IV. PRINCIPLES AND DEFINITION OF DLT'S INTERMEDIATE LANGUAGE (IL).
===

1. Principles.

In this section, we will focus on the subject of structural or
syntactic ambiguity of IL surface representations. A common
belief and fear of those first confronted with the idea of
using an Esperanto-based language is that such a thing would
be much too 'natural' and therefore too ambiguous to serve as
IL.
In part this criticism is true, and even the creator of Esperanto,
Zamenhof, and other eminent Esperantologists [Waringhien, 1980]
have pointed out that, despite all attempts for regularity and
explicitness, human language is something else than mathematics,
and requires a certain degree of flexibility which tends to
raise ambiguity.
But, to the extent that this criticism is justified, it can
also be turned in favor of a 'natural' IL: how little success-
ful have not been the experiments with purely 'logic' abstract
ILs? Moreover, as we saw in Chapter III, ambiguity is a relative
concept in a translation system.

The choice and design of DLT's IL is strongly inspired by
considerations of contrastive linguistics. As already pointed
out in section III.3.3.3.2, the struggle against ambiguity
should not exceed the demands for translation. There's no
point in trying to resolve ambiguities which are common
to all SL and TL candidates, if not universal, e.g.

(1) "The detective and his profession".

For this 'intersectional' semantics (mainly used in titles)
of the normal conjunction 'and', which has its parallel in
most European languages and Esperanto, one could
theoretically introduce a separate conjunction in the
IL, but this would only complicate, not facilitate the
translation process.

Another example of 'over-disambiguation' would be to
reserve one and only one preposition for time point
references. Actually, it is not attractive to reduce
the existing supply of three such prepositions in
Esperanto, because these largely reflect the practice
of other languages even where it is ambiguous whether
mere time or an event or social habit is referenced (as
in the lower line of Table IV-1).

These examples are meant to remind that the concern
about ambiguity should never surpass the practical
purpose of the exercise: high-quality translation
between natural languages.

Esp.	English	French	German	Russian	used for
je	at	à	um	в	hour
	on	ø	an	в, ø	day, date
en	in	en	in, ø	в	month, year
ĉe	at, on	à	bei	за, на	breakfast, wedding, ...

Table IV-1. Time point reference prepositions
            in several languages (simplified).

To outline the scope of section IV.1, a few points must
be made clear: the topic here is not the process of SL
disambiguation itself. This has been dealt with in Chapter
III, and includes interleaved basic semantics, separate
advanced semantics and interactive dialogues [see also fig.
III-10].

The topic of this section is how to record safely the result
of the relatively expensive disambiguation process, i.e. how to
distinctly represent, within the IL, each potential interpreta-
tion of an SL ambiguity (notice however, that only one interpreta-
tion will actually be recorded as a result of each completed
disambiguation!). In other words, we have to prove that our IL
itself does not show any of the ambiguities which under DLT
will be judged worth resolving during SL disambiguation.
As an introductory example, notice the formal ambiguity of
modifier scope in:

(2) Pregnant women and children.

which once was machine-translated into French (by SYSTRAN
[Pigott, 1982]) as follows:

(3) Des femmes et des enfants enceintes.

The correct interpretation selection (following the
detection of a formal ambiguity during SL analysis) is
a matter of knowledge of the world or context, and
outside the scope of this paper. What we are interested
in here is the alternative representation of each
interpretation in the IL:

(4a) Gravedaj virinoj kaj infanoj.
(4b) Gravedaj virinoj _kaj infanoj.
     (Pregnant women   and children.)

In 4b, an extra space (denoted here by an underscore)
before the conjunction limits the scope of the modifier
to the first noun.

Another example is a well-known sentence of Wilks [Wilks, 1979]:

(5) The soldiers fired at the women and they fell on the
    ground.

where the pronoun 'them' presents a problem.
Regardless of how its antecedent is selected (by AI or
human interaction), the selection made during SL
analysis should be propagated via the IL in such a
manner that the problem will not present itself anew at
a TL synthesis module. Because of the presence of
French among the TL's, the IL should therefore
discriminate between two pronouns, analogous to 'ils'
and 'elles'. This motivates the inclusion of 'iŝi' in
the IL ('iŝi' had been proposed as the 3rd person
plural feminine pronoun decades ago in Esperanto
circles, but never gained ground there).

After two introductory paragraphs [1.1 and 1.2] we will give
[in 1.3] a more systematic overview of the representational
power of our IL, i.e. the capability to distinctly represent
each alternative sentence hidden behind an SL surface string.
The lexical-ambiguity aspects of the IL will be dealt with in
section IV.4.

1.1. IL design criteria.

As in any design work, several criteria must be taken
into account at the same time. For the grammar of our
IL, the following properties of its sentences determine
its usefulness:

    i. Unambiguity.
       This is defined here as: fully automatically, rapidly
       and unambiguously parsable, primarily by syntactic
       processing. Because the boundary between syntax and
       semantics is difficult to draw, a few more words about
       the IL parser requirements must be added to our definition
       of unambiguity:
       'Deep' semantics or knowledge-of-the-world, e.g. far-
       going lexical subcategorization (such as "edible",
       "mental activity" etc.) and macrocontext-based logic
       inference algorithms are definitely excluded from the
       parser. Even lexical subcategories that are quite common
       in natural language processing systems (and sometimes
       referred to as 'shallow' semantics), such as "human",
       will not be relied upon.
       On the other hand, the IL-parser will make extensive use
       of syntactic IL valency information contained in the IL

lexicon (verb valency with regard to specific prepositions,
noun and adjective valency with regard to infinitival
supplements, etc.).

ii. **Compactness.**
This is the average number of bits required to encode
an IL sentence. Each IL morphem, including punctuation,
the space and any special sign one might wish to add,
has its internal computer code: a bit pattern, varying
in length from just a few bits for the most frequent
IL elements (closed class of function words) to 16 bits
for the open class of lexical elements.
Given a set of language elements with their frequency
distribution, a statistically based Huffman coding
technique as chosen for the IL offers maximum compaction
[see also section IV.5].
But as long as a language is still under design, the
introduction of new elements (morphems, special signs)
should be pondered carefully as to their effects on the
overall frequency distribution. This is the case with
the IL.

iii. **Inspectability.**
During long-term development and maintenance of the
(multilingual) DLT system, the IL will be the main
pivot and interface between the various subsystems. The
quality of a French SL analysis module cannot reliably
be judged on the output of a German TL module. The
presence of the intermediate IL stage in between
provides an opportunity to measure or verify the
translation performance of SL- and TL-modules indepen-
dently of each other [see III.5.2].
But only an IL which is easily and quickly accessible
on a daily-routine basis will make its inspection
acceptable and even attractive to linguistic personnel.
This requires an IL which is practically 'readable' as
natural language, and which also can quickly be
'written down' manually as test input to a TL-module.
Whether automatically or manually generated, any IL
sentence will be filtered by the IL recognizer [see III-
4.1], to ensure error-free IL throughout a DLT network.

The first two criteria of the above trio generally tend to
be in conflict with each other. Considering this fundamental
fact, our choice of an error-free natural-language type IL
seems an adequate compromise. Still, we must be careful not
to give away compactness too readily by the addition of
disambiguating elements to the language.

The criteria ii. and iii. therefore limit the range of
methods for obtaining unambiguity to:

a. Rearrangement and strict prescription of WORD ORDER.

b. Use of several ALTERNATIVE function words.

c. Addition of only a small number of PARTICLES and function
   words.

d. A consistent use of punctuation (COMMAs, hyphens, etc.).

e. Insertion of a few special disambiguating elements (to
   indicate otherwise ambiguous word group boundaries,
   modifier scopes etc.), of which the EXTRA SPACE is the
   most important one.

As can be seen from this list, space-consuming and obtrusive
extralingual elements, such as labeled brackets, are forbid-
den. Instead, 'natural' textual elements are used. Apart
from word order prescription, all the above provisions more
or less affect compactness, either by adding morphems or by
changing the statistical distribution of the existing ones.
Regarding provision e, it should be remarked that in the
variable-length (Huffman) code scheme adopted for DLT,
the space (blank), as the most frequent element, consumes only
1 bit. This makes its choice as major 'separator' optimal
from a compactness point-of-view. Whether the same can be
said from the readability point-of-view is subject to
experimentation, but it should be noted that nothing
prevents the mapping (under a simple homomorphism) of these
extra spaces onto other signs (e.g. vertical bar, underscore,
etc.) for ease of inspection.

1.2. IL composition as a modified Esperanto subset.

To state that we use Esperanto as IL would be, in fact,
a loose formulation and gross simplification. What actually
happens is that we make careful use of several 'layers' of
alternative or additional syntactic elements and construc-
tions, on top of the basic layer of common Esperanto [see
fig. IV-1].

This basic layer represents the language as it exists in
word and print, and as it is taught and used in practice.
This current Esperanto largely overlaps with the official
Esperanto, which started with the so-called 'Fundamento'
(a concise grammar kernel, unaltered since its publication
in 1887), and which has intermittently been supplemented by
publications of the Esperanto Academy.
Here we find several basic properties, which already make

```
IV    ┌─────────────────────────────────┐
      │  MODIFICATIONS                  │
      │      FOR                        │
      │      DLT                        │
      │                                 │
      │                                 │
      │   ┌─────────────────────────────┼───────┐
III   │   │ MODIFICATIONS               │       │
      │   │ PROPOSED                    │       │
      │   │ BY OTHERS                   │       │
      │   │                             │       │
      │ ┌─┼─────────────────────────────┼─────┐ │
II    │ │ │ PIV ⎫                       │     │ │
      │ │ │ PAG ⎬ DISAMBIGUATION        │     │ │
      │ │ │     ⎭    RECOMMENDATIONS    │     │ │
  ┌───┼─┼─┼───────────────────────────┐ │     │ │
  │ I │ │ │ PRACTICE                  │ │     │ │
  │ N │ │ │   ┌───────────────────────┼─┼───┐ │ │
I │ C │ │ │   │                       │ │   │ │ │
  │ O │ │ │   │  AKADEMIO             │ │   │ │ │
  │ R │ │ │   │                       │ │   │ │ │
  │ R │ │ │   │                       │ │   │ │ │
  │ E │ │ │   │  FUNDAMENTO           │ │   │ │ │
  │ C │ │ │   └───────────────────────┼─┼───┘ │ │
  │ T │ │ └───────────────────────────┼─┘     │ │
  │   │ └─────────────────────────────┘       │ │
  │   └───────────────────────────────────────┘ │
  └─────────────────────────────────────────────┘
```
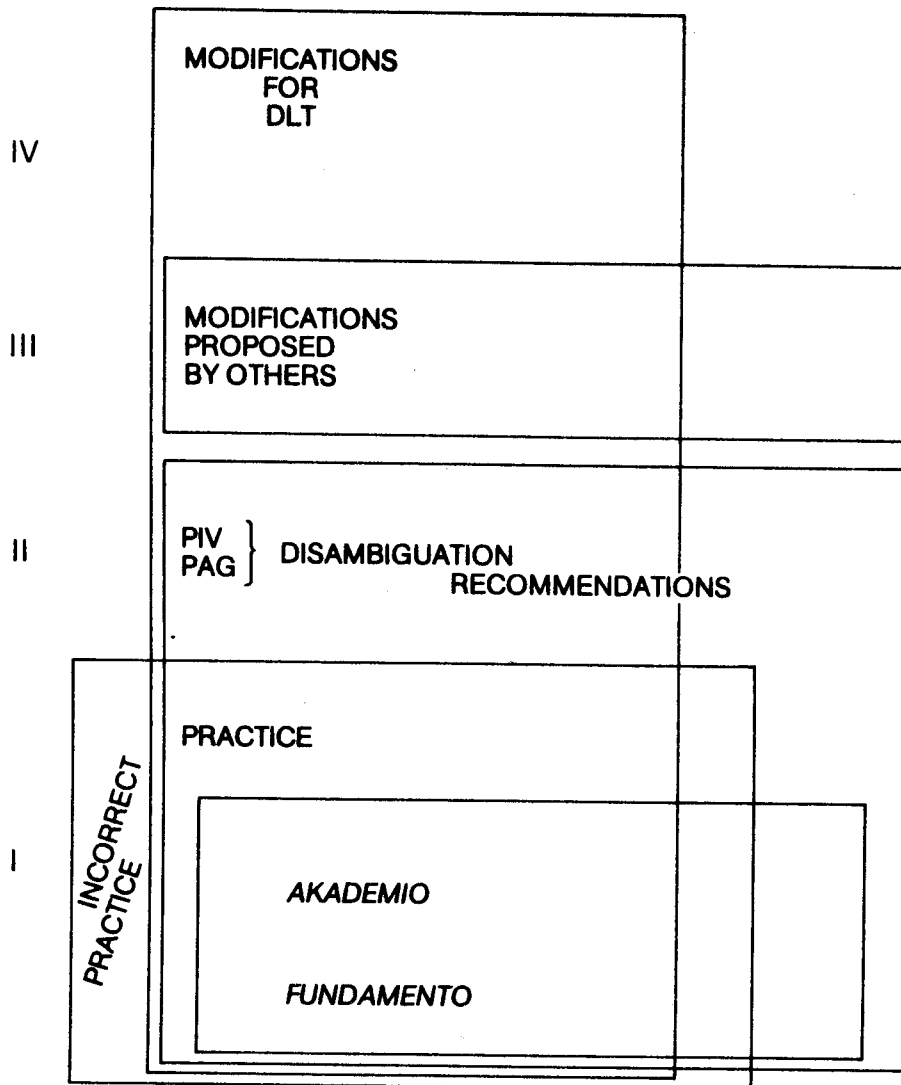
Fig. IV-1. Composition of DLT's IL out of four 'layers' of
           linguistic material. Layer I consists of the common
           Esperanto as it is currently used and supported by
           the Esperanto Academy and the so-called 'Fundamento'.
           In layer II, PIV and PAG are comprehensive standard
           reference works (lexicon and grammar), reflecting 8
           to 9 decades of experience with the language. Layer
           III contains contributions from various authors in
           the area of Esperantology and interlinguistics.
           The partial overlap of the rectangles indicate that
           DLT does NOT adopt everything from the layers below
           it. The relative proportions of the rectangles are
           very cursory.

Esperanto superior to many other languages regarding struc-
tural transparency, e.g.:

- explicit part-of-speech marks, in the form of
  grammatical word endings (-o, -a, -as, -i, -e),
  distinguish nouns, adjectives, finite and infinite
  verbs, as well as derived adverbs;

- the existence of an accusative case (ending -n) helps
  to distinguish the direct object (from the subject,
  but also from an object's predicative complement);
  in connection with locative prepositions, it
  denotes motion towards;

- more precise demarcation of prepositions,
  corresponding to semantic functions: instrumental
  ('per'), sociative ('kun'), tractative ('pri') etc.

The second layer [fig. IV-1] consists of recommendations
made in the two comprehensive and authoritative standard
works of Esperanto, known by the acronyms PAG (the
grammar) and PIV (the dictionary). These contain advice
on how to avoid possible ambiguities by more stringent
or alternative constructions, addition of particles,
etc. Of course, the users of the language do not syste-
matically follow these recommendations (many of which
are not even known to them), because potential ambiguities
often are unproblematic in human communication. For the IL
however, these recommendations provide a treasure of useful
ingredients, e.g.:

- number agreement scheme for the coordination of
  premodifiers and nouns (plural ending is -j):

  (6) Miaj frato   kaj fratino.
      (My   brother and sister.)

  (7) La itala   kaj la franca lingvoj.
      (The italian and the french language.)

  (8)   Juna kaj alloga   virino.
       (A young and attractive woman.)

  which helps in the analysis of premodifier
  scope (6) or the number of entities referenced:
  whereas (7) is an elliptic noun coordination,
  (8) is a descriptive adjective coordination.

- a special preposition ('far') to denote the
  agent of the passive;

- particles to disambiguate between time/space
  semantics of certain prepositions ('antau',
  'post'):

  (9a) Iam antau...
       (Before...   Fr.: Avant...)

  (9b) Ie antau...
       (Before...   Fr.: Devant...)

- other combinations of function words:

  (10a) Krom...
        (Except for...)

  (10b) Krom....ankau...
        (Besides....)

Some of the recommendations for disambiguation cause a
slight difference of stylistic content, which makes
them less attractive as systematic elements of a
translation system. Therefore, layer I and II [fig. IV-1],
are not quite sufficient for our IL design. They form
however the solid base of legitimate and well-tested
language material.

With the third layer, we enter into the realm of mere
proposals and suggestions for alternative syntactic
constructions. For the IL design, we limit our interest
to those alternatives which further contribute to
unambiguity. The difference between this and layer II is
that things found here have not been legitimated. Also,
they have not been used in practice. But in fact, the
boundary between II and III is not easy to draw [Lo Jacomo,
1981]. Some of the not (yet) legitimated proposals
(including much of the small print of PAG itself) come from
prominent authorities, e.g.:

- retaining the accusative (-n) despite the
  presence of a preposition ('de') for the direct
  object of a nominalized verb:

  (11) La manĝado de la anason.
       (The eating of the duck.)

- a synthetic instead of analytic passive:

  (12) Li edukatas.
       (He is [being] educated.)

- extension of the set of 3rd person pronouns with
  an explicitly masculine ('hi','ihi') and a plu-

ral feminine ('iŝi') pronoun (apart from a
contribution to an anti-sexist trend, this helps to
reduce anaphoric ambiguity).

Layer IV consists of additional constructions and
arrangements worked out by the author and his team in
the framework of the DLT project. Some of these have a
far-reaching effect on the overall syntactic look of
the language; at the other end of the scale we have
modifications affecting only one function word. E.g.:

- canonical word group order (basically SVO) for
  independent and dependent clauses, infinitival
  clauses, participial clauses etc.;

- use of extra spaces as disambiguating elements
  throughout the language (in this paper, each such
  extra space will be denoted by an underscore);

- obligatory use of the preposition 'je' in
  combination with the "accusative of measure":

  (13) Li kaptis   je duonan metron longan fiŝon.
       (He caught a   half   meter long   fish.)

  whereas 'je' with nominative will be used for
  time point referencing (table 1) and replace the
  naked "accusative of time point" of common
  Esperanto;

- introduction of new function words:
  the conjunction 'dum-kiel' (whereas),
  the preposition 'kromau' (besides).

In part, the contents of layer IV are restrictive
prescriptions within the framework of layer I and II,
i.e. they consist of forbidding certain elements which
are legalized in the human use of the language.

The four-layered structure can be most appropriately
summarized as "a modified subset of Esperanto" :
despite significant changes and restrictions, it still
is Esperanto in the first place, and will profit from
the vast lexical coverage and productivity of that
language.

In what follows, we will present the subset as one
homogeneous language (which in fact it is), without
bothering from which layer a construction or element
stems.

1.3. The IL's 'ambiguity-resistance' - contrastive overview
        of structural ambiguities.

In this section, we will give a characteristic of the
IL by presenting an overview of typical ambiguities as
they appear in English, German etc. Next to each SL
ambiguity, its two (sometimes even more) IL translations
will be given. These are not as it were paraphrases, but
strict translations of the alternative interpretations of
the ambiguous SL expression. In all these cases, there is no
single IL equivalent for the SL surface string, i.e. the IL
does not share the syntactic ambiguity.

From the theory of formal languages, it is known that the
unambiguity of a grammar cannot be proved.
In order to empirically check for coverage of an ambiguity-free
IL design, we have had to obtain a comprehensive inventory
of syntactic ambiguities. To that end, we have relied heavily
on Agricola [Agricola, 1968] and on a variety of other
sources, partly publications in the MT field, partly general
linguistics literature [Waringhien ', 1959; Weinreich 1966;
Kooij, 1971; Brandt Corstius, 1978; Pitrat, 1980; Melby,
1981; Pigott, 1982b].

For descriptive purposes, we have divided syntactic ambiguity
in 9 groups, which will be shown below [a tenth group, the
ambiguity of coordination and modifer scope, will be discussed
in section 2.4.5]. The groups overlap each other to some
extent.

1.3.1. Part of speech (open-class words).

Apart from the classic and sometimes comical confusions
between noun and verb, adverb and adjective, as are so
typical of English:

(14)   Army demands change.
(14a) Armeo postulas ŝanĝon.
(14b) Armeaj postuloj ŝanĝiĝas.

we also have the frequent ambiguity between an
adjectival participle and a finite verb:

(15)   Britain has raised tariffs.
(15a) Britujo havas altigitajn tarifojn.
(15b) Britujo altigis tarifojn.

(16)   The door is closed.
(16a) La pordo estas fermita.
(16b) La pordo fermajtas.

(17)   His voice is relaxing.
(17a) Hia voĉo estas trankviliganta.
(17aa) Hia voĉo estas trankviligema.
(17b) Hia voĉo trankviliĝas.

By virtue of a rigid system of distinctive grammatical
endings, the IL does not have these problems. This system
includes the endings '-as' (present tense) and '-is' (past
tense) for the active, and '-ajtas' for the passive voice.
Adjectival participles can have the ending '-anta' (active)
or '-ita' (passive). The verb root suffixes '-ig' and '-iĝ'
denote a transitive and intransitive meaning respectively.

Notice that 16b and 17b describe an event, whereas 16a
and 17a are mere qualifications [17aa is a variant formed
with the lexical affix '-em', denoting disposition or
tendency].

1.3.2. Function words.

Among function words, part-of-speech ambiguity also occurs,
e.g. between articles (or demonstratives) and relative
pronouns:

(18)   All endeavour that man can carry out.
(18a) Ĉiuj penoj   'kiujn la homo povas plenumi.
(18b) Ĉiuj penoj   kiujn tiu viro povas plenumi.
(18c) Tiu viro povas plenumi ĉiujn penojn ĵe.

(19)   Das bedeutet Regeln in Betracht ziehen, die
       Konstituenten weglassen, ...
(19a) Tio signifas konsideri regulojn, kiuj
       forlasas konsistaĵojn, ...
(19b) Tio signifas konsideri regulojn, forlasi la
       konsistaĵojn, ...

In the IL, these problems do not occur. This is because
of the absence of reduced relative clauses (18b) and the
canonical word group order (in 18c, the particle 'ĵe'
preserves the emphasis on the object).
Other cases involve the distinction between adverbs
and prepositions, as in:

(20)   Around fifty trees...
(20a) Ĉirkau kvindek arboj...
(20b) Ie-Ĉirkau kvindek arboj...

Whereas in 20a, 'ĉirkau' is an adverb to the numeral,
the prefix 'ie-' in 20b indicates that it serves as a
locative preposition to the noun phrase there.

Under syntactic ambiguity we also cover semantic category
or subcategory differences of function words, to the
degree to which these differences are relevant for
translation between European languages.
This is certainly the case with the meaning of prepositions,
the precision of which will largely determine the quality
of the IL. An important example is the time/space
distinction, for which the IL uses hyphenated prefixes:

(21)   Between the exhibition and the meeting...
(21a) Iam-inter la ekspozicio kaj la kunveno...
(21b) Ie-inter la ekspozicio kaj la kunveno...
(21c) ...inter la ekspozicio kaj la kunveno...

Of these, 21a refers to time, 21b to a spatial and
21c to an abstract relationship (the latter being
verb or noun valency bound). Note that the
time/space distinction may not only be necessary for
the translation of the function word itself (as in
samples 9a and 9b), but also for generating a
stylistically correct word group order in the TL.

Beyond a certain degree however, further differentiation
does not contribute to the translation process. This is
the 'over-disambiguation' against which we have already
warned in the Introduction (example 1 and table 1). In his
work on ambiguity in natural languages [Kooij, 1971],
Kooij points out that in a phrase like

(22)   He missed his train.
       (Hi maltrafis sian trajnon.)

the possessive 'his' expresses "some vague
relation", and "any paraphrase here is more awkward
and less precise than the expression itself".

On the other hand, where an English genitive may
include the meaning of 'agent' (author, publisher,
printer etc.), the IL provides for an explicit
indication of this meaning category via the
preposition 'far':

(23)   Chomsky's new book.
(23a) La nova libro de 'Chomsky'-o.
(23b) La nova libro far 'Chomsky'-o.

just as in:

(24)   Music by Beethoven.
       (Muziko far 'Beethoven'-o.)

[Proper names are inserted in the IL as character

strings, to which a grammatical ending must be
attached: e.g. -on for accusative, -a for adjective
('Chomsky'-a = Chomskian) etc.]

1.3.3. Subject, object, Spc and Opc.

This group of ambiguities includes the subject, the
direct object and their respective predicative
complements, denoted as Spc (Subject predicative
complement) and Opc (Object predicative complement).
Also the indirect object can be involved.

Confusion between subject and object typically
arises where the normal word is reversed, as in
relative (25) or interrogative (26) clauses:

(25)   Die Frau die Trudeau liebte.
(25a) La virino kiu amis 'Trudeau'-on.
(25b) La virino kiun amis 'Trudeau'-o.

(26)   Welche Aufgaben erwarten Sie heute?
(26a) Kiuj taskoj hodiau atendas vin?
(26b) Kiujn taskojn hodiau vi atendas?

It also occurs in connection with participles, as in
the following English examples:

(27)   Visiting relatives can be tiresome.
(27a) Vizitantaj parencoj povas esti enuigaj.
(27b) Viziti parencojn povas esti enuiga.

(28)   A program checking strategy.
(28a) Programo, kontrolanta strategion.
(28b) Programon kontrolanta strategio.
(28c) Program-kontrolanta strategio.

The IL resists the ambiguity by its nominative-
accusative contrast, and by the fact that its
participles can only function as adjectives (27a).
Adjectives may have their own supplements, including
an object, and may be placed either after or before
the noun (28a and 28b). In case of the former, a
comma is obligatory (otherwise an ambiguity would
arise if the head noun itself were in the
accusative). A lexicon-based variant is suggested by
28c.

In the next example, ambiguity between a participle
and a finite verb is mixed with direct-indirect
object ambiguity:

(29)   He showed the plan to his colleagues,
       interested friends and...
(29a) Li montris al siaj kolegoj,
       interesitaj amikoj kaj... la planon.
(29b) Li montris al siaj kolegoj la planon,
       interesis amikojn kaj...

An important class of ambiguities are those involving
predicative complements. In the IL, both the Opc and the
Spc are in the nominative. This makes the Opc clearly
marked off from the object (30a).
However, in wh-questions ('ki'-questions in the IL) about
the Spc or Opc, the complement comes in front of the
subject (31 through 34). Unambiguity of the syntactic
structure is then assured by insertion of a separator (an
extra space, denoted here by the underscore in 31b and
32b), unless the subject has a determiner (33 and 34):

(30)   The witch made the prince a frog.
(30a) La sorĉistino faris la princon rano.
(30b) Por la princo la sorĉistino faris ranon.

(31)   They don't know how good meat tastes.
(31a) Ili ne scias, kiel-do bona viando gustas.
(31b) Ili ne scias, kiel-do bona _viando gustas.

(32)   ..., how long new programmers make their
       programs.
(32a) ..., kiel-do longaj novaj programistoj faras
       siajn programojn.
(32b) ..., kiel-do longaj _novaj programistoj faras
       siajn programojn.

(33)   Kia   viro hi estas?
       (What kind of man he is?   )

(34)   Kiu rango la kolonelo nomumis   sian fraton?
       (What rank   the colonel nominated his brother?)

[The suffix '-do' in dependent questions distinguishes
the interrogative meaning of the 'ki'-words from
their relative or correlative meaning.]

Another type of ambiguity which we assign to this
group is the one between a free adjunct and subject
or object:

(35)   Jede dritte Stunde ist eine Pause.
(35a) Je ĉiu tria horo estas pauzo.
(35b) Ĉiu tria horo estas pauzo.

(36)   He studied the whole year.
(36a) Dum la tuta jaro li studis.
(37b) Li studis la tutan jaron.

In the IL, free adjuncts are put in front of the SVO
backbone and can only be adverbs, PP's or subordinate
clauses.
[35a is a case of subject inversion (word order VS
instead of SV), whereas in 35b the verb is used as a
copulative.]

