ABSTRACT

My paper will deal with the theoretical and practical questions of the <u>co-textual analysis</u> of texts.

The <u>theoretical</u> frame is contained by the chapters titled Introduction and Conclusion. My aim here is to show that the automatization of the text analysis expects the accumulated experiences of the general linguistics and the documentation-theory to be summed up in <u>one co-herent theory</u>. The linguistics has to examine the problems of documentational thesauri, abstracting, indexing from its own aspect, and has to insert these means and methods of the documentation in the system of the text-analysis.

The <u>practical</u> questions, certain concrete phases of the co-textual analysis of texts are dealt with in the chapters 2 and 3. In connection with certain steps we hope our demonstrative examples to contain solutions that will necessarily become simpler further on. Our aim was, first of all, the presentation of the complete process on a special type of text. Not all the parts of this complete process can be automatized yet. But the presentation will always happen with respect to the automatization.

The next step should be the getting up of an analysing system that is built on the thesauristic elaboration of a relatively not big material. Then we have to solve such special problems of certain phases of the analysis that we could not even mention in this flow chart-like survey.

Beside and after the theoretical analysis, only this can be the test of the usefulness of the system.

+ + +

1.
INTRODUCTION

1.1 Before I should analyse the problems of the
technical means and a method for the text analysis,
I should like to touch a more general question. I wish to
give a short sketch of the connection between the method
to be presented and the problems of the modern theory of
grammar and theory of documentation.

To see the connections of the theory of grammar
clearer, let us start from Figure 1.

Every message is an element of a net of connections
/context/, determined by time, geographical place, cultural
environments etc. The linguistic analysis can set the only
aim of discovering the inner /co-textual/ linguistic
structure of the text with the demand of completeness.
/It can draw near to the contextual connections only to
an extent that is made possible by the structure of the
dictionary having used at the analysis./

The co-textual analysis --and the immediate con-
stituent and sentence analysis that are organic parts of
it-- can be done on different ways.

The Figure 1 shows the components of Chomsky's
generative theory. Chomsky's generative process /G/ starts
from the 'base component' /B/. The deep structure of the
sentence /DS/ comes to existence from the sentence-basis
by filling it up with lexical elements in the 'lexicon'
/L/. The surface structure /SS/ comes to existence by
using different kinds of transformations in the 'trans-
formational component' /T/. /The situating of the lexicon
and the transformational component in Figure 1 wishes to
demonstrate that we may need turning to the lexicon even
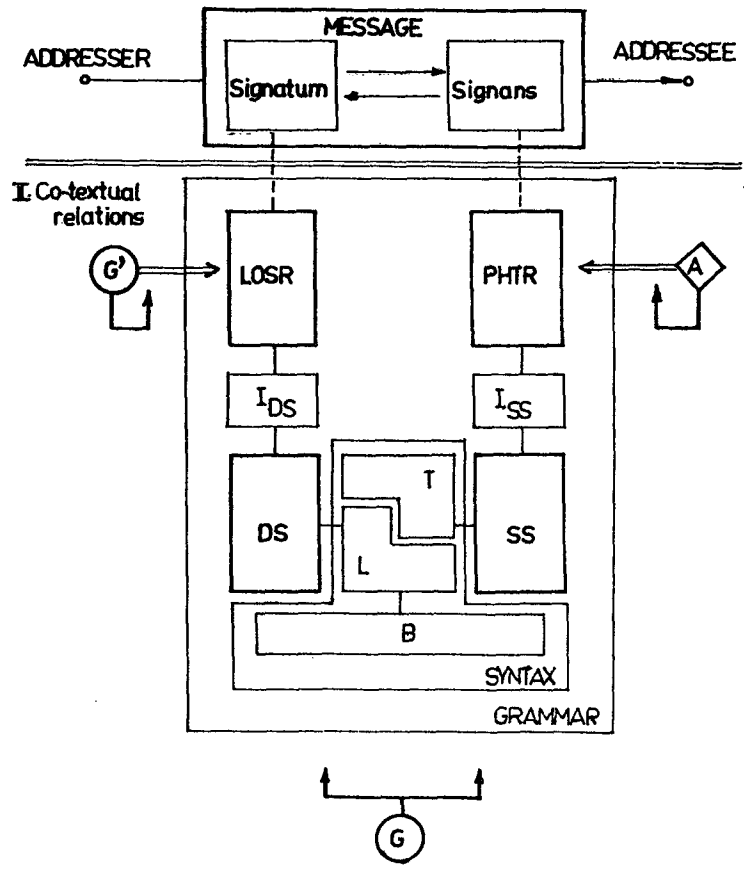
I. Contextual
   relations

Figure 1.

during dispatching the transformations./ The proper
interpretating components /I_{DS} and I_{SS}/ provide the
surface and deep structure with phonetic and semantic
interpretations. These interpretations, to a certain
extent, relate to the extralinguistic reality already.
The phonetic interpretation contains the basic elements,
too, necessary to the real pronouncing of the sentence,
beside the phonological interpretation of the surface
structure /PHTR = phonetic representation/. /The elements
originating in the speaker's subjective interpretation
either settle on or colour this./ The semantic inter-
·pretation --beside the establishing of the semantic
character of the immediate constituents-- trends to the
discovering of logical connections independent from
language, hidden behind the given verbal representation
/LOSR = logical semantic representation versus LISR =
linguistic semantic representation/.

A part of the investigations included by the
generative linguistic theory, like other examinations of
other kinds, raise the possibility of generating sentences
on another way, too.[1] This generative process /G'/ starts
from the universal logical construction /LOSR/. /The
universal logical construction is a net of connections
containing 'notion-words'[2]./ We have to order a linguistic
structure / DS/ --reflecting the characteristics of the
given language-- to this construction, and make a
concrete sentence of it / SS/.

The way of the sentence-analysis /A/ is of the
opposite direction to both. At the same time, our
sketched analysis is built on the theoretical basis,
which has developed during the investigation concerning
the problems of the generative processes /G and G'/
summarized above. But, as the text analysis goes beyond the

boundaries of the sentence, we have to widen the sentence-centered theoretical basis, as well.[3]

1.2   When analysing the notion 'text-structure' we can
make a difference between the linguistic and the
sound-textural components of the text. Both can be
linearly and hierarchically patterned.

The 'linear patterning' is a net of the recurrences
of elements that interweave the whole text. The
'hierarchical patterning' means the way as the text as a
whole is built up from the basic units of the structure
through the levels of the composition units of different
complexity.[4]

Of course there are texts where the sound-textural
component is the organizer of the structure, but as it
is not my intention to deal with this now, I shall ignore
their problems at present.

A great part of the essays dealing with text analysis
considers making lists of the structural elements of
similar construction. /It mostly remains among sentence-boundaries, so the use of a sentence-centered theory is
enough./ These lists are undoubtedly essential
characteristics of a text, but they reflect only one
aspect of the text structure. This is what I called
'linear patterning'.

