

APPENDICES

Tables and ATNs concerning the IL	1
Summary in French	16
Summary in German	22
Summary in Dutch	28
Summary in Danish	35
References	43
Glossary of abbreviations	58

Esperanto at a Glance

The Alphabet of Esperanto

A a B b C c Ĉ ĉ D d
 ah bo tso cho do
 E e F f G g Ĝ ĝ H h
 eh fo go Joe ho
 Ĥ ĥ I i J j Ĵ ĵ K k
 hho ee yo zho ko
 L l M m N n O o
 lo mo no oh
 P p R r S s Ŝ ŝ T t
 po ro so ŝo toe
 U u Ŭ ŭ V v Z z
 oo woe vo zo
 No Q, W, X, or Y.

A, E, I, O, U have approximately the vowel sounds heard in *last, make, marine, not, bull*.

C is not sounded like S or K, but like *ts* in *tses-ty, wits*.

J has the sound of *y* in *Yak*.

The sounds of Ĉ, Ĝ, Ĥ, Ĵ, and Ŭ are heard in *CHurCH, Join, loCH, aZure, SHow,* and *moUnt*.



ESPERANTO IS PHONETIC.

All letters sounded: one letter one sound.

ACCENT or STRESS falls on the last syllable but one.

No IRREGULARITIES. No EXCEPTIONS.

THE GRAMMAR is based upon SIXTEEN FUNDAMENTAL RULES, which have no exceptions.

THE PARTS OF SPEECH are formed from Root-Words by the addition of appropriate Letters.

O is the ending for all names of things (NOUNS) **A** ADJECTIVES (descriptive words) end in

<i>fakto</i>	<i>gluo</i>	<i>evidenta</i>	<i>freĝa</i>
<i>distanco</i>	<i>fajro</i>	<i>longa</i>	<i>furioza</i>
<i>piano</i>	<i>tasko</i>	<i>granda</i>	<i>simpla</i>

NOUNS and ADJECTIVES form PLURALS by adding **J**

<i>evidentaj</i>	<i>longaj</i>	<i>grandaj</i>
<i>faktoj</i>	<i>distancoj</i>	<i>pianoj</i> (<i>aj, oj</i> sound as in <i>my boy</i>)

THE SIMPLE VERB HAS ONLY SIX ENDINGS.

INFINITIVE	PRESENT	PAST	FUTURE	CONDITIONAL	IMPERATIVE
I	AS	IS	OS	US	U
ESTI	<i>estas</i>	<i>estis</i>	<i>estos</i>	<i>estus</i>	<i>estu</i>
LERNI	<i>lernas</i>	<i>lernis</i>	<i>lernos</i>	<i>lernus</i>	<i>lernu</i>
HELPI	<i>helpas</i>	<i>helpis</i>	<i>helpos</i>	<i>helpus</i>	<i>helpu</i>

N marks the ACCUSATIVE (*direct object*). **E** ADVERBS end in

<i>Mi (I) helpas lin (him)</i>	<i>energie</i>
<i>Li (he) helpas min (me)</i>	<i>entuziasme</i>
<i>Ŝi lernas Esperanton</i>	<i>diligente</i>

DLT's IL, which is entirely based on Esperanto, has an internal (BCE) and an external representation (CCE). For the latter, the above alphabet and diacritic signs will be used without changes, utilizing the extended ASCII (ISO/2022) character code and modern display hardware.

[In this report, the diacritic sign on the 'u' has been left out, because in fact this never leads to confusion].

Esperanto/IL one-word adverbs as opposed to multi-word units,
such as frequently occur in French, English and German.

aldone = en plus / par-dessus le marché # in addition @ zusätzlich
 aliflanke = de l'autre côté # on the other hand @ anderseits
 aliloke = autre part # elsewhere @ (irgend)wo anders / sonst irgendwo
 aliparte* = d'autre part # for the other part @ anderenteils
 alivorte = en d'autres mots # in other words @ mit anderen Worten /
 anders gesagt
 ankaŭe* = aussi # also @ auch
 cetere = du reste # in other respects / besides @ Übrigens / im
 Übrigen
 ĉi-konekse = à ce propos # in this connection @ anschliessend
 ĉi-rilate = à ce propos # in this respect @ in dieser Beziehung
 ĉi-sence = dans ce sens # in this sense @ in diesem Sinne
 ĉi-spite = nonobstant cela # nonetheless @ dagegen
 ĉiaokaze = en tout cas / quoi qu'il en soit # in any case / at any
 rate @ jedenfalls / auf jeden Fall
 ĉiuokaze = en tout cas / à tout propos # in any case @ jedenfalls
 do = alors/donc # then @ also
 dume = en attendant # meanwhile @ inzwischen
 ekzemple = par exemple # for example @ zum Beispiel
 escepte = par exception # by way of exception @ ausnahmsweise
 fine = enfin # finally @ schliesslich/endlich
 inverse = au contraire # on the contrary @ umgekehrt
 konklude = par conséquent # in conclusion / to conclude @
 konsekvence = par conséquence # in consequence @ folglich/demzufolge
 kontraste = par contraste # by contrast @ im Gegensatz dazu
 kontraŭe = au contraire # on the contrary @ im Gegenteil
 krome = en outre # moreover @ ausserdem
 laste = en dernier lieu / finalement # finally/lastly @ schliesslich
 male = au contraire # on the contrary / quite the opposite @ im
 Gegenteil / dagegen
 malgraŭe = néanmois # nevertheless @ dennoch/desungeachtet/trotzdem
 miaŭice = à mon tour # in my turn @ meinerseits
 nome = à savoir / c'est à dire # that is to say / i.e. @ nämlich
 nu = or/soit/alors # well @ nun
 plie = en plus # furthermore @ noch mehr
 plue = pour poursuivre # furthermore / to continue @ dazu/ausserdem
 poste = ensuite/puis # afterwards/then @ dann
 resume = en résumé / en bref # in summary / in brief @ kurz
 zusammengefasst
 rezulte = par conséquence # as a result / consequently @ folglich
 same = de même # in the same way @ auf gleiche Weise
 samokaze = par la même occasion # on the same occasion @
 sed = pourtant # however/but @ aber
 sekve = par suite # consequently @
 siaflanke = de son côté # for its/his/her part @
 siaŭice = à son tour # in its/his/her turn @ seinerseits
 simile = pareillement # similarly @ ähnlicherweise
 tamen = pourtant/cependant/néanmoins/toutefois # nevertheless/however
 @ trotzdem
 tiakaze = en tel cas # in such a case @ in solchem Falle
 tiamaniere = de telle façon # thus / in such a way @ auf solche Weise
 tiel = ainsi # thus / in this way @ auf diese Weise / so
 tiucele = à cette fin # to this end @ mit diesem Zwecke
 tiukaze = en ce cas # in that case @ in dem Fall / gesetzt den Fall
 tiukondiĉe = à cette condition # with this condition @ unter dieser
 Bedingung

	ĉirkaŭ NUM ĝis NUM NUM ĝis NUM inter NUM kaj NUM super NUM po NUM pli ol NUM malpli ol NUM	minimume NUM maksimume NUM proksimume NUM preskaŭ NUM apenaŭ NUM almonaŭ NUM ekzakte NUM certe NUM	antaŭ NUM NUM antaŭtaŭ NUM ne NUM NUM de NUM ne NUM sed NUM ne nur NUM sed NUM eble NUM eble NUM ajne NUM ajne NUM	ankoraŭ NUM ankoraŭ foje NUM nur NUM eble NUM ĉiutage NUM ĉiokaze NUM ĉiokaze NUM nepre NUM preferi NUM kiohence NUM
ankau	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ankoraŭ	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ankoraŭ ne	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ĵam	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ĵam ne	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ĉe	- + - - - +	- - - - - +	- + - - - +	- - - - - +
ĵe	+ + + + - +	+ + + + - +	+ + + + - +	+ + + + - +
ĉre	- - - - - +	- - - - - +	- - - - - +	- - - - - +
ĉro	- - - - - +	- - - - - +	- - - - - +	- - - - - +
pli	- - - - - +	- - - - - +	- - - - - +	- - - - - +
plej	- - - - - +	- - - - - +	- - - - - +	- - - - - +
malpli	- - - - - +	- - - - - +	- - - - - +	- - - - - +
nur	+ + + + - +	- - - - - +	+ + - - - +	- - - - - +
multe	- - - - - +	- - - - - +	- - - - - +	- - - - - +
kelke	- - - - - +	- - - - - +	- - - - - +	- - - - - +
sufiĉe	- - - - - +	- - - - - +	- - - - - +	- - - - - +
mal multe	- - - - - +	- - - - - +	- - - - - +	- - - - - +
neniom	- - - - - +	- - - - - +	- - - - - +	- - - - - +
ĉiom	- - - - - +	- - - - - +	- - - - - +	- - - - - +
iom	- - - - - +	- - - - - +	- - - - - +	- - - - - +

Compatibility matrix for numeral supplements and preceding intensifiers by which they can be modified.

	ne	ankau ec	mem je	tre tro pli plej	preskau apenau	nun nur	ankorau ankorau n	jam jam ne dume ofte	parte tute certe fakte fine leble	multe bone rapide malfacile	sufice
ne	+	-	-	+	-	+	-	+	+	+	+
ankau	+	•	-	-	-	+	+	+	+	+	+
ec	+	•	+	+	+	+	+	+	+	+	+
mem	-	-	•	-	-	-	-	-	-	-	-
je	-	-	-	•	-	+	-	+	+	+	-
tre	+	-	-	•	-	-	-	-	+	+	+
tro	-	-	-	•	-	-	-	-	+	+	+
pli	-	-	-	•	-	-	-	-	+	+	+
plej	-	-	-	•	-	-	-	+	-	+	+
preskau	+	-	-	+	•	-	+	-	+	-	+
apenau	+	-	-	+	•	-	+	-	-	-	+
nun	+	+	-	-	-	•	-	-	-	-	-
nur	+	-	-	-	-	•	-	-	+	-	+
ankorau	+	-	-	+	+	•	-	+	+	+	+
ankorau n	-	-	-	+	+	•	-	+	+	+	+
jam	+	-	-	+	+	•	-	+	+	+	+
jam ne	+	-	-	+	+	•	-	+	+	+	+
dume	+	-	-	-	-	-	-	•	-	-	-
ofte	+	-	-	-	-	-	-	•	-	-	-
parte	+	+	-	+	-	+	+	-	•	+	+
tute	+	-	-	-	-	-	-	+	•	+	+
certe	+	-	-	-	-	-	-	-	•	-	+
fakte	+	-	-	-	-	-	-	-	•	-	+
fine	+	-	-	-	-	-	-	-	•	-	+
leble	+	-	-	-	-	-	-	-	•	-	+
multe	-	+	-	+	-	-	-	-	-	•	-
bone	-	+	-	-	-	-	-	-	-	•	-
rapide	-	+	-	-	-	-	-	-	-	•	-
malfacile	-	+	-	-	-	-	-	-	-	•	-
sufice	-	+	-	+	-	-	-	-	+	+	•

Compatibility matrix for intensifiers and other adverbs, indicating to what extent these words can act as premodifiers for each other.

TEMPORAL	
<u>Time since</u>	de kiam iam-de komence komence da
<u>time till</u>	antaŭ bio, ke fine de ĝis ke iam-antaŭ iam-ĝis ĵus kiam foreverne
<u>time after</u>	antaŭ ol openaŭ ke baldaŭ kiam iam-post post kiam post ol post bio, ke ĉiutempe tuj kiam venontjare
<u>others</u>	alifoje aliam ankoraŭfoje anticipo ĉe ĉe bio, ke ĉi-foje ĉiufoje kiam ĉiufoje se deŭre de de tempo- al tempo dum dum ke dum bio, ke dum de foje iam iam-ĉirkaŭ iam-ekster iam-inter iam-tra ĉe ĵus kiam kiam ajn. kiel longe multfoje nokte potage se taje tagokte bel longe dum ke ĉiel longe kiel tuj unufoje unuavide.
LOCAL	
<u>inside / centre</u>	centre de ie-ĉirkaŭ ie-en ie-inter interne de kerne de meze de
<u>at top of</u>	super supraĵe, de supren sur
<u>under</u>	funde de malsupren sub sube de.
<u>direction</u>	al al kio, ke al kiu, ke al tru, kiu direkte al kien tien, kien
<u>others</u>	alflanke alie aliloke aparte apud ĉe ĉi-flanke cis de kie dekstre de bio, de kie, de kie dis ekstere de ol flanke de ĝis kie ĝis bio, ĝis kie ie ie-antaŭ ie-dal- ie-ekster ie-ĝis ie-kontraŭ ie-post ie-tra je (ĉiome) kie kie, ke laŭlonge de maldekstre nenie prober tie, kie trans unuflanke unuparta vidalvide al
PURPOSE	cele al favore al honore al kompenso al memore al por por ke por tro, ke prepare al renkonto al responde al ĉiucele
REASON	ĉar ke ial konsidere al miaflanke pro pro bio, ĉar ke proprave tial, ĉar ke tial ke
CAUSE	dank' al dank' al bio, ke al bio, ke el bio, kio hazarde helpo de bio, ke helpo de tru, kiu kaŭzo de konsidere al per per tro, ke pere de pere de bio, ke principe sekve de
AGENT	far fare de
WAY/MEANS	ĉiel ĉiuriske ĉiokore ĉiamaniere ial ĉiel-tiel interŝange kiel kiel ajn ĉiel, tiel kiela kiel, ke komunkonsente laŭ laŭ tio, ĉu laŭ tio, ke laŭ tio, kiel laŭ tio, kio laŭlibere man-en-mano parkere rabate surpapere tio, kiel ĉiamaniere ĉiamaniere, ke tiel tiel, ke ĉiom, kel tutkore vere
ACCOMPANIMENT	kun kun bio, ke kun bio, kun kio kun tru, kiu sen sen ke sen bio, ke
TRACTATIVE	concerne pri pri kio, ke pri kiu, ke pri bio, ĉu pri bio, ke pri bio, kiam pri bio, kiel pri bio, kio pri bio, pri kio
CONCESSIVE	kvankvam malgraŭ malgraŭ ke spite al spite ke.
IDENTIFYING	ekzemple estiel jene jensekve ke ke, se (-vo) kvazaŭ ke tio, ke
DEGREE	apenaŭ ekstermezure ĝis ekstreme grandparte iagrade iom iugrade ju pli, des malpli kiel longe kiam kiam ajn kiom de kiugrade laŭpove maksimume maltra minimume ol plejaltgrade plejparte pli - malpli tiagrade ĉiom, ke ĉiom, kiam ĉiom same, ke ĉiom same, kiam ĉiom same multe, kiam ĉiomgrade ĉiugrade tro
CONTRASTING	ĉetero kontraŭ male

(continued from page 5)

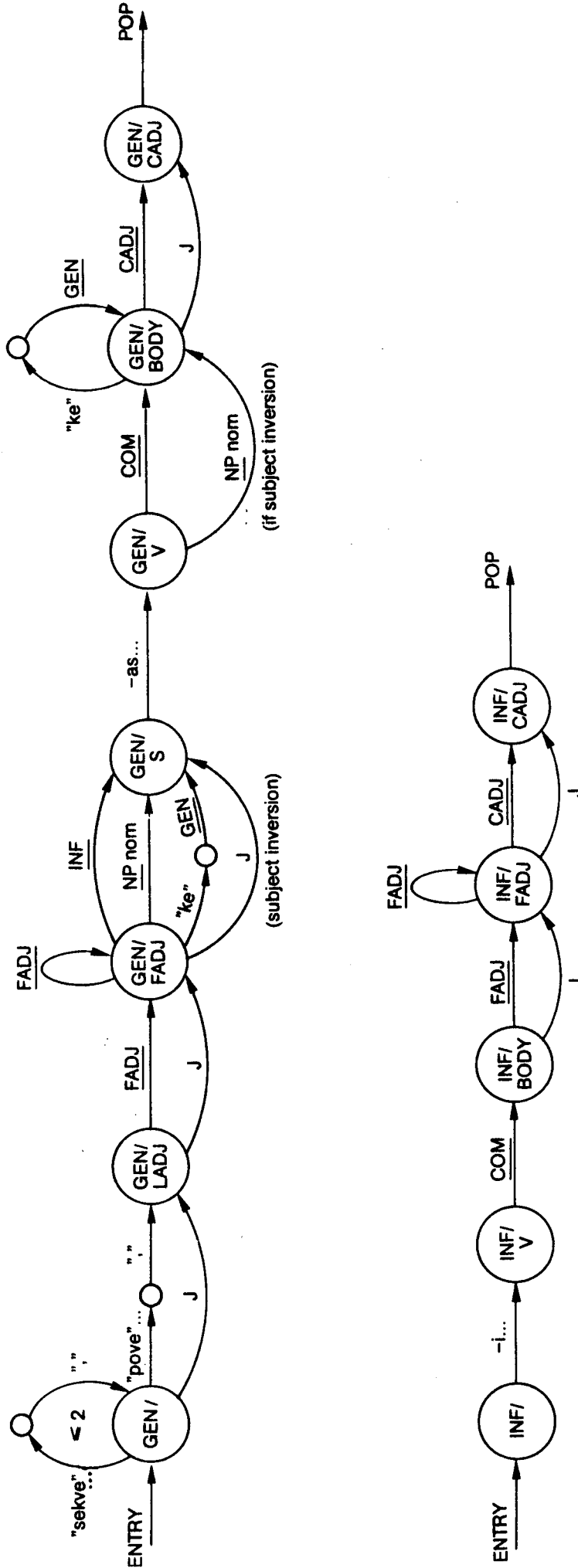
<u>CIRCUMSTANCE</u>	ĉe ĉeeste de ĉiukaze ĉiukaze espereble foreste de helpe de konsiderante krome malutile al manke de okaze de protektite de unuopa
<u>CONDITION</u>	ĉokaze ĉiufoje se ĉu, aŭ ĉu, ĉu eĉ se escepte se kaze de kiam (-us) kiel, se kiel, kiam (-us) kondiĉe ke krom se multkaze nenikaze neniam tiel, kiel se rezerve de se supoze ke tiukaze
<u>RELATION</u>	aliparte ĉi-riate ĉu de tiu, kiu de tio, ke de tio, kio je kio, ke kia kia ĝin kia-ĉu kies kies-ĉu kio kio-ĉu kiu kiu-ĉu komisia de konforme al konsekvence de kontraŭe de nome de riata al tio, ke tia, kia tio, kio tiu, kiu
<u>MODAL</u>	aparte ĉionkonsidere fakte kompreneble pretekste de rondcifere sendube senprokaste tra versojne butcerbe tute ne verdire.
<u>OTHERS</u>	anstataŭ anstataŭ tio, ke ekskluzive de escepte de escepte de tio, ke inkluzive de je tio, ke an tio, ke krom krom tio, ke kromaŭ kromaŭ tio, ke

ATN Arc Annotation Convention.