1.3.4. PP-ambiguity.

Prepositional phrase (PP) ambiguity is a notorious
phenomenon and has caused much trouble to MT system
experimenters [Bruderer, 1977]. To tackle this
unruly thing in a systematic way, we distinguish:

- cases where only the grammatical function of the
  PP at clause level is involved;
- cases where also the subordinate relation of the
  PP with a noun phrase is involved;
- cases where only the subordinate relation of the
  PP with noun phrases is involved.

The grammatical function candidates for an ambiguous PP
are:

- the free adjunct (FADJ), an adverbial constituent
  of time, place, manner, instrument etc.,
  determining the clause as a whole;
- the verb valency bound object (Ovv), an indirect
  object in the form of a PP, the preposition of
  which is governed by the verb and recorded in
  its dictionary entry; in the IL, the Ovv-concept
  is extended to include locative PP's with verbs
  of situation and motion;
- the agent of the passive (Ag); in the IL, the
  preposition 'far' has been exclusively reserved
  for this.

Two dominant IL clause patterns (out of a canonical
set of several dozen) help further explanation:

$$[FADJ]^s \quad [S] \quad V \qquad [Ovv] \quad [O]$$

and: $[FADJ]^s \quad [S] \quad Vpass \quad [Ovv] \quad [Ag]$

(Vpass = passive verb group; square brackets denote
optionality, the superscript the allowed repeat
factor).

This fixed word group order is an important IL
weapon against PP-ambiguity. Of the first type of
cases, (38) shows the confusion between FADJ and
Ovv. Somewhat different is (39), one interpretation
of which involves a discontinuous verb, but
discontinuities do not exist in the IL:

(38)   These islands were named after Ascension Day.
(38a) Post Ĉielirotago tiuj-ĉi insuloj nomajtis.
(38b) Tiuj-ĉi insuloj nomajtis lau Ĉielirotago.

(39)   She called up the stairs.
(39a) Supren ie-lau la ŝtuparo ŝi vokis.
(39b) Ŝi telefonis al la ŝtuparo.

[The locative prefix 'ie-' in 39a prevents confusion with
the other meaning ("according to") of the preposition 'lau'].

Confusion between FADJ and Ag is, apart from distinctive
word group positions, prevented by the exclusive use of
the preposition 'far' for Ag:

(40)   The ice was piled high by the river.
(40a) Apud la rivero la glacio alte amasigajtis.
(40b) La glacio alte amasigajtis far la rivero.
In infinitival or participial clauses, which have a somewhat
different word order than general clauses, 'far' may become
the only distinctive element, e.g.:

(41)   The ice, piled high by the river, began to melt.
(41a) La glacio, alte amasigita apud la rivero,
       ekfluidiĝis.
(41b) La glacio, alte amasigita far la rivero,
       ekfluidiĝis.

[in 'ekfluidiĝis', the prefix 'ek' indicates the
ingressive meaning].

The preposition 'far' also prevents confusion between the
Ovv and the Ag (their word order does not provide certainty:
Ovv and Ag follow each other, but both are optional):

(42)   In 1633 wurde Konstanz von den Schweden befreit.
(42a) Iam-en '1633'-o 'Konstanz'-o liberigajtis de la
       svedoj.
(42b) Iam-en '1633'-o 'Konstanz'-o liberigajtis far la
       svedoj.

The second type of cases concerns the distinction
between a clausal and a phrasal function of the PP.
The former can be a FADJ, an Ovv or an Ag, as shown

above, the latter applies to the PP as a NP supplement:

(43)   He saw the girl with the binoculars.
(43a) Hi vidis la knabinon kun la binokloj.
(43b) Per la binokloj hi vidis la knabinon.

(44)   Sie hat sich die Adresse von einem Freund geben
       lassen.
(44a) Ŝi lasis la adreson de amiko donajti al ŝi.
(44b) Ŝi lasis la adreson donajti al ŝi far amiko.

[notice the use of the infix '-ajt-' to denote a passive
infinitive in 44a and 44b].

Both the canonical word order and the more precise use
of prepositions in the IL contribute to unambiguity of
the PP-function here. In the following example (45),
the phrasal PP (as supplement to 'talk') becomes an Ovv
(45a), as the analytic verbal expression is translated
into a synthetic one:

(45)   He gave a talk on the new 'Mistral' express train.
(45a) Li paroladis pri la nova rapidtrajno 'Mistral'-o.
(45b) En la nova rapidtrajno 'Mistral'-o li paroladis.

Sometimes, the word order is the only distinction, as
in the next example, where 46b contains an Ovv:

(46)   He freed the men from the king of Spain.
(46a) Hi liberigis la virojn de la reĝo de hispanujo.
(46b) Hi liberigis de la reĝo de hispanujo la virojn.

Instead of as an Ovv, the PP may function as an
adjectival supplement in one of the readings (47b):

(47)   ...dieser Fortschritt der Entwicklung des Landes
       nützlich zu machen...
(47a) ...tiun progreson de la evoluado de la lando
       fari utila...
(47b) ...tiun progreson fari utila al la evoluado de la
       lando.

In the third type of cases, different analyses within
the same NP are involved:

(48)   The dog of my neighbour with that long tail.
(48a) La hundo de mia najbaro kun tiu longa vosto.
(48b) La hundo de mia najbaro _kun tiu longa vosto.

In 48a, it is the neighbour, not the dog, that has the
long tail!

Notice that the IL uses morpho-syntactic means (an extra
space, denoted in 48b by the underscore), NOT knowledge-
of-the-world, to differentiate between two such inter-
pretations. This is a general property of the IL, not
confined to PP-ambiguity alone.
In 48b, even replacement of the demonstrative 'tiu' by
the possessive 'ĝia' (a possessive which can only refer
to non-human possessors) would not make the disambiguating
extra space (before 'kun') unnecessary: without knowledge-
of-the-world, one could easily assume that the neighbour
bears his dog's tail on his hat.

More about PP-ambiguity can be found in section 2.4.5.

1.3.5. Infinitives and participial constructions.

Uncertainty whether an infinitive functions as a
syntactical object or as a free adjunct, sometimes
occurring in the SL, does not exist in the IL.
In the first place, IL free adjuncts can only be
adverbials (adverbs or PP's), not naked infinitives.
Secondly, free adjuncts have their own place in the IL
canonical word order, clearly separated from objects.
For example in:

(49)   He loves to live.
(49a) Hi ŝatas vivi.
(49b) Por vivi hi amas.

(50)   He prays to be saved.
(50a) Hi petegas savajti.
(50b) Por savajti hi preĝas.

one of the interpretations (49b, 50b) is a free adjunct
(of goal).

Also note the following cases, in which a participle is
involved:

(51)   She stopped laughing.
(51a) Ŝi Ĉesis ridi.
(51b) Ridante ŝi Ĉesis.

(52)   These people don't love standing in queues.
(52a) Tiuj-Ĉi homoj ne ŝatas stari en vicoj.
(52b) Starante en vicoj tiuj-Ĉi homoj ne amas.

The IL does not use (51a, 52a) participial objects as
in English.

[Notice the effect of the following slight variation of
the SL examples 51 and 52:

(53) She stopped, laughing.
(53a) Ŝi, ridanta, ĉesis.

(54) Standing in queues, these people don't love.
(54a) Tiuj-ĉi homoj, starantaj en vicoj _ĵe, ne amas.

with the particle 'ĵe' (in 54a) keeping the emphasis on
the subject adjunct.]

Neither does the IL use participial subjects:

(55) Flying planes can be dangerous.
(55a) Pove, flugigi aviàdilojn estas dangera.
(55b) Pove, flugantaj aviadiloj estas dangeraj.

Nor does the IL allow ACP ('Accusativus cum Participio')
constructions:

(56)   The president did not consent to any of the documents
        leaving the office.
(56a)   La prezidento ne konsentis, ke iuj ajn da la dokumentoj
        eliru el la oficejo.

[By use of the particle 'da' in 56, 'dokumentoj' remains
the head of the NP, and 'iuj ajn da' ("any of") is
treated as a complex determiner.
The verb ending '-u' in 56a denotes volitional mode,
as opposed to the indicative.]

The ACI ('Accusativus cum Infinitivo') is allowed in the IL,
and may be used to translate an English ACP in certain cases
(when the IL-verb's lexicon entry indicates ACI-valency).

In some cases, insertion of an extra space as a
disambiguating element is needed in order to distinguish
an ACI construction from an infinitive functioning as
noun supplement, e.g. in:

(57a) Mi vidas ŝancon _veni.
       (I see a chance come/coming.)
(57b) Mi vidas ŝancon veni.
       (I see a chance to come.)

where, according to IL dictionary entries, the verb
('vidi') allows an ACI and the noun ('ŝanco') an
infinitival supplement. This, obviously, is not a case
of SL ambiguity but merely prevention of IL ambiguity,
whereas in the next example:

(58) Er hat die Kunst zu lieben vergessen.
(58a) Hi forgesis ami la arton.
(58b) Hi forgesis la arton ami.

the SL ambiguity is resolved in'the IL by word order
and by the absence of an ACI-valency-mark in the
dictionary entry of 'forgesi'.

As other examples in this group, we include SL
infinitival use for expressing obligation or capacity:

(59) Das Problem ist zu übersetzen.
(59a) La problemo estas traduki.
(59b) La problemo devas tradukajti.
(59c) La problemo povas tradukajti.

(60) Sein Zweck ist zu erläutern.
(60a) Lia celo estas klarigi.
(60b) Lia celo devas klarigajti.
(60c) Lia celo povas klarigajti.

For the sake of completeness, we end this group with
a remark on the use of participial constructions:
As one might expect, the IL does not allow "absolute"
adjuncts, such as in French or English:

(61) The contract having been signed, the company's
      president gave a big party.
(61a) Post ke la kontrakto subskribajtis, la prezidento
      de la kompanio tenis grandan feston.
(61b) Pro ke la kontrakto subskribajtis, la prezidento
      de la kompanio tenis grandan feston.

Instead, the IL uses subordinate⁄clauses as free adjuncts,
which (by their subordinate conjunctions) explicitly show
their nature (temporal, causal, etc.). Also, SL constructions
like 62 are not matched by participles in the IL (except for
the phrasal noun postmodifier interpretation of 62b):

(62) The problem was solved by the employee paying the
      bill.
(62a) Per ke la dungito pagis la fakturon, la problemo
      solvajtis.
(62aa) Per la pagado de la fakturon far la dungito, la
      problemo solvajtis.
(62b) La problemo solvajtis far la dungito, paganta la
      fakturon.

1.3.6. Verb nominalization.

Verb nominalization in SL often blurs the distinction between
the thematic agent and patient that might both be associated

with the nominalized verb:

(63)   The accusation of the broker.
(63a) La akuzado de la makleriston.
(63b) La akuzado far la makleristo.

(64)   Das Studium der Frauen.
(64a) La studado de la virinojn.
(64b) La studado far la virinoj.

(65)   The eating of the mouse by the cat.
(65a) La mangado de la muson far la kato.

(66)   The eating of the mouse.
(66a) La mangado de la muson.
(66b) La mangado far la muso.

As these examples show, the distinctive morpho-
syntactic features of the verbal object (the accusative
'-on' ending) and the verb's agent (the preposition
'far') are preserved in the IL under verb
nominalization.
If an NP is headed by a nominalized verb, its structure
still corresponds to that of the underlying verbal clause:

   [DET] [ADJ] -ado [PP] [de -on] [far -o] [PP]

([Adv]   V   [Ovv] [O]   [FADJ]
                                         [Ag]   )

[FADJ is placed out of the canonical word order (for non-
infinitival clauses) here, to show its correspondence
with the NP's trailing PPs; any PP corresponding to
the Ovv (verb valency bound prepositional object) comes
immediately after the head noun.]

1.3.7. Verb semantics.

In this group, we consider SL ambiguities which the IL
solves by a finer differentiation of available verbs in
the lexicon:

(67)   He felt a fool.
(67a) Hi sentis sin frenezulo.
(67b) Hi palpis frenezulon.

(68)   The shooting of the hunters.
(68a) La pafado far la ĉasistoj.
(68b) La mortpafado de la ĉasistojn.

The superiority of the IL in clearness of verb semantics

is partly due to a rather regular system of word formation,
using suffixes with a well-defined meaning (e.g. '-ig'
for causative, '-iĝ' for medial verbs):

(69)   Eine abkühlende Ammoniaklösung.
(69a) Malvarmiĝa amoniaksolvajo.
(69b) Malvarmiĝanta amoniaksolvajo.

(70)   Boiling champagne is interesting.
(70a) Boliĝi ĉampanon estas interesa.
(70b) Bolanta ĉampano estas interesa.

(71)   The boy rolled down the slope.
(71a) La knabo rulis suben ie-lau la taluso sin.
(71b) La knabo ruliĝis suben ie-lau la taluso.

(72)   He cut himself with a knife.
(72a) Per tranĉilo hi tranĉis sin.
(72b) Per tranĉilo hi tranĉiĝis.

Notice the IL differentiation between an intentional,
voluntary act of the subject (corresponding to the
'deep agent': 71a, 72a) and a non-voluntary one
(corresponding to the 'experiencer' role of the
subject: 71b, 72b). Compare also:

(73a) Per   ŝtono'mi rompis la fenestron.
       (With a stone I broke the window.)

(73b)   Ŝtono rompis la fenestron.
       (A stone broke the window.)

(73c) Per   ŝtono la fenestro rompiĝis.
       (With a stone the window   broke.   )

(73d) La fenestro rompajtis far ŝtono.
       (The window   was broken by a stone.)

(73e) La fenestro estis rompita far ŝtono.
       (The window   was   broken-by-a-stone.)

Of these five variants, 73a and 73b show the active
voice (with and without instrumental), 73c the medium
voice, and 73d the passive. In contrast to 73d, which
indicates an act or process, 73e merely describes a
state: it has a copula and a predicative complement,
composed of an adjective and its supplement.

Special attention deserve the semantics of modal verbs.
The IL distinguishes obligation from logical inference,
by using a modal verb ('devi') for the former (74a),
and an adverb ('deve') for the latter (74b):

(74) He must go to Paris now.
(74a) Nun li devas iri ie-al parizo.
(74b) Deve, nun li iras ie-al parizo.

Also, assumption or rumour (75b), with which obligation
could be confused in German surface strings, is
represented distinctly in the IL:

(75) Er soll noch studieren.
(75a) Hi ankorau devas studi.
(75b) Onidire, hi ankorau studas.

The leading adverbs of 74b and 75b are considered
speaker-oriented adjuncts, which always have their
place at the beginning of the IL clause.

Likewise, the IL makes a difference between capacity (75a),
possibility or probability (75b), and permission (75c):

(75) Il peut venir.
(75a) Hi povas veni.
(75b) Pove, hi venos.
(75c) Hi darfas veni.

1.3.8. Compound noun strings.

Whereas compound noun strings are abundant in English
(especially in technical and scientific texts), they
are practically absent in the IL, which only allows
concatenation of a noun and a literal (" prezidento
'Reagan' "). Of course, a great deal of trouble in
English is caused by the lack of morphological
distinction between a word's substantival and adjectival
use. As we saw in Group 1, this confusion does not
exist in the IL. Besides, the IL has several other
differences from English to prevent the latter's
typical ambiguity in this group.

The IL never omits prepositions with indirect objects:

(76) We sent the general reports.
(76a) Ni sendis la generalajn raportojn.
(76b) Ni sendis al la generalo raportojn.

In some cases, the nominative-accusative opposition
provides unambiguous representations in the IL:

(77) Two electrical passenger carrying vehicles are
      missing.
(77a) Du elektrajn pasagerojn transportantaj veturiloj
      mankas.

(77b) Du elektraj pasaĝerojn transportantaj veturiloj
      mankas.

[If the whole NP of 77 were itself in the accusative,
insertion of the dummy article 'le' before 'pasaĝerojn'
would be applied to separate it from 'elektrajn'.]

The problem of 77 could also be solved via the (IL)
lexicon, if this could be made to include the composite
adjective 'pasaĝerojn_transportanta' (note the use of
the underscore, here instead of a space, to indicate
a lexical unit). Compare:

(78)   The criminal lawyer.
(78a) La kriminala juristo.
(78b) La kriminala_juristo.

Apart from lexical units, the IL uses the hyphen to
indicate a stronger tie (relating to an underlying
verb phrase):

(79)   The German students.
(79a) La germanaj studentoj.   -
(79b) La germana-studentoj.

As for partitives, the IL always uses the particle
'da':

(80)   ...dass dieses Fass Bier für die Maurer ist.
(80a) ...ke tiu-ĉi barelo da biero estas por la
      masonistoj.
(80b) ...ke tiu-ĉi barelo estas biero por la
      masonistoj.

Finally, the IL makes strict use of commas and
apostrophes:

(81)   Strebsame junge Künstler.
(81a) Diligentaj junaj artistoj.
(81b) Diligentaj, junaj artistoj.

[81a shows a restrictive ("cumulative"), 81b a descriptive
("parallel") modifier chain.

(82)   The term insurance.
(82a) La periodo-asekuro.
(82b) La termino 'insurance'.
(82c) La termino "asekuro".

[the apostrophed string of 82c is translated, the one
of 82b is not.]

1.3.9. Anaphorics.

Unambiguous pronominalization of nouns referring to
humans is facilitated by separate pronouns for
explicitly masculine or feminine entities, and for
entities (functionaries, professions etc.) of which
the sex is unknown or irrelevant:

(83)   After the judge had heard the man's story about
        his ex-wife, he decided that he did not have to
        pay her alimony.
(83a) Post ke la juĝisto audis la rakonton far la viro
        pri sia eksedzino, li decidis, ke hi ne devu pagi
        al ŝi alimenton.
(83b) pri lia eksedzino, li decidis, ke li ne devu pagi
        al ŝi alimenton.

In sample 83, the antecedents of "his" and of "he" (2nd
occurrence) are ambiguous: its unclear whether the
judge's or the story-teller's ex-wife is involved, and
who of the two is exempt from alimony.
In the interpretation given by 83a, the story-teller's
divorce is dealt with. Notice the use of the possessive
adjective 'sia', which always refers to the nearest
grammatical subject, in this case 'viro' (the
preposition 'far' indicates that the NP headed by
'rakonton' is a nominalized verb phrase). Further,
of the human nouns in the sentence, 'viro' is the only
explicitly masculine one, and is therefore referred to
unambiguously by 'hi'.
In both 83a and 83b, 'juĝisto' is the only sexless
human noun, and 'li' and 'lia' refer to it; 'eksedzino'
is explicitly feminine, and referred to by 'ŝi'.
[The interpretation of the judge's ex-wife in
83b does not make the word 'juĝisto' masculine: this
would imply using knowledge-of-the-world, which is
beyond our objective of making the IL unambiguously
parsable without 'deep' semantics.]

A further pronoun ('ĝi') exists for non-human entities
(institutions, concepts, things, animals).

As for the plural 3rd person, 'ili' refers to "sexless"
or "mixed", 'ihi' to explicit masculine and 'iŝi' to
explicit feminine entities:

(84)   The town councillors refused a permit to the
        women, because they advocated revolution.
(84a) La komunumaj konsilantoj rifuzis al la virinoj
        permeson, ĉar ili advokatis revolucion.
(84b) permeson, ĉar iŝi advokatis revolucion.

Though the differentiation of personal pronouns
contribute to achieving anaphoric unambiguity, it is
not always sufficient. The IL therefore has other means
at its disposal, such as a strictly regulated use of
demonstrative pronouns to refer to preceding nouns:

(85)   Le père emmena l'enfant chez le docteur Dupont,
       parce qu'il avait des boutons.
(85a) La patro alvenigis al la kuracisto 'Dupont' la
       infanon, Ĉar hi havis aknojn.
(85b) infanon, Ĉar tiu havis aknojn.
(85c) infanon, Ĉar tiu-ĉi havis aknojn.

The pronoun 'tiu' refers to the farther, 'tiu-ĉi' to
the nearer of two preceding nouns [notice that 'li'
could not be used in the above sample, because
there are two "sexless" entities to which it can
refer].
The IL anaphoric system also has distinctive pronouns
(demonstrative and relative) for entities on the one
hand ('tiu','kiu') and whole clauses ('tio','kio') on
the other hand:

(86)   His results are connected with a non-standard
       model, which was first shown...
(86a) Liaj rezultoj estas ligitaj kun nenorma modelo,
       .   kiu unue montrajtis...
(86b) kio unue montrajtis...

[in 86b,'kio'refers to the preceding clause as a
whole].

For verbs without a logical subject (statements about
the weather are a classical example), the IL article 'lo'
serves as a grammatical dummy subject. This prevents the
following case of ambiguity:

(87)   Schmiede das Eisen, solange es warm ist.
(87a) Forĝu la feron, dum ke ĝi estas varma.
(87b) Forĝu la feron, dum ke lo estas varma.

1.4. IL syntactic categories.

In accordance with the IL's design principles, its syntactic
categories are the result of a practical compromise between
language-contrastive considerations, adherence to existing
Esperanto and fast parsability. It should be reminded that the
aim of the exercise is language translation (via an Esperanto-
based IL), NOT language investigation.

We distinguish between syntagmatic and syntactic function
categories:

Syntagmatic categories:

GEN   general clause;
INF   infinitival clause;
DEP   dependent clause;
REL   relative clause;

NP   noun phrase;
PP   prepositional phrase;
ADOC noun phrase formed by a nominalized verb
      ('-ado' type of noun, the supplement of which
      shows similarity with a clause complement);
ADJ1 adjectival clause, placed before the head of an NP;
ADJ2 adjectival clause, placed in the tail of an NP,
      or functioning as Spc or Opc [see below];

Syntactic function categories:

S   subject;
O   (direct) object;
Ovv · indirect or prepositional object, according to
      verb valency;

V   finite verb form;
Vinf infinitive verb form;
Vpass passive verb form;

Ag   agent of the passive (always introduced by the
      preposition 'far');
Spc   subject predicative complement;
Opc   object predicative complement;

FADJ free adjunct (place, time, manner, cause, etc.);
LADJ context-linking or speaker-oriented adjunct;
CADJ consecutive adjunct, only occurring at the end
      of a clause.

As this overview shows, the IL syntactic categories, apart
from terminology differences, are roughly a cross-section
of syntactic distinctions made in other languages [e.g. Quirk,
1980; Van Dam, 1972; Grevisse, 1964]. More details can be
found in section 2.1.2 (Canonical word order) and in the
Appendix [p. 7], where also basic word categories are listed
(prior to the ATNs in which they occur). For the use of the
above categories in the labelling of IL tree nodes, the reader
is referred to section 3.3 of this chapter.

2. Descriptive definition.

After the IL design principles have been explained and
illustrated (in section IV.1), a systematic description of the
IL grammar is the aim of this section. The coverage is not
exhaustive: an estimated two third of the IL morphology and
syntax, including a number of important and characteristic
elements, will be dealt with. Lexical aspects will be
discussed in a separate section [IV.4].
The illustrated IL grammar description given below is in
accordance with the formal - and therefore more abstract -
IL definition (by means of ATNs, tree structures etc.)
contained in section IV.3. Whereas the latter will serve
mainly computational linguists and MT experts, the former is
directed at a less computer-oriented readership of linguists
and translators. To suit readers with knowledge of Esperanto,
the IL's deviations from it will be summed up first, in a
seperate subsection [2.1].
On the following pages we will often refer to the IL by its
more specific name BCE (Binary Coded Esperanto). Strictly
spoken, this designation only applies to the internal (bit
string) representation of the IL, and its external (readable)
representation used for human inspection and descriptional
purposes is in fact known as CCE (Character Coded Esperanto)
in DLT. Frequent references will be made to the PAG (Plena
Analiza Gramatiko [Kalocsay, 1980]), the comprehensive
standard work on the grammar of Esperanto (in these references,
the abbreviation PAG will be followed by the number of the
relevant paragraph).

2.1. Deviations from Esperanto.

This subsection lists the differences between BCE and common
Esperanto. Though the general character, structure and
vocabulary of Esperanto are preserved in BCE, significant
limitations and modifications were necessary in order to
arrive at a syntactically unambiguous and fast parsable IL.

2.1.1. Deviations from the 'Fundamento'.

The modifications also affect - to some degree - 7 rules
out of the so-called 'Fundamento', the original 16-rule grammar
proclaimed in 1887 by Zamenhof, the inventor of Esperanto.
This 16-rule kernel is still being considered as the founda-
tion of the language. We will sum up these deviations, using
the rule numbers of the 'Fundamento' (indicated by 'F1', 'F4',
etc.), and illustrate them with BCE examples:

F1. The article.

a. Introduction of the dummy article 'le' (for all
   genders, numbers and cases). This is NOT an indefinite
   article, but merely a disambiguating element to
   indicate the beginning ('left bracket') of a noun
   phrase under certain circumstances, e.g.:

   (88) La flughaveno posedas ok elektrajn le pasaĝerojn
        transportantajn veturilojn.
           (The airport owns eight electrical passenger-
             carrying vehicles.)

   (89) Kiel bona le dana fromaĝo gustas?
        (How good does Danish cheese taste?)

   Without this provision, confusion with the following
   readings would exist [cfr. example 77 in this chapter]:

   (88a) The airport owns eight electrical-passenger
         carrying vehicles.

   (89a) How does good Danish cheese taste?

b. Introduction of a second definite article, 'lo', to be used
   for nominalized adjectives (including superlatives and
   ordinals) not referring to an individual entity , e.g.:

   (90) Lo unua, kion ni povas fari, estas esploro.
        (The first [thing] what we can do is an investigation.)

   (91) Lo plej bona de tiu-ĉi lando estas, ke ĝi ne konas
        krimecon.
        (The best [thing] of this country is, that it does not
        know crime.)

   (92) Lo malfacila ĉe tiu-ĉi afero estas eltrovi kiu-do
        prenis tian decidon.
        (The difficult [thing] in this affair is to find out
        who took such a decision.)

   This should not be confused with certain elliptic
   cases allowed in BCE, where the adjective refers to a
   nearby entity:

   (93) La plej bona el tiuj ŝraubiloj.
        (The best one of those screw-drivers.)

   (94) Tiu-ĉi kalkulilo estas la plej malgranda, kiun mi
        iam vidis.
        (This calculator is the smallest one I ever saw.)

(95) Je la sesa la buso jam alvenis.
     (The bus arrived already at six [o'clock].)

F4. Numerals.

The cardinal 'unu' is inflected, whether used as premodifier or not:

(96) La komisiono ricevis nur unun proponon.
     (The commission received only one proposal.)

(97) Japanio fabrikis unun el la plej rapidaj
     komputeroj.
     (Japan manufactured one of the fastest
     computers.)

F5. Personal pronouns.

The distinctive use of pronouns as to the natural gender (actual sex) of the entities to which they refer, is as follows:

a. For the feminine 3rd person plural, 'iŝi' will be used.

b. Explicitly masculine 3rd person pronouns are introduced 'hi' (singular) and 'ihi' (plural). These must be used (by analogy with the feminine pronouns) when referring to implicitly or explicitly masculine entities ('viro', 'tradukistico', 'aktorico').

c. The pronoun 'li' will be used with humans when their sex does not matter or when it is unknown (e.g. referring to 'persono', 'studento', 'delegito', 'politikisto'); 'ĝi' remains for human institutions, concepts, things and animals; 'ili' serves as plural for both 'ĝi' and 'li'.

F6. The verb.

The passive voice is formed synthetically instead of analytically:

(98a) La leĝprojekto aprobajtas.
      (The bill is approved.   )

(98b) La leĝprojekto aprobajtis.
      (The bill was approved.   )

(98c) La legˆprojekto aprobajtos.
       (The bill will be approved.)

(99) Ŝi igis sian edzon mortigajti.
      (She had her husband killed.)

No compound verb forms remain. The verb 'esti' only
serves as situational or copulative verb:

(100) La auto estas ĉe la garaĝo.
       (The car is at the garage.)

(101) Si estas edukita kaj alloga.
       (She is educated and attractive.)

The preposition 'far' (instead of 'de') always
indicates the agent of the passive:

(102) Si edukajtis far sia onklo.
       (She was educated by her uncle.)

F8. Case government of prepositions.

On the rule that every preposition governs the nominative
case, there are two important exceptions:

a. Accusative following 'de', to designate the deep object of
   a nominalized verb [PAG 135, Rim.]:

   (103) La manĝado de la muson.
          (The eating of the mouse.)