The other --and, from a certain point of view more
important-- aspect is the 'hierarchical patterning' of
the text. Its analysis means much more problems.

The problems are different, according to the types
of texts. Even if we consider only the most homogeneous
texts --the scientific didactic prose or the prose told
only in the third person not containing indirect speech--

even then we meet the following basic problem: grammatical /syntactic, semantic/ connections can be discovered only among sentences constructing the so-called 'paragraphs'. /At the same time, no one has yet established the types of it, neither has anyone described its rules as a system./ Among larger text-units than 'paragraphs' only such connections can be shown that are carried by the 'content structure'. Thus, beside the examination of the grammatical construction of the 'paragraphs' we also have to look for the means to help us to discover the connections of this kind.

1.3   On this field we can get the greatest help from the documentation theory. During the last ten years the documentalists wrought out means and methods that will probably prove useful at the analysis of not scientific texts as well.

I should like to mention the thesaurus as the most significant means. The thesaurus —as known— is a notion-dictionary serving for the normalization of the indexing and searching language, which provides the different connections among the notions. It is this feature, that I think makes this means very useful to the linguistics for its own purposes, too.

From the methods those of the abstracting and automatic indexing may become significant in our analysis. In the text analysis there is a 'great obstacle', that the texts are too extensive. For the sake of making the analysis easier and the hierarchical structure clear-cut it is both useful and necessary to replace the larger structure-units by their abstracts. When speaking of abstracting —though abstracts made by statistic methods may mean a good help temporarily— I do not think of

statistic abstracts alone.

In a broad outline these aspects of the documentation theory mean the wider frame, in which certain questions of the following analysing method can be examined.

2.
ON THE MEANS OF THE CO-TEXTUAL ANALYSIS OF TEXTS

In my opinion, the following means are necessary for the co-textual analysis of texts:

1. a thesaurus including different sectors;
2. a rule-system working on sentences:
   - a/ phonological, morphonological rules,
   - b/ syntactic rules,
   - c/ semantic rules

     to the linguistic semantic interpretation
     to the logical semantic interpretation;
3. a system of the syntactic-semantic rules of the basic composition units;
4. a rule-system of an abstracting process;
5. a rule-system of a process that is able to establish the thematic connections of a 'text' consisting of abstracts.

In this chapter I want to deal only with the questions of the structure of the thesaurus in details. The work of all the rule-systems is based on the below-sketched thesaurus structure.

2.1 The thesaurus

The base of the linguistic analysis has to be a thesaurus that unites the structures of the thesaurus made for the purpose of documentation and that of the lexicon having developed during the general linguistic

investigations.

Such a thesaurus consists of two main parts: the
sector of definitions and that of classifications.

2.1.1 On the sector of definitions

The headwords of the documentational thesaurus —
combining the structural characteristics of the more
important thesauri[5]— contains the following constituents:

```
headword /descriptor/
    SY        synonyms
    EQ        equivalent terms
    TR        translation
    DEF       definition
    SN        scope note
    SF        semantic factors are
    ISF       is semantic factor of
    FIELD     field, thematic group
    CAT       category /material, feature, process/
    BT        broader terms
      BT-LOG  broader terms - logical
      BT-WH   broader terms - whole
      BT-CON  broader terms - connected
              /terms to which the headword is connected
                but not as a logical broader term or a
                whole/
    NT        narrower terms
      NT-LOG
      NT-PT
      NT-CON
    COL       collateral terms
      COL-LOG
      COL-PT
      COL-CON
    ASC       associated terms
      ASCR    associated reflexive
      ASCT    associated to
      ASCF    associated from
    EC        empirically connected
```

Let us demonstrate these constituents by the
thesaurus-entry of two concrete words:

```
madár /bird/
            DEF
            SY       'poultry'
            ISF      'migratory bird',...
            FIELD    'animals'
            BT-LOG   'vertebrate'
              -WH    'living being'
            NT-LOG   'singing-birds','birds-of-prey'
              -PT    'beak', 'wing'
              -CON   'migratory birds'
            COL-LOG  'mammals', 'reptilia'
            EC       'nest', 'air', 'tree', 'water'
            ASCR     'flying'
            ASCT     'song'
            ASCF     'chirping', 'shrieking'

    tenger /sea/
            DEF
            SY       'ocean'
            ISF      'seaside', 'sea level'
            FIELD
            BT-LOG   'still waters'
              -WH    'Earth'
            NT-LOG   'North Sea','East Sea'
              -PT    'bay'
              -CON   'sweet water', 'salt water'
            COL-LOG  'lake', 'tarn'
              -PT    'land'
            EC       'coast', 'island', 'harbour',
                     'ship', 'buoy', 'lighthouse',
            ASCR     'infinitude'
            ASCT     'waves', 'storm'
            ASCF     'gulls'
```

/We shall deal with DEF —that contains SF and CAT
— separately later./

We did not meet TR and SN in these examples, though
in certain cases they may be necessary also in non
technical texts;

E.g. at foreign words and geographical names
related to the given language

```
        archipelagus /archipelago/
                     TR        'insular world'
```

Duna

SN   'the main river of Middle-Europe
      /and Hungary/'

Stockholm

SN   'the capital of Sweden'

The DEF /the lexical definition/ has to contain
the following informations on the base of the results of
the linguistic theoretical research referring to a lexical
unit /LU/:[6]

PHM		phonological matrix
SYI		syntactic informations
	USC	the undifferentiated syntactic category of the given LU
	USC-CH	the undifferentiated syntactic category-chain marking the micro-syntactic structure of the given LU
	SYMS	the ordered sets of syntactic markers of the given LU
	MOI	morphological informations
SEI		the ordered sets of semantic markers of the given LU
COI		informations referring to the syntactic semantic conditions and consequences of the use of the given LU
OTH		other informations: the origin, the stylistic value, etc. of the given LU

Thus the DEF of a lexical unit is of the following
structure:

érzés /feeling/

PHM /é//r//z//é//s/
SYI USC   N
    USC-CH V + n
    SYMS   +common
           -count
           $\pm$abstract
    MOI   ....
SEI   physical
           or mental
           awareness
COI /V: +animate_X/
OTH   ....

kisér /'he' accompanies/

PHM /k//i//s//é//r/
SYI USC   V
    USC-CH -
    SYMS   +transitive

    MOI   ....
SEI   go with;
           happen or do
           at the same time
COI / X _ Y /
OTH   ....

Nevertheless, in the thesaurus we have to find not only the entries of the lexical units but those of the grammatical morphemes /GM/ as well.[7]

The structure of the DEF of the grammatical morphemes is simpler, though it shows analogous features to the DEF of lexical units:

PHM	phonological matrix
SYI USC	the undifferentiated syntactic category of the LU that the given GM can join
SEI	the grammatic meaning of the given GM
COI	informations referring to the possible immediate context of the given GM
OTH	informations referring to the rules that realize the morphonological changings that are connected with the use of the given GM

Thus the DEF of a grammatical morpheme is of the following structure:

-e		-t	
PHM	/e/	PHM	/t/
SYI USC	N	SYI USC	V
SEI	possessive personal suffix 3. person singular	SEI	sing of 'past'
COI	root or formative suffix /N/__ suff.	COI	root or formative suffix /V/__ suff.
OTH	......	OTH	......

