van rekenaargesteunde vertaling te staak (vgl. 2.1). Die METAL-projek is dus in 1974 beëindig.

In 1978 het steun van die Amerikaanse lugmag (USAF) se Rome Air Development Center die projek laat herlewe en 'n jaar daarna het die Siemens-maatskappy van Wes-Duitsland met die Universiteit van Texas 'n ooreenkoms gesluit om die alleenborg van die projek te word. Siemens het vir eie bedryfsdoeleindes 'n Duits-Engelse vertaalprogram vir telekommunikasie- en dataverwerkingstekste nodig gehad. Op grond van die Texas-groep se vorige navorsing, juis op 'n Duits-Engelse stelsel, kon Siemens voorstel dat 'n prototipe binne 'n jaar of twee gelewer word. Die prototipe het in die tydperk 1980-84 toetsing en wysigings ondergaan. Sedertdien is die operasionele stelsel ontwikkel, en tans word 'n kommersiële weergawe vir bemaking voorberei (vgl. 3.13.5).

### 3.13.2 Oorsig van die METAL-stelsel

Die hoofkenmerke van die stelsel kan soos volg saamgevat word:

#### (a) Dokumentgebaseerde benadering

In die vroeë dae van navorsing oor rekenaargesteuende vertaling, dus ook in die eerste fase van die METAL-projek, was die aandag op die teksgedeeltes van dokumente toegespits. Diagramme, tabelle, titels, opskrifte, voetnotas, verwysings, en formatteringsinligting (inkepings, kursiefgedrukte woorde, spasiëring tussen onderafdelings, en so meer) moes afsonderlik bygewerk word. Hierdie invoegings- en herformatteringswerk is so tydwend dat dit in groot mate die voordele van rekenaargesteuende vertaling ondermyn. Daarbenewens kan foute maklik insluip.

Die METAL-stelsel vir Siemens moes uit die staanspoor voorsiening vir die hantering van die hele proses van dokumente-afwerking maak. Dit is dus 'n vertaalprogram wat in 'n sogenaamde vertalerswerkstasie geïntegreer is (dieselfde begrip onderlê SUSANNAH, vgl. 3.12.6). Die proses begin met die program se analise van die dokument se geledinge, die uithaal van tabelle, diagramme, en so meer, en die registrering van formatteringsinligting. Na afloop van die vertaling van die teksgedeeltes, word die dokument outomaties heropgestel en, met behulp van die naredigeringsfasiliteit, afgewerk.

#### (b) Modulariteit en uitbreikbaarheid

Die beginsel van modulêre stelselontwerp, met die oog op verbeterings, vervangings en uitbreikbaarheid (vgl. die besprekings in 3.11.3 en 3.12.2), word in METAL een stap verder gevoer deurdat daar doelbewus ook beplan is vir die aanwending van onderdele van die METAL-stelsel vir toepassings soos natuurliketaal-raakvlakke vir databasisse, en semantiese analise vir die ontwikkeling van inferensiesisteme in kennisgebaseerde databasisse.

(c) Die METAL-intertaal

METAL volg die gewone patroon van oordragstelsels, naamlik 'n ontledingsmodule, 'n oordragstruktuur en 'n genereringsmodule. Stukbou-algoritmes, transformasionele reëls en kasusrame speel, soos in ander vertaalprogramme, 'n groot rol. 'n Onderskeidende kenmerk van METAL is egter dat die ideaal van 'n doeltaal-onafhanklike invoer na die oordragstruktuur, en 'n brontaal-onafhanklike afvoer daarvan, uitdruklik as 'n nie-haalbare doelwit bestempel word. Taalpaarspesifieke elemente word dus waar nodig gebruik.

3.13.3 Die METAL-grammatikas

Lehmann en Bennett (1985), Bennet (1982), Slocum (1984), Schneider (1987) en Hutchins (1986) verskaf fyn besonderhede van die taalreëls wat vir METAL gebruik is. Terloops, dit kan ter ere van die Texas-groep gesê word dat hulle uitstaan as navorsers wat met groot openheid dokumentasie oor hul navorsing aan belangstellendes beskikbaar stel.

Hier word volstaan met 'n kort samevatting van opvallende aspekte van die taalkundige komponent in METAL. Eerstens, daar is doelbewus gepoog om die aantal taalreëls te beperk, byvoorbeeld deur reëls nie by te voeg sonder om moontlike oortollighede wat deur die byvoeging veroorsaak kan word, te ondersoek nie. Die reëltotaal staan op 500-600. Tweedens, die reëls word vlaksgewys geïndekseer en toegepas. Sodoende kan eenvoudige strukture op grond van laevlakreëls ontleed word. Hoërvlakreëls hoef slegs by meer ingewikkelde strukture toegepas te word. Hierdie benadering is elegant ekonomies.

Die METAL-ontleder bevat ook voorkeurmeganismes wat tussen ontledingspaaie onderskei. Die ontleder volg naamlik 'n aantal paaie, al is die eerste een suksesvol, dan selekteer dit volgens waarskynlikheidsaan-  
duiders (wat gedeeltelik op reëlordening berus) die mees gepaste een. Indien daar nie 'n aanneemlike vertolking van 'n sin verkry word nie, gaan die program na 'n deelsstakende roetine ('fail-soft routine') oor en dit verskaf vertalings van onderdele van die sin. Hierdie afvoer is volgens Schneider (1987) verrassend dikwels steeds 'n korrekte vertaling van die oorspronklike sin.

'n Verdere innovasie in die stelsel is die saambinding van stukbou-, konstruksie-, transformasie- en oordragreëls in reëlbondels. Die toepassing van hierdie saamgestelde reëls spoor sommige foute in die ontledingspad van die een module na die ander op. Hutchins (1986) wys egter daarop dat hierdie reëls teen die beginsel van modulariteit indruis.

#### 3.13.4 Die METAL-woordeboeke

Die METAL-woordeboeke word in modulêrgestruktureerde databasisse opgebou, met aparte leksikons vir elke taal, en 'n oordragleksikon vir taalparing. Op dieselfde beginsel van 'kortpadtoegang' wat in die reëlkomponent gebruik word (waar laevlakreëls eers toegepas word en, indien die ontleding slaag, daar dan daarmee volstaan word), is die leksikons volgens frekwensie-data en vakgebiedhierargieë ingedeel. Daarvolgens bevat die toeganklikste modules funksiewoorde, dan word daar beweeg na frekwente algemene woordeskat, dan na algemene tegniese woordeskat en daarna kom vertakkings na vakgebiede en subdissiplines.

Alle woordeboekinskrywings word eenvormig geformateer. Afgesien van die gewone inligting wat in woordeboeke nodig is, gebruik die METAL-sisteem 'n aantal kodes om semantiese kenmerke en verhoudings aan te dui. Konsepnummers groepeer semanties-verwante vorms (bv. *compute*, *computation*, *computational*, *computable*). Onderwerpsvelde word spesifiek met die oog op poliseemprobleme volgens 'n semantiese sisteem ingedeel, en naamwoorde en werkwoorde word op basis van 'n stel semantiese merkers geannoteer.

Volgens Lehmann en Bennett (1985) word die koste van woordeboekaanvulling so laag moontlik gehou, enersyds deur METAL se koppeling aan die TEAM-termbank, vanwaar verwerkte terminologie oorgeneem kan word, en andersyds deur die DEFAULTER-program, wat op grond van 'n minimum inligting met betrekking tot 'n bepaalde item die inskrywing outomaties kan voltooi. DEFAULTER is so doeltreffend dat die proeflesers min wysigings hoef aan te bring.

#### 3.13.5 Resultate en vooruitsigte

Die prototipe van METAL is in die tydperk 1980-84 op meer as 1 000 bladsye teks getoets. Die gehalte van die vertalings is aan die hoeveelheid teks wat na vinnige naredigering geen tekortkominge op morfologiese, sintaktiese of semantiese vlak getoon het nie, gemeet. Die slaagsyfer het na gelang die ingewikkeldhede van die tekste tussen 45% en 85% gewissel. In een bepaalde toets in 1980, op 'n teks van 50 bladsye, is gevind dat 83% van die sinne korrek ontleed is, 7% nie korrek nie, 9% is nie ontleed nie, hoewel tentatiewe ontledings van sinsgedeeltes verskaf is, en slegs 1% kon glad nie 'verstaan' word nie (Hutchins 1986:254).

Volgens Schneider (1987) vertaal METAL tans ongeveer 'n woord per sekonde. Dit kom op meer as 200 bladsye per dag (8 uur) neer. Ten einde die voordele van hierdie vinnige omset ten volle te benut, word die vertaalde teks onderverdeel en na 'n aantal mikrorekenaars versprei waar naredigeerders dit afwerk. Die tempo waarteen naredigering gedoen

word, het in die afgelope paar jaar van 10 bladsye per dag per naredigeerder tot 30 gestyg.

Die volgende is 'n voorbeeld van 'n ongeredigeerde METAL-vertaling:

Bronteks

EINTEILUNG DES PLATTENSPEICHERS

BLOCKSTRUKTUR Die kleinste adressierbare Informationseinheit ist ein Block = 1 Sektor. Zu jedem Block gehoert ein Header. Der Header enthaelt die gesamte Adresse sowie Angaben ueber den Zustand des Blockes (Benutsbarkeit!). Zur Sicherung der Header-Information und der Daten befindet sich am Ende des Headers und des Datenfeldes ein Pruefzeichen von 16 Bit.

Vor dem Headerfeld befindet sich eine Praeambel von 42 Byte Laenge fuer den Ausgleich aller Toleranzen.

Vor dem Datenfeld befindet sich eine Praeambel von 5 Byte Laenge zur Aufsynchronisierung der Leseverstaerker. Vor und hinter dem Datenfeld Befindet sich eine Luecke.