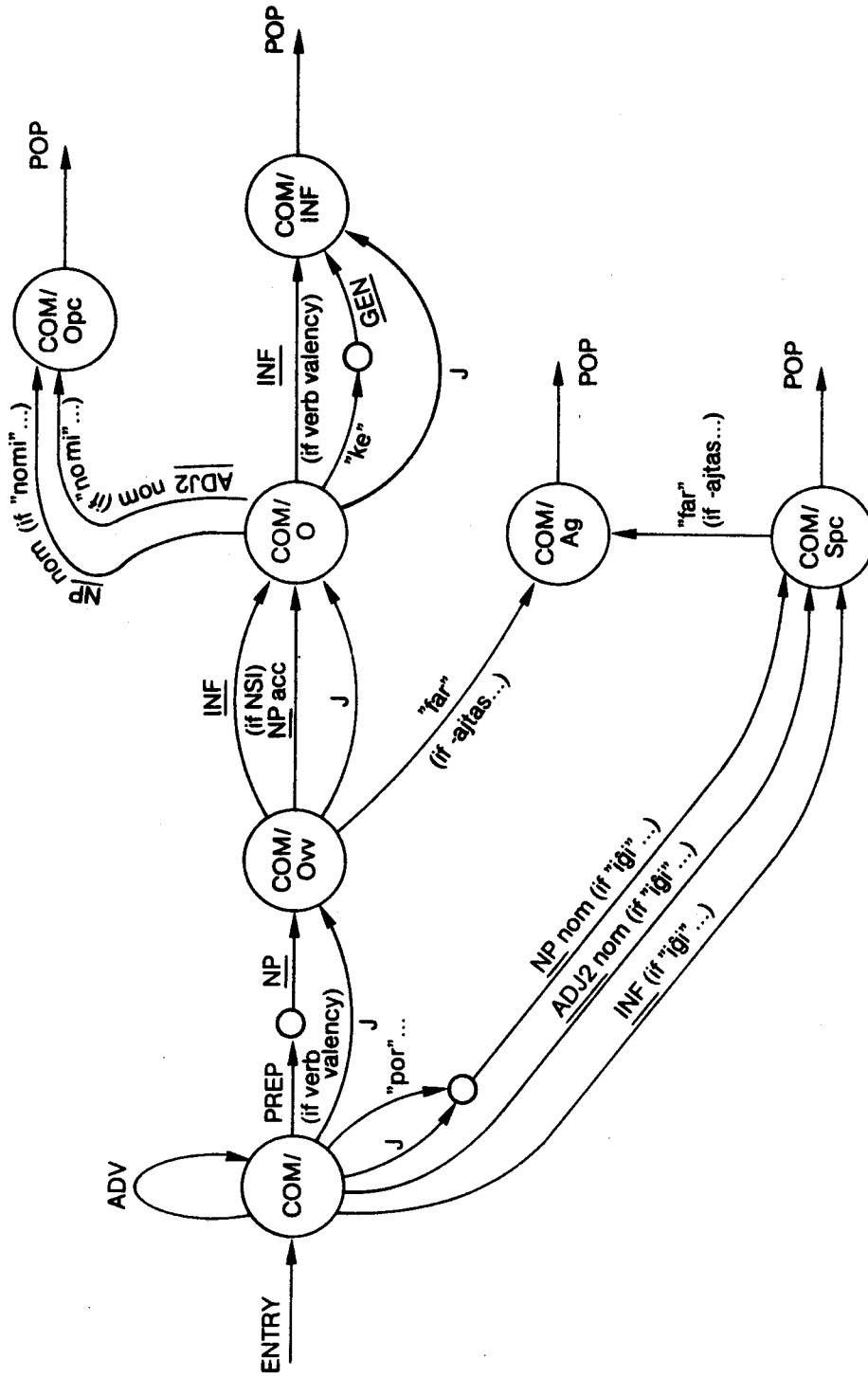
The following pages (pp. 8-15) contain the set of Augmented Transition Networks (ATNs) that finally define DLT's IL (one ATN, fig. IV-6, is contained in the body of the report.

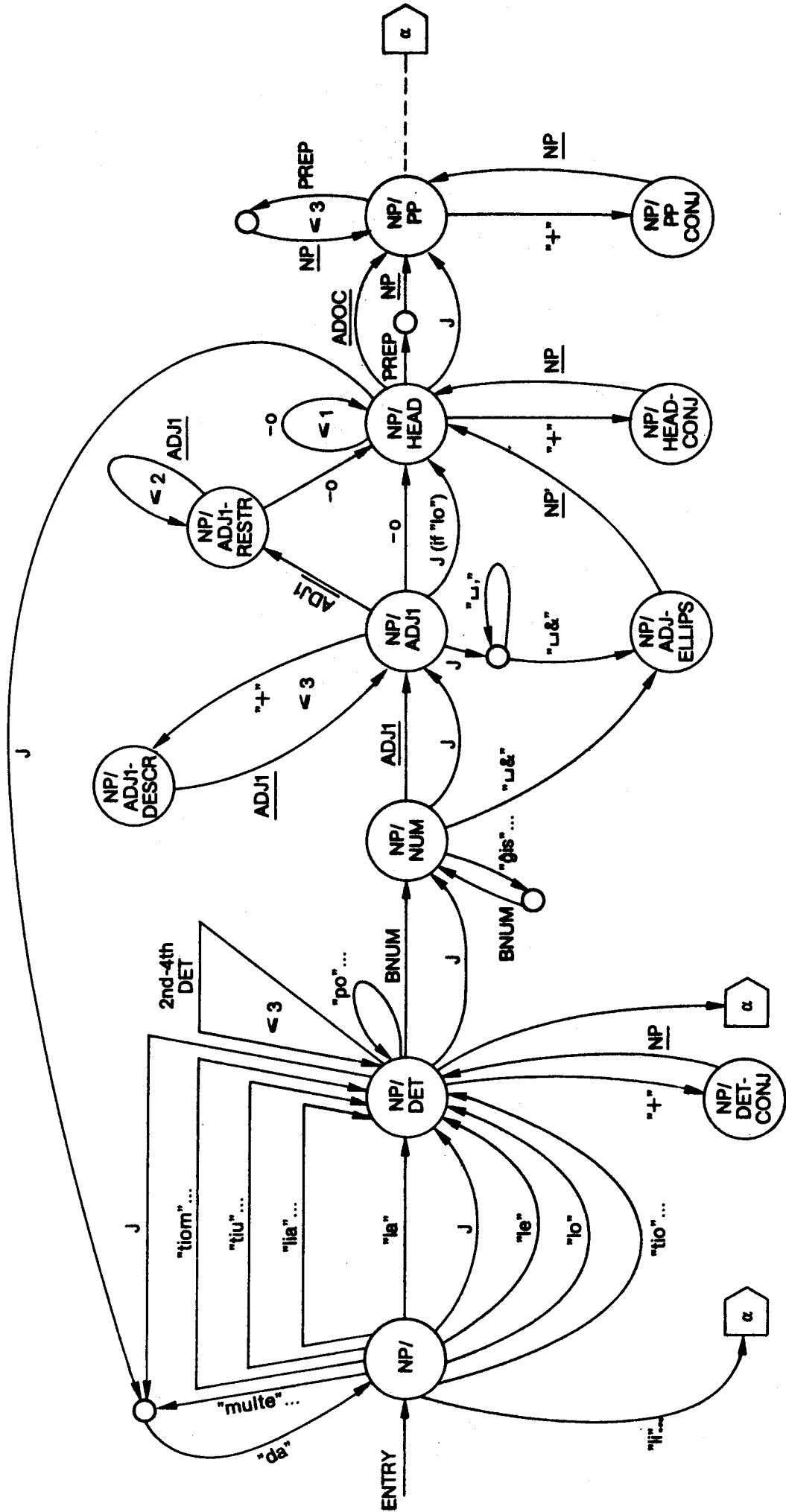
XYZ	CAT XYZ
-as	instead of the name of a grammatical category (CAT), its corresponding grammatical ending is indicated: -as = finite verb; -i = infinitive; -o = noun; -a = adjective; -e = adverb.
-as...	a list of grammatical endings (instead of CAT).
"xyz"	WRD xyz.
"xyz"...	MEM list "xyz"...
<u>XYZ</u>	PUSH XYZ (jump to subnetwork).
<u>XYZ'</u>	PUSH XYZ, with restrictions.
<u>XYZ_{nom.}</u>	PUSH XYZ for nominative case only.
<u>XYZ_{acc}</u>	PUSH XYZ for accusative case only.
(if "xyz")	condition referring to WRD, CAT etc. previously processed.
J	JUMP
"␣"	WRD ␣ ; this is the extra space, serving as a disambiguating element.
"&"	WRD & ; this stands for a conjunction (AND, OR, ...).
"+"	WRD + ; this stands for a conjunction, including a comma (, , AND, OR).
"␣&"	WRD ␣& ; a conjunction preceded by an extra space.
"␣,"	WRD ␣, ; a comma preceded by a space.
⏪2	up to 2 transitions after each other.

For the abbreviations of IL syntactic categories see p. IV-27.



ATNs for the general and the infinitival clause. Objects and complements are covered by the common subnetwork (COM) on the next page.