I have dealt with the construction of the lexical entries and grammatical morphemes in the thesaurus above. I should like to add the following to what I told so far.

1. By dividing the entry of the thesaurus into a lexical definition /LDEF/ and a thesauristic information block /a thesauristic definition: TDEF/, the verbally

- 11 -

manifestating knowledge can be separated from the so-
called encyclopedic knowledge.

2. The syntactic informations of the lexical
definitions naturally contain only the informations
relevant in the given language.

3. The semantic informations are demonstrated here
only with a lexical definition. Of course this is not the
form that I find suitable but the definitions written in
a semantic meta-language.

4. The thesauristic informations are the results of
a multihierarchical classification. For expressing them
we have to elaborate the own 'notional language' of the
thesaurus.

5. The information FIELD points to the groups and
fields into which the notion represented by the given
word is classified. The group and field structure has to
unit the virtues of the thematic classifications of the
linguistic thesauri /Roget, Dornseiff/, the documentation
thesauri and the illustrated dictionaries. But its
realisation can be built only on a classification-
theoretical basis.

6. For reaching a satisfactory or suitable degree
of the division of the informations BT NT and COL we have
to analyse it further.

7. When defining the informations ASC we —of
course— think only of the minimal informations that can
be defined with a relatively big probability. We wanted to
mark the indefiniteness of the definition by selecting
it from the others with a dotted line.

2.1.2 On the sector of classifications

The sector of classifications first of all has to contain an <u>alphabetic index</u> that includes all the roots of words and grammatical morphemes, with reference to the form by which they can be found in the sector of definitions.

/Beside the basic alphabetic index we may construct an a-tergo list and a morphematic KWIC list as well. This latter one reflects the connections SF and ISF./

Different special classifications can be made on the basis of the lexical and the thesauristic informations.

<u>Lexical classifications</u>:
The classes of lexical units having
    the same <u>syntactic</u> categories,
                <u>semantic</u> markers,
                <u>conditions</u>,
                syntactic semantic <u>consequences</u>,
                <u>origin</u>,
                ......

The classes of
                <u>derivative endings</u>,
                <u>suffixes</u>,
                ......   of the same character.

<u>Thesauristic classifications</u>:
                <u>groups</u>
                <u>fields</u> of the elements having the
same FIELD marker
    Different   <u>hierarchical</u>
      and/or   <u>associative</u> nets.

We can provide the classifications with

identificators, and we can order the identificators to
the thesaurus entries.

Thus the structure of a thesaurus-unit is as follows:

ENTRY

L-DEF	/lexical definition /	
T-DEF	/thesauristic definition/	
IDENT	/identificators/	

2.2 As I have already mentioned, I do not wish to deal
here with the other means of the co-textual text
analysis. We can feel the character of their structure
at the presentation of the way of analysing. Their real
elaboration can be realized only on the basis of a
profound analysis of the mutual influences of all the
means.

3.

ON THE METHOD OF THE CO-TEXTUAL ANALYSIS OF TEXTS

3.0 The algorithm of the analysis

The process of the co-textual analysis of texts
consists of the following larger units:

1. the segmentation of the texts,
2. the discovery of the grammatical structure of
   the individual sentences,
3. compiling of special text-thesauri,
4. the discovery of the net of connections of the
   communication units, the establishment of the
   basic composition units,
5. making the abstracts of the basic composition
   units,

6. the thematic analysis of the text consisting
    of abstracts --in more steps.

3.1 The segmentation of the text

The notion of 'text' is interpreted on the basis
of extra-linguistic criteria. The text is always given to
us. The 'whole' is obvious owing to typographical reasons
or --mainly in case of spoken texts-- the beginning and
the end are determined another way.

For illustrating the method of analysis we have
chosen a written, 'simple narrative text', one that
contains only communications in the third person and maybe
direct speech in the first person beside. The boundaries
of the segments of the primary classification --as we have
a written text-- are marked by the author's punctuation-
marks.

The text chosen to the demonstration and its informativ
English equivalent are as follows. /The bigger line-spaces
sign the boundaries of the paragraphs located by the
author./

Madarak a tenger felett	Birds above the sea
1. A Duna hídján, a kor-látnál, emberek állnak, felnőttek és gyermekek vegyesen.	There are people standing on the bridge of the Danube, at the rail, adults as well as children.
2. Nézik a folyó felett hintázó s a vízre le-leszálló sirályokat.	They are looking at the gulls that are swinging above the river and flying down to the water from time to time.
3. A mi Duna-szakaszunkon is gyakran látom ezeket a víz felett lebegő ko-rállpiros csőrű madara-kat	I often see these coralline-beaked birds hovering above the water at our reach of the Danube as well.
4.1 Néha magasra felcsapnak,	Sometimes they dart up high

4.2	s a szemem elveszti ő-ket a kékségben vagy a ködben.	and my eyes lose them in the azure or in the mist.
5.	De tenger felett is láttam sirályokat.	But I have seen gulls above the sea, too.
6.1	Akkor már téliesre for-dult az idő,	Then the weather had already turned to winter-like,
.2	sárgán világítottak a sziget nyírfái	the birches of the island shone yellow
.3	s az Északi-tenger ha-ragosan csapkodta grá-nit-partjait.	and the North Sea dashed against its granite coasts with anger.
7.1	Hajóm finn kikötőből indult el	My ship had started from a Finnish harbour
.2	s a svéd archipelaguson haladt Stockholm irá-nyában.	and was leaving for Stockholm on the Sweedish archipelago.
8.1	Amerre haladtunk, gyak-ran felbukkant egy-egy sötét gránitsziget,	Wherever we passed, there were often appearing some dark granite islands,
.2	rajta fenyő és nyír,	with pines and birches,
.3	fenyő és nyír.	pines and birches on it.
9.1	Magányos világítóto-rony emelkedett néme-lyiken,	There were lonely lighthouses towering on some of them,
.2	s el kellett hinnem, hogy ott ember él.	and I had to believe that there was a man living there.
10.	Ember, aki talán lélek-ben is magányos, mert azzá nevelte az Északi-tenger nagy mélabúja.	A man, perhaps lonely in soul too, because that is what he was performed by the great melancholy of the North Sea.
11.1	Süvöltött a fedélzeten a szél,	The wind was wuthering on the deck,
.2	s a hullámok fenyegető-en csapkodták a hajó oldalát.	and the waves threateningly dashed against the side of the ship.
12.1	Már télikabát volt rajtam,	I had my winter-coat on already,
.2	s egy ideig künn marad-tam a fedélzeten.	and I stayed out on the deck for a time.
13.1	"My native land --good night "--küldtem Byron szavaival a búcsúszót az eltűnt finn partok felé,	"My native land --good night" I sent farewell to the dis-appeared Finnish coasts with Byron's words,
.2	pedig nem is szülőföl-demtől búcsúztam.	though it was not my native land I said farewell to.
14.	De olyan otthonos ér-	But I had such a homely