---

Ongeredigeerde vertaling:

DIVISION OF DISK STORAGE

BLOCK STRUCTURE The smallest addressable information unit is a block = 1 sector. A header is part of every block. The header includes the entire address, sowie specifications about the state of the block (usability!). A check character of 16 bits is found for the saving of the header information and the data at the end of the header and the data field.

A preamble of 42 byte length for the adjustment of all tolerances is found in front of the header field.

A preamble of 5 byte length is found in front of the data field for the synchronization of the read amplifier. A gap is found in front of and behind the data field.

Siemens het die METAL-stelsel in die eerste plek vir gebruik in die maatskappy self bedoel. Dit word egter binnekort ook kommersieel aangebied onder die naam LITRAS. Tans vertaal die stelsel slegs uit Duits in Engels, maar die volgende taalpare is reeds in verskillende stadiums van ontwikkeling: Duits-Spaans, Engels-Duits, Engels-Spaans, en Duits-Chinees. Eerste aanvoerwerk aan 'n Franse en aan 'n Nederlandse program is ook aan die gang.

Siemens en die bemarker van LITRAS, Computer Gesellschaft, Konstanz, stel dit in die vooruitsig dat komponente van METAL gebruik sal word vir toepassings soos outomatiese ekserpering, kennisgebaseerde inligtingsherwinning en rekenaargesteuende onderrigprogramme. Vooruitsigte vir die verskaffing van werklik hoë gehalte vertaling, wat heeltemal geen naredigering vereis nie, bly egter skraal.

### 3.14 SAMEVATTEND

Met die beskrywings in hierdie hoofstuk is daar getrag om op 'n bondige wyse inligting wat vir besprekings oor moontlike ontwikkelingstappe in Suid-Afrika tersaaklik kan wees, byeen te bring. Vir beplanners en besluitmakers op verskillende vlakke, vanaf beleidsbesluitneming tot op die vlak van detailbesluite oor taalreëlsisteme, bied oorsese projekte 'n waardevolle inligtingskat. Die verskillende wyses waarop befondsing geskied, die spektrum benaderings tot beplanning, die uiteenlopende toepassings vir vertaalprogramme, die debatte oor stelselontwerp, die taalmodelvariasies, en die strategieë om tussen navorsingsprojekte en operasionale stelsels 'n brug te slaan, is almal sake wat deeglike oordenking verdien.

Die volgende hoofstuk bied 'n standbeskrywing van die vertaalwese in Suid-Afrika en 'n bespreking van die koste-effektiwiteit van rekenaargesteuende vertaling.

## HOOFSTUK 4: STANDBESKRYWING EN KOSTE-IMPLIKASIES

### 4.1 DIE VERTAALWESE IN SUID-AFRIKA

'n Onderzoek na die stand van rekenarisering in die Suid-Afrikaanse vertaalwese is einde 1987 uitgevoer. Die ondersoek het benewens navraag na rekenarisering ook aandag gegee aan ander kenmerke van die bedryf wat in berekening gebring sal moet word indien rekenaargesteunde vertaling as 'n ontwikkelingsgebied in Suid-Afrika oorweeg gaan word. Die inligting wat ingewin is, kan onder die volgende rubrieke ingedeel word:

- (a) Die omvang van die vertaalmark
- (b) Die aard van die stof wat vertaal word
- (c) Die tale wat vertaal word
- (d) Die stand van rekenaargebruik in die vertaalwese

#### 4.1.1 Die omvang van die vertaalmark

~~Vanweë die samestelling van die vertaalbedryf is daar-in-die-onderzoek uit-die-staanspoor-besef dat 'n presiese vasstelling van die omvang van die vertaalmark onmoontlik is. Eerstens, 'n aansienlike hoeveelheid van die vertaalwerk in Suid-Afrika word deur vryskutvertalers uitgevoer en deur individue wat tussendeur hul hooftake ook vertaling doen. Hierdie sektor van die vertaalmark is moeilik omlynbaar of sistematies ondersoekbaar. Tweedens, in teenstelling met die infrastruktuur wat ten opsigte van vertaaldienste vir die amptelike tale bestaan (bv. die Staatstaaldiens, die Parlement se taalkantoor, die Poskantoor, die Weermag, Vervoerdienste, UNISA se vertaalafdeling, die SAUK se taaldienste vir Engels en Afrikaans), vind vertaalaktiwiteite ten opsigte van die Afrikatale op 'n verspreide wyse plaas, byvoorbeeld in kantore in administratiewe liggame in die swart state, in howe dwarsdeur die land, by die verskeie radiostasies en by TV2 en 3, by opleidingsentra, en in die handel en nywerheid. Dit is dus 'n sektor wat hom nie tot sistematiese ondersoek leen nie.~~

Vir die doeleindes van die ondersoek kon daar egter twee afleidings met betrekking tot die bogenoemde onomlynde sektore gemaak word. Eerstens, daar kan aangeneem word dat rekenaargesteunde vertaling nog nie in hierdie sektore gebruik word nie, want daar is in die ondersoek gevind dat selfs die groot taalkantore in die land nog nie verder gekom het as die rekenarisering van woordelyste en die gebruik van die rekenaar gewoon as woordverwerker nie. Tweedens, as gevolg van faktore soos die snelle uitbreiding van die swart verbruikersmark, van die media in Afrikatale en van die deelname van die swart bevolking in administratiewe en ander liggame, kan daar aangeneem word dat die vertaalmark in Suid-Afrika snel toeneem. Voorheen was Suid-Afrika wat betref die geskrewe taal in 'n groot mate 'n tweetalige land, maar nou word dit multitalig, gevolglik brei die aanvraag na vertalings uit.

Selfs by die groot taalkantore was dit in sommige gevalle moeilik om presiese gegewens in te win, veral ten opsigte van die jaarlikse omset. Die 25 taalkantore waar gestruktureerde onderhoude gevoer is, sluit die hoof staatstaalkantore (Staatstaaldiens, die Weermag, die Poskantoor, die Parlement, Vervoerdienste), 'n aantal semi-staatsliggame, provinsiale en munisipale kantore, nywerhede, finansiële instansies en die SAUK in.

Uit Figuur 4.1 hieronder kan daar afgelei word dat die 25 kantore - wat 'n snit van die Suid-Afrikaanse vertaalmark verteenwoordig - saam tussen 500 000 en een miljoen bladsye jaarliks vertaal (dalk meer, omdat die totale omset by die SAUK 'n groot onbekende faktor is). Konserwatief bereken, verteenwoordig hierdie snit van die vertaalmark 100 000 mandae, dus ongeveer 5 000 manjare per jaar. Selfs 'n besparing deur rekenaarondersteuning van 10-20% sou 'n mannekragbesparing van 500-1 000 manjare per jaar meebring (net in hierdie snit van die bedryf). Mannekragbesparing is in die vertaalsektor van kernbelang, eendersyds omdat die bedryf feitlik uitsluitlik uit gegradueerdes bestaan (vgl. Pinchuck 1986), andersyds omdat stygende loonkoste vertaaldienste ál duurder maak. Bogenoemde syfers is vanselfsprekend projeksies van ordegroottes.



Figuur 4.1

Taal- kantore volgens grootte	Aantal ver- talers	Gebruik ook vryskut	Bladsye per jaar	Hooftale	Ander tale	Aard
1	±500	JA	Onbekend	A-E,E-A	Afrikatale Europatale	1. Algemeen 2. Tegnies
2	50	JA	70 000+	A-E,E-A	Afrikatale Europatale	1. Tegnies 2. Algemeen
3	50	JA	±15 000	A-E,E-A		Tegnies
4	±30	Onbekend	24 000	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
5	22	JA	16 000	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
6	19	NEE	±100 000+	A-E		Algemeen
7	15	NEE	Onbekend	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
8	12	JA	Onbekend	F-E	Europatale	Tegnies
9	12	JA	Onbekend	A-E,E-A	A-E,E-A Europatale	1. Tegnies 2. Algemeen
10	11	NEE	20 000	A-E,E-A	Europatale <del>Afrikatale</del>	1. Algemeen <del>2. Tegnies</del>
11	10	JA	13 000	A-E,E-A	Afrikatale Europatale	1. Algemeen 2. Tegnies
12	7	NEE	Duisende	E-A	Europatale	Tegnies
13	7	NEE	Onbekend	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
14	7	NEE	8 500	A-E,E-A		Algemeen
15	7	NEE	2 000	A,E,D		1. Algemeen 2. Tegnies
16	5	JA	4 000	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
17	5	NEE	4 000	A-E,E-A		1. Tegnies 2. Algemeen

18	5	NEE	Onbekend	A-E,E-A		1. Tegnies 2. Algemeen
19	4	JA	Onbekend	E-E	E-A	1. Tegnies 2. Algemeen
20	4	NEE	3 500	Europa- tale-E	A-E,E-A	1. Tegnies 2. Algemeen
21	2	NEE	Onbekend	A-E,E-A		1. Algemeen 2. Tegnies
22	2	NEE	4 000	A-E,E-A		1. Tegnies 2. Algemeen
23	1	NEE	Onbekend	A-E	E-A	Algemeen
24	1	JA	3 500	A-E,E-A	Europatale	1. Tegnies 2. Algemeen
25	1	NEE	Onbekend	A-E,E-A		1. Tegnies 2. Algemeen

#### Afkortings

A=Afrikaans

E=Engels

F=Frans

D=Duits

#### Nota

A-E,E-A beteken dat die kantoer meer uit Afrikaans in Engels vertaal as andersom (E-A,A-E)

#### 4.1.2 Die aard van die vertaalmark en die tale wat figureer

Uit die gegewens van Figuur 4.1 blyk dat vertaling tussen die twee amptelike tale die grootste gedeelte van die vertaalomset verteenwoordig, en dat meer uit Afrikaans as uit Engels vertaal word. Daar kan egter nie uit die oog verloor word dat die 25 kantore met die uitsondering van UNISA en vier finansiële instansies almal aan staats- of staatsverwante liggame gekoppel is nie, en dus hoofsaaklik die amptelike tale gebruik. Die indruk dat die Afrikatale in die vertaalmark 'n onbeduidende rol speel (behalwe by die SAUK - kantoer 1 op die tabel) is waarskynlik misleidend.