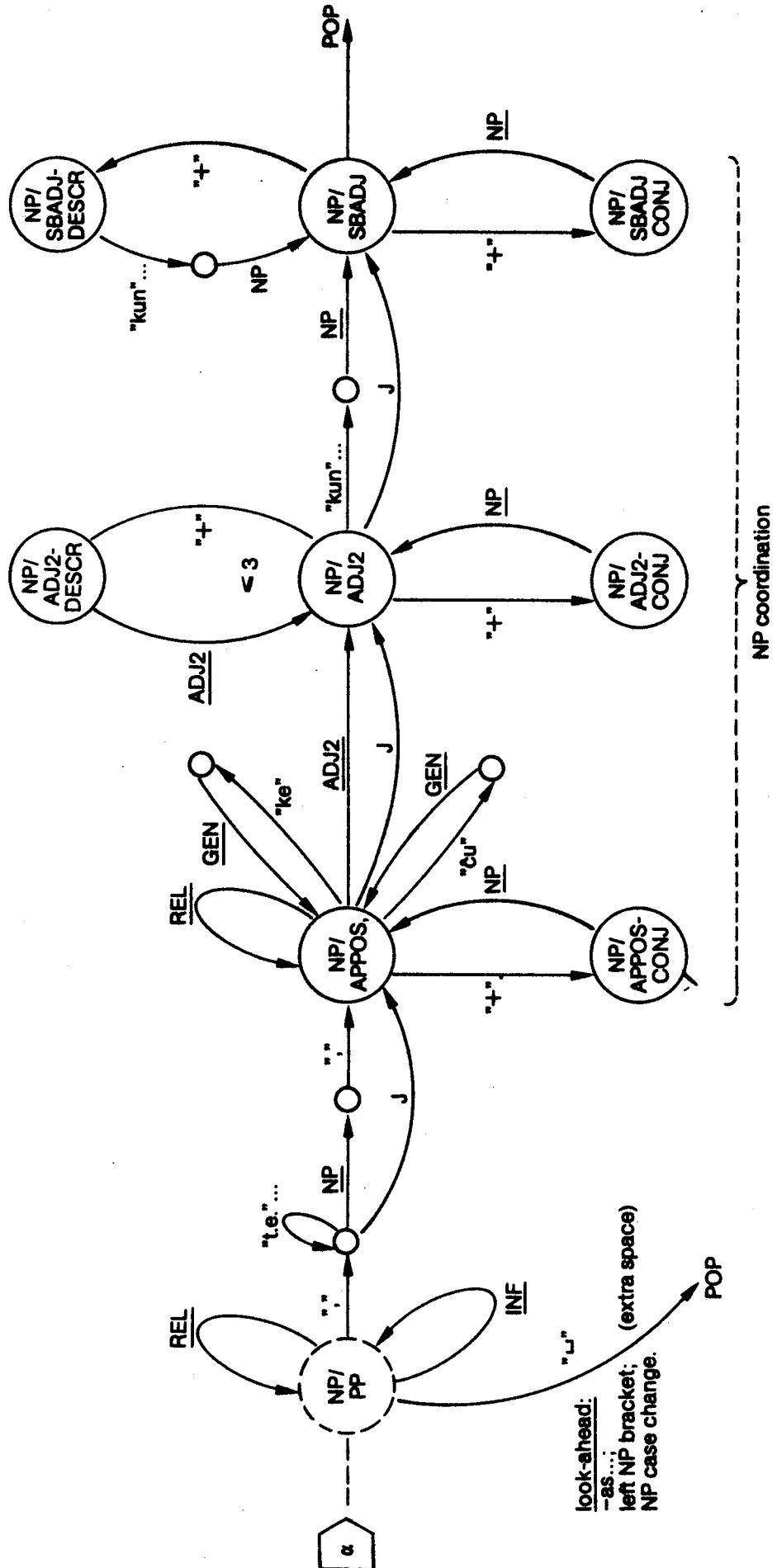


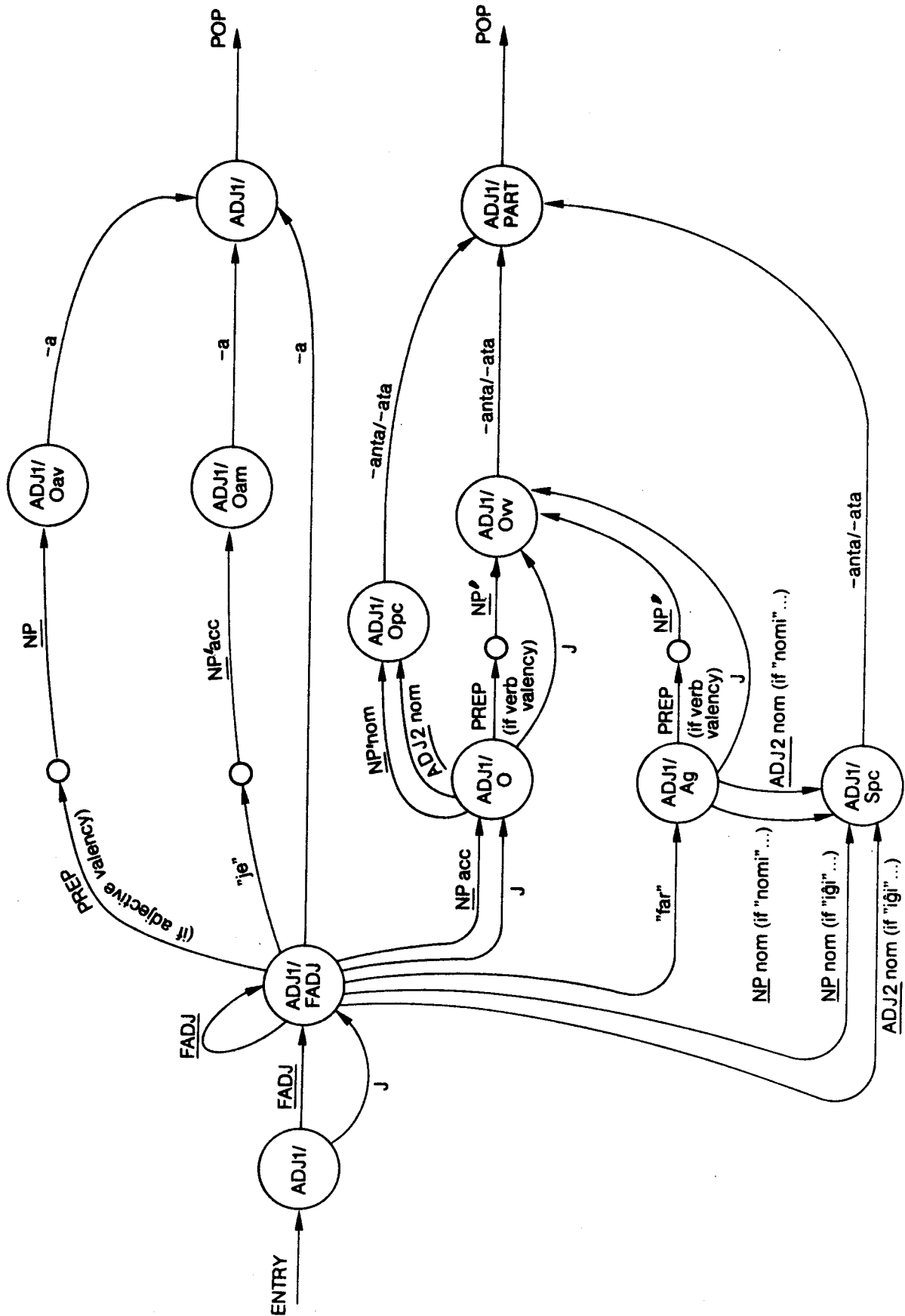
DETERMINERS

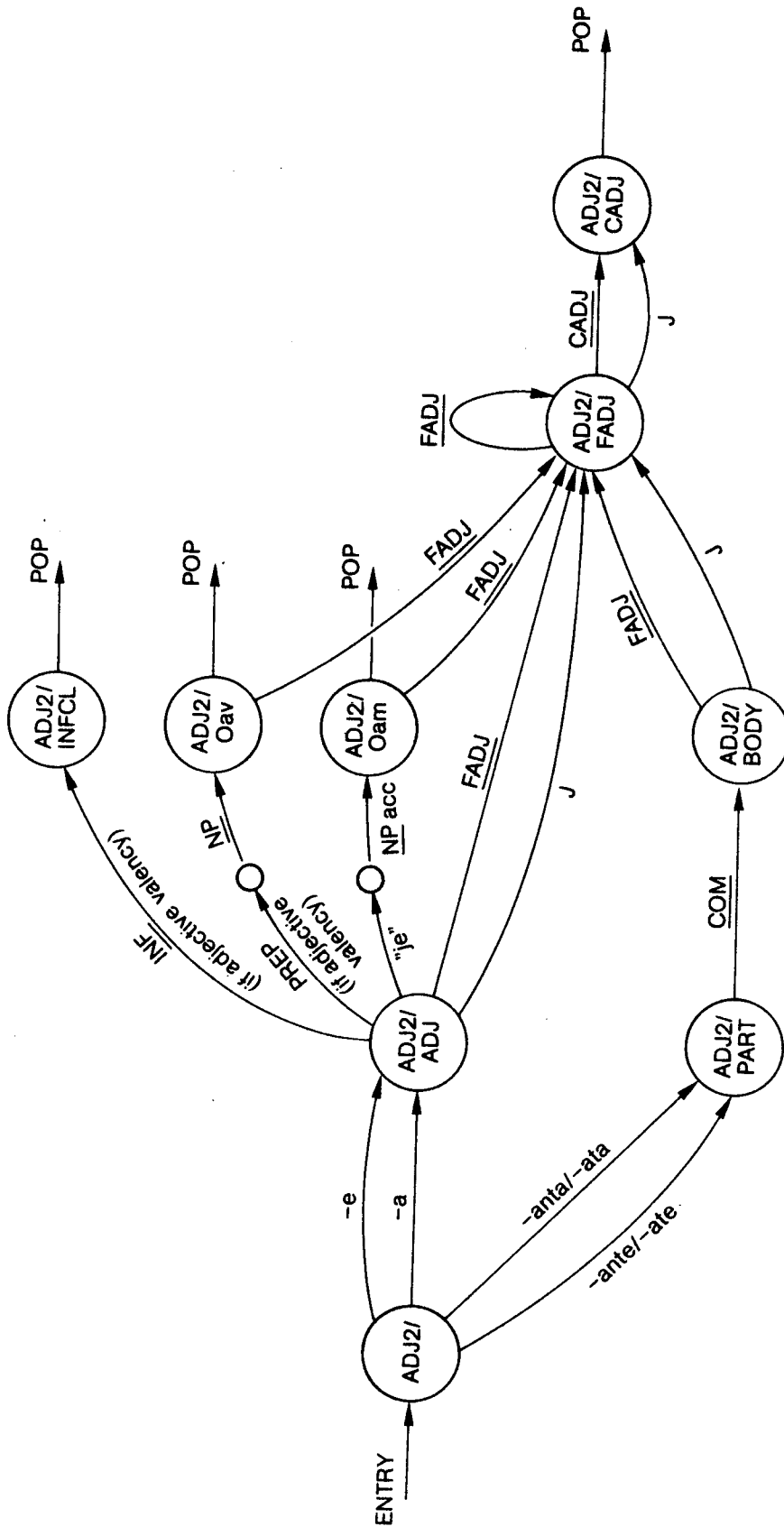
NUMERALS

ELLIPTIC NOUN COORDINATION

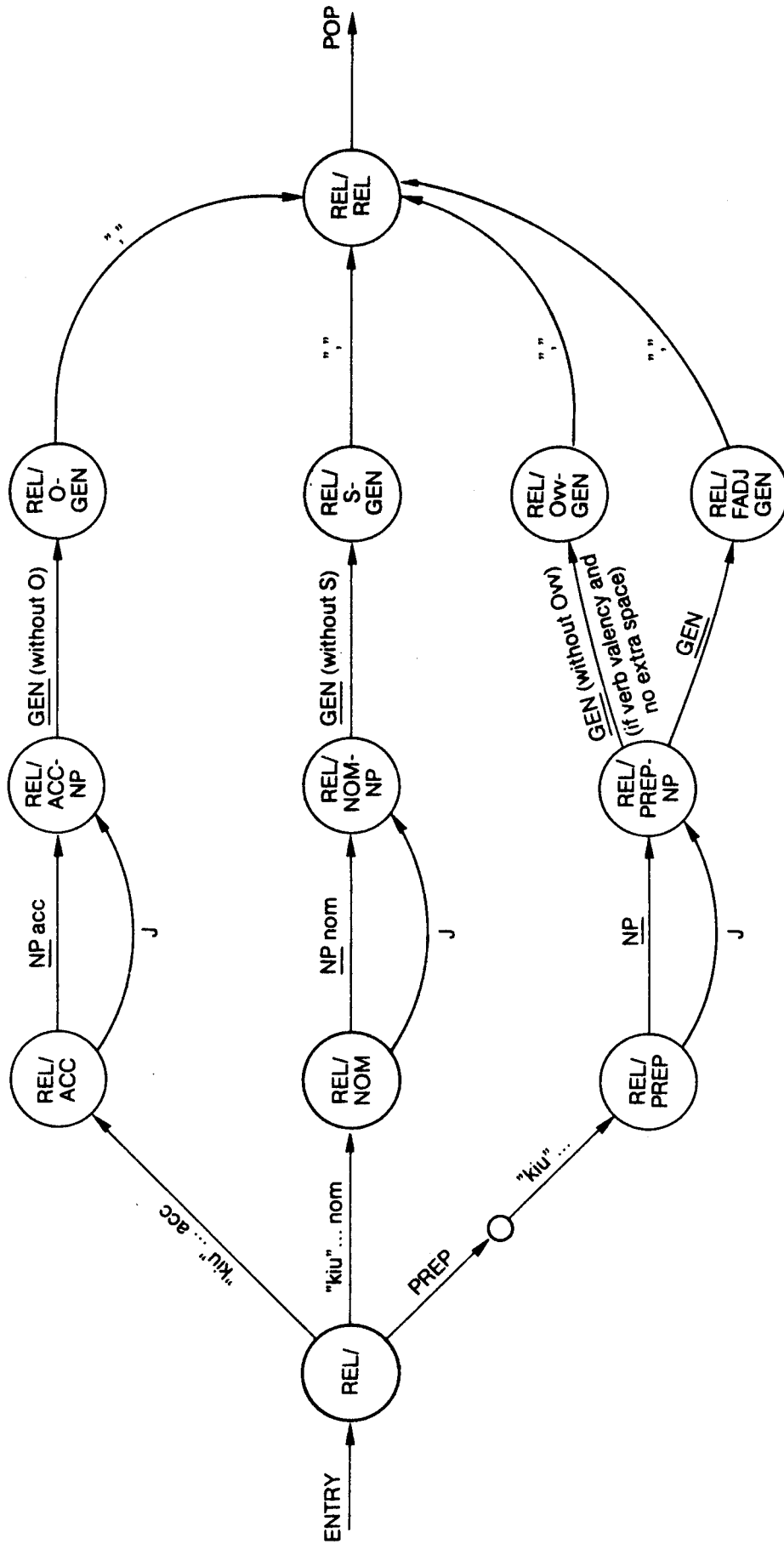
NP COORDINATION



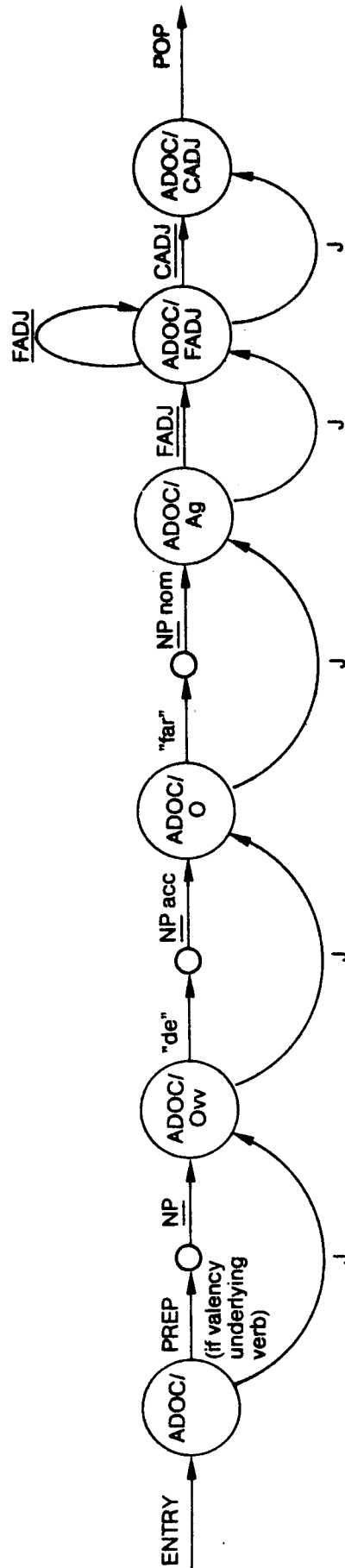




ATN for postmodifying adjectival clauses (including those serving as predicative complements). In addition, the ATN includes adverbial clauses.



ATN for relative clauses called from the ATN for NPs.



ATN for the NP supplement in case of nominalized verbs. Note the imposed ordering of PPs and the similarity with clausal structure. The acronym ADOC should be read as "ADO"-complement (in Esperanto, the '-ado' ending characterizes a nominalized verb).

ETUDE DE FAISABILITE DU SYSTEME DLT 1982-1983RÉSUMÉ

Nous présentons dans ce rapport le résultat d'une étude de 14 mois, destinée à la Commission Européenne, sur la faisabilité du DLT (Distributed Language Translation). Le système DLT proposé ici a pour but la traduction semi-automatique de langues naturelles écrites. Il a été conçu et étudié au départ par la société de service informatique BSO/Automation Technology b.v. des Pays Bas, de 1979 à 1982. L'étude de faisabilité présentée ici est la conclusion de la phase d'étude "sur le papier".

Le projet DLT se décompose en tranches successives étalées dans le temps. Il vise à traduire au moindre coût diverses langues européennes (en commençant par le français, l'allemand et l'anglais); mais il semble pouvoir être facilement étendu à d'autres langues (japonais, hébreu, arabe). Les textes qui seront traités peuvent se ranger dans la catégorie "textes d'information" allant des manuels d'instructions techniques jusqu'aux résumés de textes scientifiques, et des rapports commerciaux jusqu'aux règlements sur l'enlèvement et le traitement des déchets nucléaires. Les effets stylistiques, connotations, et autres subtilités ("lire entre les lignes") ne pourront pas d'une manière générale être conservés. Ceci mis à part, et au risque de refléter plus ou moins la structure et les formulations du texte de langue-source (LS), les traductions produites par le système DLT seront fiables et grammaticalement correctes.

Environnement opérationnel

La meilleure façon de décrire le système DLT est de le considérer comme un ensemble cohérent de systèmes intégrés destinés à faciliter la génération et l'affichage des textes, c'est à dire qu'il sera utilisé en connections avec des systèmes de traitement de textes et des terminaux videotex. DLT n'est pas du type "système batch", et n'est donc pas particulièrement adapté à la traduction de textes d'archives entrés par procédé de lecture optique OCR (optical character recognition).

Cependant il présente un grand intérêt pour les réseaux de distribution d'information et pour les services internationaux connectés (publics aussi bien que privés), et il est en outre compatible avec les techniques modernes de transmissions de données synchrones et de commutation par paquets.

Les essais de traduction de langues naturelles par ordinateur, rêve poursuivi depuis la 2e guerre mondiale, n'ont donné que des résultats partiels et modestes. Le domaine spécialisé de la TA (traduction automatique) a progressé lentement et n'a jamais connu de percées spectaculaires. Le système DLT doit être replacé dans le contexte créé d'un côté par le regain d'intérêt des 6 dernières années en Europe, en Extrême Orient et en Amérique du Nord, et de l'autre par la diversification des "services multilingues".

Les services informatiques multilingues d'aujourd'hui vont des machines de traitement de textes avec dictionnaires en langue étrangère aux vastes systèmes centralisés produisant des traductions brutes en mode batch. Les premiers peuvent être dénommés TAO (traduction assistée par ordinateur), les seconds TAA (traduction automatique autonome). En TAO, le travail de traduction et d'édition est fait par l'homme, qui ne fait appel à l'ordinateur qu'en cas de besoin. Il existe cependant un troisième cas de figure, que nous appellerons TSA (traduction semi-autonome), où le travail est fait par l'ordinateur, lequel est assisté par un être humain qui n'intervient que si l'ordinateur le demande. C'est dans ce type de système, où à un processus automatique s'ajoute un dialogue interactif homme-machine déclenché par le programme, que le système DLT trouve sa place.

Dans la conception TSA l'ordinateur compte sur l'aide humaine pour éclaircir le sens de mots ou de constructions ambiguës pour l'ordinateur (mais non pour l'homme, qui est censé connaître le contexte de la phrase en LS). Outre le fait que TAO et TSA se différencient par la nature de la relation maître-esclave, la seconde est plus proche de la TAA que la première. Mais alors que la TAAHQ (traduction automatique autonome de haute qualité) restera, de l'avis général, un but inaccessible pour les cinq décades à venir, la TSAHQ (traduction semi-automatique de haute qualité) apparaît comme un objectif réaliste (surtout pour une société informatique dotée d'une vaste expérience dans la mise au point de dialogues homme-machine). Telles sont les raisons qui nous ont incités à lancer le projet DLT.

Architecture du processus de traduction

Si on regarde les principales configurations de systèmes de traduction (sans intermédiaire, avec transfert, ou via interlangue) il faut souligner que le système DLT a été conçu, tant du point de vue lexical que grammatical, comme un système à interlangue. A cette fin, nous utilisons comme langue intermédiaire (LI) un sous-ensemble modifié de l'espéranto, et une grande partie de ce rapport est consacré à la description et à la définition de la grammaire de cette interlangue.

L'architecture du système à interlangue implique un processus comprenant deux étapes principales: analyse de la LS jusqu'à l'arrivée en LI, synthèse de la langue-cible (LC) à partir de l'interlangue. Ceci convient parfaitement à l'environnement opérationnel externe (répartition du processus de traduction entre expéditeur et destinataires dans un réseau d'information). La LI doit être considérée comme une passerelle, un échangeur condensé d'information entre modules de LS et modules de LC, transitant par des réseaux de télécommunication (tarifés en fonction du volume).

Si l'on compare le système DLT avec une approche concurrente en cours de développement - l'impressionnant projet EUROTRA -, on constate une différence notable entre la LI du premier et la structure d'interface du second: EUROTRA a donné à sa représentation intermédiaire la forme d'une arborescence à étiquettes complexes, incluant des variables sémantiques aussi bien que syntaxiques et morphosyntaxiques, c'est à dire des marqueurs abstraits. La LI de système DLT, au contraire, consiste essentiellement en une chaîne linéaire d'éléments lexicaux.

Dans ces deux approches, la structure intermédiaire doit avoir une "valeur ajoutée" par rapport aux données initiales de la LS, elle doit être dépouillée des particularités de la LS, et pouvoir ensuite être traitée par les modules de synthèse de LC. En particulier, elle doit être exempte d'ambiguïtés.

Alors qu'EUOTRA, semble-t-il, a tendance à stocker, dans la structure d'interface, de plus en plus d'informations abstraites, le système DLT a cherché à atteindre le but exposé plus haut en définissant soigneusement sa LI à base d'espéranto. Il utilise, ce faisant, l'expérience d'une langue semi-artificielle déjà existante (morphèmes invariants et autonomes, index d'agglutination de Greenberg = 1.00, transparence de la structure grammaticale, système de prépositions relativement précis). On pourrait dire que l'espéranto modifié utilisé pour le système DLT possède sa propre structure arborescente avec toutes les étiquettes morphosyntaxiques requises (désinences grammaticales, particules et mots grammaticaux). L'information liée aux relations de valences est préservée dans les entrées-dictionnaire de l'interlangue.

Tandis qu'un système de transfert comme EUROTRA essaie (pour des raisons économiques évidentes) de limiter la taille de la partie transfert LS-LC à une pure et simple substitution de lexèmes (les mots de la LS sont remplacés par les mots de la LC), un système à interlangue véritable, comme l'est le système DLT, tire profit de l'existence, au niveau de la LI, de dictionnaires très complets, situés aussi bien en amont (côté LS) qu'en aval (côté LC). Dans le DLT, la traduction peut s'appuyer dans une large mesure sur le niveau des relations de valence, ce qui compense l'absence d'étiquettes abstraites concernant les relations sémantiques. De plus, on conserve les avantages d'un développement modulaire du système par des équipes différentes pour la LS et la LC: il s'agit alors de se familiariser avec la grammaire et le lexique de la LI et non plus de chercher à se mettre d'accord sur une interface à étiquettes abstraites.

Le fait de limiter la structure intermédiaire du système DLT à une chaîne linéaire d'éléments lexicaux a deux avantages pratiques qui déterminent la forme générale du système: possibilité de contrôle rapide (pour le développement et la maintenance) et compacité (pour une transmission à faible coût).

Non-ambiguïté de l'interlangue

La préoccupation principale sous-jacente aux recherches a été la non-ambiguïté de la LI du DLT à base espéranto, condition indispensable à une étape entièrement automatique de la traduction, LI-LC. A cette fin, on a défini plus précisément la "non-ambiguïté" comme la possibilité d'analyser la LI à l'aide d'un analyseur simple, n'incluant ni "sémantique profonde" ni "connaissance-du-monde", mais fondée sur des informations (morpho-)syntaxiques et lexico-grammaticales (ces dernières fournies par le dictionnaire d'interlangue). La grammaire de la LI, décrite dans ce rapport, a été établie en ajoutant à l'espéranto commun de base trois niveaux de modification, chacun contribuant à la non-ambiguïté de l'interlangue. Les modifications comprennent: des règles strictes pour l'ordre des mots et groupes de mots, l'introduction d'un nombre limité de mots grammaticaux nouveaux et de particules nouvelles, un usage cohérent de la ponctuation et l'insertion d'un séparateur universel. On a pris grand soin d'éviter les éléments extralinguistiques encombrants et importuns qui pourraient nuire à la compacité et à la transparence de la LI.