   (104) La evoluigo de sistemon.
          (The development of a system.)

b. Accusative following 'je', to indicate measure (cf. the
   classic 'Accusativus Mensurae'; in this context, not so much
   the accusative, but rather its combination with the preposi-
   tion 'je' is a deviation from the 'Fundamento'):

   (105) Por tiu raketlanĉado, du je naudek metrojn altaj
          muntadoturoj estos necesaj.
          (For that rocket launch, two ninety meters high
          assembly towers will be necessary.)

   (106) Tiu-ĉi saksofono estas je cent dolarojn pli kosta
          ol tiu.
          (This saxophone is hundred dollars more expensive
          than that one.)

With the nominative, 'je' will be used for time point
references of date and hour:

(107) Je la dua de augusto _je la sesa matene la raketo
      lanĉajtos.
      (On the second of August, at six [o'clock] in the
      morning, the rocket will be launched.)

F13. Accusative of direction.

An accusative of direction is only allowed in combination
with a preposition. Instead of:

(108) * Morgau la ĉefministro iros parizon.
         (Tomorrow, the premier will go to Paris.)

we must use:

(108a) Morgau la ĉefministro iros ie-al parizo.

As in common Esperanto, the accusative will only be used
with certain prepositions (those that can express location
as well as direction).

F14. The 'default' preposition.

Instead of 'je' (the new use of which is defined under
F8.b), the particle 'pe' will be reserved as preposition
for those cases where no other preposition appears to be
appropriate. It is virtually intended as a last resource:
for dictionary-defined valency connections with verbs,
nouns and adjectives, other prepositions ('pri', 'por'
etc.) must be employed as much as possible:

(109a) Kredi pri iu.
       (To believe in someone.)

(109b) Aspiro al io.
       (Aspiration for something.)

(109c) Plena de eraroj.
       (Full of errors.)

2.1.2. Canonical word order.

One of the features of common Esperanto is the flexibility
that it allows in the order of the main functional word groups
(subject, object, etc.) within a sentence [Wells, 1978]. In
the IL, this freedom of word group order has been reduced to a
strictly prescribed set of sentence patterns, mainly variations
on the SVO (subject-verb-object) sequence [Table IV-2]. This

Table IV-2. CANONICAL WORD GROUP ORDER IN BCE.

1. The GENeral clause:

1.1	[LADJ,][FADJ] [S]	V	[Ovv] [O]		[CADJ]		active
1.2	[LADJ,][FADJ] [S]	V	Spc		[CADJ]		copulative
1.3	[LADJ,][FADJ] S	V	O	Opc	[CADJ]		copulative, assigning
1.4	[LADJ,][FADJ] [S]	Vpass	[Ovv] [Ag]		[CADJ]		passive
1.5	[LADJ,][FADJ] S	Vpass	Spc	[Ag]	[CADJ]		copulative, passive
1.6	[LADJ,][FADJ] V	S					
1.7	[LADJ,][FADJ] V	Spc		Ske	[CADJ]		subject-verb inversion
1.8	[LADJ,][FADJ] V	[Ovv] [O]		Ske	[CADJ]		
1.9	[LADJ,][FADJ] Vpass	[Ovv] [Ag]		Ske	[CADJ]		
1.10	[LADJ,][FADJ] S	V	[O]	INF	[CADJ]		Accusativus cum Inf.
1.11	[LADJ,][FADJ] S	V	Ovv	INF	[CADJ]		Dativus cum Inf.

2. The INFinitival clause:

2.1	Vinf	[Ovv] [O]	[INF] [FADJ][CADJ]		active
2.2	Vinf	Spc	[FADJ][CADJ]		copulative
2.3	Vinf	O	Opc	[FADJ][CADJ]	copulative, assigning
2.4	Vinf-pass [Ovv] [Ag]		[FADJ][CADJ]		passive
2.5	Vinf-pass Spc	[Ag]	[FADJ][CADJ]		copulative, passive

3. The DEPendent clause:

3.1	ke GEN ,	declarative, volitive
3.2	ĉu GEN ,	interrogative

(continuation of table IV-2)

4. ADJectival and participial clauses:

4a. Placed before the noun (ADJ1):

4a.1	[FADJ] [Oam]		-A	
4a.2	[FADJ] Oav		-A	
4a.3	[FADJ] [O]	[Ovv]	-ANTA	active
4a.4	[FADJ]	Spc	-ANTA	copulative
4a.5	[FADJ] O	Opc	-ANTA	copulative, assigning
4a.6	[FADJ] [Ag]	[Ovv]	-ATA	passive
4a.7	[FADJ] [Ag]	Spc	-ATA	copulative, passive

4b. Placed after the noun (ADJ2):

4b.1	, -A [Oam]	[FADJ][CADJ] ,		
4b.2	, -A Oav	[FADJ][CADJ] ,		
4b.3	, -ANTA [Ovv] [O]	[FADJ][CADJ] ,	active	
4b.4	, -ANTA Spc	[FADJ][CADJ] ,	copulative	
4b.5	, -ANTA O Opc	[FADJ][CADJ] ,	copulative, assigning	
4b.6	, -ATA [Ovv] [Ag]	[FADJ][CADJ] ,	passive	
4b.7	, -ATA Spc [Ag]	[FADJ][CADJ] ,	copulative, passive	
4b.8	, -A INF ,			

Legend [see also IV.1.4]:

-A	adjective, with ending '-a';
-ANTA	active participle functioning as adjective, with ending '-anta', '-inta', or 'onta';
-ATA	passive participle functioning as adjective, with ending '-ata', '-ita', or '-ota';
Oam	prepositional object, being accusative of measure [cfr. IV.2.1.1: F8.b], with preposition 'je';
Oav	prepositional object, according to adjective valency;
Ske	subject in the form of a dependent clause, introduced by 'ke'.

CADJ

CONSECUTIVE ADJUNCTS

SBADJ

SUBJECT/OBJECT-BOUND ADJUNCTS
– PPs

accompaniment, exclusion, replacement,
illustration, comparison, identification...

Ov

VERB-VALENCY BOUND
PPs

– dependent on the
verb's lexicon entry

dative, ablative, tractative,
origin, state, result...
also: complements of
verbs of position and motion

FADJ

FREE ADJUNCTS

– adverbs
– PPs
– subordinate clauses

place, time,
manner, instrumental,
measure, cause, goal...

LADJ

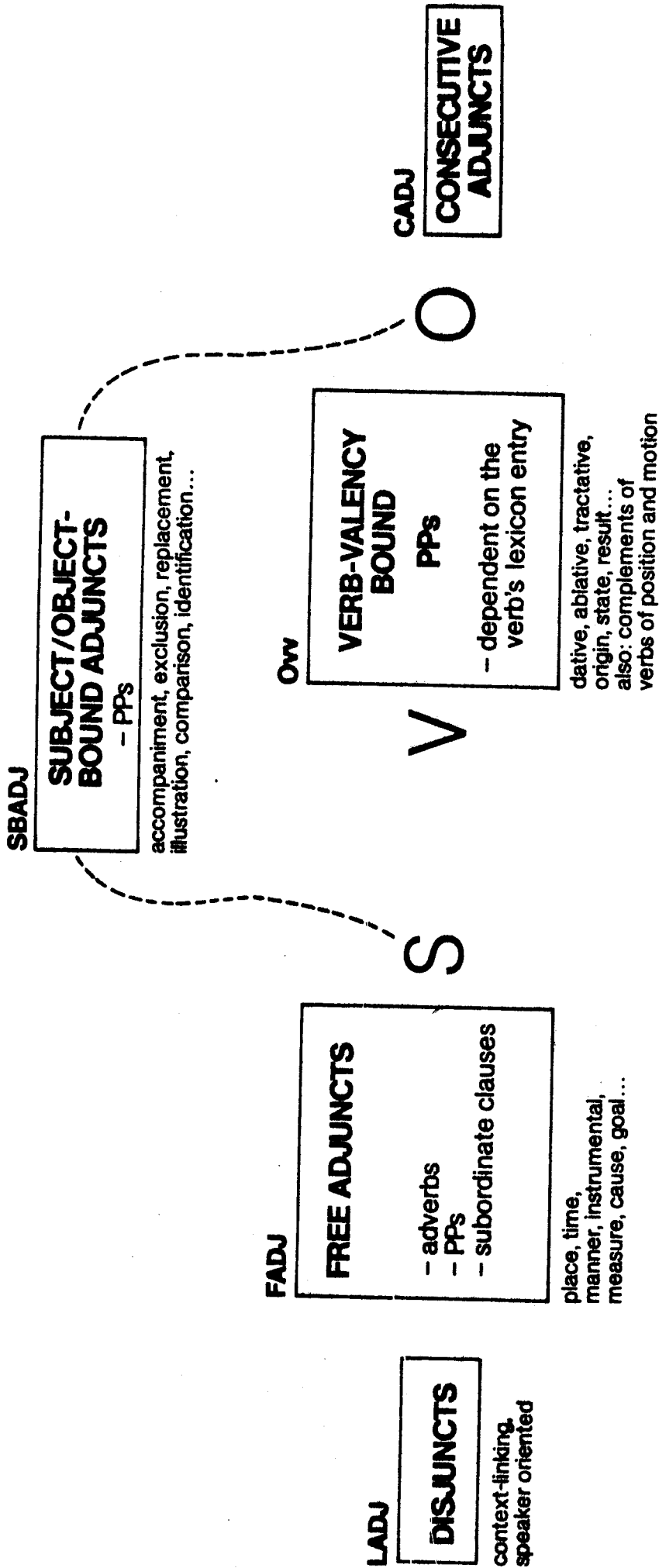**DISJUNCTS**

context-linking,
speaker oriented

Fig. IV-2. Arrangement of functional groups, illustrated for the
SVO pattern 1.1 of Table IV-2. Notice the separation
between PPs functioning as free adjunct and PPs bound
to the verb by valency (the latter are kept close to
the verb). In addition, the subject and object may
contain (postmodifying) PPs themselves.

is one of the factors which makes the IL a subset of common
Esperanto.

Word order regulation is an effective instrument in the struggle
against syntactic ambiguity. Moreover, it is an instrument which
does not affect compaction, as other disambiguation provisions
(addition of prefixes, particles etc.) do. In line with the IL
design principles, the canonical word-group sequences of Table
IV-2 assure structural unambiguity and fast parsability, without
sacrifying readability.
The canonical patterns have deliberately been chosen in such a
way, that they fit the most frequeht actual sentence patterns
and ambiguities (notably PP-ambiguity [see 1.3.4]). In particular,
the resulting IL word-group order largely coincides with the
dominant sentence patterns of French and English, in the text
type anticipated for DLT [see section V.2.1].

Though the prescription of word order [Table IV-2] has an
undeniable 'streamlining' effect on the IL, a rich representa-
tional power is maintained (including subject-verb inversion,
cleft sentences, participial clauses as pre- or postmodifiers
of a noun, etc.).
On the other hand, some concessions had to be made with
respect to natural-language similarity. Typically, the size
and complexity of a functional constituent (e.g. the direct
object) will not influence the IL word order, in contrast to
natural languages (where a sequencing in order of increasing
complexity can be observed [Dik, 1978: 212]). This conceded
lack of flexibility contributes to a somewhat 'mechanical'
appearance of BCE.

Only ordering relations at clause level are listed in this
section, i.e. sequencing patterns for the functional consti-
tuents of a clause. Remarks on phrase-level word-ordering are
contained in the subsequent sections on particular categories
[e.g. section 2.4 on NPs]. All word ordering is defined by and
can be derived from the ATNs [see section 3.2 and Appendix pp.
7-15].
A schematic explanation of the most fundamental sentence pattern
(the first pattern listed in Table IV-2) is given in fig. IV-2.
Subject and object can take the form of an NP, but also of an
INF or DEP clause [see the ATN on p. 8 of the Appendix].
As for the CADJ (consecutive adjunct), we can give the following
examples [see also the small ATN in section IV.3.2; the CADJ
is not covered in a separate section]:

(110)   La politikisto tiel defendis sian proponon, ke li
        forgesis ĝian celon.
        (The politician defended his proposal in such a way,
        that he forgot the purpose of it.)

(111)   La firmao petos tiom da monon, ke nur kelkiuj povos
        aĉeti tiun patenton.
        (The firm will ask so much money, that only a few will
        be able to buy that patent.)

2.1.3. Synopsis of other deviations.

The following is a quick overview of IL deviations from common
Esperanto, other than deviations from the 'Fundamento' or
word-order differences. Most of the items listed here will be
more fully treated in subsequent sections.

   - Adjective instead of adverbial ending in case of an Spc
     (adverbial predicatives as in common Esperanto [PAG 169,
     Rim. I] do not occur):

     (112)   Viziti parencojn estas enuiga.
             (Visiting relatives is tiresome.)

     (113)   Estas stranga, ke ...
             (It is strange that ...)

     (114)   Lo estas varma.
             (It is warm.)

   - Accusative of time [PAG 193] is not allowed: prepositions
     must be used to express time point references or duration:

     (115)   Je la unua de februaro...
             (On the first of February ...)

     (116)   Dum la tuta nokto ni studis.
             (We studied the whole night.)

   - The particle 'da' is not regarded as a preposition, but
     as a special determiner [see 2.4.3], after which an
     accusative as well as a nominative can follow:

     (117)   Malgranda nombro da homoj havas multe da monon.
             (A small number of people have a lot of money.)

   - Introduction of several new function words: 'aunome'
     [see 2.4.4], 'kau' [see 2.4.5], 'kromau' and 'dum-kiel'
     [see 1.2, p. IV-9].

- Prefixing of certain prepositions with 'ie-' and 'iam-',
  corresponding to whether they relate to PLACE or TIME,
  e.g.:

          ie-antau
          iam-antau

  (in contrast to examples 9a and 9b at the beginning of
  this chapter, BCE requires the particles to be connected
  by hyphens, to distinguish their mere disambiguating
  function from any incidental stylistic use).

- Distinction between the relative and the interrogative
  pronoun by means of the suffix '-do' [cfr. PAG 230, Rim.
  II]:

  (118a) La demando kiu venos ...
          (The question that will come ...).

  (118b) La demando kiu-do venos ...
          (The question who will come ...).

  This regulation removes a structural ambiguity, typical
  for common Esperanto (and connected with the fact that
  'kiu', unlike the English 'who', can refer to inanimate
  objects as well as to persons).

- A more general form of subordinating conjunctions (based
  on more systematic derivation from prepositions):

        BCE: Common Esperanto:

      dum ke'   dum
      ĝis ke   ĝis
      post ke   post kiam
      antau ke   antau ol

- Use of adverbs instead of (modal) verbs, to express certain
  modalities [see also 2.2.4 and example 74]:

          'deve'   (LOGICAL INFERENCE)
          'pove'   (PROBABILITY)

- Introduction of the particle 'ĵe', to indicate emphasis
  (special or contrastive focus) on the preceding constituent
  [see also example 54]:

  (119) Morgau ĵe la delegitaro alvenos.
         (The delegation will arrive tomorrow.)

2.2. Verbal constructions.

2.2.1. Tense and aspect.

Principally, only 3 simple tenses are used:

    PAST   '-is'
    PRESENT   '-as'
    FUTURE   '-os'

Compound tenses are not included in BCE, at least not in its
very first development stage. They could be added later, after
experience with the pilot project [Chapter VII] has been built
up. In that case, any compound tenses in BCE will be realized
synthetically instead of analytically, e.g.:

(120)   Antau ke ili komencos, ni skribintos la raporton.
        (Before they will begin, we will have written the report.)

in which the BCE-form 'skribintos' replaces the common Esperanto
construction 'estos skribinta'. Similar replacements would be:

    mi laborintas   (= mi estas laborinta),
    mi laborantis   (= mi estis laboranta),
    mi laborintis   (= mi estis laborinta).

For expressing 'verbal aspect' (e.g. the English progressive
vs. non-progressive, the Slavonic perfective vs. imperfective,
etc.), BCE does NOT have any formal variations of its verbal
forms. On the syntactic level, adverbs must be used for this
purpose (a new adverb 'ante' has been considered to convey the
progressive aspect). Lexical units may profit from prefixes
such as the ingressive 'ek-':

    ekbruli   (to catch fire)
    ekfluidiĝi   ([to begin] to melt)
    eksidi   (to sit down)
    ekmoviĝi   (to begin to move)

but the same prefixes may not be used for syntactic purposes:

(121)   Hi komencis legi.
        (He began to read.)

The above limitations result in BCE-to-English divergent mapping
patterns as illustrated below. During the DLT translation
process, the selection of the proper TL form will be guided by
the presence of certain non-verbal elements in the IL sentence,
such as preceding subordinating conjunctions (e.g. 'post ke') and
adverbs ('jam antaue', 'tuj', etc.) [see also section III.3.3.3.2
and fig. III-21].

```
mi parolis      I spoke
                I was speaking              [mi parolantis]
                I have spoken              [mi parolintas]
                I have been speaking       [mi parolintas]
                I had spoken               [mi parolintis]
                I had been speaking        [mi parolintis]


mi parolas      I speak
                I am speaking              [mi parolantas]


mi parolos      I will speak
                I will be speaking         [mi parolantos]
                I will have spoken         [mi parolintos]
                I will have been speaking [mi parolintos]
```

[The forms between square brackets denote theoretically
possible BCE-extensions, derived from analytical forms in
existing Esperanto].

2.2.2. Voice (active/passive).

The active voice includes the reflexive:

(122a) Mi rulis suben ie-lau la monteto la barelon.
        ·(I rolled the barrel down the hill.   )

(122b) Mi rulis suben ie-lau la monteto min.
        (I rolled myself down the hill.   )

(123)   Mi eraris.
        (Ich irrte mich.)

(124)   Mi ne memoras tion.
        (Je ne me rappelle pas ça.)

The same applies to the 'middle voice' (French: 'construction
moyenne'), which we can regard as a lexical variant or derivative
of the verb:

(125)   La projekto glate evoluas.
        (The project develops smoothly).

(126)   Mi ruliĝis suben ie-lau la monteto.
        (I rolled down the hill.   )

(127)   Tiuj-ĉi pomoj tre bone vendiĝas.
        (These apples sell very well.   )

The passive is formed synthetically instead of analytically.
To avoid the on-going Esperanto discussion on how the verbal

qualities 'dynamic' vs. 'stative' [Quirk, 1980: 46-47] affect
the passive, we decided on the choice of a new infix '-ajt-'
instead of the possible use of '-at-' or '-it-'.

An example of a dynamic verb:

    mi vokajtis   (= mi estis vokata)
    mi vokajtas   (= mi estas vokata)
    mi vokajtos   (= mi estos vokata)

A stative verb:

    mi amajtis   (= mi estis amata)
    mi amajtas   (= mi estas amata)
    mi amajtos   (= mi estos amata)

As in the active, only the 3 simple tenses (and no verbal
aspects) will in principle be realized in the first stage of
BCE development, resulting in a mapping pattern such as:

    mi vokajtis   I was called
                  I was being called
                  I had been called   [mi vokitis]

[the form in square brackets indicates a theoretically possible
possible BCE-extension].

In this way, the syntax of a verbal act or process can be
clearly separated from copulative clauses in which a
participle functions as predicative adjective to describe
a state:

(128a)   ACT: La konkurso fermajtas.
               (The competition is [being] closed.)

(128b) STATE: La konkurso estas fermita.
               (The competition is closed.)

Notice that the introduction of a synthetic passive also
removes the general irregularity in 'number agreement' that
exists in common Esperanto between a verbal construction and its
subject (the passive voice does, but the active voice does not
have number agreement there). In BCE, neither of the two
voices has subject-verb number agreement.

The agent of the passive is always formed by the preposition
'far' (instead of 'de'):

(129) La barelo rulajtis suben ie-lau la monteto far mi.
      (The barrel was rolled down the hill by me.   )

The place of the agent at the back of the sentence is in

accordance with BCE canonical word order [see Table IV-2 and
Appendix p. 9].

Also the infinitive for the passive is formed synthetically,
and (unlike in English, German etc.) is strictly applied in
BCE [cf. also example 99]:

(130)   Hi lasis la permeson donajti al hi far konato.
        (Er hat sich die Genehmigung von einem Bekannten
        geben lassen.)

(131)   Hi lasis la soldatojn mortigajti.
        (He let the soldiers be killed.   )

2.2.3. Auxiliaries.

Because the passive voice and any (possibly later required)
compound tenses will be formed synthetically, the verb 'esti'
loses its classic role of auxiliary verb.

In BCE, the following uses of 'esti' remain:

SITUATIONAL:

(132)   La kongreso estas ie-en parizo.
        (The congress is in Paris.   )

(133)   Mi vidis la katon esti sur la tegmento.
        (I saw the cat [be] on the roof.   )

EXISTENTIAL:

(134)   Mi pensas, do mi estas.
        (I think, therefore I am.)

COPULATIVE:

(135)   Tio estas necesa.
        (That is necessary.)

(136)   La senlaboreco estas grava problemo.
        (The unemployment is a big problem.)

(137)   La celo de la ekonomio estas kontentigi la
        bezonojn de la homoj.
        (The economy's aim is to satisfy the needs of
        the people.)

Use of 'esti' as in example 133 [see also 2.3.2] can be
regarded as a new auxiliary function. Apart from this, BCE is

marked by the absence of 'primary' auxiliary verbs (comparable to the English 'do' and 'have').

2.2.4. Modals and mode.

Esperanto has three <u>basic</u> <u>modals</u>:

    'devi'   (= must)
    'povi'   (= can, may)
    'voli'   (= want)

Usage of the first two of these verbs is strictly regulated in BCE, according to the following scheme:

OBLIGATION (internal or external, moral or material):

(138) Mi devas iri ie-al vieno.
       (I must go to Vienna.   )

(139) Tiu-ĉi afero devas klarigajti.
       (This affair must be clarified.)

LOGICAL INFERENCE:

Instead of a modal verb, the adverb 'deve' must be used (as a speaker-oriented or context-linking adjunct, placed at clause beginning):

(140) Deve, nun hi estas ie-en parizo.
       (He must be in Paris [by] now.   )

CAPACITY:

The modal verb 'povi' keeps its function here:

(141) Bedaurinde, hi ne povas veni.
       (Unfortunately, he cannot come.)

(142) Libroj povas legajti ie-en la librejo.
       (Books can be read in the library.   )

POSSIBILITY, PROBABILITY:

Instead of a modal verb, the adverb 'pove' must be used:

(143) Pove, morgau hi venos.
       (He may come tomorrow.)

PERMISSION:

A different verb ('darfi', 'rajti') must be used for this.

Notice that BCE does not use Esperanto forms such as:

(144)   Tiu-ĉi afero estas klarigenda.

(145)   Libroj estas legeblaj ie-en la librejo.

at least not as verbal auxiliary constructions: 139 and 142
must be used instead. If forms like 144 and 145 appear in BCE,
they are understood as copulative clauses with a special kind
of predicative adjective.

As for mode (subjunctive, imperative), the verbal ending '-us'
will be reserved for irreality and conditionality, as opposed
to the reality indicated by '-as', '-is' and '-os':

(146a) Ili difinos la kvaliton de la produkto tuj kiam estos
         ebla.
         (They will define the quality of the product as soon
         as possible.)

(146b) Ili difinus la kvaliton de la produkto se tio estus
         ebla.
         (They would define the quality of the product if it
         were possible.)

The verbal ending '-u' is used for commands, either directly
(with or without verbal complements):

(147)   Atendu! Inkrementu! Montru la rezulton ...
         (Wait!   Increment! Show the result ...   )

or indirectly, within a dependent clause, when the main clause
expresses some form of command:

(148a) La politikisto diris, ke la registaro diskutas pri tiu
         afero.
         (The politician said that the government discussed that
         matter.)

(148b) La politikisto diris, ke la registaro diskutu pri tiu
         afero.
         (The politician said that the government should discuss
         this matter.)

2.3. Infinitival constructions.

Infinitival constructions form a important element in the
syntax of languages, and therefore deserve special attention
in BCE. This section of the IL description is concerned
primarily with the relation between an infinitival clause
(INF) and its syntactic surroundings.
The internal structure of the infinitival clause is shown in
Table IV-2 (canonical patterns 2.1 through 2.5) and in the ATN
on p. 8 of the Appendix. It can range from a sole infinitive
to a long clause with complex objects and adjuncts (and even
nested infinitival clauses again).

2.3.1. Infinitival clauses.

First of all, infinitival clauses can be subordinate to a
noun, i.e. be embedded within an NP (cf. 57b, 58b):

(149a) La neceso investi.
        (The need to invest.)

(149b) La neceso investi tri milionojn por novaj maŝinoj.
        (The need to invest three million for new machines.)

Secondly, adjectives - when placed after the noun or when
functioning as a predicative complement - can have an
infinitival clause as a supplement [cf. pattern 4b.8 of
Table IV-2]:

(150)   'John'-o estas avida plaĉi.
        (John is eager to please.)

(151)   La robotoj estas pretaj veldi la tubojn.
        (The robots are ready to weld the tubes.)

[The type of pattern illustrated in 150 should not be confused
with the one shown in examples 175-177.]

In all these cases, the valency for an infinitival supplement
must be indicated in the lexicon entry of the IL noun or
adjective. Clearly, only a limited number of words will have
this valency ('preta', 'kapabla', 'libera', 'ema', 'inklina',
'feliĉa', 'avida', ...). This subset will largely coincide
with corresponding subsets in English, French, German etc.,
which demonstrates the pivotal elegance of the IL.

Instead of being embedded within an NP, infinitival clauses
can also replace an NP:

(152)   Investi estas postvivi.
        (To invest is to survive.)

(153)   Pove, ĉe reumatismo movi la brakojn estas dolora.
        (With rheumatism, moving the arms can be painful.)

(154)   Per senlace labori la esploro povis finajti iam-antau junio.
        (By working hard, it was possible to end the investigation
        before June).

As these examples illustrate (and as the ATNs in this report
confirm), the NP replaced by an infinitive can be the subject
(152, 153), its predicative complement (152), or an NP enclosed
within a free adjunct (154). More examples (49-52 and 58-60)
can be found in section 1.3.5.
An infinitival clause can also occur in a verb-valency preposi-
tional object (Ovv), or as a direct object (O) itself:

(155)   La direktoro kredas pri investi.
        (The director believes in investing.)

(156)   Bona politiko konsistas el antauvidi kaj preni decidojn.
        (A good policy consists of looking ahead and taking
        decisions.)

(157)   Bone investi postulas antauvidi.   /
        (To invest well requires looking ahead.)

(158)   Mia frato instruas programi.
        (My brother teaches programming.)

Care should be taken not to confuse the latter two examples
with the sentential infinitive constructions treated below
[2.3.2]. Here, the infinitives do not have a subject, neither
explicit, nor implicit. Their subject is NOT the subject of
the finite verb, and they can be replaced by nominalized
verbs (nouns with '-ado' ending) practically without change
of meaning:

(158a) Bone investi postulas antauvidadon.
        (To invest well requires looking ahead.)

(159a) Mia frato instruas programadon.
        (My brother teaches programming.)

However, infinitives remain allowed in this position because
not all of them can be replaced by '-ado' nouns without intro-
ducing ambiguity (e.g. 'movado' = 1. the moving, 2. the [poli-
tical] movement).
These infinitives, which replace the direct object and which
have no assumed or named subject at all, will be referred to
as NSI (Neutral Subject Infinitive), to distinguish them from
the infinitives treated in 2.3.2.   An IL verb's dictionary
entry will include a separate valency mark indicating whether
or not the verb can have an NSI. For some verbs, this valency

will be conditional and will depend on the occurrence of an
Ovv, e.g. compare 158 with:

(158b) Mia frato instruas al hi programi.
       (My brother teaches him to program.)

Whereas 158 is an NSI, 158b is a sentential infinitival
construction (with 'hi' as implied subject of 'programi'),
with which we enter the next subsection.

2.3.2. Sentential infinitival constructions.

Though consisting of various types, the main characteristic
of this category is: the assumèd subject of the infinitival
clause is at the same time either the subject (S) or the object
(O or Ovv) of the main clause. This particular syntactic
connection between a main and a subordinate clause explains
the term 'sentential construction'.