Waar daar in die figuur na die tale van Europa verwys word, gaan dit (in dalende volgorde) hoofsaaklik oor Portugees, Frans, Duits, Spaans en Italiaans, maar daar is ook 'n klein aanvraag na vertalings uit ander tale, byvoorbeeld Russies, Japannees, Chinees, Koreaans, en Indiese tale.

Ten opsigte van die rubriek 'tegnies', toon besonderhede in die opname dat, met die uitsondering van die SAUK, SABS, UNISA, die SAW, die-RGN en die Staatstaaldiens, die kantore se tegniese vertalings tot die bepaalde vakgebied waarmee die instansie hom besig hou, beperk is. Hierdie bevinding is van belang omdat ervaring in die buiteland getoon het dat vertaalprogramme wat op een bepaalde vakgebied afgestem is die suksesvolste funksioneer (bv. CULT, vgl. 3.7). (Nota: die WNNR se vertaalkantoor is nie in die ondersoek ingesluit nie as gevolg van die herstrukturering wat ten tye van die RGN-ondersoek plaasgevind het.)

Die rubriek 'algemeen' in die figuur verwys in die breë na die taal wat in korrespondensie, administratiewe en ander bedryfstukke gebruik word, met die uitsondering van die Parlement (kantoor 6 op die tabel) waar dit feitlik uitsluitlik gaan oor die vertaling van gesproke taal. Vertaalprogramme kan algemene taal doeltreffend hanteer mits dit dokumente van roetine aard is. Volksraadsdebatte is egter van die moeilikste stof om te vertaal omdat die inhoud wydreikend is, die registerspektrum groot is, en veral omdat die sprekersdoel dikwels oorreding eerder as inligtingsverskaffing is. Vanselfsprekend is die sintaksis van besprekingstaal - onvoltooide sinne, herhalings, spronge, en so meer - 'n groot struikelblok vir meganiese ontleding.

#### 4.1.3 Die stand van rekenarisering in die Suid-Afrikaanse vertaalwese

Die 25 kantore is gevra of

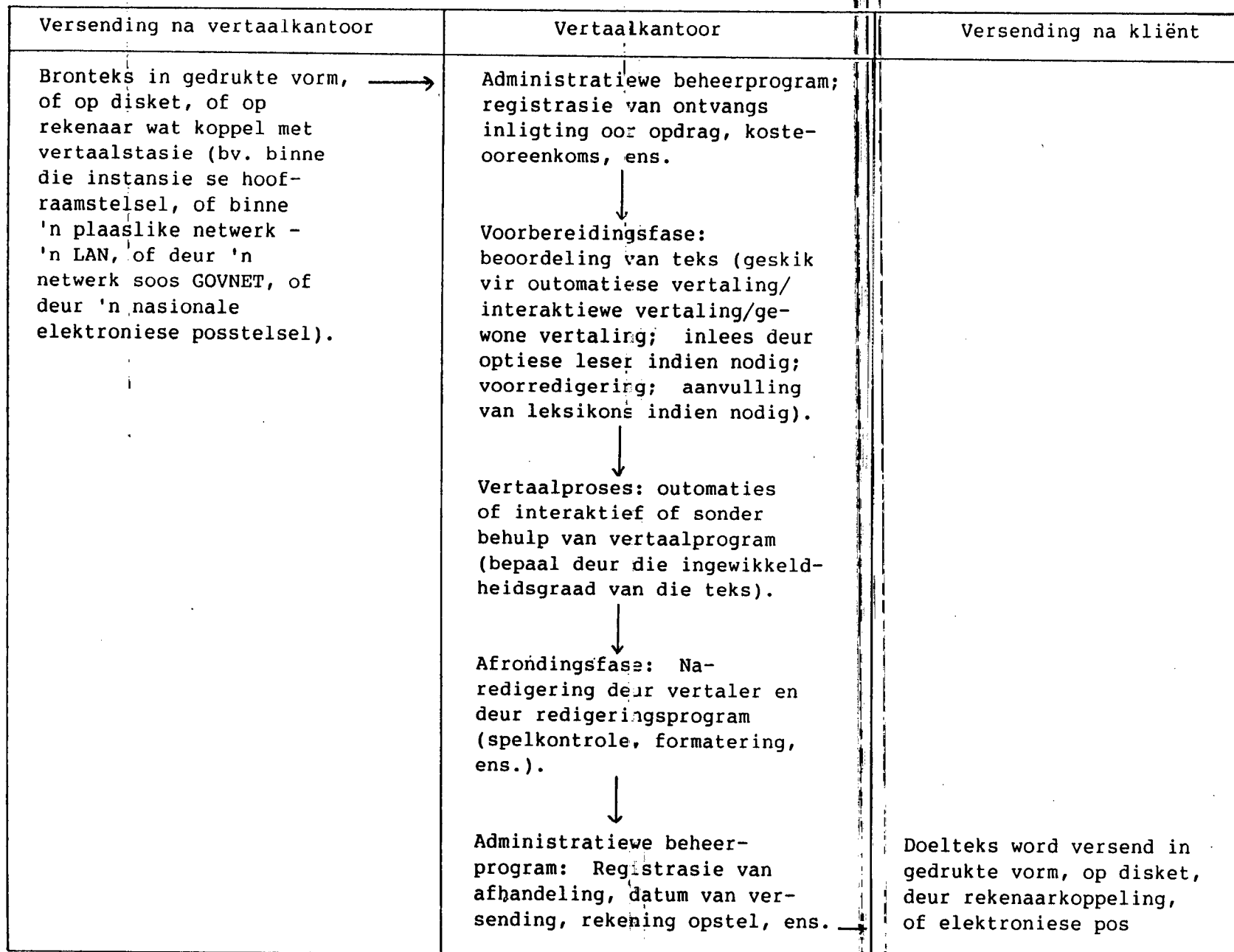
- (a) vertaalprogrammatuur hoegenaamd in die kantoor gebruik word;
- (b) die vertalers vertalings op rekenaarskerm doen, en, indien ~~nie, watter werkswyse gevolg word;~~
- (c) woordelyste gerekenariseer word;
- (d) die kantoor aan 'n plaaslike of groter netwerk gekoppel is;
- (e) die kantoor in beginsel die idee van rekenaargesteunde vertaling sou ondersteun.

Volgens die response gebruik geen kantoor tans vertaalprogrammatuur nie, maar by drie kantore is daar reeds oorweeg om ondersoek na beskikbare programmatuur in te stel. Woordverwerkers word by 13 van die 25 kantore gebruik terwyl 6 van diktering gebruik maak of met die hand skryf, en by 6 kantore word alle vertalings met die hand gedoen. Gerekenariseerde woordelyste word by 13 van die kantore opgebou; 5 kantore is reeds gekoppel, of sal binnekort gekoppel wees aan 'n netwerk, twee aan interne netwerke, drie aan GOVNET. Op een uitsondering na, steun die kantore geesdriftig die idee van rekenaargesteunde vertaling; die uitsondering is 'n een-vertaler kantoor waar daar blykbaar gevoel word dat die omset nie rekenarisering regverdig nie. Bostaande bevindinge toon dus dat

- (a) vertaalprogrammatuur nog nie in vertaalkantore gebruik word nie, maar dat daar 'n positiewe belangstelling daarin bestaan;
- (b) die rekenaaromgewing, wat noodsaaklik is rondom die gebruik van vertaalprogrammatuur, wel in vertaalkantore aan die ontwikkel is; netwerke, wat noodsaaklik is vir die spoedige hantering van dokumente, en vir woordeskat hulp (bv. koppeling aan 'n termbank), word daargestel.

Figuur 4.2 gee 'n voorstelling van die gerekenariseerde diens wat rondom die gebruik van vertaalprogrammatuur nodig is om te verseker dat die voordele verbonde aan rekenaargesteuende vertaling nie as gevolg van vertraging in die kantoorstelsel prysgegee word nie.

Figuur 4.2 SKEMA VAN 'N GEREKENARISEERDE VERTAALDIENS



Tydens besoeke aan 'n aantal vertaalkantore is daar vasgestel dat daar oor die algemeen betreklik min aandag aan die daarstelling van doeltreffende vloeisisteme bestee word. Verder blyk dit duidelik dat beskikbare fasiliteite onderbenut word, byvoorbeeld stukke vir vertaling word in gedrukte vorm deurgestuur terwyl dit regstreeks van die bronkantoor na die vertaalkantoor oorgespoel kon word. In die buiteland geniet hierdie aspek van vertaaldienste groot aandag, in dié mate dat selfs vryskutvertalers toenemend van elektroniese invoer en afvoer (deur elektroniese posstelsels) gebruik maak.

Die laaste deel van die ondersoek na die stand van sake met betrekking tot rekenaargesteunde vertaling in Suid-Afrika was 'n vraelysopname aan die universiteite om vas te stel of daar enige navorsing of ontwikkeling op die gebied onderneem word. Daarvolgens blyk dat hoewel vertaalprogrammatuur as sodanig nie ontwikkel word nie, daar wel in enkele departemente (by drie universiteite) aan ontleders (*parsers*) gewerk word, waarvan een, 'n morfologiese ontleder vir Afrikaans, sodanig omvattend is dat dit moontlik as komponent in 'n volwaardige vertaalstelsel vir Afrikaans sou kon dien.