Une part importante de ce travail a été de vérifier la non-ambiguïté de l'interlangue par une étude contrastive des ambiguïtés syntaxiques dans les LS (anglais, allemand, etc..;). Dans ce rapport, plus de 100 exemples illustrent la capacité de la LI à représenter de façon distincte les différentes interprétations, dans de domaines variés d'ambiguïté structurelle comprenant les traquenards bien connus de la TA: partie-du-discours, mots grammaticaux, ambiguïté des prépositions, nominalisation, anaphore, etc... Bien que nous admettions qu'une énumération, même détaillée, ne prouve pas que la LI soit totalement dépourvue d'ambiguïtés, la démonstration n'en est pas moins faite qu'il est possible d'accroître la résistance de cette LI à l'ambiguïté. Un algorithme pour l'insertion automatique d'un séparateur, expliqué à cette occasion, garantit un traitement sûr des ambiguïtés syntaxiques accidentelles (et donc difficiles à prévoir). Le même algorithme préserve la LI de l'ambiguïté systématique qui se rencontre fréquemment dans les cas de coordination et de détermination (grâce, à l'occasion, à un dialogue interactif de levée d'ambiguïté).

Avantage d'une approche en phases successives

La possibilité d'une "Interlangue-Espéranto" non ambiguë ayant été effectivement démontrée par le projet d'interlangue décrit dans ce rapport, une traduction LI-LC entièrement automatique devient réalisable. La construction de modules de synthèse d'une langue naturelle (à partir d'une interface non ambiguë) est fonction du degré d'avancement des travaux de recherche sur la question, et ne constitue pas en soi une percée technologique. Cependant, dans le cas présent, la nature particulière de la LI a une signification pratique indéniable: dérivée d'un langage de communications humain, sa commodité et sa limpidité en font non seulement un excellent moyen de développement, mais aussi un langage d'entrée de données parfaitement adapté à certaines applications sur des bases de données multilingues (telles que: les environnements contrôlés d'un fournisseur d'information videotex, un département de production de manuel techniques, l'établissement d'un système expert, etc...).

La phase suivante prévue pour le projet DLT est donc le développement d'un système pilote qui inclut au moins un module LI-LC (avec le choix de l'allemand comme première LC, et un travail estimé à 12 hommes-années sur une période de 2 ans). Quant à la qualité de la traduction, il faut s'attendre à ce que la sortie en LC présente des imperfections dans l'utilisation des articles et des temps passés (bien que l'étude indique éventuellement comment surmonter ces lacunes). En outre, il faudra s'accommoder d'un style quelque peu monotone, car la variation naturelle des schémas de phrase en fonction du thème (thème/rhème) est difficile à réaliser. Pour obtenir une traduction correcte des lexèmes, laquelle, vu les divergences lexicales entre LI et LC, est fonction de la capacité à formaliser et à augmenter le dictionnaire IL-LC de procédures de choix de termes, il suffit de limiter le système pilote au registre "dénotatif" (c'est à dire aux textes "techniques") et à un champ d'application spécifique.

L'économie et les affaires internationales ainsi que le droit international ont été retenus comme champ d'application pilote. La raison en est, mis à part leur très grand intérêt commercial, l'existence d'une terminologie espéranto (donc LI) adéquate et déjà considérable. Bien que des travaux ultérieurs soient nécessaires, la quantité actuelle de termes parfaitement adaptés constitue une base solide pour asseoir un système de traduction orienté vers l'application immédiate et démontre à nouveau la valeur pratique d'une LI authentique, diversifiée et aisément accessible.

Perspectives à long terme

Le délai nécessaire à la mise en circulation d'un système DLT multilingue complet et câblé (avec au moins une LS et deux LC) est d'environ 5 ans, qui s'ajoutent aux deux années à venir prévues pour le système pilote.

Le système DLT se distingue quelque peu de ses rivaux actuels par un trait important, qui est la réalisation d'un module d'analyse semi-automatique de LS vers LI, capable d'interroger à bon escient le pupitreur humain. Pour le moment, seuls les caractères généraux du projet d'analyse de la LS peuvent être révélés:

- analyse par intervalles, en fonction des données et en une seule fois, de gauche vers la droite, synchronisée avec l'entrée des mots au pupitre (intégration du système DLT à la machine de traitement de textes);
- création progressive et quasi-parallèle d'arborescences de LS orientées vers la LI, en nombre limité, constituant aussi peu de voies d'analyses indépendantes. On préfère combattre l'indétermination par le parallélisme des solutions plutôt que par des retours en arrière. Ce choix est justifié par la lenteur relative de la frappe manuelle (qui laisse des plages de temps libre aux systèmes de traitement dédiés) et les projets de commercialisation de micro-ordinateurs à grande capacité de stockage.

Quant au dialogue de désambiguation interactif, il ne peut être commencé avant qu'une unité complète de traitement (c'est à dire une phrase) ait été entrée en machine, et seulement après qu'un essai de désambiguation automatique ait échoué. Le système doit générer des questions qui mettent en évidence la présence de diverses interprétations, sans utiliser un jargon linguistique ni une langue autre que la LS (le pupitreur, qui peut être aussi l'auteur du texte d'entrée en LS, est supposé ne connaître que le contenu du texte et la LS). De plus, le module d'analyse devra contenir un algorithme optimisant l'ordre des questions, de manière à en réduire le nombre et alléger le travail du pupitreur. Ces tâches assignées au module d'analyse de LS sont loin d'être négligeables et demandent une étude plus détaillée dans le cadre des échéances fixées pour le projet pilote. Une simulation du dialogue interactif devrait faire partie de cette étude.

A plus longue échéance, le système DLT offre la possibilité d'une amélioration progressive de la correction stylistique du produit en LC, en procédant à des transformations syntaxiques à l'intérieur de la LI (les premières sorties en LC, bien que grammaticalement correctes, reflèteront la structure de l'entrée de LS). En outre, un allègement très progressif du dialogue interactif peut être réalisé par l'implémentation d'algorithmes de désambiguation du type intelligence artificielle orientés vers le macrocontexte, lesquels opéreraient sur la LI, en relation avec une banque de données stockée en LI. Ce développement à très long terme bénéficiera de la richesse (lexiques, terminologie) et de la structure morphémique de la LI (dont le codage interne non orthodoxe accélère toutes les opérations d'appariement de chaînes); il est par conséquent étroitement lié au projet DLT tel qu'il est présenté ici.

Conclusion finale

Le problème fondamental de cette étude - possibilité de créer une interlangue non ambiguë fondée sur l'espéranto - a reçu une réponse positive. Même si on tient compte de l'investissement considérable et des problèmes spécifiques qui restent à maîtriser pour développer le système DLT final (de langue-source à langue-cible), les enseignements qu'on peut tirer d'une phase de réalisation partielle (d'interlangue à langue-cible) justifient déjà la poursuite du soutien accordé au projet pilote.

Studie zur Durchführbarkeit von DLT (1982 -1983)Zusammenfassung

Der vorliegende Bericht enthält die Ergebnisse einer vierzehnmonatigen Studie für die KEG bezüglich der Durchführbarkeit des Projektes DLT (Distributed Language Translation - Verteilte Sprachübersetzung). DLT ist ein System, das zur halbautomatischen Übersetzung von geschriebenen Texten in natürlicher Sprache vorgeschlagen wird. Es wurde innerhalb des niederländischen Softwarehauses BSO/Automation Technology b.v. im Zeitraum 1979 - 1982 konzipiert und erstmals untersucht. Die vorliegende Studie zur Durchführbarkeit des Vorhabens dient als Bestandsaufnahme der Ergebnisse der Planungsphase.

Das DLT-Projekt ist als ein in Realisierungsphasen eingeteiltes langfristiges Unternehmen auf wirtschaftliche Übersetzungen zwischen europäischen Sprachen (zunächst Französisch, Deutsch und Englisch) ausgerichtet, verspricht jedoch hervorragende Möglichkeiten zur Ausweitung auf andere Sprachen (Japanisch, Hebräisch, Arabisch). Die bearbeitbaren Textsorten können als "informativ" gekennzeichnet werden, mit einer Bandbreite von Handbüchern für Wartung und Benutzung technischen Geräts bis zu wissenschaftlichen Abstracts und von Geschäftsberichten bis hin zu Bestimmungen über die nukleare Entsorgung. Besondere stilistische Effekte, Konnotationen und andere Feinheiten (etwa solche, die man "zwischen den Zeilen" liest) können i.a. nicht wiedergegeben werden. Abgesehen hiervon werden die DLT-Übersetzungen verlässlich und grammatisch korrekt sein, wobei sich allerdings die Struktur des quellsprachlichen Textes in der Übersetzung widerspiegeln wird.

Die operationelle Umgebung

DLT muss als integraler Bestandteil der Texterzeugungs- und Textausgabe-Hardware angesehen werden, d.h. DLT wird in enger Kopplung mit textverarbeitenden Prozessoren und Videotext-Bildschirmgeräten verwendet. DLT ist somit typischerweise kein Stapelsystem und deshalb auch weniger zur Übersetzung von klarschriftgelesenen archivierten Texten geeignet. Sein Einsatzbereich liegt vielmehr in den Gebieten der verteilten Informationsnetze und der internationalen (Öffentlichen wie auch privaten) on-line-Dienste. Der Einsatz von DLT ist verträglich mit der modernen synchronen Datenübertragung und den packet-switching-Techniken.

Die Bestrebungen, natürliche Texte mit Hilfe von Rechnern zu übersetzen, sind ein Traum, an dessen Verwirklichung seit dem Zweiten Weltkrieg gearbeitet wird, der aber nur teilweise und bescheidene Realität geworden ist. Das Fachgebiet der MÜ (Maschinelle Übersetzung) entwickelt sich bisher nur langsam fort und ein rechter Durchbruch ist bislang nicht gelungen. Ein wiedererwachendes Interesse in den vergangenen sechs Jahren - sowohl in Europa als auch im Fernen Osten und in Nordamerika - zusammen mit der Verbreitung multilingualer technischer Hilfsmittel bildet den Hintergrund, vor dem DLT zu bewerten ist.

Multilinguale computergestützte Hilfsmittel reichen heute von Textsystemen mit fremdsprachlichen Wörterbüchern bis zu Grossrechnern, die Rohübersetzungen im Stapelbetrieb herstellen. Ersteres kann als CAT (Computer Aided Translation - rechnergestützte Übersetzung) und letzteres als FAT (Fully Automatic Translation - vollautomatische Übersetzung) bezeichnet werden. Bei der rechnergestützten Übersetzung führt der Humanübersetzer oder Editor die Arbeit aus und verlangt nur dann Unterstützung vom Rechner, wenn er/sie das wünscht. Es gibt jedoch noch eine dritte Art der Mensch-Maschine-Kooperation, die wir SAT nennen wollen (Semi-Automatic Translation - halbautomatische Übersetzung): Der Rechner führt hier die Arbeit aus und verlangt wenn nötig erforderliche Informationen vom Bediener. In genau einem solchen Systemrahmen, in dem ein automatischer Prozess durch eine vom Rechner angestossene menschliche Interaktion ergänzt wird, ist DLT anzusiedeln.

Im SAT-Konzept verlässt sich der Computer auf menschliche Hilfe bei der Klärung von Wörtern und Konstruktionen, die für die Maschine (nicht aber für den Benutzer, von dem das Verstehen des quellsprachlichen Textes vorausgesetzt wird) mehrdeutig sind. Abgesehen von dem Unterschied zwischen CAT und SAT hinsichtlich der Verteilung der Rollen für "Herr" und "Sklave" steht SAT wesentlich näher an FAT, als das für CAT der Fall ist. Aber während die FAHQT (Fully Automatic High Quality Translation - vollautomatische Übersetzung von hoher Qualität) nach allgemeiner Übereinstimmung in den nächsten fünfzig Jahren ein unerreichbares Ziel bleibt, stellt die SAHQT (Semi-Automatic High-Quality Translation - halbautomatische Übersetzung von hoher Qualität) eine realistische Zielsetzung dar (insbesondere für ein Softwarehaus mit reicher Erfahrung auf dem Gebiet des Mensch-Maschine-Dialogs), und aus diesem Grunde wird mit DLT auch die genannte Richtung eingeschlagen.

Die Architektur des Übersetzungsprozesses

Betrachten wir die gängigen Architekturen von Übersetzungssystemen, nämlich direkte, transferorientierte und interlinguale Konzepte, dann müssen wir betonen, dass DLT von vornherein als interlinguales System angelegt ist, und zwar sowohl lexikalisch als auch grammatisch. Zu diesem Zweck benutzen wir eine modifizierte Teilmenge des Esperanto als IL (Intermediate Language - Zwischensprache, Interlingua); der Beschreibung und der grammatischen Definition dieser Interlingua wird daher im vorliegenden Bericht breiter Raum gewidmet.

Die interlinguale Konzeption hat einen zweistufigen Übersetzungsprozess zur Folge (quellsprachliche Analyse mit IL-Ergebnis und zielsprachliche Synthese auf der Grundlage des IL-Textes), der sich nahtlos in die operationale Umgebung einbinden lässt. (Verteilung des Übersetzungsprozesses über Sender und Empfänger in einem Informationsnetz). Die IL ist als schmale Brücke zu sehen, als ein kompakter, sich über Telekommunikationsnetze erstreckender Informationsaustausch zwischen quellsprachlichen und zielsprachlichen Modulen, wobei sich die Kosten aus dem Umfang des Datentransports ergeben.

Vergleicht man DLT mit einem derzeit laufenden Konkurrenzunternehmen, nämlich der beeindruckenden Entwicklung von EUROTRA, dann lässt sich ein bemerkenswerter Unterschied zwischen der IL von DLT und der Interfacestruktur von EUROTRA aufzeigen: EUROTRA hat als Zwischenstruktur eine Baumdarstellung mit komplexen Etiketten festgelegt, die semantische wie auch oberflächensyntaktische und morphosyntaktische Variable, d.h. abstrakte Formative abdecken. Die Interlingua von DLT besteht dagegen grundsätzlich aus einer linearen Kette lexikalischer Formative.

In beiden Ansätzen braucht die Zwischenstruktur gewissermaßen einen "zusätzlichen Wert" im Vergleich zur ursprünglichen quellsprachlichen Eingabe: Sie darf keinerlei Besonderheiten oder Idiosynkrasien der Quellsprache enthalten und muss darüber hinaus von den zielsprachlich orientierten Moduln verarbeitet werden können. Insbesondere soll sie frei sein von Ambiguitäten.

Wo EUROTRA dahin tendiert, die Interfacestruktur immer mehr mit abstrakten Informationen zu füllen, versucht DLT das oben genannte Ziel durch einen sorgfältigen Entwurf seiner auf Esperanto basierenden IL zu erreichen, indem dabei die linguistischen Eigenschaften einer schon lange existierenden und erprobten halb-künstlichen Sprache ausgenützt werden (unveränderliche und autonome Morpheme, Greenbergs Agglutinationsindex 1,00, Transparenz der grammatischen Struktur, ziemlich präzises Präpositionalsystem). Man könnte sagen, dass die für DLT verwendete Esperanto-Modifikation Baumstrukturen einbezieht, die komplett mit morpho-syntaktischen Etiketten (grammatischen Endungen, Partizipien und Funktionswörtern) attribuiert sind. Die Valenzbindung ist in den IL-Lexikoneinträgen festgehalten.

Während ein mit einem Transferteil arbeitendes System wie EUROTRA (aus offensichtlichen ökonomischen Gründen) den Umfang des Transfers (zwischen Quell- und Zielsprache) möglichst gering zu halten versucht, indem nur eine Ersetzung der quellsprachlichen durch zielsprachliche Lexeme angestrebt wird, zieht ein voll interlinguales System wie DLT seinen Nutzen aus dem Vorhandensein reichhaltig ausgestatteter Wörterbuchspalten auf beiden Seiten der Wasserscheide zwischen Quell- und Zielsprache. In DLT kann sich die Übersetzung umfassend auf die Ebene der Valenzbindung verlassen, wodurch das Fehlen semantischer Relationen kompensiert wird. Dennoch bleibt der Vorteil einer modularen Systementwicklung durch getrennte quell- und zielsprachliche Arbeitsgruppen gewahrt. An die Stelle von Absprachen und Bemühungen zur Vereinheitlichung einer allgemeinen abstrakten Schnittstelle tritt in DLT die Gewöhnung an Grammatik und Lexik der IL.

Die Beschränkung der DLT-Zwischenstruktur auf eine lineare Kette von lexikalischen Formativen hat zwei praktische Vorteile, die die generelle Form des Systems recht stark festlegen: die Möglichkeit einer raschen Inspektion von (Zwischen-) Ergebnissen (wichtig für Entwicklung und Wartung) und die Kompaktheit (niedrige Übertragungskosten).

Die Eindeutigkeit der IL

Das Hauptergebnis der zugrundeliegenden Studie war die Eindeutigkeit der auf Esperanto basierenden IL von DLT, eine notwendige Vorbedingung für den vollautomatischen Übersetzungsschritt von der IL zur Zielsprache. Zu diesem Zweck wurde der Begriff der Eindeutigkeit eher im Sinne der Analysierbarkeit der IL durch einen einfachen Parser definiert, ohne Berücksichtigung einer tiefen Semantik oder des Weltwissens, jedoch unter Beachtung der (morpho-) syntaktischen Informationen und der (im IL-Wörterbuch verzeichneten) Valenzangaben.

Die im Bericht beschriebene IL-Grammatik wurde erstellt durch Hinzufügung von drei "Modifikationsschichten" zu der normalen Esperanto-Basis, von denen jede zur Vereindeutigung der IL beiträgt. Die Modifikationen beinhalten: eine strikte Vorschrift bezüglich der Abfolge von Wörtern und Wortgruppen, Einführung einer beschränkten Anzahl neuer Funktionswörter und Partikeln, eine durchgängig definierte Zeichensetzung sowie die Einführung eines universellen Separators (eines Trennelements zur Abgrenzung von Wortgruppen). Besonderes Gewicht wurde auf die Vermeidung von platzverschwendenden und aufdringlichen aussersprachlichen Elementen gelegt, die die Kompaktheit und Inspizierbarkeit der IL beeinträchtigen könnten.