In one of the 3 types shown below, the infinitival clause is
apparently an additional constituent of the main clause (type 2:
S V O INF). In the other types it seems to replace the direct
object, but one could also regard these construction types as
allowing an object (O) or infinitival clause (INF) as alternative
complements. Anyhow, the above-stated main characteristic
remains the criterion for inclusion in this category (instead
of in 2.3.1).

Type 1:   V INF.

In this type, no direct or indirect object is enclosed
between the finite verb (V) and the infinitival clause (INF).
Throughout this type, the following rule applies:

                Subject INF = Subject V.

First of all, we have the three **basic modals** ('devi', 'povi',
'voli'), which can occur as the finite verb. In BCE, there is no
such thing as a compound verbal constituent. In a sentence like:

(159) Mi devas iri ie-al parizo morgau.
       (I must go to Paris tomorrow.   )

the IL constituent structure is not:

(159a) * (mi) (devas iri) (ie-al parizo) (morgau)
        [S]   [V]   [Ovv]   [FADJ]

but rather (in accordance with canonical patterns 1.10 and 2.1
of Table IV-2):

(159b) (mi) (devas) ((iri) (ie-al parizo) (morgau))
        [S]   [V]   [Vinf]   [Ovv]   [FADJ]

This derivation, based on the concept of fully equipped
infinitival clauses, allows DLT some choice in the placement
of free adjuncts (FADJ): these can either be put at the front
of the sentence (as main clause constituents) or at the end
(as infinitival clause constituents). It also causes a
pleasant variation in the otherwise monotonous sentence
patterns of BCE.
Further, the way in which the basic modals are treated here
does not create any (confusing) opposition with other finite
verbs that can be used with an infinitival complement:

            hi volas   iri morgau,
            hi deziras iri morgau,
            hi esperas iri morgau,
            hi preferas iri morgau.

But in contrast to these other verbs, the three basic modals
cannot have nominal objects in BCE.

Secondly, the finite verb can belong to the following mixed
group of verbs which can either have a nominal object or an
infinitival complement [PAG 181.2] (it is hard to demarcate
this group semantically):

        ami   preferi   kutimi
        bezoni   rifuzi   lerni
        celi   riski   pensi
        decidi   sukcesi   projekti
        deziri   fini   profi
        emi   forgesi   revi
        esperi   honti   scii
        eviti   imagi   ŝati
        intenci   konsenti   timi
        komenci   kuraĝi

(160) La komisiono decidos publikigi la raporton.
       (The commission will decide to publish the report.)

(161) La ministro sukcesis konvinki la parlamenton.
       (The minister succeeded in convincing parliament.)

Thirdly, the finite verb can belong to a small group of
intransitive verbs [PAG 182.1]:

        heziti   ĝoji   peni   klopodi
        ĉesi   ĵuri   rapidi

(162) La piloto klopodis surterigi sian aviadilon.
       (The pilot attempted to land his aircraft.   )

(163) Hi rapidis trafi sian trajnon.
       (He hurried to catch his train.)

Type 2:   V O INF.

This type is marked by the presence of a direct object (O),
which stands in between the finite verb (V) and the infinitival
clause (INF). The rule which applies to this type is:

                 Subject INF = Object V.

Again, different subtypes can be distinguished, the first of
which consists of the verbs 'igi' and 'lasi':

(164) Ili lasis hin telefoni al sia oficejo.
       (They let him phone his office.   )

(165) Hi igis siajn porkojn manĝi.
       (He made his   pigs   eat.)

Note that a construction as 165 is obligatory if the factitive is
syntactical, i.e. not lexicon-based. Only lexical factitives may
be expressed by the suffix '-ig' in a compound form [cf. a
similar remark in section 2.2.1]:

(165a) Hi manĝigis siajn porkojn.
       (He fed   his   pigs.)

(166) Tiu procezo purigas la akvon.
       (That process purifies the water.)

Another subtype is formed by the verbs of perception [PAG 172]:

                 vidi   rigardi   senti
                 audi   auskulti

The infinitival construction with these verbs:

(167) Ni audis hin veni.
       (We heard him come.)

is an example of the classic 'Accusativus cum Infinitivo'
(ACI). An interesting detail of BCE is the use of 'esti' (in
the situational sense) with verbs of perception. This enables
distinctive representation of an ACI-related 'snapshot'
locative PP (133, 168a), a phrasal PP serving as the object's
stative epithet (168b) and a PP functioning as free adjunct
in the main clause (168c):

(168a) Ŝi audis la roboton esti ie-en la kuirejo.
       (She heard the robot [be] in the kitchen.)

(168b) Ŝi audis la roboton ie-en la kuirejo.
       (She heard the kitchen robot.   )

(168c) Ie-en la kuirejo ŝi audis la roboton.
       (In the kitchen, she heard the robot.)

Parallel to 168a (in agreement with pattern 2.1 of Table IV-2)
we can also have:

(169) Ŝi vidis ĝin lavi pladojn zorge.
       (She saw it wash dishes carefully.)

in which the adverb (free adjunct) at the end is part of the
infinitival, NOT of the main clause.

Verbs of desire and demand ('deziri','postuli','atendi' etc.
and also the basic modal 'voli') combine with 'esti' in the same
way as the verbs of perception:

(170) Mi volas raporton esti sur mia skribotablo morgau.
       (I want a report on my desk tomorrow.   )

Type 3:   V Ovv INF.

Here, a verb-valency prepositional object (Ovv) stands between
the finite verb (V) and the infinitival clause (INF). The Ovv
must be formed with the preposition 'al' (finite verbs with
other prepositional object valencies cannot be used here). For
this reason, we call this type of construction DCI ('Dativus
cum Infinitivo'). We distinguish two subtypes, depending on
the group to which the finite verb belongs:

Subtype 3a: Subject INF = Subject V.

This concerns only a small group of verbs ('promesi', 'garanti',
'komplezi', 'minaci'). An example is:

(171) Mi promesis al ŝi veni.
       (I promised here to come.)

Subtype 3b: Subject INF = Object V.

This concerns a much larger group, including the verbs of
command and communication:

komandi	konsili	diri
ordoni	admoni	indiki
devigi	persvadi	skribi
komisii	instrui	telefoni
trudi	helpi	teleksi
permesi	malhelpi	respondi
cedi	peti	

(172) La sergento ordonis al la soldatoj pafi.
        (The sergeant ordered the soldiers to fire.)

(173) Ni helpis al ili fini siajn programojn.
        (We helped them to finish their programs.)

For all the verbs listed under type 3, the DCI (i.e. the use
of the preposition 'al') is obligatory in BCE, also for those
verbs that have the mark 'alternaj objektoj' [PAG 179.B] in
common Esperanto. So we must use:

(174) Instrui al iu naĝi.
        (To teach somebody to swim.)

(175) Persvadi al iu aĉeti ion.
        (To persuade somebody to buy something.)

Also note that, instead of an infinitival complement, these
verbs may have either a direct object (O) or a second indirect
object (Ovv), e.g.:

(174a) Instrui al iu ion.
        (To teach somebody something.)

(175a) Instrui al iu pri io.
        (To teach somebody about something.)

But the following is forbidden in BCE:

(174b) * Instrui iun pri io.
         (To teach somebody about something.)

i.e. the thematic recipient or beneficiary of the action must
be marked by the preposition 'al', whether an infinitival
construction follows or not. This regulation is a typical
example of how IL-ambiguity is prevented without the need for
a 'deep semantics' IL-parser: the latter will be able to tell
the recipient from the topic (of teaching) by means of simple
morphological distinctions.

At the end of this section on sentential infinitival construc-
tions, let us briefly show a few specimens of another interesting

pattern:

(175)   Plaĉi al 'John'-o estas facila.
        (John is easy to please.   )

(176)   Solvi tiujn-ĉi problemojn estas malfacila.
        (These problems are hard to solve.   )

(177)   Estas verŝajna, ke hi venos.
        (He is likely to come.   )

These examples demonstrate the morphological explicitness
with which a direct or indirect object is distinguished
from a subject (cf. example 150; also compare 176 with 59-
60).

2.4. Noun phrases (NPs).

2.4.1. General structure.

The Noun Phrase (NP) takes a dominant place in the IL's
syntagmatic structure. Its frequency will be high, due to
the 'informative' text type to be processed, which is
characterized by the intensive use of nominalizations [see
Chapter V].
The size of an NP can range from a two-letter function word
to several lines of text, including dependent clauses and
embedded NPs again (BCE's nesting depth limitations are
more or less in accordance with human cognitive bounds).
The noun (N) which usually heads an NP will be referred to
here as 'head noun'. Its place can be taken by a pronoun
or a substantivized adjective. In contrast to English, BCE
does not allow compound noun strings, with an exception for
proper names or literals, e.g. (see also examples 82b and 82c):

(178)   La ĉefministro 'Thatcher' renkontos la prezidenton 'Reagan'.
            (Premier Thatcher will meet president Reagan.   )

[If a sole literal acts as a head noun, it receives a gramma-
tical ending, as in examples 23-25].
The head noun (N) can be preceded by one or more determiners
(DET), a numeral (NUM) and adjectives (ADJ), collectively
referred to as premodifiers:

    [DET] [NUM] [ADJ] N

(179)   Iliaj tricent novaj misiloj.
            (Their three hundred new missiles.)

[notice the plural agreement between noun, adjective and
determiner].

After the head noun, quite a variety of constituents can occur
as part of the NP. In the first place, we can have: one or more
prepositional phrases (PPs), appositions (APPOS) and again
adjectives (ADJ), which can all be referred to as postmodifiers:

     N [PP] [APPOS] [ADJ]

Notice that PPs cannot occur in front of the head noun in BCE.
The apposition (APPOS) includes forms with the 'equivalence'
particles 'aunome', 'nome', 'alinome' and 't.e.' [PAG 131]
(the IL-form 'aunome' replaces the equivalence sense of the
common-Esperanto 'au'):

(180) Oni signas la azoton aunome nitrogenon per 'N'.
       (One indicates azote or nitrogen by 'N'.   )

[this implies that 'au' in BCE never connects two different
names for the same entity].
Adjectives (ADJs) can be simply one-word, but they can also
consist of an adjectival clause (with embedded NPs). To
distinguish between premodifying and postmodifying ADJs
(which obey different grammar rules in BCE), we will denote
the former as ADJ1 and the latter as ADJ2 [as in sections 1.4,
2.1.2, and the ATNs on pp.12-13 of the Appendix]. More on pre-
and postmodifying adjectives can be found in section 2.4.4,
where also the subject-object bound adjunct (SBADJ), another
type of postmodifying NP-constituent, will be discussed.

In addition, a head noun dependent infinitive and various
types of dependent clauses can occur in the tail of an NP:

(181) La preteco de la soldatoj batali.
       (The readiness of the soldiers to fight.)

[Note that a PP may stand in between a noun and its dependent
infinitive; the above example is unambiguous because the IL
lexicon entry of 'soldato' contains no valency-indication for
infinitives, otherwise a separator should have been inserted
(cf. example 57) to relate 'batali' to 'preteco'].

(182) La opinio de la publiko, nia bazo, ke la ŝtato
       respondecas pri senlaboreco ...
       (The opinion of the public, our basis, that the state
       is responsible for unemployment ...   )

(183) La sciigo, ke ili volas vendi sian fabrikon ...
       (The news that they want to sell their factory ...)

(184) La demando, ĉu ne estas pli justa ...
       (The question, whether it is not more right ...

These examples (and also 118b) are characteristic for only a

limited group of nouns   (denoting a mental or communication
process), labelled with the appropriate valency-indication in
the IL lexicon, and largely coinciding with similar groups in
English, French, German etc.
Apart from this, every NP can have a relative clause, either
restricted (without a comma) or unrestricted (with a comma):

(185a) La parlamentanoj kiuj neniam legas tiujn dokumentojn ...
       (The MPs who never read those documents ...   )

(185b) La parlamentanoj, kiuj neniam legas tiujn dokumentojn ...
       (The MPs, who never read those documents ...   )

[Note that punctuation in the IL follows strict rules (commas
occur as input specifications on ATN-arcs)].

Directing our attention back to the head noun, there are three
special forms which it can have besides the 'regular' noun on
'-o': the substantivized adjective, the nominalized verb and -
notably - the pronoun.

The substantivized adjective has the adjectival '-a' ending,
but is marked by the article 'lo' [examples 90-92, 186a] instead
of 'la'. However, if the adjective refers to an entity in the
nearby context, then 'la' must be used [93-94, 186b]:

(186a) Lo plej bona estas tiu-ĉi ŝraubilo.
        (Am besten ist dieser Schraubenzieher).

(186b) La plej bona estas tiu-ĉi ŝraubilo.
        (Der beste [Schraubenzieher] ist dieser Schraubenzieher.)

If the NP governed by an adjective contains a relative clause,
then the use of 'lo' corresponds with 'kio', whereas 'la'
corresponds with 'kiu'. The use of 'la' without a noun can be
considered as a case of ellipsis, and includes phrases for
clock time [example 95] and calendar date:

(187) Je la sesa de januaro.
       (On the sixth [day] of January.)

['je' with nominative is reserved for time point references;
the occurrence of the name of a month disambiguates between
clock time and calendar date.]

The nominalized verb is characterized by the '-ado' ending,
but has no 1-1 relationship with it (as shown by 189 and 190b):

(188) La daura uzado de tiun-ĉi metodon.
       (The lasting usage of this method.)

(189) La uzo de tiun aparaton.
       (The use of that apparatus.)

(190a) La movado de la brakojn.
       (The moving of the arms.)

(190b) La socialisma movado.
       (The socialistic movement)

Nominalized verbs form a separate category because of the
special rules that apply to the form and order of their
postmodifying PPs: these still reflect the underlying verbal
clause structure, with verb-valency object (Ovv), direct
object (O) etc. [see also section 1.3.6, and the special ATN
on p. 15 of the Appendix]:

(191) La pensado pri sia reelektado far la prezidento.
       (The thinking of the president about his reelection.)

Notice that the Ovv stays close to the (nominalized) verb, as
in the clause structure. Another interesting detail is that
the reflexive pronoun 'sia' refers to the nominalized verb's
deep subject, indicated by the preposition 'far' (Ag).

2.4.2. Pronouns.

As in all languages, a variety of pronouns can replace an NP's
head noun: personal pronouns, demonstrative pronouns, relative
pronouns, interrogative pronouns etc. [see also example 86]:

(192) Nia komputero estas pli rapida ol <u>tiu</u> de la
       universitato.
       (Our computer is faster than the one of the university.)

If pronouns govern an NP, there are important restrictions on
the occurrence of determiners and other NP constituents, similar
to conditions in most other languages. These restrictions have
been laid down in the NP-ATN [Appendix, p. 10] and related
compatibility matrices. Notice that many pronouns can serve as
a determiner as well, without a morphological distinction (in
order to decide on this, the IL-parser has to look ahead or to
consume subsequent words of the input string).
In this section, we will point out a few deviations of the IL
with respect to common Esperanto.

<u>Personal pronouns</u> make very concise NPs. For DLT, the 3rd
person pronouns are the most interesting ones [cf. section V.2].
Their significance as unambiguous anaphorics has already been
illustrated in 1.3.9 (examples 83-85). Table IV-3 summarizes the
IL grammar on this issue, which is in accordance with various
proposals, old and recent ('iŝi' [PAG 46 Rim. I]; 'hi', '-iĉ'

'hi'/'ihi' is used   1. with nouns that are implicitly
                        masculine:
                              viro, patro, filo, knabo etc.

                     2. with nouns that are explicitly
                        masculine (see proposal mentioned in
                        [Lo Jacomo, 1981: 195-198]):
                              profèsoriĉo, instruistiĉo,
                              stultuliĉo
                        etc. using the suffix '-iĉ'.

                     3. to refer to the non-mentioned
                        masculine member of a couple,
                        with antecedents like:
                              gepatroj, geezdoj, geknaboj
                        etc., mainly in contrastive
                        alternation with 'ŝi'.

'ŝi'/'iŝi' is used   in referencing nouns that are
                        explicitly feminine, marked by the
                        suffix '-in':
                              virino, profesorino,
                              instruistino.
                        (implicitly feminine nouns in
                        Esperanto are rare, e.g. 'matrono');
                        further, it can be used in cases
                        parallel to 3. above.

'li'/'ili' is used   1. when sex does not matter or when it
                        is unknown, including references to
                        names such as:
                              juĝisto, policano,
                              prezidanto, infano, homo;
                     2. when referring to mixed groups or
                        couples in plural.

'ĝi' is used   1. in referencing institutions,
                        concepts, things or animals:
                              registaro, ministerio,
                              sperteco, suno, bovo;
                     2. with collective names in singular:
                              nacio, oficistaro, publiko,
                              grupo;
                        the plural form of 'ĝi' is 'ili'.

Table IV-3.   "The sexual revolution". Rules for the use of
              3rd-person personal pronouns (singular/plural)
              in BCE, based upon proposals from Layers II and
              III [cf. fig. IV-1].

[Lo Jacomo, 1981: 195-198]). Though DLT, by this extension, offers the best prospects for avoiding sexist language (in line with modern trends), the main motivation for it has been the improvement of referential precision and correspondence with certain existing language elements, such as the French 'elles'.

Another IL element based upon a proposal by [Lo Jacomo, 1981: 199-206] is the inclusion of 'aliu' into the set of so-called 'tabelvortoj' [PAG 85]:

(193a)   Aliu venos.
         (Someone else will come.)
(193b)   Aliu viro venos.
         (Some other man will come.)

(193c)   Alia viro venos.
         (A different man will come.)
(193d)   Alispeca viro venos.
         (A different kind of man will come.)

(193e)   Alie hi faros tion.
         (He will do this elsewhere.)
(193f)   Aliloke hi faros tion.
         (He will do this at a different place.)

(193g)   Alien hi iros.
         (He will go elsewhere.)
(193h)   Aliloken hi iros.
         (He will go to a different place.)

(193i)   Aliel hi laboras.
         (He works differently.)
(193j)   Aliele hi laboras.
         (He works differently [slightly more emphasis].)
(193k)   Alimaniere hi laboras.
         (He works in a different way.)

(193l)   Alio okazis.
         (Something happened.)
(193m)   Alies propono priparolajtos.
         (Somebody else's proposal will be discussed.)

At the same time, these examples demonstrate the IL's ability to express various graduations of emphasis (cf. also the existence of 'tiela' as intensified form of the demonstrative 'tia').
As to the reciprocal pronoun, the IL has a compact, hyphenated form when the direct (O) or indirect object (Ovv) is concerned:

(194a)   Ili renkontis unu-aliun.
         (They met each other.)

(194b)   Ili dependas de unu-aliu.
         (They depend on each other.)

(195c)   Ili sendas al unu-aliu informojn.
         (They send each other informations.)

(195d)   Hi metis ie-en unu-aliun la skatolojn.
         (He put the boxes into each other.)

In subject- or object-bound adjuncts (SBADJ), an analytic
as well as a compact form is possible, which can be used to
transfer stylistic graduations:

(196a)   Ili, la unu kun la aliu, iris ie-al parizo
         (They, the one with the other, went to Paris.)

(196b)   Ili, kun unu-aliu, iris ie-al parizo.
         (They went with each other to Paris.)

When occurring separately, 'unu' as well as 'aliu' are
inflected (in the same way as 'tiu' [see also exaples 96-97]).

2.4.3. Determiners and numerals.

Various classes of pronouns (demonstrative, interrogative
etc.) can act as determiner, or have determiners as their
derivatives (by appending the adjectival '-a' ending to a
personal pronoun, it becomes a possessive determiner).

As in other languages, an NP can be introduced by a combination
of determiners. In BCE, the maximum number of determiners is
four, taking into account the fact that quantifiers ('multaj',
'kelkaj' etc.) and words like 'aliu' and 'ambau' are considered
as determiners too. But of course, only certain sequences of
determiners are allowed [cf. Wells, 1978: 48], again reflecting
conditions common to the languages that surround us. For DLT,
the mutual compatibility of determiners is laid down in a matrix,
to be used in conjunction with the ATN on p. 10 of the Appendix.

As in common Esperanto, there is no indefinite article in BCE.
The rules for the definite article' largely coincide with those
in French (except for the partitive, see below). The introduction
of two new articles, 'le' and 'lo' (for formal disambiguation and
substantivized adjectives) has been explained in section
2.1, F1 [examples 88-92]. In addition, 'lo' is used for verbs
without subject [87b].

Unlike French, BCE does not have the simple partitive, but omits
the article in such a case:

(197) Ŝi aĉetos fromaĝon.
       (Elle achetera du fromage.)

However, BCE does have a partitive construction in combination
with adverbs, correlatives and nouns that denote a quantity or
subset. This construction uses the particle 'da' as the linking
element, and has been inspired by the mentioning of 'multe da',
'tiom da', 'kiom da' etc. in the determiners section of the PAG
[PAG 84].
Whereas common Esperanto regards 'da' as a preposition (always
followed by the nominative), the IL treats this particle as
the concluding element of a composite determiner:

> pli da   iom da   litro da
> tro da   kiom da   glaso da
> multe da   tiom da   barelo da
> multo da   ĉiom da   skatolo da
> kelke da   aliom da   autobuso da
> kelko da   neniom da   grupo da
> sufiĉe da   ĉiom-do da   speco da

[in fact an open class, as the third column suggests].
The result of this treatment is that the head noun keeps
its distinctive case ending (nominative or accusative), which
importantly contributes to syntactic unambiguity:

(198a) Kloromicetino entenas pli da kloron ol da klorofilon.
       (Chloromycetine contains more chlorine than chlorophyll.)
(198b) Kloromicetino entenas pli da kloron ol klorofilo.
       (Chloromycetine contains more chlorine than chlorophyll
       [does].   )

The particle 'da' can also be preceded by a numeral, including
the nominal numerals ('miliono' etc.):

(199a) Tri da tiuj leĝprojektoj pritraktajtis.
       (Three of the bills were dealt with.)

(199b) Tiuj-ĉi tri da tiuj dek nomoj ...
       (These three out of those ten names ...)

(199c) Malsatego minacas milionojn da homojn en afriko.
       (Famine threatens millions of people in Africa.)

[Notice that constructions with 'da' may be followed by other
determiners (199a). Also note that that the noun preceding
'da' agrees in case with the head noun of the NP.]

Talking about numerals, BCE uses the following endings when
they appear in the text as literal number strings:

```
'-b'   for basic numerals (cardinals):              '250'-b,
'-a'   for ordinals (adjectival numerals):           '28'-a,
'-o'   for autonomous use, such as:      iam-en '1983'-o.
```

If a numeral is preceded by a 'floater' (one of the particles
'ankau', 'ne', 'ec' [see also 2.5.4]), the latter's scope is
made unambiguous by insertion of a separator, indicated here
as an underscore [see p. IV-5, item e]:

(200a) Ne tri viroj, sed kvar.
       (Not three men, but four.)

(200b) Ne _tri viroj, sed virinoj.
       (Not three men, but women.)

(201a) Ni povas pagi ankau kvin mil guldenojn.
       (We can pay also five thousand guilders.)

(201b) Ni povas pagi ankau _kvin mil guldenojn, anstatau frankojn.
       (We can pay also five thousand guilders, instead of francs.

[The underlined words have emphasis because of their relation
with 'ne' and 'ankau'. Notice that the accusative is preserved
in the object-bound adjunct at the end of sentence 201b].

Several prepositions and adverbs ('ĉirkau', 'ĝis', 'ekzakte',
'almenau' etc.) can precede a numeral. Some of these supplements
can again be modified themselves, by certain preceding
intensifiers and other particles ('tre', 'eĉ', 'ankorau' etc.),
according to a so-called compatibility matrix [see p. 3 of the
Appendix]. An example is:

(202) Ĉiuj tiuj je jam pli ol cent jarojn agaj arboj ...
       (All those already more than hundred years old trees ...)

in which 'je' separates the determiners ('ĉiuj tiuj') of the
head noun ('arboj') from the numeral supplement contained in
the 'Accusativus Mensurae'. See also the more simple numeral
supplement in example 20a.

2.4.4. Pre- and post-modifiers.

This section contains remarks about adjectives and adjectival
clauses. In addition, subject-object bound adjuncts (SBADJs)
will be covered. More on NP modifiers, especially in connec-
tion with modifier scope and coordination, can be found in the
next section [2.4.5]. PPs are also discussed in section 1.3.4.

Adjectives can be divided according to their origin.
On the one hand, we have the 'qualitative' adjectives, which

denote a quality directly, e.g.:

>   granda   varma   fleksebla

On the other hand, there are the adjectives that have been
derived from a noun or adverb. They are called 'relational',
because they relate the indicated quality via something else:

>   japana   hodiaua   subtegmenta

Relational adjectives are incompatible with certain adverbs,
especially with intensifiers and comparatives. They include
material source adjectives ('stala', 'nilona' etc.).
With respect to compound relational adjectives, Esperanto
appears to be more productive than other languages [cf. section
4.2], and this will of course have its influence on the IL. An
adjectival premodifier in BCE will therefore often correspond
with a postmodifying PP in other languages, e.g.:

(203) La subtegmenta konstruo ...
       (The construction of the attic ...)

Sometimes, the same adjective can have a qualitative as well as
a relational meaning. This is solved by the use of a connecting
underscore in case of the latter [example 78, p. IV-24]. The
same device can be used for any compound term included in the DLT
lexicon:

(204) Ekscesa_aĉetpovo kaj troa_elspezado.
       (Excess purchasing power and overspending.)

(205) Pogranda_komerco.
       (Wholesale trade.)

[The connecting underscore here should not be confused
with the separating underscore used elsewhere in this report.]

Around the adjective, an adjectival clause can be formed,
especially if the adjective is a participial one (of the -ANTA
or -ATA type, see p. IV-34, which also shows the canonical
word order for ADJ1 and ADJ2 adjectival clauses):

(206a) La far la profesoro uzataj novaj fakvortoj ...
       (The new technical words used by the professor ...)

(206b) La novaj fakvortoj, uzataj far la profesoro, ...
       (The new technical words used by the professor ...)

(207a) La jam iam-en '1965'-o ie-en parizo amplekse priparolitaj
       proponoj ...
       (Die bereits in 1965 in Paris ausführlich erörterten
       Vorschläge ...   )

(207b) La proponoj, priparolitaj iam-en '1965'-o ie-en parizo
       amplekse, ...
         (Die Vorschläge, bereits in 1965 in Paris ausführlich
         erortert ...   )

[In most languages, postmodifying clauses (ADJ2s) lend themselves
better to long and complex supplements, particularly in technical
texts.]
But not only participial adjectives can have a supplement. We
have also seen the "accusative of measure" [examples 13, 105,
106, 202], and the adjective's lexicon entry can indicate
valency for a prepositional object (Oav). Moreover, a free
adjunct (of place, time etc.) can supplement an adjective:

(208a) Tiu al li agrabla fakto ...
         (That to him enjoyable fact ...)

(208b) La dum la ekzameno trankvila lernanto ...
         (The during the examination quiet pupil ...)

Subject- or object-bound adjuncts (SBADJs) can occur at the
very end of an NP. In contrast to free adjuncts, they specify
the subject or object in particular, instead of the clause as
a whole [see also fig. IV-2]. They are introduced by one of
the following simple or compound prepositions:

        kun   krom   escepte de
        sen   kromau   inkluzive de
        kiel   ekster   ekskluzive de
        kvazau   anstatau
        estiel   ekzemple
        kvankam

[For 'kromau', see p. IV-9. For 'estiel' see PAG 200 Rim. II.
also note that 'kvankam' is a preposition in BCE (the
corresponding subordinating conjunction is 'kvankam ke'.]
In addition, an SBADJ can consist of the reciprocal adjunct
'kun unu-aliu' or 'la unu kun la aliu' [see example 196].

Another type of constituent which can occur at the back of an
NP is the comparative supplement correlating with 'pli',
'tro' etc. in an adjectival clause, e.g.:

(209) Ŝanĝo, tro granda por akcepti ĝin ...
         (A change too big to accept ...   )

(210) Ni havas proponon, multe pli bonan ol la ĵus prezentitan.
         (We have a proposal, much better than the one just presented.)

An adjectival clause can also have a consecutive adjunct (CADJ)
[cf. 2.1.2, examples 110-111].