Die RGN se Afdeling Leksikologie het onlangs 'n vertaalprogram wat tot die subtaal van weerberiggewing beperk is, voltooi. Die klein woordeskate, en betreklike eenvoudige sinsbou in weertaal bring mee dat 'n volwaardige sintaktiese ontleder nie nodig is nie. Hierdie vertaalprogram is alreeds vanaf Julie 1988 by die Suid-Afrikaanse Weerburo geïnstalleer. Dit vertaal weerberigte uit Engels in Afrikaans. Die volgende is 'n voorbeeld van 'n ongeredigeerde vertaling:

1. PRETORIA, WITWATERSRAND EN DIE OOSTELIKE HOEVELD:-  
MOOIWEER EN WARM MAAR WORD GEDEELTELIK BEWOLK MET ENKELE DONDERBUIE OOR DIE SUIDELIKE DELE VANMIDDAG EN VANNAG. DIT SAL BAIE WARM WORD OOR DIE PRETORIA RUSTENBURG GEBIED MORE.

2. WES EN SUID-WES TRANSVAAL:-  
MOOIWEER EN WARM MAAR WORD GEDEELTELIK BEWOLK VANMIDDAG MET ENKELE DONDERBUIE OOR SUIDWES-TRANSVAAL. DIT SAL MORE WARM WEES.

3. ORANJE-VRYSTAAT:-  
GEDEELTELIK BEWOLK EN WARM MET ENKELE DONDERBUIE VANMIDDAG EN VANNAG. DIT SAL MORE OPKLAAR.

Ander programme is by die RGN in 'n ontwikkelingstadium.

Opsommend kan daar na aanleiding van die ondersoek na die vertaalwese in Suid-Afrika gesê word dat

(a) die bedryf omvangryk is en uitbrei;

- (b) 'n aansienlike hoeveelheid van die stof wat vertaal word, van tegniese of roetine aard is, dit wil sê die kategorie wat die beste deur vertaalprogrammatuur hanteer word;
- (c) rekenaargestesteunde vertaling nêrens in die Suid-Afrikaanse vertaalwese voorkom nie, maar die rekenaaromgewing wat rondom vertaalprogrammatuur wenslik is, is by die groot vertaalkantore aan die ontwikkel;
- (d) houdings teenoor rekenaargestesteunde vertaling is positief;
- (e) buiten 'n embrio-program by die RGN word daar blykbaar nog nie in Suid-Afrika vertaalprogrammatuur ontwikkel nie (maar die moontlikheid dat daar oor hierdie saak dalk geheimhouding bestaan, is nie uitgesluit nie).

#### 4.1.4 Slotopmerkings

In die geheel beskou, tref dit dat daar belangrike ooreenkomste tussen die Suid-Afrikaanse taal- en vertaalsituasie en die situasie in die EEG (vgl. 3.10 en 3.11) bestaan. Hoewel die totale gemeenskap van die 9 lidlande van die EEG die bevolking van Suider-Afrika tans ver oortref, is die multitalige aard van die situasies vergelykbaar en so ook die feit dat die taal- en staatsgrense nie in 'n een-tot-een verhouding tot mekaar staan nie. Leersame insigte kan ten opsigte van kommunikasie- en samewerkingsbehoefte en -moontlikhede uit die EEG-voorbeeld verwerf word.

In die laaste hoofstuk van hierdie verslag word enkele aanbevelings met betrekking tot die moontlike ontwikkeling van vertaalprogrammatuur in Suid-Afrika aan die hand gedoen.

## 4.2 DIE KOSTE-EFFEKTIWITEIT VAN REKENAARGESTEUNDE VERTALING

Daar is 'n aantal veranderlikes wat in berekening gebring moet word om die koste-effektiwiteit van rekenaargestesteunde vertaling vas te stel. Wanneer 'n instansie byvoorbeeld beweer dat produksie met X% gestyg het en koste met Y% gedaal het, kan die stelling nie sonder bykomende inligting geëvalueer word nie. Hoe goed was die vorige produksie-tempo? Watter ander aspekte van die vertaaldiens het ook verander toe vertaalprogrammatuur in gebruik geneem is? Dit is bekend dat die oorskakeling na rekenaargestesteunde vertaling in vertaalkantore dikwels gepaard gaan met heropleiding en met veranderings in die diensvloei-sisteem. Dit kan ook gebeur dat die kantoor vir sommige doeleindes (bv. interne dokumente) vertalings net vinnig nareidigeer (vgl. 2.3), en sodoende kan produksie 'abnormaal' styg.

Die drie hoofaspekte van die finansiële implikasies van rekenaarge-steunde vertaling wat vervolgens bespreek word, is

- Ontwikkelingskoste
- Implementeringskoste
- Bedryfskoste

#### 4.2.1 Ontwikkelingskoste

Die beskrywing van oorsese stelsels in Hoofstuk 3 toon dat die ontwikkeling van vertaalprogramme besonder duur is. EUROTRA (vgl. 3.11), wat weliswaar verreweg die grootste projek is, kos meer as 40 miljoen rand in die eerste fase (7 jaar vir die ontwikkeling van 'n 9-talige prototipe, wat dan uitgebrei sal word). LOGOS (vgl. 3.8), wat relatief gesproke nie ingewikkeld is nie, en streng taalpaarspesifiek is, het teen 1987 reeds 600 manjare aan ontwikkelingswerk gekos, met as teenprestasie vier programme: Engels na Duits, Frans en Spaans, en Duits na Engels. Die ALPS-maatskappy (vgl. 3.4) berig dat die basiese ontwikkelingswerk vir sy interaktiewe stelsel 450 manjare gekos het. Die TAUM-METEO-stelsel in Kanada (vgl. 3.2), wat slegs weerberigte hanteer, is oor twee jaar ontwikkel, deur 'n span wat egter reeds ongeveer tien jaar van universiteitsnavorsing oor verskeie aspekte van die rekenarisering van taal agter die rug gehad het.

Vir Suid-Afrika is die eerste vraag wat ontstaan of programme van die buiteland ingevoer kan word, sodat basiese ontwikkelingswerk vermy of verminder kan word. Aangesien daar in die buiteland nie programme vir Afrikaans (of Afrikatale) bestaan nie, sou 'n Engels-Duitse program in beginsel die nuttigste aankoop wees, met die moontlikheid om dan die Duitse komponent met 'n Afrikaanse program te vervang. Die programme wat in die handel beskikbaar is, word egter nie verkoop nie. Dit word uitgehuur, onder streng voorwaardes ten opsigte van kopiereg, instandhouding deur die vervaardiger, en so meer. Die gebruiker verkry nie toegang na die binne-stelsel nie en mag die program nie wysig nie (afgesien van toevoegings tot die leksikons). Die uitsondering hierop is SYSTRAN, maar dit is 'n stelsel met groot nadele in 'n multitalige situasie (vgl. 3.10.4).

Een van die medewerkers in die RGN-ondersoek het by die ALPS-maatskappy die moontlikheid van die ontwikkeling van 'n Afrikaanse program bespreek. Die maatskappy se (informele) voorstel was dat Suid-Afrika ontwikkelingskoste van etlike honderdduisende rand voorsien, en op eie koste 'n paar Afrikaanse taalkundiges vir ongeveer 'n jaar na Switserland stuur om aan die ALPS-span die taalkundige inligting te verskaf. Die produk, wat ALPS-besit bly, kan dan deur Suid-Afrikaanse instansies gehuur word. In ag genome die relatiewe eenvoudigheid van die program, lyk dit verstandiger om fondse wat moontlik beskikbaar gestel kan word, plaaslik vir die opbou van eie kundighede en produkte aan te wend. Waar dit reeds kleinskaals gedoen word (vgl. 4.1), is die eerste resultate belowend. Een van die groot nadele van die ALPS-aanbod is die feit dat die Suid-Afrikaanse taalkundiges slegs taalkundige



insette sou lewer, sonder om insae in die 'binnewerkinge' van die ALPS-stelsel te verkry.

By nie-kommersiële ondernemings in die buiteland kan daar egter heelwat geleer word. Tydens oorsese besoeke het die skrywer van hierdie verslag gevind dat sommige sentra bereid is om gasnavorsers by hul projekte te betrek en ook om dokumentasie en programmatuur op informele vlak uit te ruil. Daar is egter gevind dat kort besoeke aan projekte nie die nodige geleentheid bied om 'n program te leer ken nie. Elke vertaalprogram het sy eie formalismes en koderings, wat eers deeglik bestudeer moet word.

Dit is noodsaaklik dat geleenthede vir gasnavorsing in die buiteland geskep word. Dit is ook belangrik dat Suid-Afrikaanse navorsers opleiding in die rekenaarlinguistiek ontvang. Die mees koste-effektiewe wyse om dit te bewerkstellig, terwyl die rekenaarlinguistiek nog nie in Suid-Afrika gevestig is nie, is deur bywoning van die Linguistics Society of America se jaarlikse uiters intensiewe ses-weeklange Summer Institute.

Benewens befondsing vir gasnavorsing en studie, sou 'n ontwikkelingsprojek vir 'n Afrikaanse vertaalprogram behels dat 'n span van 5-10 (afhangend van die omvang van die beoogde program) minstens vir 'n periode van vyf jaar geborg moet word. Hierdie projek is konserwatief gebaseer op spansituasies soos by METAL (vgl. 3.13), SUSY (vgl. 3.12) en EUROTRA (vgl. 3.11).