Die Überprüfung der resultierenden Eindeutigkeit anhand einer kontrastiven Übersicht mit syntaktischen Mehrdeutigkeiten der Quellsprachen (Englisch, Deutsch usw.) war ein wichtiger Bestandteil der Studie. Im Bericht veranschaulichen mehr als 100 Beispiele die Mächtigkeit der IL, alternative Lesarten mit unterschiedlichsten strukturellen Mehrdeutigkeiten unter Einschluss der notorischen Stolpersteine der MÜ klar zu unterscheiden: Wortklassen, Funktionswörter, Mehrdeutigkeit der Anbindung von Präpositionalphrasen, Nominalisierung von Verben, anaphorische Bezüge usw. Obwohl zugestandenermaßen eine solche umfassende Aufzählung noch nicht die völlige Eindeutigkeit der IL beweist, wird doch der Beweis für die Ausbaufähigkeit der IL in Bezug auf Eindeutigkeit geliefert. Ein Algorithmus für die automatische Einsetzung der Separatoren, die in diesem Zusammenhang erklärt wird, gewährleistet eine sichere Behandlung zufälliger (und daher kaum vorhersagbarer) syntaktischer Mehrdeutigkeiten. Derselbe Algorithmus schützt die IL vor der systematischen Mehrdeutigkeit hinsichtlich des Skopus von Konjunktionen und Modifikatoren (eventuell nach einem interaktiven Disambiguierungsdialog).

Der Nutzen des Stufenplans

Nachdem die Realisierbarkeit einer eindeutigen "Esperanto"-IL durch den im vorliegenden Bericht vorgestellten IL-Entwurf aufgezeigt ist, liegt die vollautomatische Übersetzung aus der IL in eine Zielsprache im Bereich des Möglichen. Der Aufbau der Moduln für die Synthese natürlicher Sprachen ist - unter Voraussetzung einer eindeutigen Schnittstelle - nach dem gegenwärtigen Stand der Wissenschaft leistbar und verlangt keineswegs einen technologischen Durchbruch. In diesem Fall ist die Besonderheit der IL jedoch von unleugbarer Wichtigkeit: Die Tatsache, dass die IL von einer der menschlichen Verständigung dienenden Sprache abgeleitet ist, sowie ihre Zugänglichkeit und Klarheit machen sie nicht nur zu einem ausgezeichneten Entwicklungsmittel, sondern auch zu einer effektiven Dateneingabesprache für bestimmte Anwendungen in mehrsprachigen Datenbanken (z.B. kontrollierte Umgebungen Bildschirmtext-Anbieters, Produktionsabteilungen für technische Handbücher, Aufbau eines Expert-Systems usw.).

Die nächste für das DLT-Projekt beantragte Phase besteht daher in der Entwicklung eines Pilotsystems, das wenigstens einen IL-zielsprachlichen Modul enthalten soll (vorzugsweise mit Deutsch als erster Zielsprache und einem geschätzten Aufwand von 12 Mannjahren innerhalb von 2 Jahren). Hinsichtlich der Übersetzungsqualität sind gewisse Unvollkommenheiten im Gebrauch der Artikel und bei den Vergangenheitstempora zu erwarten (unsere Studie zeigt jedoch mögliche Wege zur Bewältigung des Problems).

Daneben muss natürlich mit einem etwas monotonen Stil in den Übersetzungen gerechnet werden, da eine natürliche Variation der Satzmuster in Übereinstimmung mit dem Thema-Rhema-Verlauf schwer zu verwirklichen ist. Die zutreffende Übersetzung von Lexemen, die durch die lexikalische Divergenz zwischen IL und Zielsprache und durch die Fähigkeit, lexikalische Auswahlprozeduren zu formalisieren und sie zum IL-zielsprachlichen Lexikon hinzuzufügen, bestimmt wird, kann durch Beschränkung des Pilotsystems auf Denotate (d.h. nur "technische" Texte) und ein einziges Fachgebiet erreicht werden.

Welthandel und Weltwirtschaft, ergänzt um Internationales Recht, wurden als Fachgebiet für die Pilotanwendung ausgewählt. Der Grund hierfür war - neben dem breiten wirtschaftlichen Interesse - die Verfügbarkeit einer bereits ausführlichen und auf neuestem Stand befindlichen Terminologie in Esperanto (und damit auch in der IL). Obwohl weitere Arbeiten erforderlich sind, bietet der existierende Wortschatz an wohldefinierten Termini eine tragfähige Grundlage zum Aufbau eines anwendungsorientierten Übersetzungssystems und zeigt wiederum eindringlich den praktischen Wert einer echten, umfassenden und leicht zugänglichen IL.

Langfristige Perspektiven

Der Zeitraum bis zur Marktfähigkeit eines vollständigen, hardware-technisch integrierten multilingualen DLT-Systems (mit wenigstens einer Quellsprache und zwei Zielsprachen) wird mit ca. 5 Jahren veranschlagt, die sich an das für die nächsten zwei Jahre beantragte Pilotprojekt anschließen.

Die Realisierung eines halbautomatischen Moduls für die quellsprachliche Analyse (mit dem Ergebnis in IL-Notation), in dem sorgfältig von der Möglichkeit der Befragung des Texterfassers Gebrauch gemacht wird, stellt einen entscheidenden und charakteristischen Bestandteil der DLT-Entwicklung dar, die sich damit leicht von existierenden oder konkurrierenden MÜ-Unternehmungen unterscheidet. Zur Zeit kann nur ein globaler Abriss des quellsprachlichen Analyseprozesses vorgestellt werden: eine segmentweise, datengesteuerte Analyse mit einem einzigen Durchlauf von links nach rechts, synchronisiert mit der Wort-für-Wort-Eingabe des Bedieners (Integration von DLT in den Wort-Prozessor); schrittweise quasi-parallele Erzeugung von IL-Orientierten quellsprachlichen Baumstrukturen entlang einer (beschränkten) Anzahl alternativen Pfade. Das Hauptgewicht liegt auf der Überwindung des Nicht-Determinismus durch Parallelverarbeitung anstelle von Backtracking. Dieser Ansatz wird durch die relativ geringe Geschwindigkeit der manuellen Eingabe (die genügend "freie" Rechenkapazität für die entsprechenden Prozessoren lässt) und die geplante Verfügbarkeit von Chips mit hoher Speicherkapazität begünstigt.

Der der Vereindeutigung dienende interaktive Dialog setzt erst nach vollständiger Eingabe einer Übersetzungseinheit (eines Satzes) ein, und zwar nur dann, wenn die automatische Disambiguierung misslungen ist. Das System muss in der Lage sein, seine Fragen bezüglich des Vorhandenseins alternativer Interpretationen ohne Verwendung linguistischer Fachbegriffe in der Quellsprache zu stellen. (Der Texterfasser, der der Autor des Eingabetextes sein kann, braucht nur den Kontext und die Quellsprache zu verstehen). Weiterhin wird der Analysebaustein einen Algorithmus enthalten, der die Reihenfolge der Fragen optimiert und dabei ihre Anzahl und damit den Arbeitsaufwand des Bedieners reduziert. Hier handelt es sich um nicht-triviale Anforderungen an den Analysemodul, die eine detailliertere Untersuchung während der Laufzeit des beantragten Projekts verlangen. Eine Simulation des Interaktionsdialogs sollte Teil der Studie sein.

Auf längere Sicht bietet DLT die Aussicht auf eine stufenweise qualitative Verbesserung in Richtung auf stilistisch korrekte Übersetzungsergebnisse an durch syntaktische Abbildungen innerhalb der IL (Die ersten freigegebenen Systeme werden Übersetzungen liefern, die zwar grammatisch korrekt sein werden, aber die Strukturen des Ausgangstextes noch widerspiegeln). Weiterhin dürfte eine sehr allmähliche Lockerung des Interaktionsdialogs erreichbar sein durch den Einsatz von makrokontext-orientierten Disambiguierungsalgorithmen aus der Künstlichen Intelligenz, die auf der IL in Verbindung mit einer in IL beschriebenen Wissensbasis operieren. Diese in die fernere Zukunft reichende Entwicklung wird ihren Nutzen ziehen aus dem Reichtum (Lexikon, Terminologie) und der kompakten morphematischen Struktur der IL (deren unorthodoxe interne Kodierung alle Stringoperationen beschleunigt). Sie ist daher eng verbunden mit dem spezifischen DLT-Konzept, das hier vorgestellt wird.

Zusammenfassung

Die zentrale Frage der vorliegenden Untersuchung, die Realisierbarkeit einer eindeutigen auf Esperanto gegründeten IL, konnte positiv beantwortet werden. Selbst unter Berücksichtigung der beträchtlichen Investitionen und der nicht-trivialen Probleme, die auf dem Wege der Entwicklung des endgültigen (quellsprachlich - zielsprachlichen) DLT-Systems zu meistern sind, rechtfertigt bereits die praktische Bedeutung von nur einer Teilphase (Synthese der Zielsprache aus IL) eine weitere Förderung in Form eines Pilot-Projekts.

TOEPASBAARHEIDSSSTUDIE DLT 1982-1983SAMENVATTING

Dit rapport kwam tot stand in opdracht van de Europese Commissie, na 14 maanden van studie op de haalbaarheid van DLT (Distributed Language Translation). DLT is een systeem-ontwerp voor half-automatische vertaling tussen natuurlijke talen, zich beperkend tot schrijftaal. De conceptie en eerste verkenning ervan speelden zich af binnen het Nederlandse softwarehuis BSO/Automation Technology b.v., in de periode 1979-1982. Met deze toepasbaarheidsstudie wordt de 'tekentafel-fase' afgesloten.

DLT is een gefaseerd en langlopend project, dat zich in eerste instantie richt op economisch rendabele vertaling tussen de Europese talen (Frans, Duits, Engels...), maar daarnaast ook uitstekende perspectieven biedt voor uitbreiding naar andere talen (Japans, Hebreeuws, Arabisch). Het te verwerken teksttype kan worden gekarakteriseerd als 'informatief', wat zo ongeveer alles bestrijkt van technische handboeken tot wetenschappelijke referaten, en van bedrijfsverslagen tot bepalingen over het lozen van kernafval. Stylistische effecten, bijbetekenissen en andere subtiliteiten (informatie "tussen de regels") zullen over het algemeen verloren gaan. Met dit voorbehoud, en ten koste van het min of meer erin doorklinken (qua zinsbouw en woordkeuze) van de brontekst, zullen DLT-vertalingen betrouwbaar en grammatikaal korrekt zijn.

Relatie met de omgeving.

DLT kan het best worden beschouwd als een op elkaar afgestemd geheel van inbouwcomponenten in moderne tekstverwerkings- en beeldschermapparatuur (word-processors, videotex terminals, etc.). DLT is verre van een batch-systeem, en daarom niet bepaald geschikt voor machinaal leesbare invoer (zoals OCR). Daarentegen is het een zeer aantrekkelijk systeem in combinatie met allerlei netwerken (openbare zowel als partikuliere) voor on-line informatie-distributie, en dan natuurlijk speciaal voor die welke op internationale schaal of in een veeltalige omgeving opereren. DLT maakt gebruik van geavanceerde transmissietechnieken (synchroon, packet-switching).

Reeds sinds de tweede wereldoorlog zijn pogingen ondernomen om de droom van het vertalen per computer in vervulling te doen gaan. Ondanks alle inspanningen konden slechts bescheiden resultaten geboekt worden. Het specialistische

vakgebied 'Machinaal Vertalen' ontwikkelde zich, maar echte doorbraken bleven uit. De laatste zes jaar hebben echter een opleving laten zien, niet alleen in Noord-Amerika maar ook in Europa en het Verre Oosten. Tegelijkertijd begon zich een heel spektrum van communicatievoorzieningen tussen talen af te tekenen, van word-processors met ingebouwd meertalig woordenboek tot complete batch-installaties voor ruwe automatische vertalingen. Tegen deze achtergrond van actuele ontwikkelingen moet DLT beoordeeld worden.

In plaats van geheel automatisch vertalen treft men tegenwoordig meer en meer het zogenaamde computergesteunde vertalen (CAT, Computer Aided Translation) aan, waarbij de regie van het vertaalproces in handen is van de menselijke vertaler. Er is echter nog een derde vorm, het half-automatisch vertalen, waarbij de computer de regie voert maar wel gebruik maakt van de aanwezigheid van de man of vrouw achter de word-processor, om deze zondig over de betekenis van de tekst te kunnen consulteren. Deze vorm, gekarakteriseerd door een computergestuurde interactieve dialoog, wordt toegepast in DLT.

Bij halfautomatisch vertalen steunt de computer op de mens voor de juiste interpretatie van allerlei dubbelzinnigheden in de uitgangstekst. Let wel, het betreft woorden en zinskonstrukties die dubbelzinnig zijn voor de computer, veelal niet voor de mens, die bewust of onbewust gebruik maakt van zijn kennis van de kontekst en de betreffende materie.

Behalve in het samenspel van mens en computer is er nog een belangrijk verschil tussen computergesteund en half-automatisch vertalen: het laatste staat veel dicht bij volledig automatisch vertalen dan het eerste. Maar terwijl iedereen het erover eens is dat volledig automatische kwaliteitsvertalingen voor de komende pakweg 50 jaar een hersenschim blijven, dient half-automatisch vertalen zich aan als een haalbare kaart, zeker wanneer de in de software-industrie aanwezige kennis en ervaring op het gebied van interactieve dialogen daarbij wordt ingezet. Daarom is met DLT voor het inslaan van deze weg gekozen.

Opbouw van het vertaalproces.

Gelet op de belangrijkste bouwwijzen van computervertaal-systemen: Talenpaar-, Transfer- en Tussentaal-systemen, stellen we nadrukkelijk dat DLT gekoncipieerd is als Tussentaal-systeem, zowel lexikaal als grammatikaal. Als tussentaal wordt hierbij een gemodificeerde subset van Esperanto toegepast, en een groot deel van de studie is dan ook aan de beschrijving en de grammatikale definitie van

deze subset gewijd.

Het principe van een Tussentaal-systeem impliceert een vertaalproces waarin 2 belangrijke stappen kunnen worden onderscheiden (Brontaal-Analyse, resulterend in Tussentaal, en Doeltaal-Synthese, uitgaand van die Tussentaal). Dit past uitstekend bij de buitenomgeving van gedistribueerde verwerking waarin DLT zal opereren: het vertaalproces wordt opgedeeld tussen de verzendende en ontvangende apparatuur van een informatienetwerk. Daarbij moet men de Tussentaal zien als een smalle brug, waarover de Brontaal- en Doeltaal-modules kompakte informatie uitwisselen. Kompaktheid is van belang in verband met de volume-afhankelijke tarifiëring in de nieuwere en toekomstige datakommunikatienetwerken.

Een indrukwekkend op stapel staand project voor automatisch vertalen is EUROTRA, en een vergelijking van DLT daarmee dringt zich op. Een opmerkelijk punt van verschil is het interface tussen Brontaal- en Doeltaal-modules: bij EUROTRA worden zinnen gerepresenteerd door complexe boomstructuren, voorzien van allerlei abstrakte variabelen (semantische, syntaktische en morfologische), de zogeheten abstrakte formatieven. Bij DLT is het interface de op Esperanto gebaseerde Tussentaal, hetgeen neerkomt op een lineaire reeks van lexikale formatieven.

Welke benadering men ook verkiest, de informatie op het interface moet 'verrijkt' zijn ten opzichte van de oorspronkelijke brontekst: de typische eigenaardigheden en onregelmatigheden van de brontaal moeten eruit verwijderd zijn, zodat directe verwerking door doeltaalgerichte modules mogelijk is. In het bijzonder mogen op het interface geen dubbelzinnigheden meer voorkomen.

Het ziet er nu naar uit dat EUROTRA het zoekt in steeds meer abstrakte informatie aan het interface toevoegen, terwijl DLT hetzelfde doel tracht te bereiken door een zorgvuldig ontwerp van zijn op Esperanto gebaseerde Tussentaal. Daartoe worden de met deze semi-artificiële taal reeds lang opgedane ervaring en haar bijzondere eigenschappen ten volle uitgebuit (invariante en autonome morfemen, een agglutineringsindex van 1.00, helderheid van de grammatikale structuur, precisie van de voorzetsels). Men zou het zo kunnen formuleren dat het voor DLT aangepaste Esperanto op zichzelf reeds een boomstructuur belichaamt, compleet met morfo-syntaktische etiketten (grammatikale uitgangen, partikels en functiewoorden). Gegevens over de valentie (van werkwoorden e.d.) worden via de artikelen van het Tussentaal-woordenboek overgebracht.

In een Transfer-systeem als EUROTRA wordt - om goede economische redenen - gepoogd de omvang van de transfer-

operatie zelf tot het uiterste te beperken, namelijk tot een simpele substitutie van lexemen (brontaalwoorden worden vervangen door doeltaalwoorden). Daarentegen kan een echt Tussentaal-systeem zoals DLT profiteren van de aanwezigheid van volledig uitgebouwde Tussentaal-woordenboekkolommen aan beide zijden van de scheidslijn brontaal-doeltaal. De vertaling kan daardoor in hoge mate steunen op valentiegegevens, hetgeen de afwezigheid van abstracte semantische etikettering bij het interface compenseert. Het voordeel van systeemontwikkeling door gescheiden teams (voor brontaal en doeltaal) kan evenwel behouden blijven. Zij zullen zich nu vertrouwd moeten maken met de grammatika en het lexikon van de Tussentaal, in plaats van te komen tot harmonisatie van een abstrakt systeem van etikettering als interface.

Door het terugdringen van de vereiste interface-structuur tot een lineaire reeks van lexikale formatieven biedt het DLT-systeem twee grote voordelen van praktisch belang: de via het interface uitgewisselde data zijn compact, en bovendien snel verifieerbaar. Het eerste speelt in op relatief hoge transmissiekosten, het tweede bevordert ten zeerste de bouw en onderhoudbaarheid van het systeem.

Ondubbelzinnigheid van de Tussentaal.