2.4.5. Coordination and modifier scope.

In this section, we will make use of a simplified NP pattern,
for sake of the explanation:

    [DET] [ADJ]   N   [PP]

For DET (determiners) and ADJ (adjectives) we will use the
collective term premodifiers, and PPs (prepositional phrases)
will be referred to as postmodifiers.

Case and number agreement between nouns and adjectives is
observed in BCE, as it is in common Esperanto. Apart from
adjectives, also certain determiners ('tiu', 'gia' etc., but
NOT the articles) agree with the noun. The agreement is
visible from the grammatical '-j' and '-n' endings, e.g.:

	SINGULAR	PLURAL
NOMINATIVE	kolektiva kontrakto	kolektivaj kontraktoj
ACCUSATIVE	kolektivan kontrakton	kolektivajn kontraktojn

If one adjective relates to only one noun, the above agreement
is often redundant with regard to formal syntactic unambiguity.
However, its role becomes much more important in case of
coordination. We will first have a look at coordinated nouns
in the presence of a premodifier (coordinated premodifiers in
the presence of one noun will follow further in this section):

(211a) Necesa ŝparado kaj investado.
       (Necessary retrenchment and an investment.)

(211b) Necesaj ŝparado kaj investado.
       (Necessary rentrenchment and investment.)

Because the first of the two nouns is singular, the adjective's
plural ending reveals its extended scope. In contrast to common
Esperanto [PAG 128 Rim. I], BCE applies this mechanism also to
determiners. Unfortunately, the mechanism does not work when
the first noun has a plural ending as well.
If two coordinated nouns denote one and the same entity, then
a special conjunction 'kau' (instead of 'kaj') is used, and
any premodifier will have the singular ending (unless both
nouns are plural themselves):

(212) La sperta direktoro kau fondinto.
       (The experienced director and founder.)

(213a) Fidela amiko kau kolego.
       (A faithful friend and colleague.)

(213b) Fidela amiko kaj kolego.
       (A faithful friend and a colleague.)

Case and number agreement helps to maintain unambiguity of
modifier scope in the presence of coordinated nouns, but in order
to fully achieve this goal, a supplementary mechanism will be
required. This is the insertion of disambiguating elements (EXTRA
SPACES, indicated in this report by underscores), as announced in
section 1.1 (p. IV-5) of this chapter. Its principles will be
explained below.

Turning our attention to postmodifiers, consider the linear
string:
                 N PP PP PP PP

which can be rewritten: N prep N prep N prep N prep N.

The question is how the PPs are logically related to each
other. The IL tree of such an NP shows these relations very
clearly (fig. IV-3a), making use of different NP levels.
In the IL tree-to-string conversion (Step 3 of the DLT
process [see section III-4.1]), the following mechanism is
applied to preserve the contents of the tree structure in
the linear string as well: if a PP does not belong to the N
immediately preceding it, but relates to an N which is k
levels up in the tree, then k extra spaces are inserted
before this PP. We illustrate this with the example:

(214)   La decido far la komisiono _pri la atribuado
        de grenojn _al la landoj de la tria mondo
        __lau la lego de '1965'-o...
        (The decision by the commission on the assignment
        of cereals to the countries of the third world
        according to the law of 1965...).

Without the extra space, the PP 'pri la atribuado...'
would relate to 'komisiono' instead of to 'decido'.
The PP's 'al la landoj...' and 'lau la lego...' both
relate to 'atribuado', and have to skip one and two
intermediate NP levels respectively.

During the reverse process, the string-to-tree conversion by
the IL-parser (Step 5 of the DLT process), each space in front
of a postmodifier will cause the ATN to pop out of its current
level of NP subnetwork processing (we therefore refer to these
spaces as "pop"-spaces). This ensures that the IL-tree used as
input for the IL-coder at the transmitter side, is completely
reconstructed at the receiver side (except for a difference in
word group order: the original SL-dependent word group order
has been replaced by the canonical IL order).

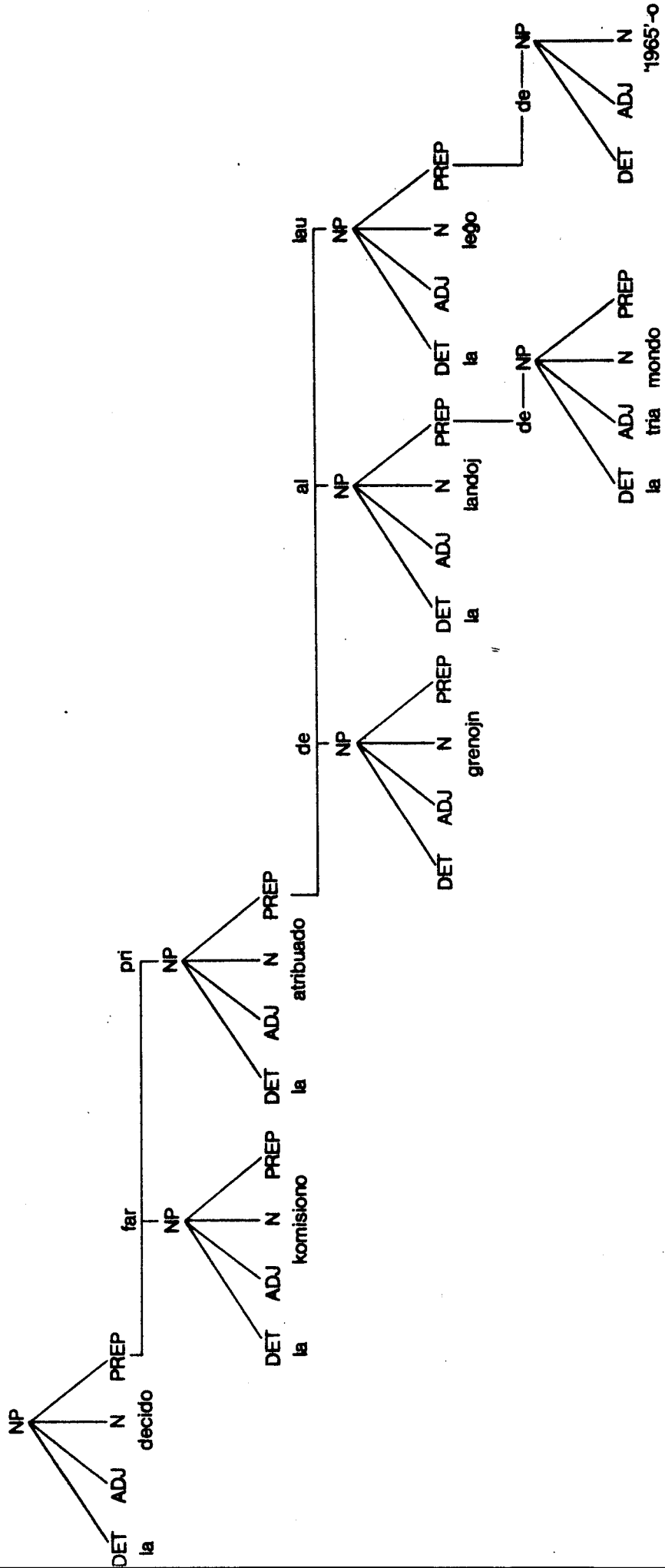Coming back to coordination of nouns, one can imagine noun

Fig. IV-3a. IL tree structure of an NP complex, showing
various levels caused by postmodifying
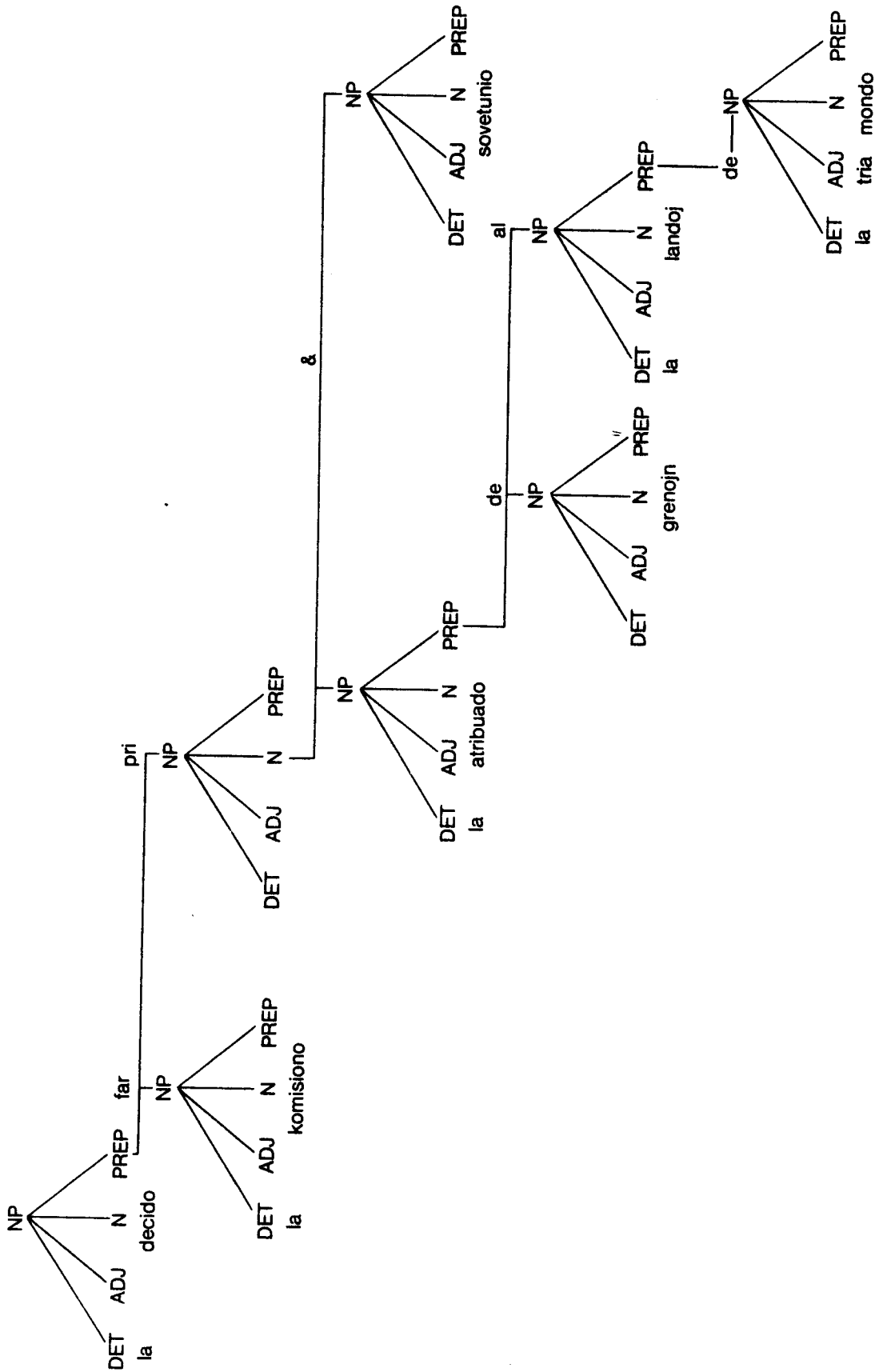(corresponding to example 214).

Fig. IV-3b. IL tree representation of noun coordination by insertion of an additional tree level (corresponding to example 214b).
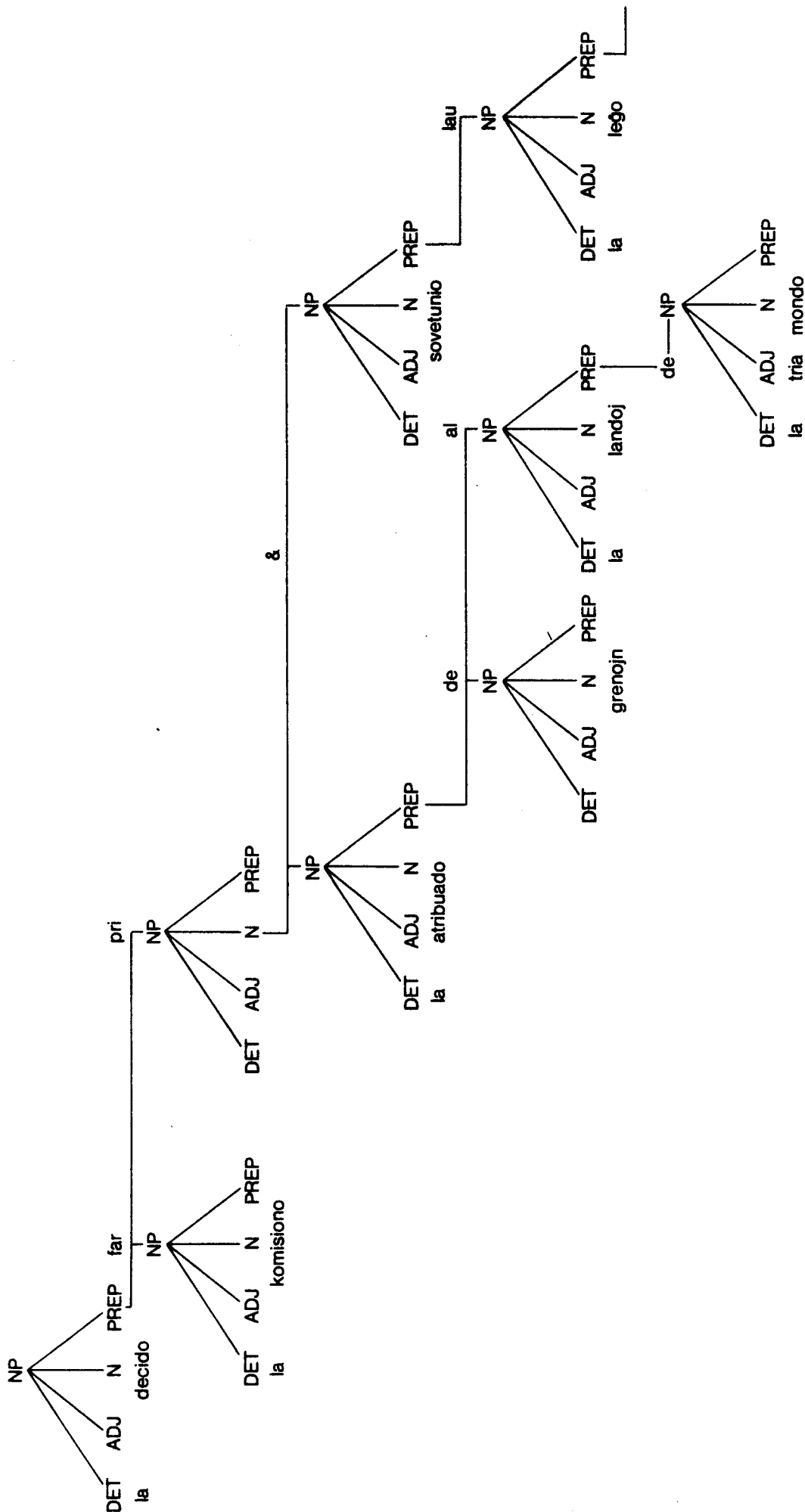
NP DET la ADJ N decido PREP far

NP DET la ADJ N komisiono PREP

pri

NP DET ADJ N PREP

NP DET la ADJ N atribuado PREP

de

NP DET ADJ N grenojn PREP

al

NP DET la ADJ N landoj PREP

de

NP DET la ADJ N tria mondo PREP

&

NP DET ADJ N sovetunio PREP

lau

NP DET la ADJ N leĝo PREP

Fig. IV-3c. The postmodifier 'lau la leĝo' applies to the immediately preceding noun.
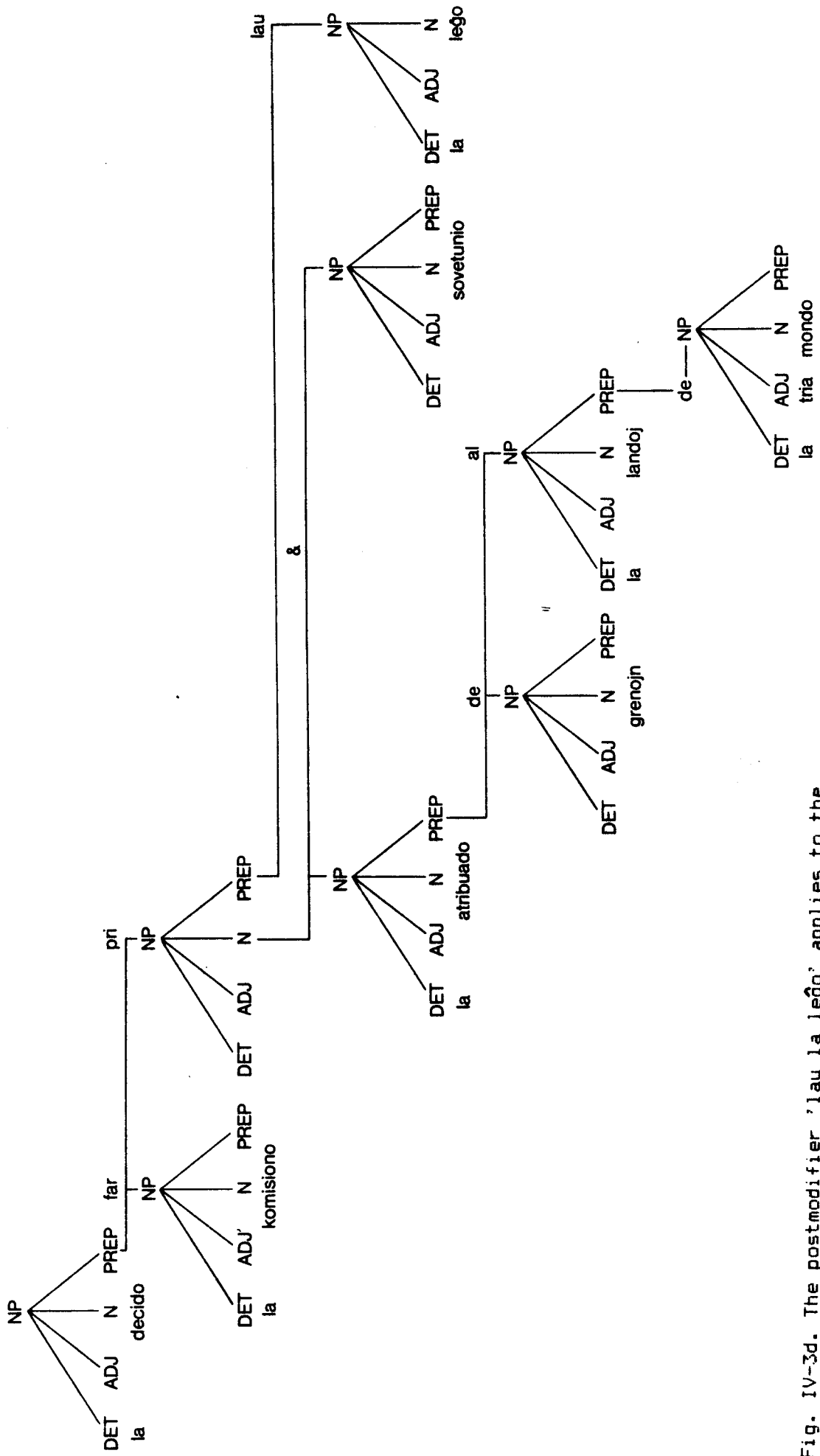
Fig. IV-3d. The postmodifier 'lau la leĝo' applies to the
coordination of the NPs 'atribuado' and
'sovetunio'.
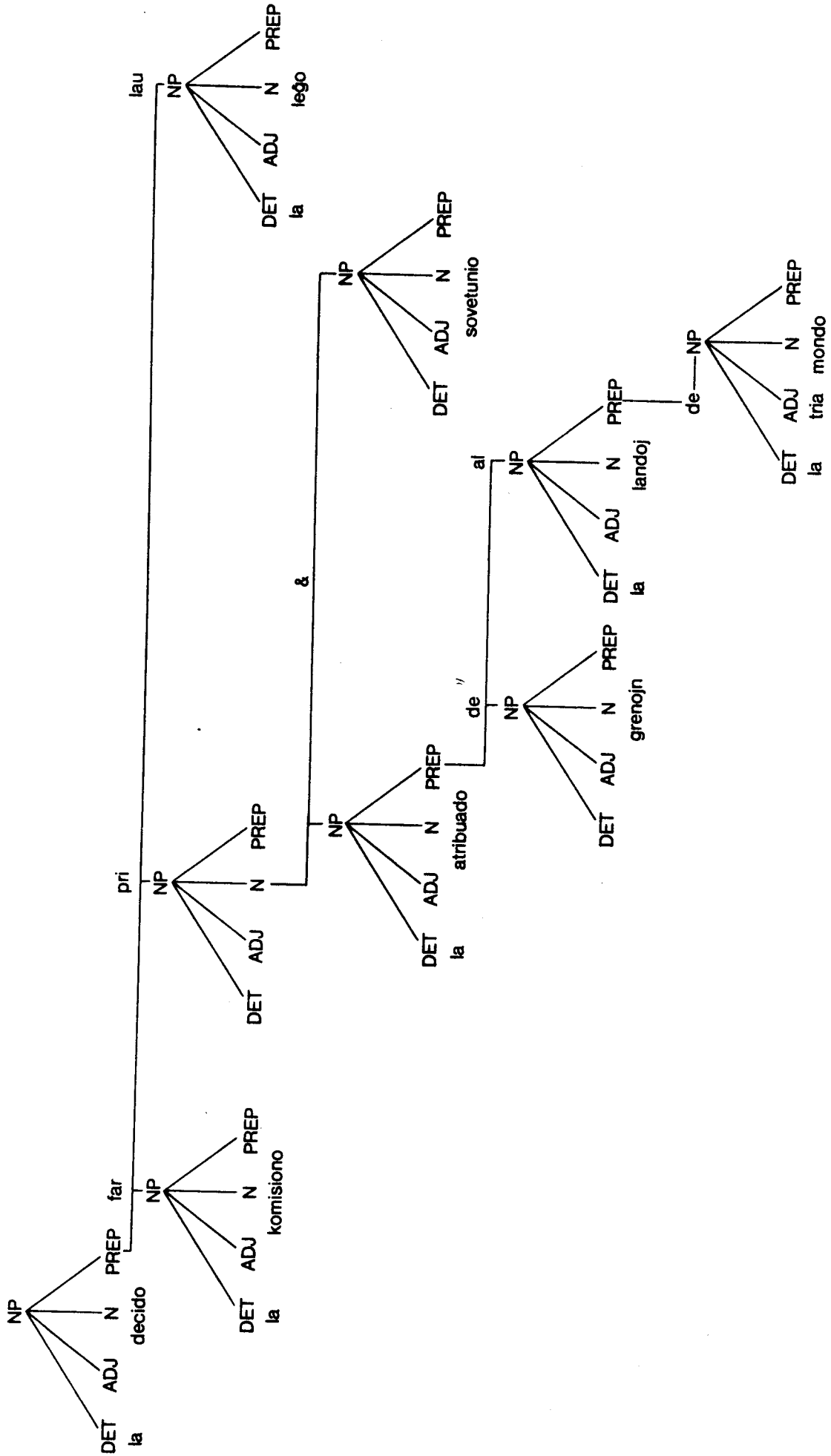
Fig. IV-3e. The postmodifier 'lau la lego' applies to the head noun of the whole complex, 'decido'.

conjunction occurring at every NP level. But how do we handle
this in a linear string? Again, in the same way as with post-
modifiers, we use extra spaces, this time in front of the
conjunction ('kaj'), if the coordination applies to a higher
NP level than that of the noun immediately preceding it. If,
for instance, we change example 214 as follows:

(214a)   ....de la tria mondo kaj sovetunio......

then the coordination is at the lowest level ('mondo'). But
if instead we have the linear IL-string:

(214b)   ....de la tria mondo __kaj sovetunio......

then the coordination is two levels higher ('atribuado'), as
laid down in fig. IV-3b, with the preposition 'pri' relating
to both 'la atribuado' and 'sovetunio'.

If, in 214b, the PP 'lau la lego' still follows 'sovetunio',
its possibilities of relating to various parts of the NP
complex will be affected by the preceding noun coordination:
'lau la lego' then can no longer relate to 'atribuado' alone,
from which it has been separated by 'sovetunio'. It can only
relate to the latter (no extra space needed, fig. IV-3c), to
the conjoin 'la atribuado kaj sovetunio' (1 extra space
needed, fig. IV-3d), or to 'decido' (2 extra spaces needed,
fig. IV-3e).

Stating this in general terms: noun coordination adds an
extra level to the IL tree structure. This level is counted as
an additional NP level when indicating the relation of post-
modifiers. In this way, postmodifier scope in the presence of
noun conjoins is strictly defined in BCE.

A noun conjoin may have more than two members, but that does
not introduce further levels. Once a coordination occurs, one
additional level is inserted into the tree, under which all
the members of that coordination can be attached (the IL trees
are by no means binary trees). This arrangement has been
inspired by the fact that any attempts to make hierarchical
distinctions in the subgrouping of the conjoined elements
appear quite irrelevant in practice [cf. Dik, 1968: 236].

An interesting issue is clause level vs. phrase level
coordination. Adopting the principle of separators ("pop"-
spaces for an ATN-based IL-parser) in front of conjunctions
(as in 214b), what will happen if we pop out of the highest
level within an NP complex? We will then simply return to the
level from which the parser descended earlier, which will
normally be the clause level.
Thus, the simple "pop"-space mechanism provides an excellent

tool for distinguishing between phrasal, clausal and even
sentential coordination. This means, for instance, that BCE
can meet the requirement [Dik, 1968: 202-215; Smith, 1969] to
distinguish between coordination within the subject (215a) and
the occurrence of two coordinated subjects (215b):

(215a) La instruisto kaj mia amiko aĉetis tiun-ĉi libron pri
        historio.
(215b) La instruisto _kaj mia amiko aĉetis tiun-ĉi libron pri
        historio.
        (The teacher and my friend bought this book about
        history.)

This provision is essential, also because BCE does not have
any verb flexion according to person and number. The pattern
of 215b (S & S V O) must be regarded as a case of clause level
ellipsis (despite the objection of [Dik, 1968: 200] against
this term) allowed in BCE.