Wat betref strategieë vir die loodsing en befondsing van navorsing en ontwikkeling op die gebied van rekenaargesteunde taalwerking (en die ~~kunsmatige intelligensie-toepassings wat daarmee saamhang~~), kan daar veral na Japan se beleid gekyk word. Japan het die terrein relatief laat betree, gedeeltelik omdat Japannees as skryftaal besondere eise aan rekenaarprogramme stel. Vandag is Japan 'n voorloper, en die grootste besteder op die gebied (Hutchins 1986:313).

Die Japannese beleid is op twee hoofbeginsels gegrond: ten einde agterstande vinnig in te haal, is die gebied van kunsmatige intelligensie en taalprosessering tot nasionale prioriteit verklaar (die sogenaamde Fifth Generation Project). Tweedens, ten einde alle beskikbare hulpbronne saam te snoer, word al die basiese ontwikkelingswerk op die begrip 'pre-competitive research' gegrond. Hiermee word bedoel dat die owerheid, die privaatsektor (onder meer die groot rekenaarmaatskappye soos Hitachi), en die universiteite 'n ooreenkoms gesluit het om tot op 'n bepaalde punt aktief saam te werk en alle navorsingsinligting te deel. Dieselfde beginsel geld in die ALVEY-program in die VK (vgl. die AILANG-verslag).

Die hoofpunte in voorgaande bespreking kan soos volg opsommend gestel word:

- Indien vertaalprogrammatuur vir Afrikaans en die ander inheemse tale ontwikkel word, sal dit plaaslik gedoen moet word.
- Met die oog op die vestiging van die gebied in Suid-Afrika, sal daar 'n tydlank op buitelandse kennis gesteun moet word; studie- en gasnavorsingsgeleentheid is dus noodsaaklik.
- 'n Ontwikkelingsprojek vir 'n Afrikaanse vertaalprogram sal na raming minstens 5 jaar duur.
- Gesamentlike navorsing op nasionale en pre-kompeterende grondslag is 'n beleid wat in Japan, en elders, slaag.

#### 4.2.2 Implementeringskoste

By die implementering van rekenaargesteuende vertaling in 'n vertaaldiens is die drie hoofkoste die aankoop van apparatuur, die lisensie-gelde vir die gebruik van die programmatuur, en die uitbouing van die woordeboeke.

##### (a) Apparatuur

Die omvang van kapitaalbelegging in apparatuur hang vanselfsprekend van die bestaande rekenaaromgewing af. Al die groot taalkantore in Suid-Afrika is byvoorbeeld in liggame wat reeds rekenaardienste bedryf, gesetel, en al hoe meer vertalers word van mikrorekenaars voorsien (vgl. 4.1). Ander apparatuur is nie nodig nie, maar netwerkkoppeling kan die vertaaldiens as sodanig verbeter (vgl. Tabel 4.1).

##### (b) Programmatuur

Soos reeds genoem in 4.2.1 verkoop kommersiële vervaardigers van vertaalprogrammatuur die produkte nie. Dit word volgens 'n lisensie-ooreenkoms verhuur. Die ooreenkoms kan voorsiening maak vir instandhouding en opdatering. Dit kan opsies bied ten opsigte van watter woordeboeke verskaf word, en dit kan ook koste per tyds- of gebruikseenheid insluit.

Aangesien daar in hierdie stadium vir Suid-Afrika se tale nie programme beskikbaar is nie, kan daar nie voorspel word wat programmatuur-lisensies sal kos nie. In beginsel word dit bereken volgens die ontwikkelingskoste, die markgrootte en 'n persentasie vir verdere ontwikkeling.

(c) Die woordeboeke

Normaalweg vind gebruikers dat hulle die woordeboeke wat met die vertaalprogram verskaf word, moet aanvul. Dit geld veral gebruikers wat op bepaalde tegniese gebiede werk. Die aanvullingswerk is tydrowend en dit kan meebring dat die vertaalprogram redelik lank nie optimaal funksioneer nie.

In die literatuur word dit beklemtoon dat die woordeboekfasiliteit in die program van kardinale belang is. Indien die toevoeging van nuwe inskrywings ingewikkeld is, kan die doeltreffendheid van die program ernstig benadeel word. Van Slype (1982) wys byvoorbeeld daarop dat terwyl 'n nuwe inskrywing in die Frans-Engelse woordeboek van die EEG-SYSTRAN-stelsel gemiddeld 15 minute in beslag neem, was die TAUM-Aviation-stelsel so ingewikkeld dat elke inskrywing meer as drie uur geneem het. Dit was dan ook een van die redes vir die staking van die projek (vgl. 3.3).

4.2.3 Bedryfskoste

Die basiese algemene formule wat in die literatuur vir die berekening van bedryfskoste in die vertaalwese voorgestel word (vgl. Datta 1985, Thouin 1982, White 1986, Pigott 1986, Van Slype 1982, Slocum 1984), bestaan uit die omsetsyfer (aantal bladsye, van ongeveer 300 woorde, jaarliks vertaal) gedeel deur uitgawes (salarisse, oorhoofse koste, waardevermindering van kapitaalitems en, waar programme gehuur word, lisensie-gelde).

Bogenoemde skrywers wys daarop dat betroubare direkte vergelykings tussen nie-rekenaargesteunde vertaling en rekenaargesteunde vertaling ~~nie getrof kan word nie. Die eerste rede hiervoor is dat min, indien enige, instansies al die veranderlikes akkuraat kan bepaal. Tweedens, betreklik min vertalers doen slegs vertaling. Redigering, terminologiewerk, die beantwoording van navrae, en dalk ook algemene kantoortake, kan 'n beduidende deel van die dagtaak uitmaak. Die rekenaar, daarenteen, kan in beginsel met 'n outomatiese program soos METAL (vgl. 3.13) deurlopend vertaal en per 24 uur honderde bladsye ruwe vertalings lewer. Indien heelwat kliënte van die betrokke vertaaldiens met ruwe, of vinnig geredigeerde vertalings tevrede is, sal die jaarlikse omset vanselfsprekend baie hoog wees. Die gehalte van die produkte kan egter nie regstreeks met vertalers se produkte vergelyk word nie.~~

Die volgende is voorbeelde van koste-vergelykings wat in die literatuur aangetref word. Ten spyte van die onvolledigheid van die inligting wat verskaf word, en dus die probleem van behoorlike evaluering, tref dit tog dat die basiese stelling dat rekenaargesteunde vertaling baie koste-effektief is, blykbaar nêrens in die literatuur bevraagteken of weerspreek word nie.

- (a) EEG-SYSTRAN (vgl. 3.10)  
Volgens Pigott (1986) kom dit voor asof die stelsel produksie kan verviervoudig en die totale koste met ten minste 50% kan laat daal.
- (b) LOGOS (vgl. 3.8)  
Volgens 'n berig in *Language Monthly* (December 1983:13) het die stelsel al met sy eerste implementering in 1980 koste met 60% laat daal. Die stelsel se werkverrigting verbeter steeds.
- (c) METAL (vgl. 3.13)  
Volgens Slocum (1984) kon 'n vertaler reeds in 1984 met die ondersteuning van die stelsel 30 bladsye per dag lewer.

Pigott (1986) skryf die florerende mark vir produkte soos ALPS, Weidner-CAT, LOGOS en SYSTRAN, en die toenemende aantal nie-kommersiële ondernemings, hoofsaaklik aan die effektiwiteit van rekenaargestesteunde vertaling toe. Hy beklemtoon ook die feit dat vertalers duurder word en vertaalprogramme goedkoper.

## HOOFSTUK 5: SAMEVATTING EN AANBEVELINGS

### 5.1 AGTERGROND

Hoofstuk 1 van hierdie verslag lig die agtergrond en doelstellings van die ondersoek toe. Die hoofpunte daarin kan soos volg saamgevat word:

- Rekenaargesteunde vertaling geniet in die buiteland groot aandag, maar in Suid-Afrika is dit blykbaar 'n feitlik onbekende gebied.
- Eerstens was die doel met die ondersoek om inligting oor oorsese projekte in te win.
- Tweedens was die doel om die potensiaal van rekenaargesteunde vertaling in die Suid-Afrikaanse vertaalwese in oënskou te neem.
- Laastens moes aanbevelings oor moontlike ontwikkelingstappe gedoen word.

### 5.2 OORSIG VAN DIE GEBIED

Hoofstuk 2 bied 'n kort geskiedenis van die gebied, 'n bespreking van die wyses waarop rekenaargesteunde vertaling die vertaler se tradisionele taakverrigting raak, en 'n aantal voorbeelde van rekenaarvertaalde tekste. Die hoofpunte kan soos volg saamgevat word:

- ~~Navorsing in die vyftiger- en sestigerjare het betreklik min~~ benutbare resultate gelewer, met die gevolg dat befondsing bykans orals gestaak is. In die laat sewentigerjare het daar 'n oplewing gekom en dit het daartoe gelei dat gerekenariseerde vertaalstelsels vandag in talle lande ontwikkel en geïmplementeer word.
- Die vertaler se taak word as gevolg van rekenaarsteun taalkundig meer gespesialiseer en bedryfsmatig meer gesofistikeer, terwyl die sleurwerk in vertaling grootliks uitgeskakel word.
- Die voorbeeldmateriaal toon hoofsaaklik dat vertaalprogramme nie hoë gehalte vertalings lewer nie, maar dat die vertalings vinnig deur naredigering op standaard gebring kan word.