zésem volt ott, hogy azt éreztem: ha nem otthon, itt tudnék élni, s talán boldogan.	feeling there, that I felt: if not at home, I could live here, and maybe happily.
15.1 Sohasem láttam Itáliát,	I have never seen Italy,
.2 mindig északra húzott a vágyam...	I have always been drawn to the North by my desire...
16.1 Sirályok szegődtek a hajó nyomába	Gulls joined the track of the ship
.2 s én azokat néztem.	and I was looking at them.
17.1 Nagyobbak, mint a dunaiak:	They were bigger than those at the Danube:
.2 tengeri sirályok.	seagulls.
18.1 Olykor egészen a fedélzetig s a kéményig emelkedtek,	Sometimes they rose quite upto the deck and the funnel,
.2 majd leszálltak a hullámok fölé,	then flied down above the waves,
.3 s a felcsapó hideg hab meg-megfürdette őket.	and the splashing cold foam bathed them again and again.
19. Mintha uszálya lettek volna a hajónak, úgy kisérték.	They accompanied the ship as if they had been a towboat to it.
20. Ritmus volt a lengésükben és lebegésükben.	There was rhythm in their swaying and hovering.
21.1 Néha felcsaptak a fejem fölé,	Sometimes they darted up above my head,
.2 bizonyosan azt várták, hogy megetessem őket.	they probably expected me to feed them.
22.1 Különös kép lehettem:	I must have been a strange picture:
.2 egyetlen fekete utas a fedélzeten,	one single black passanger on the deck,
.3 s körülöttem a vijjogó, rikácsoló, követelődző tajtékfehér sirályok.	and screeching, crying, peremtory, spray-white gulls around me.
23. Ha nem kisértek volna, kibírhatatlanul nyomasztó lett volna az a zord és viharos tengeri komorság.	If they had not accompanied me, that severe and stormy sea mournfulness would have been unbearably depressing.
24.1 Milyen jó, hogy kisérnek,	How good it is, that they are accompanying,
.2 milyen jó, hogy ezt rikácsolják: "A mi országunk a tenger.	how good it is, that they are crying: "Our country is the sea.
25. Sirályország..."	The country of gulls..."
26. Csak az övék?	Only theirs?

27. Egyszerre panaszos, magas madárhang ütötte meg a fülemet.	Suddenly a plaintive, high bird-voice stroke my ears.
28. Több is.	More.
29. Olyan hang volt, amilyent ősszel hallok, mikor nyírfáimra száll egy fájdalmasan síró kismadárcsapat.	It was a voice that I hear in autumn when a group of painfully crying little birds flies on my birches.
30. Csupa i, csupa i, de olyan szomoru i, hogy az ember szíve megfájdul tőle.	All i-s, all i-s, but such sad i-s, that one's heart begins to ache.
31. A hajó felett szállt át a síró madárkák csapata.	The group of the crying birdies was flying over the ship.
32. Álmélkodva néztem utánuk.	I was looking after them amazed.
33. Apró madarak a tenger felett északon?	Tiny birds above the sea in the North?
34. Költözőmadarak volnának?	Would they be migratory birds?
35. Tudom, a fecske is, a sárgarigó is, a gyurgyalag is, a fülemüle is vállalja a nagy utat, ha ösztöne megszólal.	I know, the swallow, the yellow-bird, the bee-eater, the nightingale all shoulder the journey when their instinct calls them.
36. De ezek talán csak vonuló madarak.	But perhaps these are only moving birds.
37. S talán a svéd szigetvilág magyarázza meg nekem hogy miért találkozom itt velük.	And perhaps it is the Sweedish insular world that explains me why I meet them here.
38. Szigetről szigetre vándorólnának?	Would they move from island to island?
39. S azért merik vállalni az utat a haragosan hullámzó tenger felett, mert a szigetek nincsenek nagy távolságra egymástól?	And do they dare to shoulder the journey above the angrily waving sea because the islands are not very far from each other?
40.1 A sirályoknak ételmaradékot dobnak ki a tengerészek	The sailors throw out scrapings to the gulls
.2 s azok mohón, szinte verekedve falják.	and they' wolf them eagerly, nearly fighting.
41. De ki eteti meg a tengernek ezeket az apró madarait?	But who feeds these tiny birds of the sea?
42. Csodálatos ez a tengeri	This recklesness that

utat vállaló vakmerő-ség.	shoulders the journey above the sea.
43.1 Biztató	Hopeful
.2 és felemelő.	and elevating.
44. Bátorító annak, akiben olyan szomorúság van, mint bennem.	Encouraging for one who has sorrow inside like me.
45.1 Mert egy drága halottat viszek magamban fel-fel-üvöltő hullámok között az Északi-tengeren,	Because I am carrying a dear deceased in myself amongst the yelling waves on the North Sea,
.2 s néha úgy sír bennem valami, mint azok az apró drága madarak sír-tak a fejem felett...	and sometimes something cries in me just as those tiny dear birds cried above my head...

3.2 The discovery of the grammatical structure of the individual sentences

The discovery of the sentence-structure happens in the following steps. /The examples show only the result of the analysis./

1. The morphematic segmentation of the sentences /its morphonological re-interpretation/:

... Duna-szakaszunkon... /at our reach of the Danube/

Duna — szakasz + unk + on
/the Danube of   reach   our   at /

2. The substitution of lexical /thesauristic/ entries for concrete words and grammatical morphemes

A	mi	Duna	–	szakasz+unk+on		is	gyakran	lát+om
T	Pr poss pl.1	N		suff poss pl.1	suff N loc	C	Adv	V suff V s.1
/the	our	Danube	of	reach	our	at	as well often	see I

ez + ek + et	a	viz	felett	lebeg +ő	koráll + piros
Pr suff N pl.	suff N acc	T	N	Postp	V	der. N adj	A
these		the water	above	hover ing	coralline

- 19 -

```
csőr +  ü      madar + ák    + at
N       der.   N        suff   suff
        adj             N      N
                        pl.    acc
  beaked            birds          /
```

The lexival entries are signed by the concrete
chaines of letters and the syntactic categories written
under them.

During the substituting it can happen that there
is a word in the text which is missing from our thesaurus.
The lexical entry of it has to be made and inserted into
the thesaurus.

Paralelly with the substitution we can make the
list of 'word forms' of the analysed text.