De studie heeft zich in het bijzonder beziggehouden met de ondubbelzinnigheid van de op Esperanto gebaseerde Tussentaal. Deze ondubbelzinnigheid is een voorwaarde voor een volledig geautomatiseerde vertaalstap van Tussentaal naar doeltaal, en is in dit verband wat preciezer gedefinieerd, in termen van 'ontleedbaarheid' door een eenvoudige zinsontleder, niet berustend op 'diepere' semantiek of kennis van de wereld, maar op (morfo-)syntactische informatie en valentiegegevens uit het Tussentaal-woordenboek. De in deze studie uitvoerig beschreven Tussentaal-grammatika is opgebouwd uit 4 lagen: een basis-laag van gewoon Esperanto, met daarboven 3 modifikatie-lagen, die elk een verdere bijdrage vormen tot de ondubbelzinnigheid van het geheel. De modifikaties omvatten: een strikt voorgeschreven volgorde op woordgroep- en zins-nivo, een beperkt aantal nieuw ingevoerde partikels en functiewoorden, een konsekvent gebruik van leestekens en insertie van een universeel scheidingsteken. Met opzet zijn storende en ruimtevretende buitentalige elementen vermeden: deze zouden de nagestreefde inspekteerbaarheid en compactheid aantasten.

De studie bevat ook een contrastief overzicht van syntactische ambigüiteiten in brontalen (Engels, Duits, etc.), naast ondubbelzinnige vertalingen in de Tussentaal. Geïllustreerd met ruim 100 voorbeelden wordt getoond hoe de alternatieve interpretaties van een ambigue brontekst onderscheidenlijk

zullen worden vertaald in de Tussentaal, dankzij het sterk oplossend vermogen daarvan. In computervertaalkringen bekende struikelblokken worden hierbij overwonnen: woordklas-toekenning, VZC-ambigüiteit, werkwoordnominalisatie, anafora, etc. Hoewel moet worden toegegeven dat zelfs een omvangrijke opsomming van voorbeelden nog geen bewijs oplevert voor het geheel ambigu-vrij zijn van de Tussentaal, wordt wel de uitbreidbaarheid van haar disambiguerend vermogen bewezen. Dit gebeurt aan de hand van een algoritme voor de automatische insertie van scheidingstekens, waarmee een veilige afhandeling van incidentele (en daardoor moeilijk te voorziene) syntaktische dubbelzinnigheden wordt gegarandeerd.

De gefaseerde uitbouw mogelijkheden.

Met de in dit rapport aangetoonde haalbaarheid van een ondubbelzinnige Esperanto-versie als tussentaal komt een volautomatische vertaalstap van tussentaal naar doeltaal binnen het bereik. Gegeven een ambigu-vrij interface, is de bouw van een doeltaalsynthese-moduul konform hetgeen bij de huidige stand van techniek mogelijk is. Op dit punt is er dus geen sprake van een bijzondere doorbraak. Evenwel levert hier het bijzondere karakter van de tussentaal een onmiskenbaar voordeel van grote praktische waarde op: haar toegankelijkheid en doorzichtigheid maken de (van een internationale hulptaal afgeleide) tussentaal niet alleen een voortreffelijk vertaalsysteem-ontwikkelingsinstrument, maar ook een doeltreffende gegevensinvoertaal voor bepaalde meertalige database-toepassingen. Men denke aan de werkomgeving van zorgvuldige tekstgenerering bij viewdata-informatieleveranciers, fabrikanten van natural language database front-ends, bedrijfsafdelingen belast met technische dokumentatie voor export-produkten etc.

Als eerste realiseringsfase van DLT wordt daarom een proefsysteem voor het traject Tussentaal-Doeltaal voorgesteld (met Duits als eerste doeltaal, een geschatte inspanning van 12 manjaar, en een projektduur van 2 jaar). De vertaalkwaliteit van dit proefsysteem zal nog beperkt worden door voorziene onvolmaaktheden in het gebruik van lidwoorden en verleden tijden (de studie geeft echter aan hoe die problemen op langere termijn kunnen worden opgelost). Bovendien moet gerekend worden op een nogal eentonige stijl: een natuurlijke variatie van zinspatronen (thema-rhema) is vooreerst niet realiseerbaar. De juiste vertaling van lexemen hangt uiteraard af van het woordenboek Tussentaal-Doeltaal: de daarin optredende lexikale divergentie en het vermogen, er geformaliseerde woordkeuze-procedures aan toe te voegen. In het kader van het proefsysteem moet het woordkeuze-probleem worden overwonnen

door een beperking tot denotatieve betekenissen en een specifiek toepassingsgebied.

Voor dit toepassingsgebied is de keuze gevallen op "internationale handel en economie", aangevuld met "internationaal recht". Doorslaggevend hiervoor was, afgezien van het grote commerciële belang, de beschikbaarheid van een moderne en goed uitgebouwde terminologie in Esperanto (dus in de Tussentaal). Hoewel verdere uitbouw en aanpassing noodzakelijk zal zijn, vormt de bestaande hoeveelheid van goed gedefinieerde termen een gezonde basis voor het opzetten van een toepassings-georiënteerd vertaalsysteem. Eens te meer wordt hiermee de niet te onderschatten praktische waarde van een echte, alles omvattende en gemakkelijk toegankelijke tussentaal aangetoond.

Vooruitzichten op lange termijn.

Niet eerder dan 5 jaar na realisatie van het (voor de komende 2 jaar voorgestelde) proefsysteem zal een compleet, in hardware gefintegreerd, meertalig DLT systeem (met minimaal 1 brontaal en 2 doeltalen) op de markt kunnen worden gebracht.

De verwezenlijking van een half-automatische brontaalanalyse-moduul, waarin een verantwoord gebruik wordt gemaakt van het (via een computergestuurde dialoog) konsulteren van de tekst-invoerende persoon, is regelrechte innovatie, en in hoge mate karakteristiek en beslissend voor DLT als volledig vertaalsysteem. Van de brontaalmoduul kunnen in dit rapport alleen de belangrijkste ontwerpfacetten genoemd worden: de ontleding verloopt in 1 doorgang van links naar rechts, en wel intervalsgewijs, het tempo volgend waarin de woorden worden ingevoerd (de DLT-brontaalmoduul wordt immers in de word-processor gefintegreerd). Daarbij vindt kwasi-parallele creatie plaats van tussentaalgerichte brontaalbomen, langs een beperkt aantal verschillende 'ontleedsporen'. Het accent ligt hierbij op parallellisme in plaats van backtracking, een taktiek die in de hand gewerkt wordt door de betrekkelijk lage snelheid van tekstinvoer via een word-processor, de aanwezigheid van dedicated processors (rekentijd die "niets kost") en de voorspelde komst van de Mbit geheugenchips tegen het eind van de jaren '80.

De interactie-dialoog is zo in het systeemontwerp ingepast, dat hij alleen kan worden gefinitieerd nadat een zin geheel is ingevoerd en een poging tot automatische disambiguering is gestrand. In deze dialoog moet het systeem de typist verschillende alternatieve interpretaties voorschotelen, zonder gebruik van linguïstisch jargon of van een andere taal dan de brontaal (de typist, al of niet de auteur zelf, wordt alleen geacht de brontaal en de kontekst te kennen). De analysemoduul zal overigens een algoritme bevatten dat de

volgorde van de vraagstellingen optimeert, waardoor hun aantal en de belasting van de typist kan worden gelimiteerd. Deze niet bepaald lichte opgave rechtvaardigt een meer gedetailleerde studie van de analysemoduul binnen het tijdsbestek van het voorgestelde proefstelsel. Een simulatie van de interactie-dialogoog zou onderdeel moeten zijn van die studie.

Voor de verre toekomst biedt DLT uitzicht op een geleidelijke kwaliteitsverbetering in de richting van stylistisch korrekte output, te realiseren door syntactische omzettingen binnen de tussentaalfase (bij de eerste systeemversies zal de structuur van de brontaalzinnen nog doorklinken in de doeltekst). Verder zal de interactie-dialogoog stukje bij beetje worden ontlast door implementatie van makrokontekstgerichte disambigueringsalgoritmen, die gerekend kunnen worden tot het terrein van de KI (kunstmatige intelligentie) en zullen opereren met een op Tussentaal gebaseerde kennisbank. Deze nog ver vooruit liggende ontwikkeling zal dankbaar gebruik maken van de rijkdom (lexicon, terminologieën) en compacte morfematische structuur van de Tussentaal. De onortodoxe interne codering daarvan betekent namelijk een snelheidsvoordeel bij alle string-matching operaties. De toekomstige uitbouw-mogelijkheden zijn daarom nauw verbonden met het nu afgeronde DLT-ontwerp.

Slotconclusie.

De brandende vraag die deze studie overheerste: "Is een op Esperanto gebaseerde, ondubbelzinnige tussentaal haalbaar?", is positief beantwoord. Om te komen tot een volledig DLT-systeem, het hele traject van brontaal tot doeltaal bestrijkend, moeten nog aanzienlijke investeringen gedaan en moeilijke technische hindernissen genomen worden. Toch is de praktische bruikbaarheid van een gedeeltelijke realisering van DLT (traject Tussentaal-Doeltaal) op zichzelf al genoeg om verdere ondersteuning in de vorm van een proefstelsel te rechtvaardigen.

DLT FEASIBILITY STUDY 1982-1983
(Undersøgelse af gennemførlighed)
RESUME

Denne rapport indeholder resultaterne af 14 måneders undersøgelse for EF-Kommissionen af gennemførligheden af DLT (Distributed Language Translation, distribueret sprogoversættelse). DLT er et forslag om et system til halvautomatisk oversættelse mellem skrevne naturlige sprog. DLT blev opfundet og først udforsket af softwarefirmaet BSO/Automation Technology b.v. i Holland i årene 1979-1982. Den foreliggende undersøgelse betegner afslutningen på 'skrivebordsfasen'.

DLT-projektet, som er et faseinddelt, længerevarende projekt, sigter i første række mod økonomiske oversættelser mellem europæiske sprog (begyndende med fransk, tysk og engelsk), men ser ud til at have vældig gode udvidelsesmuligheder til andre sprog (japansk, hebraisk, arabisk). De teksttyper, der kan behandles betegnes som 'informative', spændende fra tekniske brugervejledninger til abstrakter af videnskabelig litteratur, og fra rapporter fra områder inden for forretningsverdenen til forordninger om deponering af atomaffald. Stilistiske effekter, konnotationer og andre spidsfindigheder (såsom 'læse mellem linjerne') kan normalt ikke bevares. Bortset herfra, og med den begrænsning, der ligger i, at kilde-sprogstekstens struktur og ordvalg mere eller mindre bevares, vil DLT-oversættelser være troværdige og grammatisk korrekte.

DLT set i teknisk sammenhæng.

Den bedste måde at betragte DLT på er som et sammenhængende sæt af indbyggede faciliteter i udstyr til tekstgenerering og tekstopsætning, dvs. det bliver brugt i forbindelse med tekstbehandlingssystemer og videotex-terminaler. DLT er typisk ikke et gruppekørselssystem og derfor ikke specielt velegnet til oversættelse af tekstmængder, der er oplagret og læses ind i større mængder, f.eks. via OCR (optisk læsbart). Det er derimod meget interessant for informationsdistributionsnetværker og internationale on-line-tjenester (of-

fentlige såvel som private). DLT er kompatibelt med moderne synkron datatransmission og packet switching teknikker.

Forsøg på at oversætte tekster i naturligt sprog ved hjælp af datamater, en drøm man har haft siden 2. verdenskrig, har hidtil kun givet begrænsede og beskedne resultater. Fagområdet maskinoversættelse (MT) har kun gjort mindre fremskridt og har ikke haft virkelige gennembrud. I de sidste seks år er der imidlertid opstået en fornyet interesse, i Europa og det fjerne Østen, såvel som i Nordamerika, og dette sammen med et mere varieret udbud af 'flersprogede faciliteter' udgør den baggrund, som DLT må ses på.

Flersprogede edb-faciliteter spænder i dag fra tekstbehandlingssystemer med adgang til fremmedsprogede ordbøger til store edb-systemer, der producerer råoversættelser ved gruppekørsel (batch). Den første type kan vi kalde CAT (Computer Aided Translation, datamatstøttet oversættelse), den anden FAT (Fully Automatic Translation, fuldautomatisk oversættelse). I CAT er det en menneskeoversætter eller tekstbehandler, der udfører arbejdet, med hjælp fra datamaten hvis og når han eller hun ønsker det. Der findes imidlertid en tredje type system, som vi kalder SAT (Semi-Automatic Translation, halvautomatisk oversættelse), hvor datamaten udfører arbejdet, men spørger det menneske der betjener den, hvis og når datamaten beslutter det. Det er i denne gruppe af systemer, hvor en automatisk proces suppleres af en datamatstyret interaktiv dialog, at DLT hører hjemme.

I SAT er datamaten afhængig af hjælp fra et menneske til at bestemme betydningen af ord og konstruktioner som er flertydige for datamaten (men ikke for mennesket, som antages at kende den kontekst, kildesprogsteksten indgår i). Udover forskellen i herre-slave-forholdet adskiller CAT og SAT sig ved at SAT er meget nærmere FAT end CAT er. Men mens FAHQ (Fully Automatic High-Quality Translation, fuldautomatisk oversættelse af høj kvalitet) - det er der almindelig enighed om - vil forblive et uopnåeligt mål i de næste 5 årtier eller noget i den retning, så fremtræder SAHQ (Semi Automatic High-

Quality Translation, halvautomatisk oversættelse af høj kvalitet) som en realistisk mulighed (især for et softwarefirma med bred erfaring i design af menneske-maskine dialogsystemer) og det er derfor vi er slået ind på denne kurs med DLT.

Arkitekturen i oversættelsesprocessen.

Der er tre hovedarkitekturer for oversættelsessystemer: den direkte metode, transfermetoden og interlinguametoden, og det må understreges at DLT fra starten er tænkt som et interlinguasystem, leksikalt såvel som grammatisk. Til dette formål bruger vi en delmængde af Esperanto som IL (Intermediate Language, mellemsprog) og en stor del af denne rapport er helliget beskrivelsen af og definitionen af grammatikken for dette interlingua.

Interlinguametoden giver en opdeling af processen i 2 hoveddele (SL-analyse (kildesprogsanalyse), der resulterer i IL, og TL-syntese (målsprogssyntese), der udgår fra IL-versionen), hvilket passer meget fint med de tekniske omgivelser (distribuering af oversættelsesprocessen over sender og modtager i et informationsnetværk). IL'et skal ses som en smal forbindelsesbro, en kompakt udveksling af information mellem kildesprogs- og målsprogsmoduler, gennem telekommunikationsnetværker (hvor afgiften er baseret på den overførte informationsmængde (volume-tariffed)).

Når man sammenligner DLT med et konkurrerende forsøg, det meget store udviklingsprojekt EUROTRA, ser man en bemærkelsesværdig forskel mellem DLT's IL og EUROTRA's overførselsstruktur: EUROTRA bruger en mellemrepræsentation som er en træstruktur med komplekse etiketter, der omfatter semantiske variable såvel som overfladesyntaktiske og morfosyntaktiske variable, dvs. abstrakte formater. DLT's IL derimod består hovedsageligt af en lineær streng af leksikalske formater.

I begge metoder må mellemstrukturen have ekstra oplysninger i forhold til det originale kildesprogsinput: den må være fri for særegenheder og idiosynkrasier ved kildesproget og den må kunne underkastes yderligere behandling af målsprogsorientere-

rede moduler. I særdeleshed må den være fri for flertydigheder.

Hvor EUROTRA synes at udvikle sig mod at lagre mere og mere abstrakt information i overførselsstrukturen, har DLT søgt at nå ovenstående mål ved et omhyggeligt design af dets Esperantobaserede IL, idet man har udnyttet de erfaringer og de lingvistiske egenskaber (invariante og autonome morfemer, Greenberg's agglutination index: 1.00, gennemskuelighed af den grammatiske struktur, et ret præcist system af præpositioner) ved et allerede eksisterende semi-kunstigt sprog. Man kan sige, at den modificerede Esperanto, der bruges i DLT, har en indbygget træstruktur, komplet med morfosyntaktiske etiketter (grammatiske endelser, partikler og funktionsord). Oplysning om valensbindinger ligger i IL-ordbogsindgangene.

Mens et transfersystem som EUROTRA (af indlysende, økonomiske grunde) forsøger at begrænse omfanget af SL-TL-transferoperationen til en ren og skær udskiftning af leksemer (SL-ord udskiftes med TL-ord), så kan et rigtigt interlinguasystem som DLT udnytte tilstedeværelsen af fuldt udbyggede IL-ordbogsindgange på begge sider af SL-TL-vandskellet. I DLT kan oversættelse baseres fuldstændigt på valensbindingsniveauet, hvilket kompenserer for fraværet af markering af abstrakte semantiske relationer. Fordelen ved modulopbygget systemudvikling med adskilte SL- og TL-arbejdsgrupper bevares og tilpasningen til IL-grammatik og -leksikon erstatter den harmonisering der ligger i en fælles abstrakt overførsel af etiketter.

Begrænsningen af DLT's mellemstruktur til en lineær streng af leksikalske formater har 2 praktiske fordele som i høj grad bestemmer hele systemets struktur: mulighed for hurtig kontrol på grund af gennemskueligheden (til gavn for udvikling og vedligeholdelse) og kompakthed (til gavn for billig transmission).

Interlinguaets entydighed.

Hovedproblemet i denne undersøgelse har været entydigheden af DLT's Esperantobaserede IL, en nødvendig forudsætning for fuldt automatisk oversættelse fra IL til TL. Til dette formål er 'entydighed' blevet mere præcist defineret her, nemlig ved at IL-sætningen kan analyseres ved en simpel parser, der ikke involverer dybdesemantik eller viden-om-verden, men bygger på (morfo-)syntaktisk og (IL-leksikon-baserede) valensoplysnin-ger. IL-grammatikken, som er beskrevet i hoveddelen af rap-porten, er blevet konstrueret ved tilføjelse af 3 modifice-rende 'lag' ovenpå almindelig Esperantos basale lag. Hvert af disse nye lag bidrager til entydiggørelsen af IL'et. Modifi-kationerne omfatter: en streng lovgivning om rækkefølgen af ord og ordgrupper, indførelse af et begrænset antal nye funk-tionsord og partikler, konsistent brug af interpunktionstegn, og indsættelse af et generelt separationselement. Vi har sær-ligt været opmærksom på at undgå pladskrævende eller overflø-dige ekstralingvistiske elementer, som kunne påvirke IL'ets kompakthed og gennemskuelighed.