### 5.3 OORSESE VERTAALSTELSELS

Hoofstuk 3 bied eers 'n kort uiteensetting van verskillende benaderings tot stelselontwerp vir vertaalstelsels, waarna 'n twaalftal oorsese programme beskryf word. Hierdie programme is van uiteenlopende aard met betrekking tot doelstellings, toepassingsgebiede en ontwerp. Dit bied ook voorbeelde van verskillende modelle vir samewerking en vir finansiering. Hier word enkele hoofaspekte van die gebied as geheel uitgelig:

#### 5.3.1 Rekenaargesteunde vertaling as ontwikkelingsgebied

- Wêreldwyd is daar intense aktiwiteit op die gebied en honderde miljoen dollar word tans daaraan bestee.
- Die tale waarvoor programmatuur ontwikkel word, neem jaarliks in getal toe. Afgesien van Russies, Engels en agt Europese tale, sluit dit reeds Arabies, Japannees en Chinees in, asook etlike 'nie-internasionale' tale soos Maleis, Hongaars, Viënamees, Grieks en Fins.
- Ontwikkelings op die gebied is snel en al meer gesofistikeerd, veral waar taalprosessering en kunsmatige intelligensie die terrein deel. Lande met 'n agterstand in kundigheid sal dit toenemend moeilik vind om die agterstand in te haal.
- Elke land sal daarop aangewese wees om self vir sy nie-internasionale tale programmatuur te ontwikkel. Ingevoerde programme kan nie gewoon aangepas word nie, soos in die geval van programme vir numeriese verwerking nie, want tussen tale verskil strukture diepliggend (dieper-liggend as wat by die aanvanklike rekenaargesteunde navorsing veronderstel is).
- Daar is tekens dat 'n internasionale talenewerke opbou. Tale wat nie tot hierdie elektroniese netwerk toegang verkry nie, sal as tegnologiese tale agterweë bly.

#### 5.3.2 Die aard en gehalte van bestaande vertaalstelsels

- Gerekenariseerde vertaalprogramme lewer oor die algemeen nie vertalings van hoë gehalte nie. Die uitsondering is stelsels wat op beperkte subtale gerig is, byvoorbeeld Kanada se weerberigprogram, TAUM-METEO, vgl. 3.2.
- Sommige programme voldoen egter uitstekend aan die doelstellings wat daarmee beoog is, byvoorbeeld ruwe vertalings vir grootskaalse inligtingherwinning in militêre kontekste (bv. SYSTRAN by die Amerikaanse lugmag, vgl. 3.10) vertaling van brontekste wat in vereenvoudigde taal geskryf is (bv. SYSTRAN by die Xerox-maatskappy, vgl. 3.1), inter-aktiewe vertaling (bv. ALPS vgl. 3.4).

- Die geskikste stof vir rekenaargesteuende vertaling is tekste in vaktiaal en roetine-administratiewe dokumente. --Dit-- word onder andere deur CULT (vgl. 3.7) en TITUS (vgl. 3.6) bewys.
- Die doeltreffendheid van vertaalprogramme kan deur inbedding in 'n sogenaamde vertalerswerkstasie verhoog word. Daarmee word bedoel dat daar rondom die program self verskeie tekshanterings- en naslaanfasiliteite ingebou word, bv. SUSANNAH, (vgl. 3.12) en METAL (vgl. 3.13).

### 5.3.3 Beplanning en finansiering

- Navorsings- en ontwikkelingswerk op die gebied van rekenaargesteuende vertaling vereis multi-dissiplinêre spanwerk, tussen linguïste, rekenaarkundiges, ingenieurs, wiskundiges, logici en kognitiewe sielkundiges.
- 'n Gebrek aan duidelike doelwitstelling, deeglike beplanning en deurlopende monitering het by sommige projekte in die buiteland tot die onttrekking van finansiering gelei (bv. TAUM-AVIATION, vgl. 3.3, en SUSY 1, vgl. 3.12).
- Finansiering: die ondersoek het uitgewys dat die vier florerende kommersiële stelsels, ALPS (vgl. 3.4), Weidner (vgl. 3.5), LOGOS (vgl. 3.8) en SYSTRAN (vgl. 3.10), almal op voorgaande navorsing, wat deur staatsgeld gefinansier is, gevestig is.

## 5.4 STANDBESKRYWING VAN DIE SUID-AFRIKAANSE VERTAALWESE

~~Hoofstuk 4.1 bied die bevindinge van die ondersoek na die Suid-Afrikaanse vertaalwese.~~

### 5.4.1 Die omvang van die vertaalmark

Die omvang van die mark kan nie presies bepaal word nie omdat vertaalaktiwiteite in 'n menigte organisasies as deel van die daaglikse bedryfsaktiwiteite plaasvind, en dus nie as 'n afsonderlike en kwantifiseerbare deel van die bedryf voorkom nie. Verder is 'n groot aantal vryskutvertalers in die land bedrywig (vgl. Pinchuck 1986). Gegewens oor die klaarblyklik toenemende hoeveelheid vertaling wat in die tuislande en die onafhanklike state gedoen word, is ook nie beskikbaar nie.

Die ondersoekgroep was 25 vertaalkantore en -eenhede in die staats- en semi-staatsektore, en enkele groot maatskappye. Op grond van hierdie beperkte snit van die vertaalmark is daar bevind dat meer as 'n miljoen bladsye teks jaarliks vertaal word. Dit verteenwoordig ongeveer 5 000 manjare per jaar van gegraduateerde mannekrag.

#### 5.4.2 Die aard van die mark

Die gegewens toon dat die oorgrote gedeelte van die stof of tegnies of roetine-administratief van aard is. Dit is die tipe teks wat vir rekenaargesteunde vertaling die geskikste is.

#### 5.4.3 Die tale wat vertaal word

Vanweë die amptelike taalbeleid is die vertaalverkeer in die sektor wat ondersoek is hoofsaaklik tussen Engels en Afrikaans. Die Afrikatale figureer sterk in die howe, en daar kan aangeneem word, op grond van die uitbreidende swart infrastrukture, media en verbruikersmark, dat dit elders ook 'n al groter rol in die vertaalmark sal speel. Europese en ander vreemde tale speel 'n klein rol in die mark as geheel, hoewel daar drie of vier groot instansies is wat invoer van vreemde tale verwerk, byvoorbeeld die SAUK.

#### 5.4.4 Die stand van rekenarisering in die vertaalwese

Die ondersoek het aan die lig gebring dat geen vertaalkantore oor vertaalprogramme beskik nie, maar woordverwerking en die rekenarisering van woordelyste neem toe. Koppeling in plaaslike of groter netwerke begin ontwikkel. Houdings teenoor rekenaargesteunde vertaling is baie positief.

Navorsing op die gebied is tot dusver in Suid-Afrika beperk tot die ontwikkeling van ontleders by die universiteite van Stellenbosch en Zululand, en by die RGN, waar daar ook onlangs vertaalprogramme vir die vertaling van weerberigte uit Engels in Afrikaans en Xhosa ontwikkel is.

### 5.5 DIE KOSTE-EFFEKTIWITEIT VAN REKENAARGESTEUNDE VERTALING

Hoofstuk 4.2 gee 'n kort uiteensetting van die koste-faktore wat by die ontwikkeling, implementering en bedryf van vertaalstelsels in berekening kom.

#### 5.5.1 Ontwikkelingskoste

Die reële koste van die ontwikkeling van 'n eerste program vir 'n taalpaar, byvoorbeeld Afrikaans-Engels, is moeilik berekenbaar omdat die enkele syfers wat in verband met oorsese projekte beskikbaar is, nie rekening hou met die aansienlike hoeveelheid basiese navorsing wat die projekte voorafgegaan het nie.

Alles in ag genome kan daar egter konserwatief geprojekteer word dat 'n span van 5-10 navorsers (afhangend van die omvang van die voorgestelde projek) minstens vir 5 jaar geborg sal moet word. Afgesien van die loonkoste van die navorsers, en moontlik besoeke aan oorsese projekte,



hoef daar nie groot uitgawes te wees nie. Die bestaande apparatuur in Suid-Afrika is voldoende.

Die ondersoek het getoon dat daar veral teen verbrokkelde navorsings- en ontwikkelingswerk gewaak moet word. Een of meer nasionale projekte, waarby kundiges uit verskillende oorde betrokke kan wees, is bo 'n reeks ongekoördineerde projekte by verskillende instansies verkieslik.

#### 5.5.2 Implementeringskoste

Die implementering van vertaalprogramme behels in beginsel drie kostekomponente, apparatuur, programmatuur en woordeboekbou.

Indien die vertaalkantoor reeds oor mikrorekenaars beskik, sal daar moontlik geen verdere kapitaaluitgawes nodig wees nie. Die koste verbonde aan programmatuur word deur verskeie faktore bepaal: die ontwikkelingskoste van die vertaalprogram, die potensiële markafset en die vervaardiger se winsmotief. Ten einde koste so laag moontlik te hou, sou dit in die Suid-Afrikaanse konteks sinvol wees as die staat- en semi-staatsektore gesamentlik die ontwikkeling van 'n program sou borg, op voorwaarde dat dit dan aan al die betrokke vertaalkantore beskikbaar gestel word.

Woordeboekbou is potensieel die duurste item. Elke kliënt moet die basiese woordeboek(e) wat met die vertaalprogram verskaf word, met instansie-spesifieke terminologie en ander woordeskat aanvul. Tensy die woordeboekinskrywingsfasiliteit baie doeltreffend is, en woordeboekaanvulling onder streng gehalte- en tydskontrolle geskied, kan hierdie aktiwiteit buitensporig duur word.

#### 5.5.3 Bedryfskoste

Ten einde koste-effektiwiteit in vertaalkantore te bereken, is die basiese formule dat die totale oorhoofse koste gedeel word deur die aantal bladsye wat gelewer word. Volgens hierdie formule kan rekenaargesteurde vertaling beslis as koste-effektief beskou word. METAL (vgl. 3.13) lewer byvoorbeeld in 'n werksdag van 8 uur ongeveer 200 bladsye, wat teen 'n tempo van 30 bladsye per nareidigeerder per dag afgewerk kan word.