3. The discovery_of_the_syntactic_relations among
the elements of the given surface structure:

$$//A \; mi \; Duna\text{-}szakaszunkon/_{NP}/_{AdvP_1} \; + \; /is/_C \; +/gyakran/_{AdvP_2}$$

$$//látom/_V \; + \; //ezeket/_{Det} \; + \; /a \; viz \; felett \; lebegő/_{AdjP_1} \; +$$

$$+ \; /koráilpiros \; csőrü/_{AdjP_2} \; + \; /madarakat/._{N}/_{NP}/_{VP}$$

It seemes necessary in the surface structure to
sign the constituents of the 'topic-comment' relation.
These signs have to be preserved in the deep structure
as well.

4. The discovery_of_the_'deep_structure' belonging
to the given surface structure -- incomplete in many
cases.[9]

The deep structure is demonstrated by Figure 2.

5. The linguistic_semantic_interpretation of the
deep structure.

At the linguistic semantic interpretation the

definition of the semantic character of the immediate
constituents happens on the basis of the syntactic and
semantic markers of the basic word and the suffixes
and/or postpositions.

Eg.	-on	Duna-szakasz	nyár
	....	....	....
	....	....	....
	+place	+place	+time
	+time	....	....
	....	....	....
		Duna-szakaszon	nyáron
		/adv. of place: at the reach of the Danube/	/adv. of time: in Summer/

In the analysed sentence:

AdvP$_1$	adverb of place
AdvP$_2$	adverb of time
Det/NP/VP	deictic pronoun /one pointing out of the sentence/
AdjP$_1$ /S$_2$/	local attribute
AdjP$_2$ /S$_1$/	qualitative attribute

The analysed sentence does not contain semantic
incompatibility.

6. The logical semantic interpretation of the deep
structure.

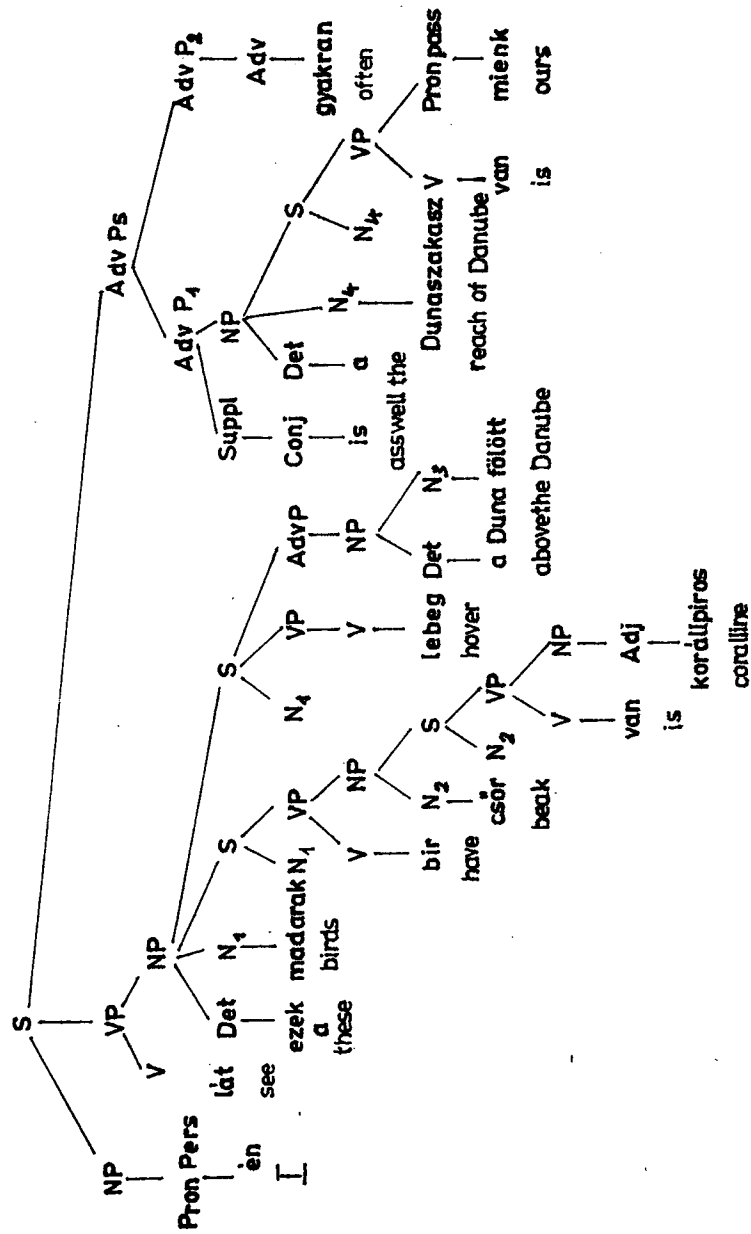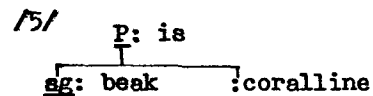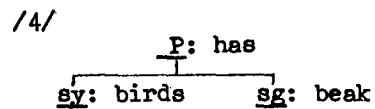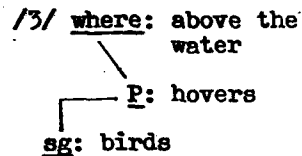The analysed sentence expresses the following
relations:

/1/ where: Reach of the Danube
    when: often
    P: sees
    sy: I   sg: birds   these

/2/   P: belongs to
    sg: reach of the Danube   sy: we
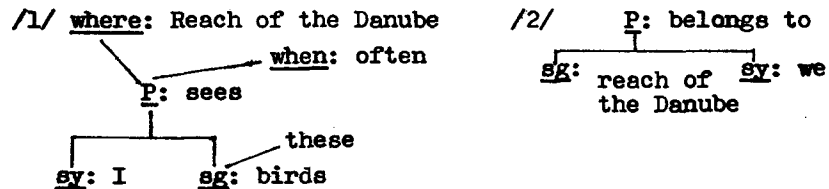
Figure 2.

The simplified form of the deep structure

/3/ where: above the
water
\
┌── P: hovers
┌
sg: birds

/4/
P: has
┌────────┴────────┐
sy: birds   sg: beak

/5/   P: is
┬
┌────┴────────┐
sg: beak   :coralline

Notes:

Roughly this succession is the way from⟨A⟩through I$_{ss}$ to ⓖ on Figure 1.

The steps from 3 to 6 happened on the basis of the lexical /thesauristic/ informations.

The 'conditions' informations /COI/ of the substituted lexical units mean a certain prediction referring to the analysis both inside and outside the sentences. For example the compulsory complement of the verbs makes it possible to look for the compulsory, possible and not complement like adverbs algorithmically.

The logical semantic interpretation of the deep structure makes it possible on one hand to establish the synonymity of sentences, and to discover the net of thematic connections of the text on the other.

3.3 The compiling of special text-thesauri

After finishing the analysis of the sentences we compile different kinds of special text-thesauri.

It is this point where the whole interpretation of the linear patterning of the text takes place on the

level of sentences. I have dealt with its problems in detail at another place. Here I should like to mention only the things that are necessary for the further investigation of the analysing process.