En vigtig del af undersøgelsen har været at teste entydighe-den af det resulterende IL ved sammenligning med syntaktiske flertydigheder på forskellige kildesprog (engelsk, tysk, osv.). I rapporten viser over 100 eksempler IL'ets evne til at give forskellige repræsentationer for alternative læsnin-ger inden for forskellige områder af syntaktisk flertydighed, omfattende flere kendte hindringer ved maskinoversættelse: ordklasse, funktionsord, PP-flertydighed (PP = Prepositional Phrase, præpositionsled), verbers nominalisering, anaforer osv. Skønt man må indrømme, at selv en så omfattende opreg-ning ikke beviser, at IL'et er fuldstændig entydigt, så bli-ver der givet et bevis for at IL'ets evne til at modstå fler-tydighed kan udvides. En algoritme til automatisk indsætning af separationsmarkører, forklaret i denne sammenhæng, garan-terer en sikker behandling af tilfældige (og derfor vanske-ligt forudsigelige) syntaktiske flertydigheder.

Den samme algoritme beskytter IL'et mod den systematiske flertydighed, der er meget almindelig, i virkefelt for kon-junktioner og bestemmere (algoritmen følges i visse tilfælde

af en interaktiv entydiggørelsesdialog).

Fordele ved faseinddeling.

Efter at det er blevet bekræftet, ved det IL-design, der er beskrevet i rapporten, at det er muligt at konstruere et entydigt 'Esperanto'-IL, er en fuldt automatisk IL-til-TL-oversættelse inden for rækkevidde. I sig selv står konstruktionen af syntesemoduler for naturlige sprog (ud fra en entydig mellemform) i et rimeligt forhold til den nuværende state-of-the-art på området og frembyder ikke noget krav om teknologisk gennembrud. I dette tilfælde har IL'ets særlige natur dog en ubestridelig praktisk betydning: afledt som det er af et sprog til menneskelig kommunikation, gør dets tilgængelighed og gennemskuelighed IL'et til et strålende udviklingsmedium, men også til et effektivt sprog til datainput til visse anvendelser i flersprogede databaser (nemlig de begrænsede emneområder inden for produktionen af videotex-information, for en produktionsafdeling for tekniske manualer, opbygningen af et eksperter-system, osv.

Den næste fase, vi foreslår for DLT, er derfor udviklingen af et pilotprojekt, som omfatter mindst et IL-til-TL modul (med tysk som det foretrukne første målsprog, og med en anslået arbejdsindsats på 12 årsværk over en 2-års periode). Med hensyn til kvaliteten af oversættelsen, forventes TL-outputtet at vise unøjagtigheder i brugen af artikler og fortidsformer (skønt undersøgelsen angiver, hvordan man i sidste ende kan klare disse mangler). Endvidere må man regne med en lidt monoton stil, fordi en naturlig variation af sætningsmønstre i overensstemmelse med fokus (thema-rhema) er svær at gennemføre. Rigtig oversættelse af leksemer, som bestemmes af leksikalske forskelle mellem IL og TL, og af evnen til at formalisere og at tilføje ordvalgsprocedurer til IL-TL-ordbogen, kan opnås ved at begrænse pilotsystemet til det denotative register ('tekniske' tekster) og til et bestemt anvendelsesområde.

International handel og økonomi, suppleret med international lov, er blevet valgt som pilot-anvendelsesområde. Grunden til

dette valg, udover emnets brede kommercielle interesse, er at der allerede findes omfattende og up-to-date terminologi på Esperanto (dvs. på IL). Skønt det er nødvendigt at arbejde videre hermed, giver den eksisterende mængde af veldefinerede termer en sund basis at bygge et anvendelsesorienteret oversættelsessystem på og viser endnu engang den praktiske værdi af et ægte, omfattende og let tilgængeligt IL.

Fremtidsperspektiver.

Tidshorizonten for at bringe et komplet, hardware-integreret, flersproget DLT-system på markedet er ca. 5 år, efter det begrænsede pilotsystem, der foreslås gennemført i de næste 2 år.

Opbygningen af et halvautomatisk SL-til-IL analysemodul, i hvilket der gøres omhyggeligt brug af at spørge tasteoperatøren, udgør en meget vigtig og karakteristisk del af DLT-udviklingen, lidt forskellig fra eksisterende og konkurrerende tiltag. Kun globale karakteristika ved designet af SL-analyseprocessen kan afsløres nu: intervalvis, data-drevet, single-pass, venstre-til-højre analyse, i trit med at tasteoperatøren taster ordene ind (DLT integreres i tekstbehandlingssystemet); trinvis quasi-parallel opbygning af IL-rettede SL-træer langs et (moderat) antal alternative analyseveje. Hovedvægten lægges på at behandle non-determinisme ved parallelle veje i stedet for ved back-tracking. Denne fremgangsmåde støttes ved at den manuelle indtastning foregår relativt langsomt (og derved giver tidsrum med 'gratis' køretid for dedikerede processorer) og ved den forventede mulighed for at kunne bruge lagerchips med høj kapacitet.

Med hensyn til den interaktive entydiggørelsesdialog vil denne ikke blive igangsat før en hel inputenhed (en sætning) er indtastet, og kun efter at et forsøg på automatisk entydiggørelse er mislykket. Systemet må generere spørgsmål, som gør eksistensen af alternative løsninger klar, uden at bruge lingvistisk fagsprog eller andre sprog end kildesproget (tasteoperatøren, som måske er forfatteren til inputteksten, forventes kun at kende dens kontekst og kildesproget). Endvidere

vil analysemodulet indeholde en algoritme, der optimerer rækkefølgen af spørgsmål og derved reducerer deres antal og maskinskriverens arbejdsbyrde. Dette er indviklede opgaver for SL-analysemodulet, som kræver et mere detaljeret studium inden for det foreslåede pilotprojekts tidsrammer. En simulering af den interaktive dialog bør indgå i dette projekt.

På længere sigt giver DLT mulighed for en gradvis kvalitetsforbedring henimod stilistisk korrekt TL-output, ved syntaktiske omformninger i IL'et (det første systems versioner vil producere TL-output, som, skønt det er grammatisk korrekt, stadig afspejler strukturen af SL-inputtet). Endvidere kan en meget gradvis nedsættelse af mængden af interaktiv dialog måske opnås ved at implementere entydiggørelsesalgoritmer, der er af makro-kontekst-orienteret AI-type (AI = Artificial Intelligence, kunstig intelligens) og opererer på IL'et sammen med en IL-baseret viden-bank. Denne lidt fjerne fremtidsudvikling vil udnytte rigdommen i udtrykket (ordforråd, terminologier) og den kompakte morfematiske struktur i IL'et (hvis uortodokse interne kodning gør alle streng-sammenlignende operationer hurtigere), og er derfor tæt forbundet med det specifikke DLT-design, der præsenteres i dag.

Endelig konklusion.

Hovedproblemet i denne undersøgelse, muligheden af at skabe et entydigt Esperantobaseret IL, er blevet besvaret positivt. Selv under hensyn til den store investering og de ikke-trivielle problemer der må klares for at udvikle det endelige (SL-til-TL) DLT-system, begrundes den praktiske betydning af en kun partiel implementering (IL-til-TL) i sig selv fortsat støtte i form af et pilotprojekt.

REFERENCES.

- AGAPEYEFF, d', A. - 1982.
An Introduction to the Fifth Generation.
In: The Fifth Generation: dawn of the second computer age.
International Conference, SPL International.
London, July 1982: pp. AA1-AA11.
- AGRICOLA, E. - 1968.
Syntaktische Mehrdeutigkeit bei der Analyse des Deutschen und des Englischen.
Akademie-Verlag, Berlin (GDR).
- AKTOJ de la Akademio 1963-1967 - 1968.
Oficiala Bulteno de la Akademio de Esperanto.
UEA, Rotterdam: No 9.
- ANDEOL, P. - 1979.
L'homme victime de son intelligence.
In: Science et Vie.
Nov. 1979, No. 746.
- ANDREYEV, N.D. - 1967.
The intermediate language as the focal point of machine translation.
In: Booth, A.D., Ed.
Machine Translation.
North-Holland Publ. Co., Amsterdam: pp 1-27.
- ASIAGRAPHICS - 1982.
Amerikanen toveren Chinese karakters op beeldscherm.
In: Computable.
Feb. 19, 1982, No.7: p. 3.
- AULD, W. - 1978.
Pri lingvo kaj aliaj artoj.
La Laguna, Antwerp.
- BAR-HILLEL, Y. - 1955.
Idioms.
In: Locke, W.N. and A.D. Booth, Eds.
Machine Translation of Languages.
Chapman & Hall, Ltd., London: pp. 183-193.
- BATES, M. - 1978.
The theory and practice of augmented transition network grammars.
In: Bolc, L., Ed.
Natural Language Communication with Computers.

Springer Verlag, Berlin (FGR): pp. 191-259.

BILLMEIER, R. - 1982.

Zu den linguistischen Grundlagen von SYSTRAN.

In: Multilingua.

Volume 1-2: pp. 83-96.

BLAAS, L. - 1951.

Statistiko de 50000 tekstvortoj.

In: Esperantologio.

Volume I, March 1951.

BLANKE, D. - 1979.

Kelkaj problemoj de la vortfarado de la germana lingvo kaj de Esperanto.

In: Blanke, D., Ed.

Esperanto. La internacia lingvo - sciencaj aspektoj.

Zentraler Arbeitskreis Esperanto, Berlin (GDR): pp.116-134.

BLANKE, D. - 1982.

Plansprache und Nationalsprache.

Akademie der Wissenschaften, Berlin (GDR).

BOURQUIN, G. - 1983.

[University of Nancy;

Personal Communication].

BRANDT CORSTIUS, H. - 1978.

Computer-Taalkunde.

Coutinho, Muiderberg (Netherlands).

BRUDERER, H.E. - 1978.

Handbuch der maschinellen und maschinenunterstützten Sprachübersetzung.

Verlag Dokumentation Saur KG, Munich.

BURSKY, D. - 1983.

Semiconductor Technology.

In: Electronic Design.

June 9, 1983: pp. 101-110.

CARBONELL, J.G. et al. - 1981.

Steps toward knowledge-based machine translation. IEEE.

Vol. PAMI-3, No. 4, July 1981.

CHARNIAK, E. - 1983.

A Parser with Something for Everyone.

In: King, M., Ed.

Parsing Natural Language.

Academic Press, London: pp 117-150.

- CHURCH, K. and R. Patil - 1982.
Coping with Syntactic Ambiguity or How to Put the
Block in the Box on the Table.
In: American Journal of Computational Linguistics.
July 1982: pp. 139-149.
- CLARK, K.L. and S.A. Tärnland - 1982.
Logic Programming.
Academic Press, London.
- COLMERAUER, A. - 1978.
Metamorphosis Grammars.
In: Bolc, L. Ed.
Natural Language Communication with Computers.
Springer-Verlag, Berlin: pp. 133-189.
- DAM, J. VAN - 1972.
Syntax der deutschen Sprache.
Wolters-Noordhoff Publishing, Groningen (The Netherlands).
- DIJK, M. VAN, Ed. - 1981.
Mieux traduire pour mieux communiquer.
Bureau Marcel van Dijk (European Commission),
Brussels.
- DIK, S.C. - 1968.
Coordination - its implications for the theory of
general linguistics.
North-Holland Publ. Co., Amsterdam: pp. 200-249.
- DIK, S.C. - 1978.
Functional grammar.
North-Holland Publ. Co., Amsterdam.
- DIK, S.C. and J.G. Kooij - 1979.
Algemene Taalwetenschap.
Het Spectrum, Utrecht.
- DRASKAU, J. - 1983.
Toward a Clarification of the Concept Jargon.
In: Fachsprache.
Vol. 1/1983: pp. 11-24.
- DRESSLER, W. - 1972.
Zur plansprachlichen Redundanz.
In: La Monda Lingvo-Problemo.
I.E.I., The Hague, vol. 4, 10, pp. 27-33.
- DUCROT, J.M. - 1983.
TITUS IV: système de traduction automatique et

- simultanée en quatre langues.
In: Taylor, P.J. and B. Cronin, Eds.
Proceedings of the EURIM 5 Conference.
Aslib, London: pp 177-195.
- EVANCZUK, S., Ed. - 1983.
Unix Special Series.
In: Electronics.
July 28, 1983: pp. 113-124.
- FILLMORE, Ch. - 1968.
The Case for Case.
In: Bach, E. and R. Harms, Eds.
Universals in Linguistic Theory.
Rinehart and Winston, New York: pp. 1-90.
- FOX, M.S., D.J. Bebel and A.C. Parker. - 1980.
The automated dictionary.
Computer (IEEE).
July 1980.
- FRITSCH, L.
Teknikaj vortaroj en Esperanto - problemoj de la
fakterminologio en Esperanto.
In: Blanke, D., Ed.
Esperanto. La internacia lingvo - sciencaj
aspektoj.
Zentraler Arbeitskreis Esperanto, Berlin (GDR):
pp. 193-199.
- FUJITANI, L. and G. Knight. - 1983.
Optical Storage: the Products.
In: Systems International.
Sep. 1983: pp. 62-64.
- GALLAIS-HAMONNO, J. - 1982.
Langage, langue et discours économiques.
Centre d'Analyse Syntaxique, University of Metz.
- GILB, T. - 1973.
Reliable EDP Application Design.
Studentlitteratur, Lund (Sweden).
- GIVOJE, M. - 1979.
Esperantonimoj.
Edistudio, Pisa (Italy).
- GONINAZ, M.D. - 1976.
In: Tišljarić, Z., Ed.
Internacia Lingvistika Simpozio.
Internacia Kultura Servo, Zagreb.

- GOODALL, A. - 1983.
Language of Intelligence.
In: Systems International.
Jan. 1983: pp. 21-24.
- GOSHAWKE, W. - 1974.
Spoken Languages Universal Numeric Translation
(SLUNT).
Unpublished paper obtainable from the Social Science
Research Council, Survey Archive, University of Essex.
- GREVISSE, M. - 1964.
Le Bon Usage. Grammaire française.
Duculot, S.A., Gembloux (Belgium).
- HENDRIX, G.G. and E.D. Sacerdoti. - 1981.
Natural-Language Processing. The Field in
Perspective.
In: BYTE.
Sep. 1981: pp. 304-352.
- HOARE, J.P. and M.A. Labode. - 1974.
Electrochemical Machining.
In: Scientific American.
Jan. 1974: pp. 34-35.
- HORECKÝ, J. - 1982.
Zu Bedeutungsbeziehungen zwischen den
terminologischen Benennungen.
In: Fachsprache.
Wilhelm Braumüller, Wien: Heft 2; pp. 50-54.
- HUNDT, M.G. - 1982.
Working with the Weidner Machine-aided Translation
System.
In: Lawson, V., Ed.
Practical Experience of Machine Translation.
North-Holland Publ. Co., Amsterdam: pp. 45-51.
- HUTCHINS, W.J. - 1978.
Progress in Documentation. Machine Translation and
Machine-aided Translation.
In: Journal of documentation.
Vol. 34., No. 2, June 1978: pp. 119-159.
- HUTCHINS, W.J. - 1982.
The evolution of machine translation systems.
In: Lawson, V., Ed.
Practical Experience of Machine Translation.
North-Holland Publ. Co., Amsterdam: pp. 21-37.
- JACOMO, LO, F. - 1981.

Liberté ou Autorité dans l'évolution de l'Espéranto.
Thèse pour le doctorat de troisième cycle.
Edistudio, Pisa (Italy).

- JOHNSON, R. - 1983.
Parsing with Transition Networks.
In: King, M., Ed.
Parsing Natural Language.
Academic Press, London: pp. 59-72.
- KALOCSAY, K. and G. Waringhien - 1980.
Plena Analiza Gramatiko de Esperanto.
UEA, Rotterdam.
- KAY, M. - 1973.
The MIND System.
In: Rustin, R., Ed.
Natural Language Processing.
Algorithmic Press, New York: pp. 111-154.
- KAY, M. - 1982.
Machine Translation.
In: American Journal of Computational Linguistics.
Vol. 8, Nov. 2, pp. 74-78.
- KELLY, E. and Ph. Stone - 1975.
Computer recognition of English word senses.
North-Holland Publ. Co., Amsterdam.
- KELLY, I.D.K. - 1978.
How Esperanto can aid Machine Translation.
In: Computer Weekly.
Series in the 19 January - 15 June issues.
- KELLY, I.D.K. - 1982.
The Machine Translation of Open-ended Answers.
Presentation at ESDMAK (European Society for Opinion
and Marketing Research); Conference, Vienna.
- KING, M. and S. Perschke - 1982.
Eurotra and its objectives.
In: Multilingua.
Mouton Publishers, The Hague: 1-1, pp. 27-32.
- KING, M. - 1982.
Eurotra: an attempt to achieve multilingual MT.
In: Lawson, V., Ed.
Practical Experience of Machine Translation.
North-Holland Publ. Co., Amsterdam: pp. 139-148.
- KITTREDGE, R. - 1982.
Sublanguages.

In: Computational Linguistics.
Apr.-June 1982, Vol. 8, No 2: pp. 79-84.

KNOWLES, F.E. - 1982.

The pivotal role of the various dictionaries in an
MT system.

In: Lawson, V., Ed.

Practical Experience of Machine Translation.

North-Holland Publ. Co., Amsterdam: pp. 149-162.

KOOIJ, J.G. - 1971.

Ambiguity in natural language.

North-Holland Publ. Co., Amsterdam.

KOSTER, C.H.A. - 1977.

CDL - a compiler implementation language.

Proceedings of the Conference on Methods of
Algorithmic Language Implementation.

In: Bolc, L., Ed.

Lecture Notes in Computer Sciences.

Springer Verlag, Berlin: Vol. 47.

KÖVES, P. - 1982.

The MPROLOG programming environment today and
tomorrow.

In: Workshop on PROLOG programming environments.

Lingköping, Sweden, March 1982.