#### 5.6 AANBEVELINGS

Op grond van die voorgaande samevatting van die hoofpunte in hierdie verslag word die volgende aanbevelings gedoen:

Aanbeveling 1 : Rekenaarondersteuning in vertaalwerk

Daar word aanbeveel dat die rekenaar veel meer in die vertaalwese gebruik word, op wyses wat nou reeds moontlik is:

- As woordverwerker, met teksredigeringsprogramme, en met verdeling van die skerm indien die bronteks in masjienleesbare vorm is.
- As woord- en termverskaffer, deur middel van gerekenariseerde leksikons (vgl. die SAW se stelsel).
- As stoorder van vertaalde materiaal waarvan gedeeltes hergebruik kan word, byvoorbeeld jaarverslae wat in sekere gedeeltes net syfer- of ander klein veranderings in die volgende weergawe bevat.
- As ontvangs- en versendkanaal vir vertaalstukke binne 'n organisasie.

Aanbeveling 2 : Rekenaargesteunde vertaling

Daar word aanbeveel dat 'n Suid-Afrikaanse navorsingsprogram vir die ontwikkeling van vertaalprogrammatuur geloods word. Die volgende voorstelle kan moontlik as riglyne dien:

- Die navorsingsprogram sou 'n nasionale program kon wees, gesamentlik beplan en uitgevoer deur navorsers uit verskillende dissiplines en van verskillende instansies.
- Binne die oorkoepelende programraamwerk sou daar beide lang- en korttermyn projekte uitgevoer kon word.
- Die baie langtermyn doelwit van 'n vertaalstelsel vir elke taal, kan in fases bereik word en onderdele van die beoogde stelsels kan deur verskillende spanne ontwikkel word.
- Korttermyn projekte sou uit die tussentydse produkte van die langtermyn projek kon voortvloei, byvoorbeeld subtaalstelsels soos CULT, TITUS en METEO, die inbou van gespesialiseerde leksikons in woordverwerkingspakette, vraag-en-antwoord stelsels, en so meer.
- Deeglike beplanning soos in die geval van EUROTRA, asook deurlopende monitering, is noodsaaklik.
- Daar kan uit oorsese projekte veel geleer word, veral ten opsigte van ontwerpaspekte soos modulariteit en uitbreibaarheid.

- Finansiering : die Japannese voorbeeld van gesamentlike befondsing van die pre-kompeterende ontwikkelingswerk deur die staat, die privaatsektor en navorsingsinstansies verdien oorweging.

### Aanbeveling 3 : Die opbou van kundigheid

Daar word aanbeveel dat kundigheid opgebou word op die volgende wyses:

- Binne die voorgestelde navorsingsprogram moet daar sover moontlik in spanne gewerk word sodat die huidige min kundigheid ten beste uitgebou kan word; die spanne sou as 'n werkgemeenskap kon funksioneer wat, byvoorbeeld jaarliks, of meer gereeld, werkswinkels hou.
- As deel van die voorgestelde navorsingsprogram moet belowende jong navorsers opleiding in die buiteland ontvang, byvoorbeeld deur intensiewe opleiding in die rekenaarlinguistiek by die VSA se jaarlikse Summer Institute. Gasnavorsing by die norsese projekte kan ook gereël word.
- Indien kundiges uit die buiteland gewerf kan word, moet die besoeke verkieslik op spesifieke taakverrigting of opleiding afgestem word, eerder as op algemene lesings en besprekings.
- Die gebrek aan opleidingsgeleenthede in Suid-Afrika op die gebied van die rekenaarlinguistiek verdien aandag van die universiteite se taaldepartemente; afgesien van die behoefte in die land aan kundigheid ten opsigte van die ~~rekenaarverwerking van natuurlike taal, bied hierdie rigting~~ aansienlike "groeimoontlikhede" vir die linguistiek as studierigting.

### 5.7 SLOTOPMERKING

In die loop van die ondersoek, en in die groter raamwerk van die LEXINET-ondersoek na rekenaargesteuende taalverwerking, het dit al duideliker geword dat die vertrekvraag, naamlik of rekenaargesteuende vertaling in Suid-Afrika aandag behoort te geniet, oorgegaan het na die vraag van hoe die land sy agterstand kan inhaal. Met inagneming daarvan dat agterstande op gevorderde wetenskaplik-tegnologiese gebiede betreklik vinnig oninhaalbaar kan word, verdien die saak prioriteits-aandag en dinamiese optrede.

Dit is op verskillende vlakke - ekonomies, tegnologies, polities, wetenskaplik, strategies - belangrik dat Suid-Afrika se tale nie van die talentewerke wat internasionaal opbou, uitgesluit word nie. Dit is moontlik nog belangriker dat Suid-Afrika die geleentheid aangryp om die voortou te neem in die uiteindelijke daarstelling van 'n Suider-Afrika

talenetwerk op die patroon van die Europese term- en vertaalnetwerke. Klaarblyklik is die middele tot grootskaalse ontwikkeling nie tans beskikbaar nie, maar vanaf die eerste stappe wat geneem word, kan daar met vrug rekening gehou word met die moontlikheid van 'n netwerkstelsel vir die sub-kontinent.

BYLAE

Chronologische tabel van die hoofmasjienvertalingsprojekte en -stelsels

Datums	Plek	Naam	Projek- leiers	Metode	Tale
49-60	U.Washington		Reifler	dir	RE, (GE, CE)
50-60	RAND Corp.		Harper	dir	RE
			Hays		
51-65	MIT		Bar-Hillel	tr	GE
			Yngve		(F, A, C, Fin)
52-63	Georgetown U.	GAT	Dostert	dir	RE
			Zarechnak		RE
		-SERNA	Toma		
		-SLC	Brown		FE
54-64	Harvard U.		Oettinger	dir	RE
			Kuno		(JE)
54-67	CLRU		Masterman	int	E, I, L
55-59	Birbeck C.		Booth	dir	FE, GE
55-67	Ramo-W.		Swanson	dir	RE
			Garvin		
55-62	U.Michigan		Koutsoudas	dir	RE
55-64?	ITVMT Moscow		Panov	dir	
			Bel'skaya	dir	ER
			Nikolaeva		RE
<del>55-66</del>	<del>MIAN-Moscow</del>		<del>Lyapunov</del>		
			Kulagina	dir	FR
			Moloshnaya	tr	ER
55-65?	Inst.Ling Moscow		Reformatskii		FR
			Mel'chuk	(int)	(HR)
57-?	1st Moscow State Ped.Inst.		Revzin		(E, R, S, R)
			Rozentsveig		
57-66	Prague		Sgall	(int)	ECz
58-67?	Leningrad		Andreev	int	mult
			Zasorina	dir	RE
58-70	ETL Tokyo	Yamato	Wada, Tadenuma	dir	EJ, (JE)
58-64	IBM	Mark I, II	King	dir	RE, (FE, CE)
58-64	UC Berkeley		Lamb	(dir)	RE, (CE)
58-72	Wayne State U.		Josselson	dir	RE
58-68	LRC Texas		Lehmann, Tosh	tr	GE(RE, F, S, J)
59-66	Milan		Ceccato	int	RE, (I, G, L)
59-63	NBS		Rhodes	dir	RE
59-66	NPL			dir	RE
61-71	U.Grenoble	CETA	Vauquois	int	RF, (GF, JF)
63-77	Patent Off. Moscow		Vasilevskii	tr, rst	ER

Datums	Plek	Naam	Projek- leiers	Metode	Tale
64-	Latsec/WTC	Systran	Toma	dir	RE EF,FE,EI,ens.
64-76	Bonn	LIMAS	Hoppe	int	EG,GE
65-81	U.Montreal	TAUM(-AV) METEO	Kittredge	tr dir,rst	EF EF
65-76	Moscow	MTM	Mel'chuk	int	ER
66-76	Leningrad		Fitialov	tr	ER
67-	Saarbr@ucken	SUSY	Maas Luckhardt	tr	RG E,F,J,Esp
67-75	CETIS Ispra	EURATOM	Perschke	tr	RE
67-75	UC Berkeley	POLA	Wang	tr	CE
69-78	Logos Corp.	LOGOS	Scott	dir	EV,etc.
69-	Hong Kong	CULT	Loh	IA,dir	CE
70-74	Stanford U.		Wilks	AI	EF
70-75	LRC Texas	METAL	Lehmann Stachowitz	int	GE,(RE)
70-79	Brigham Y.U.	ITS	Lytle	IA,int	mult
70-	Paris	TITUS	Ducrot	rst	E,F,G,S
71-	U.Grenoble	GETA -ARIANE	Vauquois Boitet	tr	RF,(E,G, C,J,P,Mal)
73-	U.Heidelberg	SALAT		int	E,F,G
74-	Moscow	AMPAR NERPA FRAP	Marchuk Marchuk	dir dir tr	ER GR FR
75-	Leningrad		Piotrowski	dir,rst	ER,JR
76-	PAHO	SPANAM ENGSPAN	Vasconcellos Leon	dir dir	SE ES
76-	Prague		Kirschner	tr	ECz
77-	Chicago	Weidner		IA,dir	A,E,F,G I,J,P,S
77-	New York	Smart	Smart	IA,rst	EF
78-	CEC	Eurotra	Perschke	tr	mult
78-	LRC Texas	METAL	Lehmann Slocum	tr	GE
78-82	Yale		Carbonell	AI	ER,ES
		MOPTRANS	Lytinen	AI	SE,FE
79-	BSO Utrecht	DLT	Witkam	int	(Esp)
80-	Provo(Utah)	ALPS		IA	EF,EG,EI,ES
80-	Kyota	TITRAN Mu	Nagao Nagao	dir,rst tr	EJ EJ,JE
81-	Hitachi	ATHENE		tr	EJ,JE
81-	Fujitsu	ATLAS		AI,IA	JE,JG
81-	Osaka			tr	EJ,JE
81-	NTT	LUTE	Nomura	AI	JE,EJ
82-	Logos Corp.	Logos		tr	GE,(EG)
83-	Colgate U.	TRANSLATOR	Tucker	AI	E,S,R,J

Afkortings

Metodes:

dir=direk, int=intertalig, tr=oordrag, IA=interaktief, rst=beperkte taal, AI=kunsmatige intelligensie

Tale:

C=Chinees, Cz=Tsjeggies, E=Engels, Esp=Esperanto, F=Frans, Fin=Fins, G=Duits, H=Hongaars, I=Italiaans, J=Japannees, L=Latyn, Mal=Maleisies, mult=veeltalig, P=Portugees, R=Russies, V=Viëtnamees

(Hutchins 1986 : 336-337)

## BIBLIOGRAFIE

ALPS-inligtingstukke. 1987.