3.3.1 First of all we make the index of the word-forms of the whole text. After the word-forms this index gives the numbers of the sentences in which the given word-form occurs. We shall neglect the complete presentation of this index, we shall just give the list of the roots of nouns, verbs and adjectives that occur more than one time. /Considering the shortness of the text, a word occurring twice can be relevant as well./

The underlined numbers in the list sign the 'implicite' occurrences of the given words. /We speak about implicit occurrences when the given word is represented by pronouns, verbal endings or demonstrative pronouns. Their identifications with the proper words have to be done already at the semantic interpretations of the sentences -- and have to be signed by a special code in the continuously compiled text-dictionary./

V	N	A	
		apró /tiny/	33 41 45
csap /dash/			4 6 11 18 21
		drága /dear/	45 45
	Duna /Danube/		1 3 17 18
	ember /man/		1 2 9 10 30
emelkedik /rise, tower/			9 18
él /live/			9 14
	észak /North/		6 10 15 33 45
	fedélzet /deck/		11 12 22
	fej /head/		21 45
		finn /Finnish/	7 13 14 14
	fenyő /pine/		8 8
	gránit /granite/		6 8 8 9 9
	hajó /ship/		7 11 16 19 31
halad /go/			7 8
		haragos /anger/	6 39

V	N	A	
	hullám /wave/		11 18 39 45
	"i"		30 30 30
		jó /good/	24 24
kisér /accompany/			19 23 24
lát /see/			3 5 15
lett /be/			19 23
	madár /bird/		3 27 29 31 32 33 34 36 37 38 39 41 45
		magas /high/	4 27
		nagy /big/	10 35 39
néz /look/			2 16 32
	nyir /birch/		6 8 8 23
	part /coast/		6 13 14 14
rikácsol /cry/			22 24
	sirály /gull/		2 3 4 4 5 16 17 18 18 18 19 20 21 21 21 22 23 24 24 25 26 40 40
sir /cry/			29 31 45
		svéd /Sweedish/	7 37
száll /fly/			2 18 29 31
	sziget /island/		6 8 8 9 9 37 38 38 39
		szomoru /sad/	30 44
	tenger /sea/		5 6 10 17 23 24 33 39 40 41 42 45
	tél /winter/		6 12
	ut /journey/		22 35 39 42
	vakmerőség /recklessness/		42 43 43 44
vállal /shoulder/			35 39 42
	viz /water/		2 3

On the basis of this list —especially in case of shorter texts— we can also collect the list of the logical relations where the most frequently occurring nouns take place.

3.3.2 Making a list of the conjugated verb-forms, personal pronouns, possessive pronouns, nouns with possessive personal suffixes reflecting the 'communicational net' of the text is an important analysing device.

Let us see for example the list of the words referring to the first person:

```
                       V /Present/      V /Past/
  3.  mi /our/
      Dunaszakaszunkon /at our reach of the Danube/
                          látom /I see/
  4.  szemem /my eyes/
  5.                                   láttam /I saw/
  7.  hajóm /my ship/
  8.                                   haladtunk /we went/
  9.                                   hinnem kellett /I
                                       had to believe/
 12.  rajtam /on me/                   maradtam /I remained/
 13.                                   küldtem /I sent/
      szülőföldemtől /from my native land/
                                       bucsuztam /I said
                                              farewell/
 14.  érzésem /my feeling/
                                       éreztem /I felt/
                      tudnék /I could/
 15.                                   láttam /I saw/
      vágyam /my desire/
 16.                                   néztem /I looked at/
 21.  fejem /my head/   megetessem /for me to feed/
 22.  körülöttem /around me/           lehettem /I must have
                                              been/
 23.  engem /me/
 24.  engem /me/
 27.  fülemet /my ears/
 29.                    hallok /I hear/
      nyirfáimra /on my birches/
 32.                                   néztem /I looked at/
 35.                    tudom /I know/
 37.  nekem /for me/
 44.  bennem /in me/
 45.  magamban /in myself/
                       viszek /I bring/
      bennem /in me/
```

3.3.3 Making a list of the words referring to the time
       and place of the narrated event means a help as well.

       Let us see here the list of the temporal 'connecting
elements':

```
       akkor   /then/   6        mindig /always/  15
       egyszerre/suddenly/  27   néha /sometimes/ 4 21 45
       gyakran /often/  3 8      olykor /sometimes/ 18
       majd /later/  18          sohasem /never/  15
       már /already/  6 12       egy ideig /for a time/ 12
       mikor /when/  29          ősszel /in autumn/  29
```

We have to make some of the special thesauri, but
the making of others can become necessary because of the
informations accumulated in the process of analysis.

3.4 The discovery of the net of connections of the communication units

3.4.1 After the analysis of the sentences and the
compilement of the special text-thesauri we have
to discover the structure of the basic composition units.

Let us see what is at our disposal at the beginning
of this process.
We have
- the 'trans-segmentated' text;
/On the basis of the sentence-analysis we can
separate, on the one hand, smaller units in certain
sentences of the author --we have written the text
in this form in the pages 15-19 -- and in certain
cases we may contract two sentences into one unit
on the other. / The units gained by the trans-
segmentation are called not sentences but
communication units already.

- the system of connections --discovered during
  the semantic interpretation-- among the separate
  communication units;

- the list of predictions determined by the
  predicates of the communication units and not
  realized in the given communication unit;

- the special text-thesauri.

3.4.2 Let us examine here the system of the connections
discovered among the communication units.
The list below contains the following conventions:

- 27 -

X/Y	X is a constituent ordered immediately under Y /if Y is the dominant sentence itself, we never write it out/,
X/Z/	Z is an information that defines the character of the constituent X more precisely -- if it is a number it stands for the communication unit that contains X as its constituent, /after a semicolon we have the concrete lexical unit that keeps the communication units in question together/,
X :: Y	Y is the definition of X,
X = Y	X is the repetition of the constituent Y,
X : Y	the missing constituent X is identical to Y,
X -: Y	the constituent X refers to Y /stands for Y/,
	The underlined constituent on the left of the signs =, :, or -: represent the proper paraphrase standing on the right.
NP, VP, AdvP	are immediate constituents,
NPo	the 'subject' is marked only by the verbal personal suffix,
NP-	the 'subject' is marked only by the adjectival predicate without copula,
Pe	personal pronoun,
Po	possessive "
Pd	demonstrative pronoun
Px	indefinite "
Det	the determinant of a NP
D	demonstrative element