We now conclude this section with an overview of how BCE
handles coordinative premodifier structures, i.e. several
adjectives or determiners relating to one noun. We distinguish
between:

> I. Genuine premodifier coordination (also referred
>   to as adjective coordination in this report), again
>   subdivided into:
>
>   a. Descriptive chains (in which the members
>   of the coordination, i.e. the adjectives,
>   are separated by commas or conjunctions):
>
>   (216) La bela, riĉa, kvieta kaj pura svislando.
>   (Beautiful, rich, quiet, and clean Switzerland.)
>
>   (217) Tiu juna, ĉarma,   edukita virino.
>   (That young, charming, educated woman.)
>
>   Fig. IV-4a shows the IL tree corresponding to 217.
>   As we will see, the descriptive chain is treated as
>   one unit regarding premodifier scope and other
>   operations.
>
>   b. Restrictive chains (in which the coordination
>   members are simply in juxtaposition):
>
>   (218) La nuna tutmonda terminologia problemo.
>   (The present world-wide terminology problem.)
>
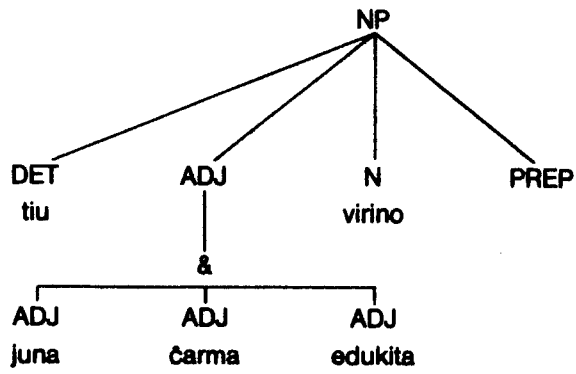>   (219) Pezaj rondaj ŝtalaj tuboj.
>   (Heavy round steel tubes.)

```
                              NP
              _____/|_____
             /            /    |         \
           DET          ADJ    N          PREP
           tiu           |     virino
                         |
                         &
                    _____|_____
                   |     |     |
                  ADJ   ADJ   ADJ
                  juna  ĉarma  edukita
```

Fig. IV-4a. Descriptive adjective coordination in the IL
            tree structure.

```
                              NP
              _____/|_____
             /            /    |         \
           DET          ADJ    N          PREP
                        pezaj   |
                                |
                               NP
                     _____/|_____
                    /       /   |         \
                  DET     ADJ   N          PREP
                          rondaj |
                                 |
                                NP
                        _____/|_____
                       /     /  |         \
                     DET   ADJ  N          PREP
                           ŝtalaj tuboj
```
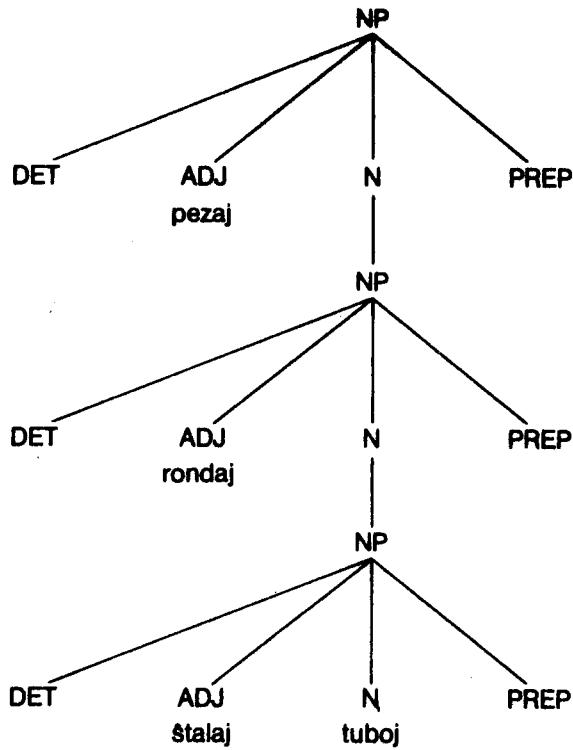
Fig. IV-4b. Restrictive adjective coordination in the IL
            tree structure
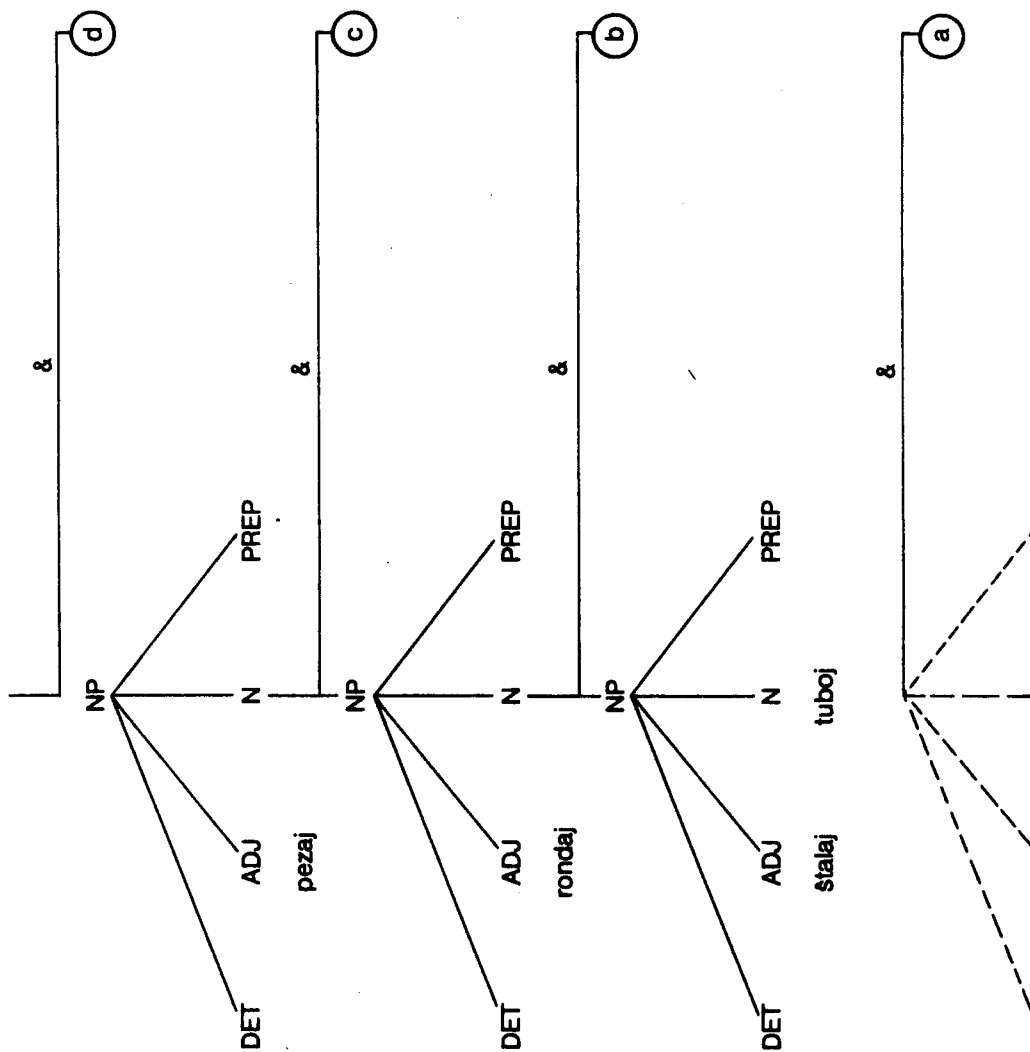
Fig. IV-4c. Attachment points for noun coordination at various levels. The attachment level determines the scope of each of the restrictive adjectives (if attachment point "a" is used, the noun 'tuboj' shifts to the dashed level).

Fig. IV-4b shows the IL tree structure. Note that
each restrictive premodifier (apart from the first)
adds another NP-level to the tree. Here we act in
agreement with the well-known principle that only
grammatical constituents may be represented by
subtrees. The linear adjective chain of 217 is a
grammatical constituent, the one of 219 is not.

As for _modifier scope_, each member of a restrictive
adjective chain has its own, and these separate
scopes may or may not coincide. This is illustrated
by the IL-handling of a noun coordination following
'tuboj' in 219:

(219a) ... tuboj kaj profiloj ...
(219b) ... tuboj _kaj profiloj ...
(219c) ... tuboj __kaj profiloj ...
(219d) ... tuboj ___kaj profiloj ...

which corresponds to different NP-levels in the tree
(attachment points a, b, c and d [fig. IV-4c]), or to:

(a) Heavy round steel (tubes and profiles),
(b) Heavy round ((steel tubes) and profiles),
(c) Heavy ((round steel tubes) and profiles),
(d) (Heavy round steel tubes) and profiles.

If, instead of noun coordination, a postmodifier
were added to 98:

(220) Pezaj rondaj ŝtalaj tuboj por usono.
      (Heavy round steel tubes for the USA.)

it would be put in the lowest-level NP-subtree
initially (lacking other information). In certain
cases [not explained here for brevity], it may move
up in the tree upon subsequent information.

In addition, mixed chains are allowed, provided that
the descriptive part precedes the restrictive part.
All these prescriptions can be derived from the ATN
for NP-structure [p. 10 of the Appendix].

II. Elliptic noun coordination.

This type of coordination is fallacious: though super-
ficially resembling an adjective coordination, it is
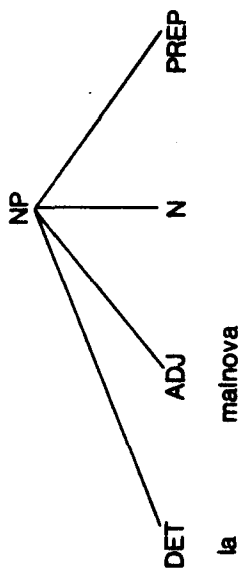in fact a coordination of nouns, one of which has been
omitted, e.g.:

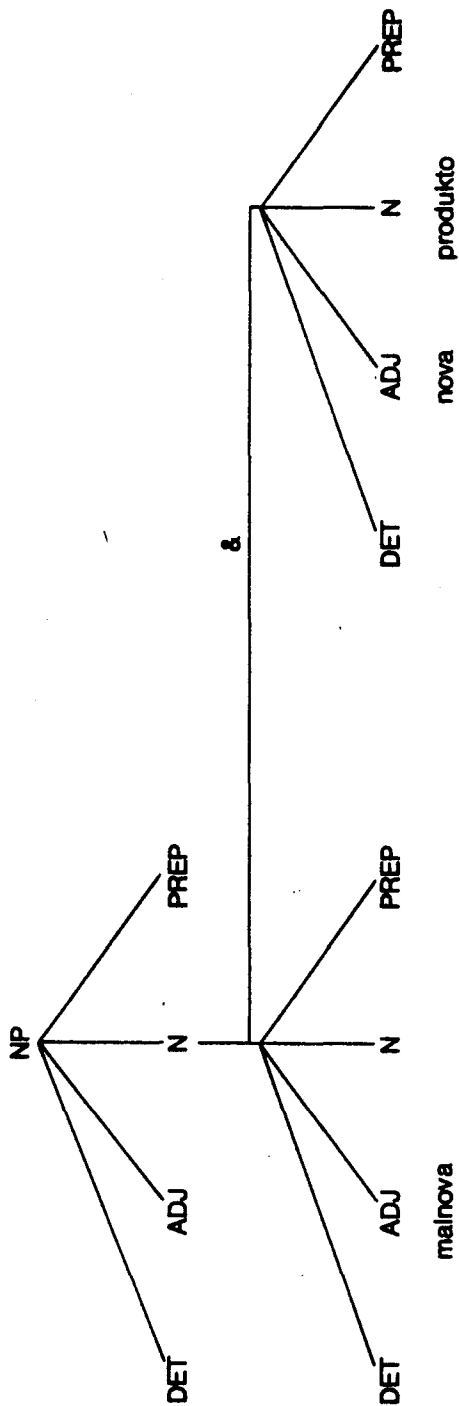Fig. IV-5a. Initial tree structure, built by IL-parser.

Fig. IV-5b. Final tree structure, built by IL-parser.

Fig. IV-5. IL tree structure in case of elliptic noun
           coordination.

(221) La malnova _kaj nova produkto.
      (The old   and new product.)

Such a BCE fragment will be processed by the IL-parser
as follows:
Initially, the determiner 'la' and the adjective
'malnova' are put in the same NP-subtree (fig. IV-5a).
As soon as the separator (1 extra space) and the
conjunction ('kaj') turn up, the parser concludes that
there is an elliptic noun coordination [cf. the ATN,
Appendix p. 10]. As a consequence, the parser not only
opens two new NP-subtrees on a lower level, but also
moves the adjective 'malnova' (whose scope is clearly
limited to the first of the two nouns) to one of these
new subtrees. The determiner 'la', on the other hand,
remains at its original level, as it will apply to the
conjoin as a whole, not only to the first noun. The
resulting tree structure is shown in fig. IV-5b.

It has been pointed out [section 1.1, item i.] that
the IL-parser works without applying 'deep' or even
'shallow' semantics. Without the formal distinction of
a separator (indicated by an underscore) in a fragment
such as 221, the IL-parser would assume genuine
premodifier coordination [type I.a above, examples
216-217]. The parser would NOT base a decision upon
the apparent mutual exclusiveness of the two 'polar'
adjectives 'nova' and 'malnova'.

2.5. Free adjuncts (FADJs).

2.5.1. Typology.

Among the IL syntactic categories, free adjuncts (FADJs) are
adverbials of time, place, manner etc. that relate to the
clause as a whole (in contrast to adjuncts bound to the
subject or object specifically). They are called 'free'
primarily to underline the distinction with PPs bound to the
clause's verb by valency [cf. fig. IV-2].

Free adjuncts have their place in BCE's canonical word group
ordering [section 2.1.2]. In the general clause, they often
act as clause introducers and can only be preceded by context-
linking or speaker-oriented adjuncts [section 2.6]. In the
infinitival clause however, they are placed near the end, an
arrangement which contributes to readability as well as
structural unambiguity. Also notice their position in adjectival
and participial clauses [p. IV-34].

The shape free adjuncts can take ranges from a single adverb
to adverbial clauses, PPs and clauses introduced by a

subordinating conjunction, e.g.:

(222)   Lastatempe oni faris proponojn pri tiu-ĉi projekto.
        (Lately, proposals have been made concerning this project.)

(223)   Legante artikolon pri nukleaj armiloj, ni vidis kelkajn
        okulfrapajn fotojn.
        (Reading an article about nuclear weapons, we saw several
        striking photos.   )

(224)   Per altigo de la socian rekonon de la senlaborecon ...
        (By increase of the social recognition of unemployment ...)

(225)   Kiel ni konstatis, la pozicio de finlando estas stabila.
        (As we ascertained, the position of Finnland is stable.)

(226)   Se konstante la sciencistoj insistas pri ĝi, ankau la
        politikistoj devos sekvi.
        (If the scientists insist on it constantly, the
        politicians will have to follow too.   )

As some of these examples suggest, a free adjunct can become
quite complex, such as a PP with dependent PPs again [224].
In addition, a clause may have a sequence of several FADJs [up
to some practical limit, still to be defined]. Therefore, BCE
must show distinction between PPs depending on one FADJ [as in
224], and a series of juxtaposed FADJs. This is done partly by
commas, partly by explicit separators [cf. p. IV-5, item e.].
A comma is always used to separate FADJs from a preceding LADJ
[section 2.6], as indicated in the ATN on p. 8 of the Appendix.

The semantic coverage of free adjuncts includes the following
categories (note the two important classes of exceptions):

        TIME
        PLACE (except: PPs with verbs of position and motion)
        MANNER
        INSTRUMENT

        GOAL
        CAUSE   } except: verb-valency bound PPs
        STATE
        MEASURE

        TRACTATIVE
        BENEFICIARY

[the term 'tractative' denotes the topic of something
(expressed in English by 'about', 'on', 'concerning' etc.,
in both BCE and Esperanto by 'pri' and 'koncerne'.]
The following pair of examples illustrates how the assignment
of a PP, to either the FADJ or the Ovv category, depends on

the valency pattern of the IL verb ('logi' does, 'renkonti'
does not have explicit valency for a place complement):

(227a) Ie-en parizo he renkontis knabinon.
       (He met a girl in Paris.   )

(227b) Hi logas ie-en parizo.
       (He lives in Paris.   )

If a FADJ is a PP or subordinate clause, its semantic category
can be established unambiguously from the preposition or
subordinating conjunction by which the FADJ is introduced.
Already in common Esperanto, the system of prepositions is
more regular than in other languages, and BCE adds even more
regularity to it (see below). In case of adverbs and adverbial
clauses, it is much more difficult if not impossible for the
IL-parser to assign the FADJ to any of the above-mentioned
categories (we remind the reader that the IL-parser does NOT
involve 'deep' semantics processing [cf. p. IV-3]). See also
examples 61-62 in this Chapter.
If a number of FADJs occur in juxtaposition, a BCE-specific
ordering scheme (e.g.: [TIME] [PLACE] [MANNER] ...) might
help [the present BCE-design does not include such a scheme].

Fig. IV-6 [in section 3.2] shows the formal definition of the
FADJ structure, in the form of an ATN. An enumeration of FADJ
introducers can be found on pp. 5-6 of the Appendix. The sub-
sections below give separate overviews of several important
classes of elements, used in FADJs but also elsewhere.

2.5.2. Prepositions.

A number of Esperanto prepositions can have a temporal as well
as a local meaning. In BCE, these two different senses are
distinguished by means of the hyphenated prefixes 'iam-' and
'ie-' respectively.
The unprefixed form of these prepositions remains available
for meanings other than temporal or local:

TIME	PLACE	(other meanings)
iam-en	ie-en	en
iam-tra	ie-tra	tra
iam-inter	ie-inter	inter
iam-ĉirkau	ie-ĉirkau	ĉirkau
iam-de	ie-de	de
iam-ĝis	ie-ĝis	ĝis
iam-antau	ie-antau	antau
iam-post	ie-post	post
iam-ekster	ie-ekster	ekster
iam-ĉe	ie-ĉe	ĉe

Besides, there is a small number of prepositions that
have a local and a non-local, but not a temporal meaning:

ie-al	al
ie-el	el
ie-lau	lau
ie-apud	apud

Borderline cases will always turn up, mainly between the local
and the 'other' meaning of a preposition, and particularly if
one thinks of metaphor. However, these cases will overlap with
those to be covered by detailed information in the DLT lexicons
(SL-IL as well as IL-TL), ranging from valency translation to
the translation of idioms, collocations etc., a comprehensive-
ness that is required anyway [Knowles, 1982: 152].
Apart from that, the above prefixing scheme provides an effective
disambiguation instrument to the non-deep IL-parser:

(228a) Iam-en la dua mondmilito.
       (In the second world war.)

(228b) Ie-en usono.
       (In the USA.   )

(228c) En certa senso.
       (In a certain sense.)

(229a) Ie-lau la rivero.
       (Along the river.)

(229b) Lau la prezidento.
       (According to the president.)

(230a) Iam-de la militista reĝimo ie-en polando usono limigis
       la eksporton de teknologio _ie-al sovetunio.
       (Since military rule in Poland, the USA limited
       the export of technology to the Soviet Union.)

(230b) Iam-de la militista reĝimo ie-en polando usono liveris
       al sovetunio malpli da teknologion.
       (Since military rule in Poland, the USA supplied
       less technology to the Soviet Union.   )

[In 230a, two locative PPs depend on an NP; in 230b, a verb-
valency bound prepositional object occurs.]
Prepositions denoting PLACE indicate POSITION if they are
followed by the nominative, DIRECTION (motion towards) if they
are followed by the accusative case:

(231a) Ie-en amsterdamo.
       (In Amsterdam.   )

TIME:	je	iam-post	dume de
	dum	iam-tra	fine de
	iam-ĉe	iam-antau	daure de
	iam-de	iam-inter	komence de
	iam-en	iam-ekster	
	iam-ĝis	iam-ĉirkau	

PLACE:	cis	ie-ĝis	funde de
	sub	ie-lau	meze de
	sur	ie-tra	flanke de
	trans	ie-apud	fronte al
	super	ie-post	centre de
	preter	ie-antau	kerne de
	ie-al	ie-ĉirkau	direkte al
	ie-ĉe	ie-inter	ekstere de
	ie-de	ie-ekster	interne de
	ie-el	ie-kontrau	laulonge de
	ie-en	sube de	supraĵe de

INSTRUMENT:	per	
AGENT:	far	
GOAL:	por	cele al
CAUSE:	pro	kauze de
ACCOMPANIMENT:	kun	sen

MISCELLANEOUS:	al	ekster	prepare al
	ĉe	kromau	rezerve de
	de	kontrau	escepte de
	el	malgrau	foreste de
	en	anstatau	kompense al
	je	kaze de	konforme al
	po	nome de	renkonte al
	ĝis	helpe de	responde al
	lau	manke de	kontraue al
	pri	sekve de	inkluzive de
	tra	spite de	konsidere de
	apud	ĉeeste de	vidalvide al
	krom	favore al	ekskluzive de
	post	memore al	konsekvence de
	antau	okaze de	
	inter	rilate al	
	ĉirkau	komisie de	
	estiel		
	kvazau		
	kvankam		
	ekzemple		

Table IV-4. Enumeration of IL prepositions.

(231b) Ie-en amsterdamon.
       (Into Amsterdam.   )

In the present BCE design, the preposition 'ie-al' (which
always indicates direction) forms an exception in this respect.
In accordance with the use of 'al' in common Esperanto, it
can only be followed by the nominative:

(231c) Ie-al amsterdamo.
       (To Amsterdam.   )

Table IV-4 gives a general overview of prepositions in BCE.
Note that 'far' can never occur in a FADJ, as it is reserved
for the agent of the passive (Ag), a separate syntactic
function category.
The table shows simple, prefixed and compound prepositions.
Simple prepositions play also an important role as elements
in the formation of subordinating conjunctions [see the next
section].
A few prepositions ('per', 'por') can introduce infinitival
clauses (INF), which is another realization of FADJs [cf. the
ATN, fig. IV-6]:

(232) Por vendi siajn produktojn oni bezonas merkatesploron.
       (To sell one's products one needs a market investigation.)

The compound prepositions listed in table IV-4 are all formed
with 'al' and 'de'. Their enumeration is not exhaustive. Besides,
there exist compounds formed with other prepositions ('kompare
kun', 'meze ie-inter'). Also, compounds with adverbialized
prepositions ('kune kun', 'sube sur'), combinations ('de ce')
and coordinated prepositions ('kun kaj sen', 'por kaj kontrau')
have been investigated as part of the feasibility study.

2.5.3. Subordinating conjunctions.

As announced in section 2.1.3 [p. IV-38], BCE has adopted a
regular scheme for the formation of subordinating conjunctions.
On the basis of this scheme, the following subclassification
applies [Note: 'ke' corresponds to the English 'that']:

a. Preposition + 'ke' :

        dum ke   pro ke   antau ke
        gis ke   sen ke   kromau ke
        lau ke   krom ke   kvankam ke
        per ke   post ke   malgrau ke
        por ke   anstatau ke

    This model replaces common-Esperanto expressions such as
    'post kiam', 'antau ol' etc. Note that the combinations are

obligatory, e.g. 'dum' without 'ke' can only act as a BCE-preposition.

b. Adverb + 'ke' :

apenau ke	tiel ke	danke ke	escepte ke
kvazau ke	tial ke	helpe ke	kondiĉe ke
		spite ke	supoze ke

Without 'ke', these words only have their adverbial meaning.

c. The conjunction 'ĉar' :

In BCE, 'ĉar ke' is the subordinating, 'ĉar' the coordinating conjunction.

d. Subordinating conjunctions, which have an explicit or implicit correlative antecedent:

[tial] kial	[tie] kie	[tiel] kiel
[tiam] kiam	[tien] kien	[tiom] kiom

In case of an interrogative subordination, the suffix '-do' is attached [Note: 'ki'-forms correspond to the English Wh-forms].

e. The subordinating conjunctions 'ke' (used in isolation) and 'se' (English: 'if'), which retain their function as in common Esperanto. Note however, that a bare 'ke' can only introduce so-called nominal clauses (dependent clauses that replace a subject, direct object or subject predicative complement), NOT adverbial clauses (FADJs).

Combination with an adverb or preposition may alter the meaning of a subordinating conjunction in a way that contributes to the translation precision of DLT, e.g.:

de kiam	since
nun kiam	now that
tuj kiam	as soon as

In addition, the Esperanto-based IL offers several shades of emphasis [cf. the list of FADJ introducers on pp. 5-6 of the Appendix]:

pro ke	because
pro tio ke	for the reason that

2.5.4. Adverbs, intensifiers and 'floaters'.

As to their form, we can divide adverbs into the following groups:

a. Regular adverbs, ending on '-e'.
   This is an open class, the members of which can be derived
   from adjectives ('rapide') but also from other word classes
   ('antaue', 'tiurilate'). This class also includes participial
   adverbs ('legante').

b. Correlative adverbs.
   The term 'correlative' in this context originates from
   Esperanto. It denotes a matrix of function words, the
   initial characters of which correlate with (roughly)
   syntactic, the final characters with semantic functions:

	PLACE	TIME	CAUSE	MANNER	MEASURE
INDEFINITE	ie	iam	ial	iel	iom
DEMONSTRATIVE	tie	tiam	tial	tiel	tiom
RELATIVE	kie	kiam	kial	kiel	kiom
COLLECTIVE	ĉie	ĉiam	ĉial	ĉiel	ĉiom
ALTERNATIVE	alie	aliam	alial	aliel	aliom
NEGATIVE	nenie	neniam	nenial	neniel	neniom

   By appending an '-n' ending to the elements of the PLACE
   column, one obtains DIRECTION. Further, the row of RELATIVE
   adverbs can be turned into INTERROGATIVE by suffixing with
   '-do' (this applies to all Wh-questions, independent as
   well as dependent). The addition of the ALTERNATIVE row is
   a deviation from common Esperanto. Note that the full
   Esperanto matrix of correlatives also includes pronouns and
   determiners ('kiu', 'kia', 'kies' etc.).

c. Adverbs ending in '-au'.
   This is a small, closed group (the Esperanto 'kvazau', when
   functioning as an adverb, is replaced by 'kvazaue' in BCE):

ankau	almenau	hierau
ankorau	apenau	hodiau
baldau	preskau	morgau

d. Other monomorphemic (simple) adverbs:

ĉu	ne	ĵus	plu	
do	ajn	mem	tre	
eĉ	for	nun	tro	ju...des...
ja	jam	nur	tuj	
ĵe	jen	pli	plej	

As to the syntactic and semantic functions of adverbs, we can
distinguish:

  i. Adverbs serving as context-linking or speaker-oriented
     adjunct [LADJ, section see 2.6]. In the IL as well as in
     Esperanto, there are many one-word adverbs that correspond
     with collocational phrases in other languages [see the list
     on p. 2 of the Appendix].
     Also the 'modal' adverbs ('deve','pove' [cf. p. IV-43]) are
     reckoned to this category in BCE.

 ii. Adverbs serving as free adjunct (FADJ) of time, manner,
     state etc. ('frue', 'rapide', 'timide'). According to BCE
     syntax, this category includes adverbs that modify a
     (participial) adjective.

iii. Intensifiers, modifying an adjective, verb or again an
     adverb:

         tre   pli   tiel
         ege   plej   tiom
         multe   preskau   parte
         treege   sufiĉe   tute
         ekstreme   tro
         eksterordinare

     The matrix on p. 4 of the Appendix gives an impression of
     the extent to which BCE intensifiers (and other adverbs)
     can modify each other.

In addition, adverbs can be used to modify prepositions ('meze
ie-en', 'ekzakte sur'), numerals and numeral supplements [2.4.3].
They can also modify a PP as a whole (functioning as FADJ or Spc):

(233) Prefere per trajno hi vojaĝas.
       (He travels preferably by train.)

(234) La parlamento estis malkaŝe kontrau tiuj planoj.
       (The parliament was openly against those plans.)

Finally, similar again to English, French, German etc., the IL has
a subclass of words for negation and emphasis, which can occur
practically everywhere in the sentence, and therefore are
called 'floaters':

    ankau   eĉ   ne
    ĵe   mem   nur

The scope of these function words is determined by their position
[the floaters have not been included in the ATNs presented in
this report; formalization of their use and their restrictions
will of course be required]. The emphasizer 'ĵe' is a BCE addition.

2.6. Context-linking adjuncts (LADJs).

In the IL grammar, the syntactic category of context-linking
adjuncts (LADJs) meets the need for those adjuncts that are
more related with the surrounding context or the opinion of
its author than with the clause in which they are actually
placed [Lehrer, 1974; Quirk, 1980]. An LADJ can only occur at
the very beginning of a general clause [cf. p. IV-33], and
must be followed by a comma. The BCE grammar permits the
occurrence of more than one LADJ [see Appendix, p. 8].

Context-linking adjuncts (in the narrow sense), also called
'conjuncts', indicate the relation between the particular
sentence or paragraph which they introduce and other
- preceding or succeeding - parts of the text, e.g.:

        aliflanke   konklude   same
        cetere   resume   tamen

In the IL (as well as in common Esperanto), many of these
LADJs are one-word adverbs [see the list on p. 2 of the
Appendix], but also multi-word expressions ('kiel dirite')
exist.

Speaker-oriented (or rather, for written texts, 'author-
oriented') adjuncts, also called 'disjuncts', express an
opinion or judgment as to the assertion contained in the
clause. The BCE modal adverbs 'deve' and 'pove' [cf. p. IV-
43], and also adverbs indicating doubt, rumour or reservation
('laudire', 'onidire') belong to this group:

(235)   Deve, nun la aviadilo estas ie-en lisbono.
        (The plane must be in Lisbon [by] now.)

(236)   Onidire, la usona intertraktanto sukcesis.
        (The American negotiator is said to have succeeded.)

Other adverbs that can serve as LADJ are:

                honeste   saĝe
        eble   konfidence   serioze
        certe   kompreneble   sincere
        evidente miaopinie   supozeble
                niaopinie   teorie
                prave   verŝajne

Note however, that some of these adverbs can as well be
integrated in the clause proper, functioning as a free
adjunct of manner, state etc. [cf. Lehrer, 1974]. In BCE,
the comma may happen to be the only formal distinction
then:

(238a) Serioze, la rusoj ne volas intertrakti.
       (Seriously [speaking], the Russians do not want to
       negotiate.   )

(238b) Serioze la rusoj ne volas intertrakti.
       (The Russians do not seriously want to negotiate.)