ARNOLD, D. 1986. General view of the design methodology. *Multilingua* 5-3 (1986), pp 136-138.

BACK, T. & F.-J.M. Wilms. 1985. Machinelle Übersetzung: SUSY - Das Saarbrücker Übersetzungssystem. *Mitteilungsblatt* 2/31 März/April 1985, pp 5-11.

BENNETT, W.S. 1982. The Linguistic Component of METAL. Working Paper LRC-82-2. University of Texas.

BENNETT, W.S. and J. Slocum. 1985. The LRC Machine Translation System. *Computational Linguistics* Vol. 11, No 2-3, April-September 1985.

CARBONELL, J., R. Cullingford and A. Gershman. 1981. Knowledge-based machine translation. *IEEE Trans. on Pattern Anal. Mach. Intell.* 3,4 (1981), pp 376-392.

CARBONELL, J.G. and M. TOMITA. 1987. Knowledge-based machine translation, the CMU approach. In Nirenburg, S. 1987. *Machine Translation. Theoretical and methodological issues* ct. Cambridge : Cambridge University Press.

CRANFIELD CONFERENCE ON MACHINE TRANSLATION. 1984. *Language Monthly* (April 1984), p 14.

DATTA, J. 1985. Machine translation: the trade-offs and the pay-backs. *Language Monthly* (May 1985), pp 8-9.

DES TOMBE, L. 1986. The < C,A >, T model for translation. In *Multilingua* 5-3 (1986), pp 139-140.

EVANS, A.D.C. 1986. SYSTRAN - The Translator's Viewpoint. *Terminologie et Traduction* 1 (1986), pp 17-23.

HABERMANN, F.W.A. 1986. Provision and Use of Raw Machine Translation. *Terminologie et Traduction* 1 (1986), pp 29-42.

HARTMANN, R.R.K.: 1983. *Lexicography: Principles and Practice*. London: Academic Press.

HOMER, J. 1986. Improving machine translation quality or eliminating the 'Invisible idiot'. *Terminologie et Traduction* 1 (1986). pp 118-145.

HUNDT, M.G. 1982. The Weidner Machine-Aided Translation System. In Lawson, V. 1982. *Practical Experience of Machine Translation*. Amsterdam: North-Holland Publishing Company.

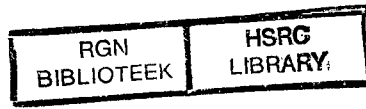


- HUTCHINS, W.J. 1982. The Evolution of Machine Translation Systems. In Lawson, V. 1982. *Practical Experience of Machine Translation*. Act Amsterdam: North-Holland Publishing Company.
- HUTCHINS, W.J. 1986. *Machine Translation: Past, Present, Future*. Chicester: Ellis Horwood.
- JASPAERT, L. 1986. The levels of representation. *Multilingua* 5-3 (1986), pp 141-145.
- KINGSCOTT, G. 1984. How many words an hour? *Language Monthly* (Nov. 1984), p 26.
- KITTREDGE, R.I. 1981. The Development of Automated Translation Systems in Canada. *Lebende Sprachen*, Vol. 26 (3) (1981), pp 100-103.
- LADD, P. 1986. Pre-editing of machine translation input: An experimental evaluation. *Multilingua* 5-4 (1986), pp 217-224.
- LAU, P. 1985. Elements of EUROTRA Today. *Terminologie et Traduction* No. 3 - 1985, pp 31-39.
- LAWSON, V. 1982. *Practical Experience of Machine Translation*. Amsterdam: North-Holland Company.
- LEHMANN, W.P. and W.S. Bennett. 1985. Human Language and Computers. *Computers and the Humanities*. Vol. 19. No 2. April-June 1985, pp 77-83.
- LUCKHARDT, H.-D. 1987. Von der Forschung zur Anwendung: das ~~Computergestützte-Saarbrücker-Translationssystem STS.~~ In Zimmerman, H.H. 1987. *Veröffentlichungen der Fachrichtung Informationswissenschaft*. Saarbrücken: Der Universität des Saarlandes.
- MAAS, H.-D. 1985. *Linguistische Arbeiten: SUSY-II-Handbuch*. Heft 14. Saarbrücken: universität des Saarlandes.
- MASTERMAN, M. 1982. The limits of innovation in machine translation. In Lawson, V. 1982. *Practical Experience of machine Translation*. Amsterdam: North-Holland Publishing Company.
- McNAUGHT, J. 1986. An appropriate environment for software construction. *Multilingua* 5-3(1986), pp 168-170.
- MORRIS, R. 1986. Putting machine translation into perspective. Paper prepared for the LEXINET seminar, 10 September 1986.
- NORWEGIAN SYSTEM READY FOR TESTING. 1987. *Language Monthly* (Nov. 1987), p 6.

- PERSCHKE, S. 1986. EUROTRA: General overview. *Multilingua* 5-3(1986), pp 134-135.
- PIGOTT, I.M. 1981. How does Systran translate? *Terminologie Bulletin* No. 40 (1981), pp 59-78.
- PIGOTT, I.M. 1983a. Systran: A key to overcoming language barriers in Europe. *Multilingua* 2-3 (1983), pp 149-156.
- PIGOTT, I.M. 1983b. Recent developments in practical MT. Commission of the European Communities, Luxembourg, Nov. 1983.
- PIGOTT, I.M. 1984. Systran Machine Translation at the EC Commission: Present Status and History, April 1984.
- PIGOTT, I.M. 1986. Current SYSTRAN Developments at the EC Commission. *Terminologie et Traduction* 1 (1986), pp 10-16.
- PIGOTT, I.M. 1987a. Systran Machine Translation at the EC Commission: Present Status and History, September 1987.
- PIGOTT, I.M. 1987b. Systran Machine Translation Processing as an example of Natural Language Understanding. Commission of the European Communities, Luxembourg, 1987.
- PINCHUCK, I. 1986. The training of translators in South Africa. Pretoria: RGN-kantoorverslag 1986-6.
- ROUVRAY, D. and G. Wilkinson. 1984. Machines Break the Language Barrier. *New Scientist* 22, March 1984, pp 19-21.
- ROUX, J.C. 1988. Kunsmatige intelligensie en die prosessering van natuurlike taal. Verslag LEXI-8. Pretoria: Raad vir Geesteswetenskaplike Navorsing.
- SCHNEIDER, T. 1987. METAL, An Operative Machine Translation System. Paper delivered at the International Conference on Chinese Information processing, Beijing, 1987.
- SIEBENALER, L. 1986. Systran for Esprit and Ecat bureau service. *Terminologie et Traduction* 1 (1986), pp 54-60.
- SLOCUM, J. and W.C. Bennett. 1982. The LRC machine translation system. Working Paper LRC-81-1. University of Texas, Austin, July 1982.
- SLOCUM, J. 1984. Machine Translation : its history, current status, and future prospects. Working Paper LRC-84-3. University of Texas.
- SLOCUM, J. 1984. METAL: The LRC Machine Translation System. Working Paper LRC-84-2. University of Texas.

- SOMERS, H.L. 1986. The research groups of the EUROTRA project. In *Multilingua* 5-3 (1986), pp 131-134.
- SUSY-INLIGTINGSTUKKE Saarbrücken: Universiteit van Saarland, Augustus 1985.
- TECHNOLOGY. 1985. *Language Monthly*, (January 1985), pp 13-16.
- THOUIN, B. 1982. The METEO System. In Lawson, V. 1982. *Practical Experience of Machine Translation*. Amsterdam: North-Holland Publishing Company.
- TRANSLATING AND THE COMPUTER. 1983. *Language Monthly*, (December 1983), pp 13-14.
- TUCKER, A.B. 1984. A Perspective on Machine Translation: Theory and Practice. In *Communications of the ACM* Vol. 27, April 1984, No. 4, pp 322-329.
- VAN EYNDE, F. 1986. The interface structure level of representation. *Multilingua* 5-3(1986), pp 145-146.
- VAN SCHALKWYK, C. en Kruger, A. 1985. *Vertaalstudies, 1985-projekte*, Vol. 2, Dept. Linguistiek, UNISA.
- VAN SLYPE, G. 1982. Economic Aspects of Machine Translation. In Lawson, V. 1982. *Practical Experience of Machine Translation*. Amsterdam: North-Holland Publishing Company.
- WALKER, P.A. 1986. The importance of word processing in the User environment. *Terminologie et Traduction* 1 (1986), pp 102-111.
- WHITE, E.N. Using Controlled Languages for Effective Communication. *The Communicator...News Supplement 67B (June 1986)*, 1-3.
- WHITE, J.S. 1985. Characteristics of the METAL Machine Translation System at Production Stage. Working Paper LRC-85-04. University of Texas.
- WILKS, Y. 1979. Machine translation and artificial intelligence. *Translating and the Computer*, B.M. Snell, Ed. Elsevier North-Holland, New York, 1979, pp 27-43.

Doc no 206788  
Copy no 206791



R 14, 78