The punctuation-marks after the numbers that sign the communication units represents the punctuation marks that end the communication units immediately preceding them.

```
1.
2.        .    NPo : NP/1 ;people/
3.        :    /Det/Pd/+NP/VP -: NP/VP/2 ;gulls/
4.1       .    NPo : NP/VP/1/
 .2       ,    D/Pe/VP -: NP/VP/1/
```

```
5.          •
6.1         •
  .2        ,
  .3
7.1⌐        •
  .2⌐              NPo ; NP/7.1 ; my ship/
8.1⌐        •
  .2⌐       ,      Adv/Pe/ -; NP/8.1 ; granite islands/
  .3⌐       ,      NP = NP/8.2 ; pines and birches/
9.1⌐        •      Adv/Px//VP -: NP/8.1/
  .2         ,     Adv/D/ -: NP/9.1 ;lighthouses/
10.         •      NP = NP/9.2/

11.1        •
  .2        ,
12.1        •
  .2        ,
13.1⌐⌐      •
  .2⌐⌐      ,      NP/VP : /13.1 ; my native land/
14.  ⌐      •      Adv/D/ -: NP/VP/13.1 ; Finnish coasts/
15.1        •
  .2        ,

16.1⌐       •
  .2⌐              NP/Pd//VP -: NP/16.1 ;gulls/
17.1⌐       •      NP=   : NP/16.1/
  .2⌐       :      NP=   : NP/16.1/
18.1⌐       •      NPo   : NP/16.1/
  .2⌐       ,      NPo   : NP/16.1/
  .3⌐       ,      NP/Pe//VP -: NP/16.1/
19.  ⌐      •      NPo   : NP/16.1/
20.  ⌐      •      Pos/N/AdvP -: NP/16.1/
21.1⌐       •      NPo   : NP/16.1/
  .2⌐       ,      NPo   : NP/16.1/
22.1        •
  .2        :
  .3⌐       ,
23.  ⌐      •      NPo/S/AdvP/ : NP/22.3 ;gulls/
24.1⌐       •.     NPo/S/NP/   : NP/22.3/
  .2⌐       ,      NPo/S/NP/   : NP/22.3/
25.  ⌐             NP-   -: VP/S/NP/VP/24.2; sea/

26.         •
27.⌐        ?
28.⌐        •            :: Det/NP/27 ;bird-voice/
29.⌐        •      NP-   : NP/27/
30.⌐        •      NP-   : NP/27/
31.         •
32.⌐        •      NP/Pe//VP -: NP/31 ; group of birdies/
```

```
33.┐ ┐        •
34.┘ │    ?   NPo : NP/33 ;tiny birds/
35.  │    ?
36.┐ ┘    •   NP/Pd/ -: NP/33/
37.┐      •   NP/Pe//VP -: NP/33/
38.┤      •   NPo : NP/33/
39.┘      ?   NPo : NP/33/
40.1┐     ?
   .2┘        NP/Pd/ -: NP/VP/40 ;gulls/
41.       •

42. ┐     ?
43.1┤     •   NP- : NP/42 ; recklessness/
   .2┤        NP- : NP/42/
44. ┘     •   NP- : NP/42/
45.       •
          •
```

/We cannot find predictions in this text.
    We presented some of the special text-thesauri
in 3.3./

3.4.3 As the compositional analysis starts with the
        analysis of communication units we build the
definitions of the higher units on the notion 'communication
unit' instead of the indefinited notion 'paragraph'.
    But we need some 'transitional notions', too:
    What we call a <u>continous communicational chain</u> /CC/
is the succession of communication units with predicates of
first person in direct speech. /In this text we can see it
in the communication units 12.2 - 15.2./

    What we call a <u>simple block of communication units</u>
/SB/ is the --mostly continuous-- chain of the communication
units all of which contains <u>one</u> explicite or implicite
reference element pointing to the same 'reference'. This
reference element has to be replaced by the 'referred
element', if we want to give the semantic interpretation
of the communication unit.

What we call a complex block of communication units
/CB/ is a —mostly continuous— chain of communication
units connected by more 'referred elements'.

/For example the communication units 16.1 - 21.2
form a SB, the communication units 8.1 - 10 form a CB./

We mean the basic units of the compositional
structure the following way:

We shall call basic composition unit /or composition
unit of the first degree/ the structure unit that forms
one thematic unit and that comes to existence immediately
from the communication units. /This always has to consist
of more than one communication unit. The 'orphan
communication unit' inserted between the composition units
forms a composition unit of the zero degree./

Generally: we shall call composition unit of the n
degree the structure unit that also forms one thematic
unit and that contains composition unit of the /n-1/
degree as well.

3.4.4 Both the simplex and the complex blocks of the
        communication units and the continuous communication
chains sign some kind of the compositional dissection
already. But the discovery of the compositional structure
demands a detailed thematic analysis.

At the thematic analysis the first fixed point is
meant by the 'referential elements'.
        When the referential elements hold not many
communication units together, this mostly means a smallest
thematic unit, too.
        Inside the longer simple blocks and in the parts of
the text outside the simple and complex blocks the
discovery of the thesauristic connections among the lexical

- 31 -

units may help us to separate the composition units.

The projection of the continuous communication chain on the text can mean a help of another kind.

/The thesauristic connections separate for example the communication units 37-39 /'insular world' — 'from island to island' — 'islands'/ as one independent composition unit.

The communication chain helps at the compositional segmentation of the text consisting of the communication units 11.1-15.2 /The simple block 13.1-14 is separated by the referential element 'Finnish coast', the 12.1-12.2 is separated from 11.1-11.2 by the Person 1./

After all we can separate the following composition units of the zero and first degree:

the serial number of the comp. unit	the number of the comm. unit	the referential elements	the character of the comp. unit
1.	1 -2	people	SB
2.	3 -4.2	Pers 1 + gulls	SB
3.	5	Pers 1 + gulls	orphan comm.u.
4.	6.1-6.3	?	orphan comm.u.
5.	7.1-7.2	my ship	SB
6.	8.1-8.3	granite island	SB
7.	9.1-10	lighthouse	CB
8.	11.1-11.2	?	orphan comm.u.
9.	12.1-12.2	?	the beginning of a CC
10.	13.1-14	Finnish coast	CB
11.	15.1-15.2	?	orphan comm.u.
12.	16.1-20	gulls + ship	SB
13.	21.1-24.1	gulls + Person 1	a CC that keeps a not continuous SB together
14.	24.2-25	gulls + sea	SB
15.	26	?	orphan comm.u.
16.	27 -30	bird-voice	SB
17.	31 -32	the group of birdies	SB
18.	33 -36	'birds'	a not continuous SB including one orphan comm.u.

19.	37 -39	birds + islands	CB
20.	40.1-40.2	gulls	SB
21.	41	?	orphan comm.u.
22.	42 -44	recklessness	SB
23.	45	?	orphan comm.u.

3.5 The process of making the abstracts of the basic composition units

The making of the abstracts of the composition units established in 3.4 happens in more steps.

__3.5.1__ We take the composition units one by one and reduce their deep structures.
This means that the constituents containing the most frequent words and the ones expressing time are left, the other constituents of completive character /that is AdvP embedded in NP/S or NP/VP/S/ are omitted.

/Naturally, the 'deep structure' means only the incomplete deep structures containing the really existing constituents of the original communication units here too./

Then we write out the surface representations one by one according to the formerly reduced deep structures.