(238c) La rusoj ne volas intertrakti serioze.
       (The Russians do not want to negotiate seriously.)

[The adverb is an FADJ of the main clause in 238b, of the
infinitival clause in 238c (cf. canonical order, p. IV-33].
As the translation of 238a suggests, an IL's LADJ will often
have several translation variants in English or other
languages, e.g.:

```
    IL                SL,TL
  sincere ———————— to be frank
          ————— to speak frankly
           ———— frankly speaking
            ——— putting it frankly
             —— to put it frankly
              — put frankly
```
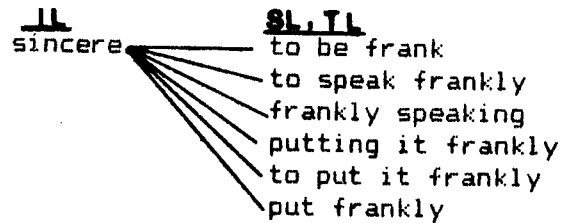
Of course, this situation contributes to the compactness of
the IL. If, on the other hand, an IL main clause is introduced
by a PP or subordinate clause expressing opinion or judgment
('lau nia nuna opinio', 'kvankam ke ni dubas pri tio') such a
constituent is treated as an FADJ, not as an LADJ, in BCE.

3. Formal definition.

This section shows how the IL is defined formally. A formal
definition of its grammar is required in order to process the IL
by computer-implemented algorithms, of which the IL-recognizer
and the IL-parser (Step 4 and 5 of the DLT translation sequence
[see section III-4.1]) are the most prominent. Besides, the
handling of the IL tree structures and their conversion to
linear strings (Step 2 and 3) will be looked at.
For actual DLT operation, these formalisms are the only effective
representation of the IL grammar. The descriptive definition
[as given in section IV.2] must be regarded as a derivative, to
be used for system documentation and explanation.

3.1. Motivation of ATN-choice.

The grammar formalism chosen for the IL (the IL-recognizer
as well as the IL-parser) is the Augmented Transition
Network (ATN), to be used with a certain degree of look-
ahead. This choice is motivated as follows:

The Esperanto-based IL is a structurally highly transparent
language, having a limited set of canonical word order patterns
and a much more stringent and clear morphology than English or
other natural languages: this characteristic lends itself well
to the use of an ATN parser.

With the IL, such a parser generally knows very well
what to expect, and the need to defer decisions is
small. A chart parser will flourish when the language
to be processed shows a lot of ambiguity at word or at
word group level: e.g. English, where the same word
can function as noun, adjective or finite verb, and
where the same NP can serve as subject, object or
adjunct. The contrary is the case for the IL, where the
part of speech of a word follows explicitly from its
grammatical ending, and where the provision of an
accusative and a more precise system of prepositions
safeguards against uncertainty of word group function.

To compare the IL with English in terms of look-ahead
requirement, we have translated 5 typical pairs of sample
sentences from [Marcus, 1980]. One of these (243) is what
Marcus calls a 'garden path' sentence, considered the most
tricky type. But also the others (239-242), most of which
require a 3-constituent look-ahead buffer in the Marcus parser,
can be processed immediately (without deferring
decisions, without quasi-parallel processing or later back-
tracking) in the IL version:

(239)   Is the block sitting in the box?
        Is the block sitting in the box red?

        Cu la bloko sidas ie-en la skatolo?
        Cu la bloko, sidanta ie-en la skatolo, estas ruga?

(240)   Have the students who missed the exam take the makeup today.
        Have the students who missed the exam taken the makeup today?

        Igu la studentojn kiuj maltrafis la ekzamenon, fari la
        kompensan ekzamenon hodiau.
        Cu hodiau la studentoj kiuj maltrafis la ekzamenon, faris la
        kompensan ekzamenon?

(241)   John had left yesterday.
        John had a book.

        Hierau 'John'-o foriris.
        'John'-o havis libron.

(242)   Who do you want to give a book to tomorrow?
        Who do you want to give a book to Sue?

        Al kiu-do vi volas doni libron morgau?
        De kiu-do vi volas, ke tiu donu al 'Sue'-o libron?

(243)   That deer ate everything in my garden surprised me.

        Ke cervoj mangis cion ie-en mia gardeno, surprizis min.

[The particle 'cu' in 239 and 240 is an obligatory introducer
and marker for yes/no questions. The suffix '-do' in 242
serves to distinguish a wh-question from a relative clause.]

It should be noted that the design of the IL, apart from
guaranteeing structurally unambiguous sentences, also
deliberately includes provisions for limiting look-ahead
needs. An example of this is the elliptic noun coordination
explained in 2.4.5 [p. 74-75]. Instead of:

(244a) La malnova _kaj nova produkto.

one could have adopted a different model (similar to what is
sometimes used in common Esperanto [see also PAG 128.I.2]),
namely a variant of the adjective-noun number disagreement:

(244b) La malnova kaj nova produktoj.

but that would have increased the IL-parser's look-ahead load.
Another example is the use of subordinate conjunctions that
have been made distinct from their corresponding preposi-

tions, by combining them with the dependent clause introducer
'ke' [cf. 2.5.3]:

	Preposition:		Subjunction:	
	dum	(during)	dum ke	(while)
	gis	(until)	gis ke	(until)
	post	(after)	post ke	(after)

Without this distinction, parsing of a constituent would
have to proceed until either a right PP parenthesis (e.g. a
comma) or a finite verb (of the subjunct clause) was found,
in order to determine the nature of the constituent:

(245)   Dum la kunveno pri la decido de la komisiono pri
        agrikulturo....
        (During the meeting on the decision of the committee
        on agriculture...)

In parsing the IL, a modest amount of look-ahead remains
required, for scanning over adverbs and for determining the
right bracket of a constituent in several cases (e.g. NPs
not followed by a special element such as an extra space,
serving as an explicit separator).
Also, notorious backtracking causers as:

        - PP-ambiguity

        - reduced relative clauses

        - compound noun strings (noun-noun modification)

        - conjunction scope ambiguity

        - modifier scope ambiguity

are absent in the IL (as has been demonstrated in the previous
sections of this chapter). This means that the backtracking as
well as the look-ahead load on the IL-parser is relatively
small. Moreover, DLT uses quasi-parallel processing instead of
backtracking: after each IL-word received by it (at a trans-
mission rate which is significantly slow compared to processor
speed), the parser will utilize the relative abundance of
available processing time till the arrival of the next word
[cf. III-4.2.1, parsing "on the fly']. This is the IL-parser's
deterministic realization of a non-deterministic system (the
ATN), and prevents a processing peak near the end of the
sentence. It makes the IL-parsing speed no point of concern.

An argument against the use of ATNs would be their lack
of ability to process incorrect sentences. But the IL-
recognizer only has to reject these (not to recover them),

and the IL-parser will never receive them [cf. III-4.1].

The above arguments for chosing ATN all underline that it is
good enough for the job. One might question whether it is not
too heavy a tool in this particular case (the IL). Rather than
by theoretical interest, our work at this point has primarily
been guided by the practical goal of building a multilingual
MT system. Out of practical considerations, we prefer to stay
on the safe side and adopt a powerful method as ATN [Bates,
1978], with which ample experience has been gained now, and
which keeps having good software support prospects in the form
of advanced programming languages [see also III-4.2.3.5a].

3.2. The ATNs making up the IL grammar.

As part of the feasibility study, ten ATNs have been drawn
to define the IL grammar, eight of them corresponding to IL
syntactic categories [IV-1.4], the other two (COM and ADO)
being mere subnetworks which resulted from practical ATN
design considerations. The set of functions of the ATN
formalism and the state labelling convention are those
described in [Bates, 1978]. The arc annotation used is largely
DLT-specific and has been defined on p. 7 of the Appendix.

Fig. IV-6 shows the ATNs for FADJ and CADJ. All other ATNs
(for GEN, INF, COM, NP, ADJ1, ADJ2, REL and ADOC) can be found
in the Appendix [pp. 8-15]. Note that some syntactic categories
(LADJ, DEP, PP) do not correspond with a separate ATN.

The ATNs presented in this report have not yet been tested out
by computer implementation. Moreover, they are not complete in
every detail [section VII-2.b lists the required completion
activities]. Conditions on the arcs have been indicated very
concisely, often referring to a specific list of function
words. Tree-building actions along the arcs still fail.
Coordination and ellipsis have been covered for the NP, but
not in the other ATNs.

The agreement between the set of ATNs and the descriptive
definition of the IL grammar [as given in section IV-2] has
carefully been checked. Nevertheless, further desk-checking
will be fruitful [readers are encouraged to report any
inconsistencies or deficiencies they might detect].

As for the DLT process steps in which these ATNs will be
employed, it should be noted that they have primarily been
drawn with the IL-parser in mind. The IL-parser will for
instance NOT apply any checks on the nesting depth limita-
tions, as the IL-recognizer does. Only in one case (the ATN
for ADJ1), restrictions on embedded NPs have been indicated
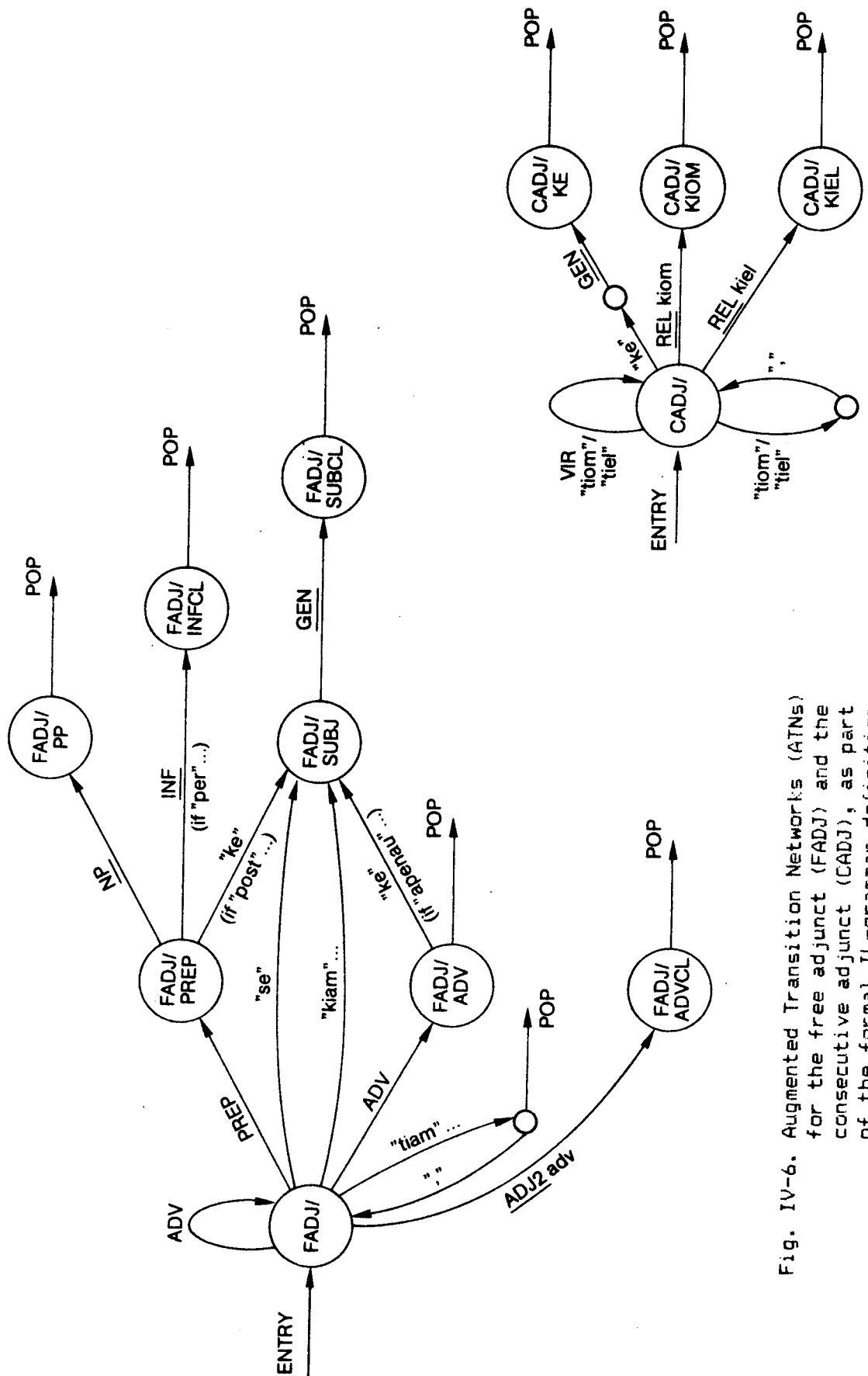[by means of an apostrophe, p. 12 of the Appendix].

Fig. IV-6. Augmented Transition Networks (ATNs) for the free adjunct (FADJ) and the consecutive adjunct (CADJ), as part of the formal IL-grammar definition.
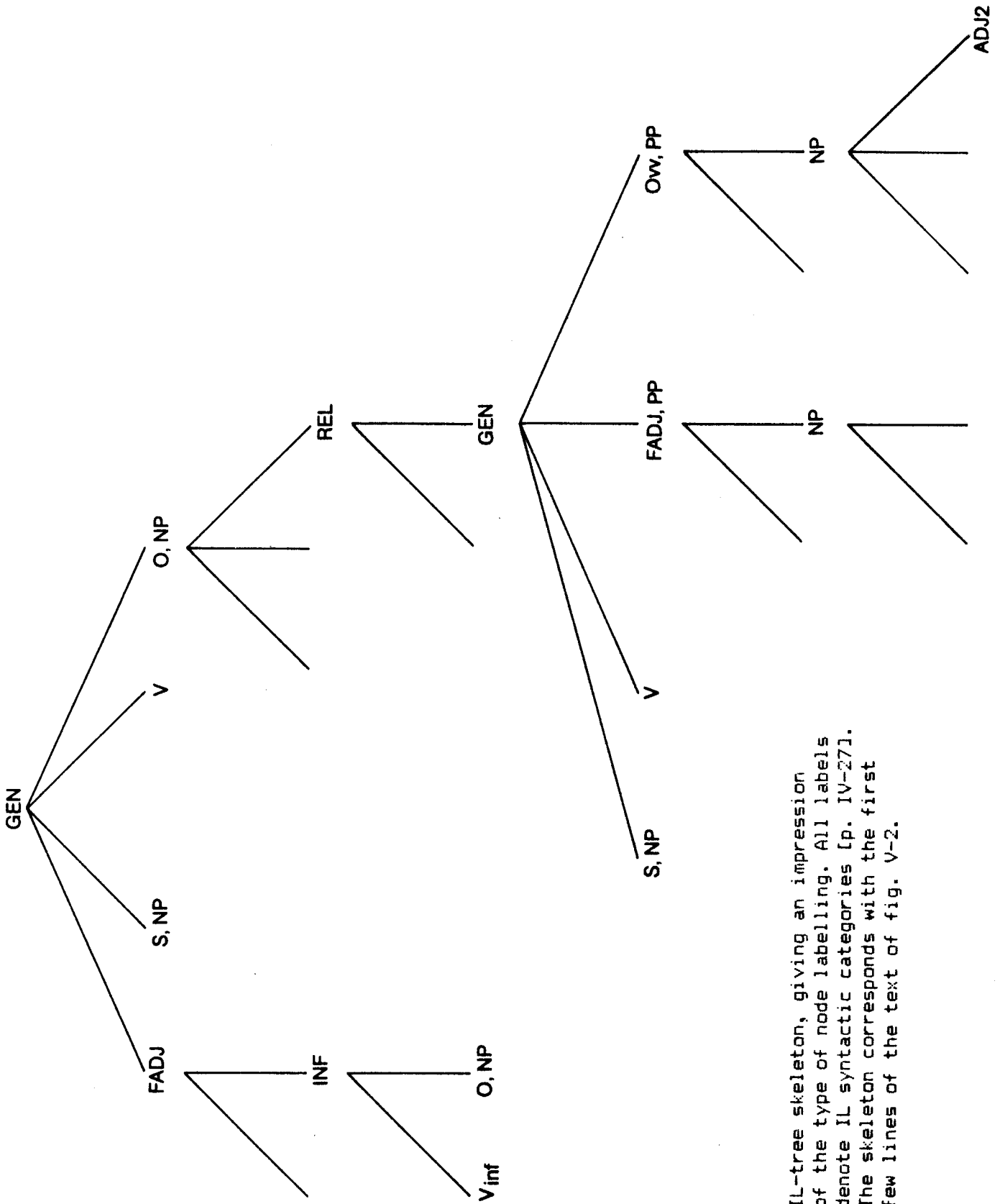
Fig. IV-7. IL-tree skeleton, giving an impression
of the type of node labelling. All labels
denote IL syntactic categories [p. IV-27].
The skeleton corresponds with the first
few lines of the text of fig. V-2.

3.3. Structure and labelling of trees.

IL-trees result from Step 1 and Step 5 of the DLT process [see
fig. III-9]. They form an intermediate stage and act as interface
between the SL- and TL-oriented steps on the one, and the
extended IL-kernel on the other hand. As they are gradually
being built during Step 1, they will carry SL as well as IL
words at their leaves [see Chapter III, figs. 11-14].

The structure of these trees reflects the IL syntactic
categories [defined in IV-1.4] and the IL canonical patterns
[pp. IV-33,34], which in turn can be derived from the ATNs.
Note that IL-trees resulting from Step 1 are unordered trees:
they conform to the IL canonical patterns except for the order
of constituents. The output of Step 5 consists of ordered
trees, fully in line with the canonical patterns and ATNs.

As an example of an unordered IL-tree, fig. IV-7 shows an
active general clause (corresponding to pattern 1.1 of Table
IV-2) at the top. The FADJ consists of an infinitival clause
(preceded by a preposition), in agreement with one of the
paths through the ATN of fig. IV-6. On its turn, the INF
constituent is realized in accordance with pattern 2.1 of
Table IV-2, and so on. Each subtree corresponds to a syntactic
constituent, and often to the call (or 'PUSH') of an ATN
subnetwork.

Apart from ordering of constituents, there is another issue
which marks the difference between actual IL-trees and canonical
patterns: in an IL-tree, whether ordered or unordered, the
subtrees will only contain those optional constituents that
are actually present. This may seem trivial, but we aim at the
need for an effective and economic representation: for certain
categories, such as the NP, the number of optional constituents
is so large, that an entirely slot- or frame-oriented solution
(as is often applied in AI research experiments) seems
unattractive. In DLT, the IL-trees - as mentioned above -
are intermediate stations on the way to and from a compact
linear string (BCE), which makes the relevance of economic
tree structures even more pronounced.
A partly slot-oriented solution (suggested by the NP-subtrees
of figs. IV-3,4,5), based at the most frequent actual realiza-
tion patterns of a constituent, could however be optimal.

As for the labelling of the tree nodes, we distinguish two
types of labels: one for the syntagmatic categories (GEN, INF,
NP etc.) and one for the syntactic function categories (S, O,
Ovv, FADJ etc.) [see also p. IV-27].

As with the ATNs, the IL-tree structure and node labelling have
not yet been worked out in every detail [for planned finaliza-
tion see VII-2.c].

The handling of modifier scope and coordination in the IL tree
structure has been illustrated for NPs [IV-2.4.5, figs. IV-3,4,5;
the branches with "&" will technically be realized by an extra
set of pointers, i.e. an additional tree dimension]. As shown,
coordination as well as multiple premodifying cause levels to
be added to the trees there. This mechanism is likely to be
extended for coordination outside NPs.

3.4. Insertion of disambiguating elements.

According to the process sequence displayed in fig. III-9, the
unordered IL trees produced by Step 1 must subsequently be
ordered (Step 2) and converted to IL proper, i.e. linear strings
(Step 3).

Syntactic unambiguity in the IL tree structure is one thing,
in the linearized BCE strings it is another. The trees are by
nature an extra-linguistic device, relying on artifacts such
as tree structure and node labels. The strings must be kept as
readable as Esperanto, a requirement which is one of the key-
stones of DLT.
As far as extralingual elements are brought into BCE, they must
therefore be modest and unobtrusive. This section deals with
the EXTRA SPACE, announced at the beginning of this chapter
[p. IV-5] and used in various examples (indicated by an under-
score).

Throughout the IL, a number of linguistic devices contribute
to ambiguity prevention (as illustrated in section IV-1, and
systematically covered in IV-2), e.g.:

    - explicit grammatical endings

    - accusative

    - precise prepositions

    - canonical word order

However, these linguistic devices are not watertight. In
particular, they do NOT cover the structural ambiguity
connected with coordination and modifier scope. Also, one
should be aware of the additional necessity to prevent
excessive look-ahead requirements [section 3.1 specifies the
use of ATNs with limited look-ahead].

We therefore have to take recourse to a more formal device,
which we can characterize as SEPARATOR insertion. Before we
discuss the formal separators, let us briefly mention two
particles that can be considered as 'pseudo-linguistic'
separators: 'le' [example 88,89] and 'je' [examples 105,106],

which on the one hand have only been included into the IL to
serve as 'left brackets', on the other hand resemble existing
particles in Esperanto and other languages, in such a way that
they function as mnemonics for the human IL inspector.

As a formal separator, the space or blank has been chosen.
From the point of view of IL-parsing, it is fully arbitrary by
which particular sign the separator is realized, as long as
the choice does not introduce new ambiguities. From the
compactness viewpoint however, the choice is important. The
space is the most compact sign in the BCE coding scheme [cf.
p. IV-5 and section IV.5], and EXTRA SPACES serving as formal
separators are at the same time unobtrusive extralingual addi-
tions to BCE. The mapping of these extra spaces onto other
devices (underscore, indentation) as a display option is again
another matter [see III-5.2].

Important to notice is that one formal separator suffices in
BCE. This one separator, the EXTRA SPACE, is used to perform
three distinct separating functions:

> I. Level popping ('POP SPACE').
> The separator enforces the IL-parser to pop out of the
> current ATN [see for instance the bottom left arc of
> on p. 11, Appendix], i.e. to climb back to a higher
> level in the IL-tree. This includes coordination and
> modifier scope, because of the way these are
> implemented in the IL tree structure.
> Enforced popping out of an ATN is required in case of
> Absent Trailing Optionals (ATROPs) in the actual
> realization of an IL constituent. The separator can
> be considered as a right bracket.
>
> II. Skipping embedded optionals ('SKIP SPACE').
> The separator serves as a skip mark, somewhere in the
> middle of a constituent, to guide the IL-parser.
>
> III. Ellipsis ('ELLIPSIS SPACE').
> This takes place with elliptic noun coordination, as
> shown in example 221 [p. IV-76]. Here, a separator is
> needed to prevent a coordinated-adjectives interpreta-
> tion, and is placed immediately before the conjunction.
> Similarly, an undue coordinated-PPs interpretation can
> be prevented by an ellipsis space immediately after
> the conjunction.

The above gives an impression of the significance of separators
during IL-parsing, i.e. string-to-tree conversion (Step 5). What
remains to be answered is how separators are inserted at the
SL-side of the extended IL-kernel [fig. III-9]. We will briefly
indicate one of the algorithms for automatic separator insertion

[in which 'Step 2' and 'Step 3' always refer to the overall
translation process sequence defined in III-4.1 (fig. III-9)]:

Automatic POP SPACE insertion:

- during Step 2, for each IL subtree, the corresponding ATN
  is traversed, as a check of the actual subtree composition
  against the IL-grammar; at the same time, the following
  operations are performed:

    * assignment of sequence numbers to the subtree's
      constituents (as part of tree ordering);

    * the generation of an ATROP-list, to be appended
      (in the form of a compact bit map) to the
      subtree's closing bracket;

- during Step 3, the IL tree is processed LR depth-first;
  whenever climbing upwards, i.e. after each (set of)
  closing brackets, the sequence of operations is:

    * match the type (syntactic category) of the next
      constituent - or any of its actual descendants,
      down to and including its leftmost terminal - to
      the ATROP-list of the closing bracket;

    * if match is positive, then insert POP SPACE;

    * if more than 1 closing bracket, then repeat
      matching procedure for each bracket (i.e. each
      ATROP-list) [Notice that sequences of two or more
      POP SPACES can occur, as shown for instance in
      example 214.] .

In contrast to pop spaces, skip spaces are automatically
inserted into the IL tree itself, along with tree ordering
(Step 2), using a device similar to the ATROP-lists.

Thus, the insertion of extralingual disambiguating elements
is arranged NOT as a collection of ad hoc measures for those
cases of syntactic ambiguity that happen to be known, but as
a systematic safeguard and watertight formalism inside the IL-
kernel.

4. The lexical component.

One of the arguments for adopting an Esperanto-based IL
has been the availability of a lexical as well as a
syntactic component [cf. III-1.3]. This section gives an
impression and overview of lexical aspects of existing
Esperanto, as well as related IL-design issues. The DLT
strategy for lexicographic consolidation and extension
is outlined in the final subsection.

4.1. Morphemic structure and word formation.

4.1.1. Esperanto as an isolating language.

The origin of its words roots (approx. 60% from Romance,
30% from Germanic and 10% from Slavonic languages) makes
Esperanto a semi-artificial language, resembling European
languages. Also its sentential syntax generally follows
European (often Slavonic) models.
However, at a less superficial level of structure,
Esperanto has been shown [Piron, 1979; Pennacchietti,1982]
to be more similar to the so-called isolating type of
languages, in particular Chinese.

The terms 'inflected', 'agglutinative' and 'isolating' are
used to classify languages in a rather traditional way. A
structural criterion for classification is the degree to
which morphemes can change their form. Morphemes are the
'atomic' elements of a language, and include lexical
roots, affixes (prefixes, infixes, suffixes) and
grammatical endings.

The highest amount of variable morphemes is found in
inflected languages, because these may alter their roots
both in case of derivation and when their grammatical
function changes. This is true of Indo-European languages
as a whole, e.g.:

sell	denken	bref	хожу
sold	dachte	brièveté	ходить
sale	Gedanke		хождение

Examples of agglutinative languages are Hungarian and
Turkish. Here the root is invariable, only prefixes and
suffixes alter their forms to some extent; e.g. the
Hungarian suffix '-ság' alternates with '-ség' according
to vocal harmony:

szabad	(free)	üde	(fresh)
szabadság	(freedom)	üdeség	(freshness)

Chinese is a well-known example of an isolating language:
not only its roots, but also its prefixes and suffixes
remain unaltered. All morphemes have an invariable form:

        gongye   (industry)   jiandan   (simple)
        gongyehua (industrialization) jiandanhua (simplification)

The term 'isolating' refers to the possibility of
independent use of affixes. In Chinese however, there is
often a meaning difference between the independent
(isolated) occurrence of a morphem and its use as an
affix, e.g.:

        jia (family)   zhengzhijia (politician)

The semi-artificial language Esperanto is characterized by
complete invariance of its morphems [Wells, 1978], with
respect to morphem meaning as well as morphem form, and
regardless whether a morphem occurs as affix or in
isolation. Thus, Esperanto appears to be even more
consistently isolating than Chinese.
On the other hand, Esperanto is marked by the obligatory,
explicit and systematic expression of grammatical
function, which is an important difference with Chinese
(where for instance no formal active-passive distinction
is made when the context is clear) as well as with other
languages (in Turkish, a plural noun ending is considered
superfluous in the presence of a numeral).

The absolute invariance of morphems and the regularity of
grammatical structure in Esperanto dominate word-formation
principles in this language (besides being the basis for a
compaction scheme [see section IV-5]).

4.1.2. Advantages and limitations of Esperanto word-formation.

The best way to discuss Esperanto word-formation is
obviously to make a comparison with other languages in a
systematic way. To that end we will use some notions from
information theory.
Lyons [1969] points out that the basic principle of this
theory lies in the fact that information value is inversely
proportional to probability of occurrence. This means that
the information value of e.g. a word which is fully
predictable ('I want .. help you') is zero. This word is
then fully redundant.
Redundancy plays a very important role in information and
communication: every medium that is used for information
transfer, regardless of what kind, has some unpredictable
physical 'noise' in its channel, which partly deforms the
message and thus leads to information loss ('channel' and

'noise' are used, as will be clear, in a wide sense, also
referring to written texts).
The optimal system for a certain channel is the one in
which there is exactly sufficient redundancy to enable the
receiver to restore the information that got lost by the
channel-noise. We will refer to this as the 'redundancy
principle'.