KOWALSKI, R. - 1982.

Logic Programming for the fifth generation.

In: The fifth generation: dawn of the second computer
age.

International Conference, SPL international.

London, July 1982: pp. RK/0-RK17.

LAPENNA, I. - 1972.

The Language Problem in International Relations.

Centre for Research and Documentation on the Language
Problem: series A / section II / No. 5.

LAPENNA, I. - 1982.

Juraj aspektoj de informperdo pro traduko.

In: Frank, H.G. et al., Eds.

Lingvokibernetiko / Sprachkybernetik.

Gunter Narr Verlag, Tübingen: pp. 82-91.

LAPENNA, I. - 1983.

[University of London, personal communication].

LEHRER, A. - 1974.

Verb and Adverb Interaction: A Problem for Semantic
Interpretation.

In: L.A.U.T. Series.

Linguistic Agency University at Trier; Trier: Series A,
Paper No 21, March 1974.

LUCKHARDT, H.-D., and H.-D. Maas. - 1983.

SUSY - Handbuch für Transfer und Synthese.

Saarland University, SBF 100, Saarbrücken: Neue
Folge, Heft 7.

LUXEMBURG. - 1977.

Overcoming the Language Barrier.

3rd European Congress on Information Systems and
Networks.

Luxemburg, May 1977.

Munich, Vlg. Dokumentation, 1977-1978. 2 Vols.

LYONS, J. - 1969.

Introduction to Theoretical Linguistics.

University Press Cambridge.

LYONS, J. - 1981.

Language, meaning and context.

The Chaucer Press, Bungay, Suffolk.

MAAS, H.-D. - 1982.

Unpublished presentation at the Saarland University
on the SUSY-system: Saarbrücken, Oct. 7, 1982.

MAAS, H.-D. - 1983.

[Saarland University, Saarbrücken.

Personal Communication].

MARCUS, M.P. - 1980.

A theory of syntactic recognition for natural
language.

The MIT Press. Cambridge (Mass.).

MARTIN, J. - 1973.

Design of Man-Computer Dialogues.

Prentice Hall, New Jersey.

MARTIN, J. - 1977.

Computer data-base organization.

Prentice Hall, New Jersey.

MASTERMAN, M. - 1979.

The essential skills to be acquired for machine
translation.

In: Snell, B.M., Ed.

Translating and the Computer.

North-Holland Publ. Co., Amsterdam: pp. 159-180.

- MAURER, H.A. and I. Sebestyen. - 1983.
Some aspects of the market penetration of 1W and 2W
videotex.
In: Electronic Publishing Review
March 1983, Vol. 3, No. 1: pp. 55-77.
- McDERMOTT, D. - 1980.
The Prolog Phenomenon.
SIGART, Apr. 1980.
- MEGADOC. - 1983.
MEGADOC Systems Description.
Philips Data Systems.
Apeldoorn (The Netherlands), 1983.
- MELBY, A.K. - 1981.
Translators and Machines - Can they Cooperate?
In: META, XXVI, 1: pp. 23-34.
- MINOR, K. - 1975.
Deutsche Redensarten in Esperanto.
Sarlanda Esperanto-Ligo, Saarbrücken.
- MOTO-OKA, T. - 1982.
Fifth Generation Computer Systems.
North-Holland Publ. Co., Amsterdam.
- MUENNICH, A. - 1976.
A common language for man-machine interactive
systems.
In: Papp, F. and G. Szépe, Eds.
Papers in Computational Linguistics.
Mouton Publishers, The Hague: pp. 493-504.
- MUNNIKSMA, F. - 1975.
International Business Dictionary.
Kluwer, Deventer (The Netherlands).
- NEWMARK, P. - 1981.
Approaches to translation.
Pergamon Press, Oxford.
- NICHOLLS, J.C. and D. van Essen. - 1974.
The nervous system of the Leech.
In: Scientific American.
Jan. 1974: p. 43.
- OPITZ, K. - 1980.
Language for special purposes: an intractable presence?
In: Fachsprache.
Vol. 1/1980: pp. 21-27.

- FENNACCHIETTI, F. - 1982.
Ne-hind-europeaj trajtoj de la Internacia Lingvo.
In: Frank, H.G. et al., Eds.
Lingvokibernetiko / Sprachkybernetik.
Gunter Narr Verlag, Tübingen: pp. 92-95.
- FIGOTT, I.M. - 1982a.
Systran - a reality of the translation world of the
1980's.
In: Lustrumboek Christiaan Huygens.
Delftsche Uitgeversmaatschappij b.v., Delft
(The Netherlands): pp. 176-191.
- FIGOTT, I.M. - 1982b.
Einführung in: das maschinelle Uebersetzungssystem
Systran.
Unpublished presentation at the Kernforschungszentrum
Karlsruhe, Sep. 23, 1982.
- PIRON, C. - 1979.
Esperanto, ĉu eŭropea au azia lingvo?
In: Esperanto-Dokumentoj.
UEA, Rotterdam: vol. 10.
- PITRAT, J. - 1980.
Les programmes qui "comprennent" le langage naturel.
In: Lingvisticae Investigationes.
IV, 2: pp. 395-414.
- REIFLER, E. - 1955.
The mechanical Determination of Meaning.
In: Locke, W.N. and A.D. Booth, Eds.
Machine translation of languages.
Chapman & Hall, Ltd., London: pp. 137-164.
- RHODES, I. - 1967.
The importance of the glossary storage in machine
translation.
In: Booth, A.D., Ed.
Machine Translation.
North-Holland Publ. Co., Amsterdam: pp. 429-449.
- RITCHIE, G. - 1983.
Semantics in Parsing.
In: King, M., Ed.
Parsing Natural Language.
Academic Press, London: pp. 199-217.
- ROLLING, L. - 1983.
Machine Translation's big ten years: 1980-1990.
In: Taylor, P.J. and B. Cronin, Eds.
Proceedings of the EURIM 5 Conference.

Aslib, London: pp. 169-176.

SAGER, N. - 1981.

Natural Language Information Processing.
Addison-Wesley Publ. Co., Inc., London.

SAMPSON, G. - 1983.

Deterministic Parsing.
In: King, M., Ed.
Parsing Natural Language.
Academic Press, London: pp. 91-116.

SCIARONE, A.G. - 1983.

[Technological University Delft, The Netherlands;
personal communication].

SEBESTYEN, I. - 1982.

The videodisc revolution.
In: Electronic Publishing Review.
Vol. 1, March 1982: pp. 41-89.

SLYPE, VAN, G. - 1982.

Conception d'une méthodologie générale d'évaluation
de la traduction automatique.
In: Multilingua.
1982, 1-4: pp. 221-237.

SLYPE, VAN, G. - 1983.

Le marché de la traduction.
In: Multilingua.
1983, 2-1: pp. 9-18

SMITH, C.S. - 1969.

Ambiguous sentences with 'and'.
In: Reibel, D.A. and S.A. Schane, Eds.
Modern Studies in English - Readings in
transformational grammar.
Prentice-Hall, Inc., Englewood Cliffs,
New Jersey: pp. 75-79.

SMITH, P. - 1982.

Japanese Input.
In: Systems International.
November 1982: pp. 81-82.

SOMERS, H.L. - 1983.

Investigating the possibility of a microprocessor-
based machine translation system.
In: ACL Proceedings of the Conference on Applied
Natural Language Processing.
Santa Monica, California: 1-3 February 1983.

- SWALES, J. - 1981.
Writing Scientific English.
Nelson, Walton-on-Thames, England.
- SZEREDI, P. - 1981.
Mixed Language Programming - a method for producing
efficient PROLOG programs.
Presented at workshop "Logic Programming for
Intelligent Systems".
Los Angeles, California, Aug. 1981: pp. 149-155.
- SZEREDI, P. and E. Sántáné-Tóth.
Prolog Applications in Hungary.
In: The Fifth Generation: Dawn of the Second Computer
Age.
International Conference, SPL International.
London, July 1982: pp. PS/0-PS18.
- TANENBAUM, A.S. - 1982.
Unpublished presentation on UNIX.
In: Computable.
May 21, 1982: pp. 13, 31.
- TIEFENTHAL, D. - 1980.
Abstract.
In: Multilingua.
Vol. 1, No. 1, 1982: p. 64.
- TOMA, P. - 1977.
Systran as a Multilingual MT system.
In: Overcoming the language barrier.
Third European Congress on Information Systems
and Networks: Luxemburg, May 1977.
Munich: pp. 569-581.
- TRELEAVEN, Ph.C. and I.G. Lima. - 1982.
Fifth Generation Computer Architecture Analysis.
In: The Fifth Generation: Dawn of the Second Computer
Age.
International Conference, SPL International.
London, july 1982: pp. PT/0-11.
- ULIJN, J.M. - 1978.
Frans als vreemde taal in een ingenieursopleiding.
Thesis, University of Nimeguen (The Netherlands) 1978.
- ULLMAN, S. - 1962.
Semantics.
Blackwell, Oxford.
- VARILE, N. - 1983.
Charts: a Data Structure for Parsing.

- In: King, M., Ed.
Parsing Natural Language.
Academic Press, London: pp. 73-87.
- WARINGHIEN, G. - 1959.
Lingvo kaj vivo - esperantologiaj eseoj.
La Laguna, Antwerp.
- WARINGHIEN, G. - 1977.
Plena Ilustrita Vortaro de Esperanto.
SAT, Paris.
- WARINGHIEN, G. - 1980.
1887 kaj la sekvo.
La Laguna, Antwerp.
- WARREN, D. and L. Pereira. - 1977.
Prolog - the language and its implementation compared
with Lisp.
In: Proceedings of the symposium on AI and Programming
Languages.
SIGPLAN/SIGART, 1977: pp. 109-115.
- WEINREICH, U. - 1966.
Explorations in semantic theory.
In: Sebeck, Ed.
Current Trends in Linguistics.
Mouton, The Hague: Vol. III: pp. 395-477.
- WELLS, J. - 1978.
Lingvistikaj Aspektoj de Esperanto.
UEA, Rotterdam.
- WHITELOCK, P.J. and K.J. Kilby. - 1982.
An in-depth study of Machine Translation techniques.
Interim Report.
Centre of Computational Linguistics. University of
Manchester. Institute of Science and Technology.
- WIELINGA, B.J. - 1982.
[Municipal University Amsterdam;
personal communication].
- WILKS, Y. - 1973.
An artificial intelligence approach to machine
translation.
In: Schank, R.C. and K. Colby, Eds.
Computer models of thought and language.
Freeman, San Francisco: pp.114-151.
- WILKS, Y. - 1979.
Machine Translation and Artificial Intelligence.

- In: Snell, B.M., Ed.
Translating and the Computer.
North-Holland Publ. Co., Amsterdam: pp. 27-44.
- WITKAM, A.P.M. and J.J. Hillan. - 1981a.
Resolving Language Barriers in International Videotex
Communication.
In: Cantraine, G. and J. Destine, Eds.
New Systems and Services in Telecommunications.
North-Holland Publ. Co., Amsterdam: pp. 143-153.
- WITKAM, A.P.M. - 1981b.
Natural Language Translation in Information
Distribution Networks.
Presentation held at COMNET, Budapest, May 1981.
- WITKAM, A.P.M. - 1981c.
Esperanto as intermediate language in computer
networks - faults and remedies (originally published
in Esperanto).
In: Bertin, C., Ed.
Unua Simpozio pri Komputiko.
CCETT, Rennes (France): July 1981: pp. 196-205.
- WITKAM, A.P.M. - 1983.
An alternative strategy for steady growth towards
high quality translation networks.
In: Taylor, P.J. and B. Cronin, Eds.
Proceedings of the EURIM 5 Conference.
Aslib, London: pp. 196-204.
- WOJNICKI, S. - 1982.
Le français de spécialité: vers une procédure moderne
de description linguistique.
In: Fachsprache.
Vol. 2/1982: pp. 54-62.
- WOOLF, F. - 1969.
English Phrases and Expressions in Esperanto.
Esperanto Press, Cardiff, Ontario (Canada).
- WUESTER, E. - 1966.
Internationale Sprachnormung in der Technik.
Sprachforum, Bonn: Beiheft No 2: pp. 85.
- WUESTER, E. - 1968.
Die Verhältnswörter des Esperanto.
Arthur E. Iltis, Saarbrücken.
- WUESTER, E. - 1975.
Konturoj de la lingvonormigo en la tekniko.
Dansk Esperanto-Forlag, Aabyhøj (Danmark).

WUESTER, E. - 1978.

Esperantologiaj studoj.
La Laguna, Antwerp.

WUESTER, E. - 1979.

Einführung in die allgemeine Terminologielehre und
terminologische Lexikographie.
Infoterm/ON, Wien.

✓ ZILINA, - 1981.

Aplikado de Esperanto en Scienco kaj Tehniko.
Conference in Zilina, CSSR, Aug. 14-17, 1981.

GLOSSARY OF ABBREVIATIONS.

Acc	Accusative case.
ACI	Accusativus cum Infinitivo. (Accusative with the Infinitive).
ACP	Accusativus cum Participio. (Accusative with the Participle).
ADJ	Adjective.
ADJ1	Adjective preceding the head-noun.
ADJ1-DESCR	Descriptive chain of ADJ1.
ADJ1-RESTR	Restrictive chain of ADJ1.
ADJ2	Adjective following the head-noun.
ADOC	Complement of a substantivized verb, also called 'ADO'-complement (used in DLT's ATNs).
Adv	Adverb.
ADVCL	Adverbial Clause.
Ag	Agent of the passive.
AI	Artificial Intelligence.
ALGOL	Programming Language of European origin.
APPOS	Apposition.
ASCII	American Standard Code for Information Interchange.
ATN	Augmented Transition Network.
ATROP	Absent Trailing Optionals.
B	Byte.
BCE	Binary Coded Esperanto.
BNF	Backus-Naur Form.
BNUM	Basic Numeral.
BOP	Bit-Oriented Protocol.

BSO	Buro voor Systeemontwikkeling. (Independent Dutch software-house; prime contractor for the DLT-project).
BSO/AT	BSO/Automation Technology. (The scientific branch of BSO).
C	Programming Language.
CADJ	Consecutive Adjunct.
CAT	Computer Aided Translation.
CCE	Character Coded Esperanto.
CDL2	Compiler Development Language [Koster, 1977].
CEC	Commission of the European Communities.
CETA	Centre d'études pour la Traduction Automatique (former MT system of Grenoble).
COM	Clause Complement (used in DLT's ATNs).
CONJ	Conjunction.
CRAL	Centre de Recherche et d'Application Linguistique.
DB	Data Base.
DCI	Dativus cum Infinitivo. (Dative with the Infinitive).
DEF	Dependent Clause.
DET	Determiner.
DLT	Distributed Language Translation. (name of the MT system presented in this report).
DOR	Digital Optical Recording.
EC	European Communities.
ECU	European Counting Unit.
EUROTRA	European Translation System. (advanced MT system currently being developed for the EC).

Esp	Esperanto.
FADJ	Free Adjunct.
FAHQT	Fully Automatic High Quality Translation.
GEN	General Clause.
GETA	Groupe d'études pour la Traduction Automatique. (later MT system of Grenoble).
GB	Gigabyte.
HLDC	High Level Data Link Control.
IL	Intermediate Language.
INF	Infinitive.
INFCL	Infinitival Clause.
IP	Information Provider.
ISO	International Standardization Organization.
J	Jump (used in ATNs).
KB	Kilobyte.
LADJ	Context Linking Adjunct.
LAN	Local Area Network.
LISP	Programming language of American origin.
LR	Left-to-right (parsing).
LSP	Language for special purposes.
MB	Megabyte.
MBM	Magnetic Bubble Memory.
METALS	Mechanical Translation and Analysis of Languages (MT system of Texas).
METEO	Canadian MT system for Meteorology.
MT	Machine Translation.
N	Noun.

NL	Natural Language.
Nom	Nominative case.
NDM	Nomination.
NP	Noun Phrase.
NSI	Neutral Subject Infinitive.
NUM	Numeral.
O	Object.
Oam	Accusative of Measure Object.
Oav	Adjective Valency Object.
OCR	Optical Character Recognition.
Opc	Object Predicative Complement.
OS	Operating System.
OSI	Open System Interconnection.
Ovv	Verb Valency Object.
PAG	Plena Analiza Gramatiko de Esperanto. (Esperanto Academy Grammar).
PART	Participle.
PC	Personal Computer.
PIV	Plena Ilustrita Vortaro de Esperanto. (Esperanto Academy Dictionary).
PP	Prepositional Phrase.
PREP	Preposition.
PROLOG	Programming language of European origin.
PROM	Programmable Read-only Memory.
RAM	Random Access Memory.
REL	Relative Clause (used in DLT's ATNs).
Rim	Rimarko (Note in PAG).

S	Subject.
SAHQT	Semi-Automatic High Quality Translation.
SBADJ	Subject Bound Adjunct (used in DLT's ATNs).
SBC	Single-board computer.
SDCL	Synchronous Data Link Control.
SIA	Structural Impact of Ambiguity.
Ske	Subject in the form of a dependent clause.
SL	Source Language.
SLUNT	Spoken Languages Universal Numeric Translation (MT system proposed by Goshawke [1974]).
SNA	Systems Network Architecture.
SOV	Subject-Object-Verb sequence.
Spc	Subject Predicative complement.
SUBCL	Subordinate clause (used in DLT's ATNs).
SUSY	Saarbrücker Uebersetzungssystem. (MT system in Germany).
SV	Subject-Verb sequence.
SVO	Subject-Verb-Object sequence.
SYSTRAN	System Translation (renown of American origin).
TGG	Transformational generative grammar.
TL	Target Language.
UNO	United Nations Organization.
V	Verb.
VDU	Visual Display Unit.
Vinf	Infinitival Verb.
Vinf pass	Infinitival Passive Verb.

VIR	Virtual arc (used in ATNs).
VLSI	Very Large Scale Integration.
VME	Bus System for the 68000 series.
Vpass	Passive Verb Group.
VS	Verb-Subject.
WFST	Well Formed Symbol Table.
WP	1: Word Processor (Chapter I; II). 2: Work Package (Chapter VII).
X25	International Standard for Communications between Computers by Packet Switching.