On the other hand, there is the 'principle of language
convenience' [Wüster 1966; 1978] also named 'the principle of
the least effort' [Lyons, 1979], as it allows the language
user to express himself with as little effort as possible.

These two principles are mutually exclusive and complementary.
Wüster mentions both principles in the context of
Esperanto. Blanke [1982] too refers to them. It is
interesting how Blanke relates Esperanto word-formation to
these principles. He refers to an investigation on
redundancy on the graphematic level of Dressler [1972] who
points out that Esperanto has sufficient redundancy on
this level (e = 0,1151). This appears to be a little less
than the redundancy of German graphemes (e = 0,15).
Blanke mentions the shortening of Esperanto words that are
(too) well explicitized as to their meaning, e.g.:

        ordigi   :   to put in order.
        ordi   :   to put in order.

        pasinteco: past.
        pasinto   :   past.

        estonteco: future.
        estonto   :   future.

Lyons [1969] refers in this connection to Zipf's law,
which states that language is more efficient (in the sense
of information theory) when the length of its elements is
inversely proportional to their predictability. The validity
of this principle for natural languages can be observed in
daily life. In Esperanto, however, this law is applied even
more generally and systematically, as the various examples in
this report illustrate. Also, the adherence to this principle
is reinforced by the language user's will to be understood, a
will that is, according to Blanke, strongly developed among
the users of Esperanto.
Blanke shows that both principles (the redundancy principle
and the principle of language convenience) are very pronounced
in the word-formation opposition:

        verbal stem + '-o' / '-aĵo' vs.
        verbal stem + '-o' / '-ado' ;

('-o' is the substantivizing suffix; '-ado' is the verb
nominalizing suffix; '-aĵo' indicates 'something
concrete'). On the one hand these three forms can function
as synonyms, on the other hand there are certain
contextual differences in meaning shades, e.g.:

    manĝi   (to eat)   :   manĝo, manĝado, manĝaĵo.
    skribi   (to write) :   skribo, skribado, skribaĵo.

Blanke compares these forms to their German equivalents
'Essen'; 'Schrift', to which one cannot simply attach a
suffix to explicitize the meaning more precisely. In
Esperanto this is readily possible and usual.

Although the Esperanto word-formation represents a system
that is regular in principle, it has of course its "gaps":
e.g. with 'trinki' (to drink), 'trinko' cannot be the
synonym of 'trinkaĵo', only of 'trinkado', whereas with
'organizi' (to organize) 'organizo' is only synonymous to
'organizaĵo'.
Some theoretically possible formations never became a lexeme
in practice, e.g. *'liberaĵo', probably because 'libero' and
'libereco' both represent the meaning 'freedom' sufficiently.
This is what one could call - analogous to phonetics - an
accidental gap.
In other cases there are semantic constraints to form
theoretically possible words. The suffix '-ino' for
instance indicates a female: 'profesorino' (woman
professor), 'posedantino' (woman possessor); there are,
however, one or two words that have this feature
inherently: 'amazono' (amazon), 'matrono' (matron). The
same applies to the indication of transitivity by means of
the suffix '-igi': 'vaporigi' (to vaporize), 'formigi' (to
take shape), whereas verbs such as 'ĉesi' (to cease),
'eksplodi' (to explode) are intransitive by themselves
already. These kinds of irregularities are (according to
Blanke) caused by the influence of other languages and analogy
with current word-formation models.

In a number of cases, the functioning of the principles of
redundancy and language convenience appears to be exactly
the explanation for accidental gaps in Esperanto.
The Esperanto Academy recognized the working of these
principles as typical of Esperanto word-formation [Aktoj,
1968]; we quote: "In the word-syntagm must be infixed all
morphems that are necessary, but no more than are sufficient,
to motify as clearly as possible and with sufficient
redundancy the meaning to be represented by the word." .

As shown, Esperanto word-formation is based on the use of
invariable morphemes. To give an idea of the combinations
of affixes with a root ('bor-': 'drill-'), we present an

example from [Wüster, 1978]:

```
'-ebl-':  borebla    : what can be drilled (through).
'-ind-':  borinda    : what is worth to be drilled.
'-em-' :  borema     : what is inclined to drill.
'-ig-' :  borigi     : to have drilled.
'-iĝ-' :  boriĝi     : to be drilled.
'-ist-':  boristo    : who drills professionally.
'-ul-' :  borulo     : who often drills.
'-il-' :  borilo     : drilling-instrument.
'-ej-' :  borejo     : place where one drills.
       :  borilejo   : place where one keeps drills.
'-uj-' :  borilujo   : object containing a number of drills.
'-ing-':  borilingo  : drill-chuck.
```

The following two examples show the Esperanto word-
formation for causative or factitive verbs, as opposed to
some natural languages [Wüster 1978]:

Esp.	German:	English:	French:	Russian:
egali	gleichen	to equal	égaler	равняться
egaligi	angleichen	to equal(ise)	égal(is)er	уравнивать
trinki	trinken	to drink	boire	пить
trinkigi	tränken	to make drink	faire boire	поить

The capacity to systematically combine morphemes appears very
clearly from the next example, in which various endings and
suffixes are appended to the prepositions 'per' (by means of,
with, through), 'anstatau', 'super', 'kontrau' and 'kun':

```
perado    : mediation
peri      : to mediate
pera      : indirect
perilo    : means
peristo   : mediator by profession
perulo    : mediator who often mediates
```

```
la anstatauado : the substituting
la superado    : the exceeding
la kontrauulo  : the adversary
la kunulo      : the companion
```

A very important feature is the discussed independent use
of morphemes, again typical of isolating languages, e.g.:

```
aĵo : thing
eco : quality
ujo : container; receptacle
ulo : fellow
```

The principles of language convenience and redundancy that are so
closely related to the functioning of Esperanto word-formation
are complementary, and therefore the fact that a rather strong
accent lies on the convenience, must lead to certain limitations
in redundancy. Blanke [1982] points out that in comparison with
German some meaning shades can indeed not be expressed:

schiessen	Schuss	Schiesserei	Geschiesse	das Schiessen
pafi	pafo	pafado	pafado	pafado

suchen	Suche	Sucherei	Gesuche	das Suchen
serĉi	serĉo	serĉado	serĉado	serĉado

fahren	Fahrt	Fahrerei	Gefahre	das Fahren
veturi	veturo	veturado	veturado	veturado

lachen	Lachen	Lacherei	Gelache	das Lachen
ridi	rido	ridado	ridado	ridado

As a conclusion, one may state that Esperanto has proved to be an
ingenious compromise of regularity and expressiveness, as witness
nearly a hundred years of successful experience. This is also
confirmed by the following table, concerning the word-formation
productivity in a number of languages, compared by Verloren van
Themaat [see Blanke, 1979]. For each of 5 languages, the table
gives the percentages of word formations that can be analyzed
transparantly as to the meanings of their composing elements:

French	:	0.16%
English	:	0.45%
Russian	:	0.61%
German	:	2.2 %
Esperanto	:	5.2 %

4.2. Word compounding.

Word compounding involves the combining of two or more word
roots. In general, word compounds may orthographically exists
of one word (typically in German), or of several words separated
by spaces (typically in English). Word compounding can be
regarded as a special case of word formation. This is especially
true of Esperanto: Wüster [1978] points out that in general,
derivation and flexion can be understood as a combination of
subordinate and superordinate terms, just as in the case of

word compounding: the rules are the same. In Esperanto, this
is shown very clearly by the fact that some suffixes are even
used regularly as roots (typical of isolating languages [see
4.1]).
Wüster mentions five kinds of relations between the terms:

disjunction :	hindoeuropeano	/ Indo-European
multiplication:	vatsekundo	/ watt-second
conjunction :	ingeniero-komercisto	/ engineer-merchant
determination :	postmarko	/ (postage) stamp
conversion :	-	/ -

[Examples of the fifth type in English and German: 'under-
ground' instead of 'underground line', 'Ober' instead of
'Oberkellner'. In Esperanto, such relations are avoided
because of their bad 'transparency'].
Of course, Esperanto too has untransparent compounds: as more
than once stressed by leading Esperantologists, including
Zamenhof himself, language is simply not such a thing as
mathematics: in human language, semantics always causes a lot
of unevitable irregularities. The particular advantage of
Esperanto lies in the avoidance of irregularities wherever it
is possible [Waringhien 1980].

Semantic irregularities exist in Esperanto as they do in
other languages. Wüster [1968, 1978] indicates that the
meaning of a compound cannot be concluded from the elementary
meanings of its composing elements. E.g. 'paperskatolo' is
analyzed as 'skatolo el papero' (box out of paper'), whereas
'paperkorbo', formally having exactly the same structure,
means 'korbo por papero' (basket for paper). As another example
of a compound with two ways of analyzing, Blanke [1979] mentions
'urbodomo', being either 'domo de urbo' (cf. Dutch: stadhuis)
or 'domo en urbo' (cf. Dutch: stadshuis).

The composing elements of a compound can be nominal, but
also verbal or adjectival. Esperanto uses the synthetic
way of compounding as in German and English, and as opposed to
French. In the case of synthetic compounding, the superordinate
term is the one on the right;the left one functions as its
determiner. Semantics again may traverse this general rule
[Goninaz 1975].
Blanke [1982] points out that word-formation devices in German
do not prescribe how to analyze compounds. In comparison with
this, analysis of Esperanto compounds follows certain rules.
In Esperanto nominal-compounds, the superordinate nominal term
substantivizes the subordinate one, and their relation can be
expressed by 'de'. This preposition can in turn be replaced by
other, more exact prepositions ('el';'pri';'por') or by 'kaj'
in the case of a copulative compound. Some examples:

```
homkoro           : koro de homo          (human heart).
ŝtonponto         : ponto el ŝtono        (stone bridge).
skribmaŝino       : maŝino por skribi     (typewriter).
scienculpedagogo  : scienculo kaj pedagogo (scientist-
                                            pedagogue).
```

The use of connecting elements tends to be somewhat
inconsistent in some languages. In Esperanto, their use is
strict: the connecting element, always one of the very limited
set of grammatical endings, can only be left out if its use is
pleonastic, e.g. 'rapidmezurilo' (tachymeter) vs.
'rapidomezurilo' (tachometer) [Wüster 1978]. In German,
these elements have rather different forms ('-s', '-es',
'-en', '-ens', '-e', etc.).

An undeniable advantage of Esperanto word-compounding is
that its isolating character [see 4.1] offers sufficient
combining possibilities to match other languages; e.g.
'ĝis' and 'nun' can easily be combined to 'ĝisnuna',
corresponding to the German 'bisherig'. Esperantists often
point to the fact that the formally unrestricted morphem
combining possibilities yield words and even concepts,
unknown in other languages. Though certainly interesting,
this seems less important in the context of a system
primarily aimed at translation.
The isolating character of Esperanto also leaves the
elements of a compound absolutely unchanged [Blanke 1982]:

```
Esp.:     blua + okulo + a    -->    bluokula
German:   blaues Auge  + "    -->    blauäugig
```

As Blanke [1979] states, in German, on the contrary, a lot
of formally possible compounding devices are highly
defective, whereas Esperanto has much wider possibilities
to fill in such formal devices.

Wüster [1978] gives an example of an alternative
construction instead of a compound: 'karbona dioksido' is
preferred to 'karbacido' (carbonic acid), because the
latter is not transparent. This is a nice example to show
that in a lot of cases combinations of adjectives and
nouns are equivalent to compounds; the same is true of
combinations of a noun, a preposition and a noun: 'Ligo de
Popoloj' vs. ' Popolara Ligo' (League of Nations) [Wüster
1978].
In fact, this lexical equivalence goes through language
boundaries and represents an important area of contrastive
linguistics: a one-word compound in one language is often a
combination of the type "adjective + noun" or "noun + noun"
in another language, e.g.:

```
reciprokeca_transakcio: transaction réciproque.
                      : Gegenseitigkeitsgeschäft.
```

                               : reprocity transaction.

kursorestariĝo   : relevement des cours.
                               : Kurserholung / Erholung der Kurse.
                               : recovery in prices.

salajro pro disponeblo: salaire d'attente.
                               : Wartegeld.
                               : retaining wage.

In a lot of cases, often involving relational adjectives
[see 2.4.4], combinations of the type "adjective + noun"
will appear to be coined terms, included as such in the
lexicon and indicated in the IL by an underscore.
All the three above types of compounding are very
productive and frequent, especially in the type of text
envisaged for DLT [see chapter V].

The importance but also the difficulty of analyzing
compounds has been recognized already in an ear-ly stage
of MT history ("... not even analyzable, let alone
mechanizable" [Rhodes, 1967]) and keeps being confirmed
[Maas, 1983].
DLT will therefore have to rely on the inclusion of as
much as possible compounds in its lexicon entries [cf.
III-3.2.5], where they can be related to their various
(compound or non-compound) equivalents in other languages.
This will prevent that the operator at the SL-terminal
were overloaded with lexical disambiguation requests
[III-4.2.4] for compounds that cannot be analyzed
automatically.

Of course, it will never be possible to have an exhaustive
coverage of compounds and their equivalents in the dictionary:
the continuing productivity of the compounding device prevents
this in the first place. Therefore, compounds that are not
included in the dictionary should be interpreted according to
the most frequent structures known from compounds analysis
research (e.g. Reifler's algorithm [Reifler, 1955; Brandt
Corstius, 1978: 128-134]). These dominating structures will
thus serve as "default" interpretations for compounds unknown
to the DLT system. Its interactive disambiguation dialogue
will then enable the human attendant to "override" the default
choice, and select a different interpretation.

Beside this apparently required pragmatic strategy for the
immediate future, strongly based on lexicon comprehensiveness
and the presence of human intelligence as a last resource, the
fact that the structure of compounds is more tansparent in
Esperanto than in other languages offers a good starting-point
to gradually improve and extend automatic-analysis algorithms
as part of a long-term strategy.

4.3. Idiom.

Before we comment on the properties of Esperanto and the IL
with regard to this subject, let us first make several remarks
on its significance to MT in general and DLT in particular.

An MT system's ability to handle idiom first of all means an
increase in the size of its dictionaries. As with compounds,
collocations and technical terms, idiom too can be provided
for in the translation system, at the cost of a heavier and
more expensive lexicon.
Of course, there is a difference: unlike collocations and
technical terms (which are strings of contiguous and
unvariable words), the elements of idioms may be interspersed
over the sentence, and one of them (usually a verb or pronoun)
may be variable [Hundt, 1982].

A complication arises when the idiom is a metaphor, which
could have a literal interpretation as well [Bar-Hillel, 1965],
such as 'to ring a bell' [Hundt, 1982] or 'to be sent to
Coventry' [Knowles, 1982]. In DLT, these idioms will
presumably be marked in the dictionary in such a way, that
their occurrence induces an interactive disambiguation
dialogue. The difficult choice between a literal and a
figurative sense can then be made by human intelligence and
understanding. Unfortunately, this comes down to an increased
work-load on the operator at the DLT SL-terminal.

Next to the interpretation problem at the SL side, there is
the problem of properly rendering an idiom at the TL side.
This does not only require the presence of a corresponding
expression in the TL, but also in the IL.

The text type aimed at for DLT [see Chapter V] is, among
other things, characterized by a low content of idioms. Even
if idioms occur in the source text, envisaged DLT applications
make a correct, non-idiomatic translation certainly acceptable,
as in the following historical example (taken from [Bar-Hillel,
1955], apart - of course - from the IL version):

(248a) Truman declared that the whole affair was a red herring.

(248b) 'Truman'-o deklaris ke la tuta okazaĵo estas erarigilo.

(248c) Truman erklärte dass die ganze Geschichte eine Finte war.

So in DLT's English-IL dictionary, the idiom 'red herring' will
correspond with the single-word IL-translation 'erarigilo'.
There will be nothing idiomatical on the other side: the IL-
German lexicon entry will give the German "Finte" as transla-
tion for 'erarigilo.

A recent account on the translation of metaphor is given
by Newmark [1981: 84-96]. It contains a warning for metaphors
that are nearly universal, but not quite, and therefore remain
potential traps, e.g.:

(249) To the letter.
      Au pied de la lettre.

(250) Water depth.
      Hauteur de l'eau.

In his article on MT dictionaries, Knowles [1982] mentions the
following categories, which all have a high degree of bonding
or predictability (once commenced):

proverbs:	"live and learn",
idioms:	"to hit below the belt",
phraseological concretions:	"to win one's spurs",
units such as:	"once and for all",
technical terms:	"small intestine".

[The last category shows an English multi-word compound, which
corresponds with a one-word compound in German ("Dünndarm"), a
phenomenon discussed in section 4.2].

As for the representational power of the IL and the underlying
Esperanto, one must make certain reservations with regard to
idiom. The practical problem is whether the available stock of
existing Esperanto idioms is adequate to match the unceasing
supply of modern idioms in French, German, English and other
languages. This is largely a codification problem. Published
materials covering Esperanto idioms appear to be old-fashioned
(such as [Minor, 1975]) or scattered and intermingled with
general vocabularies (such as [Woolf, 1969]).
Compact IL constructions could be designed for those cases,
where common Esperanto does not have a proper idiomatic
expression. But there is another problem:
The theoretical problem is that any idiomatic IL constructions
must be so unambiguously, and that an incidental literal
interpretation of the same SL original should give rise to a
different IL version. A solution for this could possibly be
found in the utilization of the underscore for another IL-
related purpose: the indication of idiomatic bonding between
words (even if not all the words of an idiom are contiguous).

4.4. Homonymy and polysemy.

In Esperanto, homonymy is confined to a remarkably limited
number of words. Givoje [1979] asserts that e.g. under the 'A'
in the 'Plena Vortaro' only 13 real homonyms are found. As one
of the reasons for this, Wüster [1978] gives the fact that
Esperanto derivation mechanisms are ambiguity-free by principle.
However, a homonym can be the result of a transfer of meaning,
e.g.:

'bobeno' :   1.bobbin.
                   2.spark coil.
'etero'   :   1.elastic fluid permeating space.
                   2. airless volatile liquid.

A proposal by defenders of unambiguity, to introduce alternative
technical terminology in such cases, remained without success.
Wüster points out that unambiguity has its limits and is not
necessary for every word (cf. the relativity of the ambiguity
concept, as repeatedly underlined in Chapter III, and the
comments on polysemy, later in this section).
Nevertheless, Esperanto has much less homonyms than other
languages. Esperanto often distinguishes two meanings by
means of orthography, e.g.:

intereso: interest   (German: Interesse)
interezo: interest   (German: Zins)
banko   :   bank   (German: Bank (pl. Bänke))
benko   :   bench   (German: Bank (pl. Banken))
pesi   :   weigh   (find the weight of)
pezi   :   weigh   (have the weight of)
proceso :   lawsuit
procezo :   process

[The difference in orthography signal the use of different
morphemes.]
In the IL, remainig cases of common-Esperanto homonymy can
be solved by creating two different forms with a strict
1-to-1 relation to the alternative meanings, e.g.:

Esp.: kubo   : 1. cube   IL: kubo   : cube
               2. die   : ludkubo   : die

    : pendoli: 1. to swing   : pendoli   : to swing
               2. to commute   : pendolveturi: to commute
    : trepano: 1. trepan   : kranitrepano: trepan
               2. drill   : mintrepano : drill

    : marko : 1. brand   : fabrikmarko : brand
               2. stamp   : poŝtmarko   : stamp

In some other cases two forms exist, but the two meanings
are not neatly divided over these forms:

Esp.: talio: 1. waist.   IL: talio: waist.
                2. thallium.

    : taliumo: thallium.   : taliumo: thallium.

A similar solution can be applied in case of incorrect or
colloquial use:

Esp.: absolvi: 1. exonerate.   IL: absolvi: exonerate.
               2.*finish   : fini: finish
                  (studies).   (studies).

    : celo: 1. cell.   : celo: cell.
                 2.*cello.   : violoncelo: cello.

An important principle and part of the DLT philosophy
illustrated by the above examples, is the deliberate use
of linguistic means and the conscious avoidance of extra-
linguistic devices [Witkam, 1981c]. The latter would mean
gradually abandoning the use of a streamlined but natural
language (with an equally natural word provision) as pivot of
DLT, in favor of a highly arbitrary and much less accessible
system of word numbering by indices, superscripts and the
like. The former exploits the unique morphemic structure
and word-formation possibilities of existing Esperanto
[see 4.1] and is more promising at the long term
[cf. III-6].

As regards polysemy, Blanke [1979] points out that this
belongs to the basic facts of language, and that it is one
of the signs of language economy: it is not necessary and
not even possible that the words of an artificial language
are unambiguous, as wrongly required by some
interlinguists [the term 'interlinguists' refers to the
field of Interlinguistics, which has artificial languages
(especially Esperanto) as main research object; Blanke can
be counted as one of the most quoted researchers in this
field].
In his thesis (on a comparison of German and Esperanto)
Blanke [1982] explains that this economy is the result of
two principles that exclude each other and at the same
time supplement each other: the 'principle of language
convenience' and the 'principle of redundancy' [see
4.1.2.].
These principles can, however, get a different importance,
as shown by Blanke's remark that German words are generally
more polysemous than Esperanto words. Comparing a dictionary
of Agricola on German linguistic usage with the PIV, Blanke
found the following numbers of different meanings:

veni	7	kommen	18
pensi	4	denken	5
opinii	1	meinen	4
pafi	1	schiessen	4
ŝipo	2	Schiff	4
danĝero	2	Gefahr	2
tero	13	Erde	5
pacienco	3	Geduld	1

At the same time, this results in a higher number of
dictionary entries for Esperanto: the 'Maschinentechnisches
Esperanto-Wörterbuch der Grundbegriffe' of Wüster contains
3363 German and 3598 Esperanto entries, i.e. 7% more Esperanto
entries [Wüster 1978].
Solutions as with homonymy (shown above) are not applicable
to polysemy. In contrast to homonymy, cases of polysemy can
hardly be split up into a number of discrete meanings. In
DLT, they therefore cannot be solved by 1-to-1 mapping onto
alternative IL-forms. Any attempt to achieve this would
unavoidably result in a so-called explosive IL [cf.
III.3.3.3.2].
By rejecting an explosive IL as a solution, DLT's IL will
preserve its identity as an autonomous (be it Esperanto-
based) language. With its attractively transparent lexical
structure, it remains fully open to future developments in
the field of semantics and AI. Ultimately this approach
will appear to be a more fruitful one than creating an IL-
word for every possible meaning shade.

4.5. Lexicographic strategy and consolidation.

As stressed in Chapter III of this report as well as in recent
MT literature, the provision of a comprehensive lexicon is of
crucial importance for the functioning of a computerized transla-
tion system. As Knowles [1982] formulates it: "However powerful
and sophisticated a particular set of parsing algorithms may
be, the successful analysis of a text into a correct meaning
representation is only possible if a commensurately 'powerful'
and sophisticated lexical data base is wedded to the dynamic
modules of the MT system." .

In the 'double' translation process of DLT, the IL dictionary
columns at both sides of the main interface [see figs. III-3
and III-4] form the bridge on which the distributed SL-to-TL
conversion relies. It has been stated in chapter III that for
DLT, the various development teams dedicated to specific SL or
TL modules are supposed to familiarize themselves with the IL
(instead of with a more abstract interface specification in
other MT systems). The IL, as has been underlined in this
section of the report, does not only consist of a grammar, but
also of a lexical part.

Being part of the interface between cooperative development
teams (probably in different countries), a high degree of
lexicographic consolidation and standardization is obviously
required.
For DLT's Esperanto-based IL, the existing Esperanto standard-
works such as the PIV [Waringhien, 1977] provide an attractive
starting point, but require considerable further extension.
As for the many existing bilingual Esperanto dictionaries,
they do not excell in the coverage of compounds. The freedom
and relative ease with which the human users of Esperanto can
form or understand word compositions out of a remarkably
limited stock of morphemes [Auld, 1978: 90] seems to have
prevented the need for such coverage.

As for the availability of special terminology, impressive
efforts have been made by a small number of specialists in the
Esperanto community, resulting in terminologic dictionaries
for various fields of application (botany, chemistry,
mathematics). However, the financial support for these
activities has always been very weak, and publications of
standing such as the "International Business Dictionary in 9
Languages" [Munniksma, 1975] (with Esperanto as the editorial
pivot language) have been rare [see also section V-3].

When one looks for general bilingual Esperanto dictionaries
with comprehensive lexical entries, enumerating the valencies,
compounds, idioms, collocations etc. in which the entry word
takes part, one notices the relatively low degree of lexical
codification in Esperanto, compared to languages with a long
literary tradition.
Besides, there appears to be no prominent body with both the
means and the authority to take the lead in the continuously
required updating of Esperanto's lexeme supply (settling
controversies such as 'komputilo' vs. 'komputero'), let alone
the approval or recommendation of new compounds.

Therefore, DLT will require a lot of additional work in the
area of Esperanto lexicography. This work will perhaps partly,
but certainly not entirely, find consensus among the
Esperanto-speaking community. The nature of the MT purpose
will no doubt demand a more exact approach than would be
required for human users.
The appropriate strategy, in line with the overall DLT
philosophy, is to make good use of what is already there.
In a number of cases, only a choice will have to be made
between two already existing alternative forms. In other cases,
the available word-formation and compounding mechanisms [4.1.2,
4.2] can be utilized.
As for the creation of new terminology, the 'naturalistic'
line will be followed [see V-3.1].

5. Compaction.

As remarked elsewhere in this report, the IL has an external
and an internal representation. The former, though often
referred to loosely as BCE, is in fact CCE (Character Coded
Esperanto): an Esperanto-based IL made visible (for system
developers) by an appropriate character set [see p. 1 of the
Appendix]. Of course, the external and internal IL representa-
tions are so strongly coupled (a straightforward 1-to-1
mapping of displayable elements to bit strings) that the
confusion of the terms BCE and CCE does not matter in a
linguistic discussion of the IL. In this section, however, we
will look a bit deeper, at the way the IL is actually coded in
a binary computer.

We are all familiar with character codes like ASCII, widely
used for the coding of texts in computer systems, including
modern videotex and teletex systems.
The entering of text into computers via keyboards and its
display on terminal screens is effectuated character-wise and
traditionally relies character codes. This also applies to
DLT: sentences of IL displayed on a screen (or, in a test
situation, entered over the keyboard) must be coded character
by character
as well. We state that for sake of IL-presentation [cf. level
VI of the OSI reference model, see fig. II-6], CCE is needed.
The particular character-code used for CCE is the extended
ASCII (ISO/2022) character set.

The IL is however, in the first place, an internal instrument
in the terminals and networks on which DLT will run. This
imposes requirements, completely different from presentation
on screens and the like. One principal requirement, evolving
from network transmission costs, is compactness; another
requirement is speed of string-matching operations, in the
intensive parsing and dictionary search activities which make
up the translation process.
An obvious decision inspired by the complete invariance of
Esperanto morphemes [section 4.1.1], is the utilization of
morphemes (instead of characters) as code units.
This decision was taken at an early stage of DLT conceptualiza-
tion [Witkam, 1981a, 1981b].
BCE proper is therefore the internal morphem-based coding
adopted for the IL. The cost of the (microprocessor-based)
conversion BCE-CCE vice versa for presentation and test entry
purposes is small and fully justifiable. The big gain lies in
compact transmission and fast manipulating of meaningful
language elements, not to speak of future AI extensions.

The BCE morphem coding scheme is a combination of variable-
length (Huffman-type) coding for a few hundreds of grammatical
morphemes (grammatical endings, suffixes, function words),

and fixed-length coding for several ten thousands of lexical (open-class) morphemes. The space is the most frequent element and is coded with 1 bit only. This explains the choice of extra spaces as disambiguating elements: the insertion of these separators hardly affects compaction.

Provisional computations, based upon available frequency statistics of common Esperanto [Blaas, 1951], gave a mean number of 18.7 bits per Esperanto word. Comparing this with conventional ASCII character coding for English SL, the BCE translation will roughly consume only _half_ the number of bits.