As the result of this step we shall get the following 'reduced text'. /Here we illustrate this process only with the composition units 1-11./

1. There are people standing on the bridge of the Danube. They are looking at the gulls.

2. I often see these birds at our reach of the Danube as well.
   Sometimes they dart up high.

3. But I have seen gulls above the sea, too.

4. Then the weather had already turned to winter-like.
   The birches of the island shone yellow.
   The North Sea dashed against its granite coasts with anger.

5. My ship had started from a Finnish harbour and was
    leaving for Stockholm.

6. Wherever we passed, there were often appearing some
    granite islands.
    There were pines and birches on it.

7. There were lonely lighthouses towering on some of them.

8. The wind was wuthering on the deck.
    The waves threateningly dashed against the side of
    the ship.

9. I stayed out on the deck for a time.

10. I sent farewell to the Finnish coasts.
    I had such a homely feeling there.

11. I have never seen Italy.
    I have always been drawn to the North by my desire.

3.5.2 Before beginning the second step of making the
       abstract we have to examine the orphan blocks
consisting of more communication units. We have to decide
if they represent one or more composition units.

       This is relatively easy in cases like that of the
composition unit 8. Its two communication units are
connected in their content by the explicite thesauristic
connection between the 'deck' and the 'side of the ship'
/both are the parts of the 'ship'/. /We can find an
implicite connection as well: "the wind was wuthering",
"the waves threateningly dashed", the question is to what
extent can this be made explicite./

       The communication units of the composition unit 11
are connected in their form by the pair of adverbs
never - always.

       The composition unit 4 raises more problems already.
We can establish from the 'birches of the island' — on
the basis of the later occurences of the 'island' and
the 'birches' /see the composition unit 6/— that it
refers to the birches of sea islands. The 'granite island'

- 34 -

'granite coast' connection also links the two last
communication units, The 'turned to winter-like' stands
only in a slack associative connection with the predicate
'dashed ... with anger', but it is not in any explicitely
demonstrable connection with the predicate 'shone yellow'.
Thus we could regard the second and third communication
units as an independent composition unit.

3.5.3 In the next step we substitute the separate
           composition units by one communication unit. This
means the choice of one communication unit /preserving
its original form/, or we have to generate a new one --
originally not occuring in this form.

To be able to carry out these tasks we must have
the complete analysis of the sentences of the reduced
text.
The choice can happen on structural and/or
statistical basis.

We choose for example the first communication unit
from the reduced text of the composition unit 16 on
structural basis because the others are such attributes
of the NP of this communication unit that can be regarded
the paraphrases of the attributes occuring inside the
communication units /'plaintive' -- 'painfully crying' --
'sad'/.

We choose for example the second communication
unit from the reduced text of the composition unit 11
on statistical basis because this is connected by'North'
to the other communication units, while the first one is
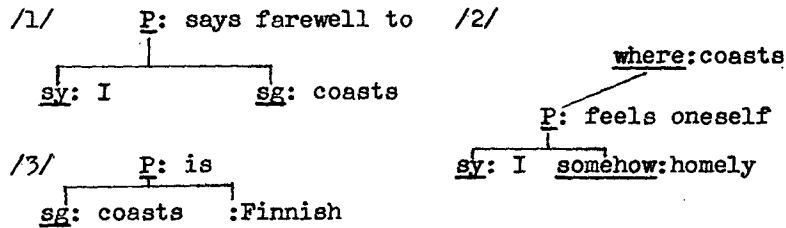not connected by 'Italy'.

We choose the last communication unit from the
reduced text of the composition unit 18 on structural
and statistical basis. Structurally because this is an

answer to the first two asking communication units, and
statistically because the last-but-one communication unit
contains NP-s that occur only once in the whole text
/'swallow', 'yellow-bird', .../.

When we _generate_ /that means a generating of (G')
type --see Figure 1/ both the structural and the
statistical aspect plays an important part. Actually it
is enough to speak of generating only because the choice
can be considered as a kind of generating by which we
make a communication unit identical with one of the
original communication units.

Let us demonstrate the character of generating by
making the abstract of the composition unit 10.

The logical deep structures of the communication
units of this composition unit are as follows:

/1/   P: says farewell to   /2/

                                        where:coasts

  sy: I   sg: coasts

                                    P: feels oneself

/3/   P: is   sy: I somehow:homely

  sg: coasts   :Finnish

Considering the paraphrase-possibilities of the /1/
and /2/ and the tenses occuring in the text we can
generate the following sentence from these:

'I said farewell to the homely Finnish coasts.'
or, holding to the structure of the text:

'I sent farewell towards the homely Finnish coasts.'

When we generate, of course it is not necessary to
condense all the informations --hidden in the logical
representation-- into one sentence.

After the generating or the choice of the sentences
we get the following 'abstract-text':

1. People on the bridge of the Danube are looking at the
   gulls.
2. I often see these birds at our reach of the Danube as
   well.
3. But I have seen gulls above the sea, too.
4. Then the weather had already turned to winter-like.
5. My ship passed from a Finnish harbour towards Stockholm.
6. Wherever we passed, there were often appearing some
   granite islands.
7. There were lonely lighthouses towering on some of them.
8. The wind was wuthering on the deck.
9. I stayed out on the deck for a time.
10. I sent farewell to the homely Finnish coasts.
11. I have always been drawn to the North by my desire.
12. Gulls joined the track of the ship.
13. If they had not accompanied me, that severe and stormy
    sea murnfulness would have been unbearably depressing.
14. How good it is, that they are crying: "Our country is
    the sea."
15. Only theirs?
16. Suddenly a plaintive bird-voice stroke my ears.
17. The group of the crying birdies was flying over the ship.
18. Maybe these are not migratory just moving birds.
19. Do they dare to shulder the journey above the angrily
    waving sea because the islands are not very far from
    each other?
20. The sailors throw out scrapings to the gulls.
21. But who feeds these tiny birds of the sea?
22. This recklessness that shoulders the journey above the
    sea is encouraging for one who has sorrow inside like
    me, because I am carrying a dear deceased in myself,
    and sometimes something cries in me just as those tiny
    dear birds cried above my head.

The informations condensed into the abstract of the
literary text are relevant with different character than
those of the scientific text. In this abstract every
composition unit has to be represented by a communication
unit.

3.6 The thematic analysis of the 'abstract-text'

Among the communication units of the 'abstract-text' we can establish very few connections with the character of those in the primary text. Here it is rather the thesauristic connections that come forward. The analysis is similar to the one told earlier. Its result can be summarized as follows.

The different thesauristic connections dissect the text of the abstract the following way:

1 - 2	'the Danube'
3 - 22	'the sea'

/In all the composition units except 1, 2, 4, 15, 16, 18 we can find 'the sea' itself or some element being in some thesauristic connection with it./

The 1-2 forms a composition unit of the second degree, its other connecting element is the 'gulls'.

We can dissolve further the unit kept together by the 'sea'. The inner 'thematic connecting elements' of the so-gained composition units of the second degree are as follows:

5 - 7	'passed' + 'islands'
9 - 11	Person 1
12 - 14	'gulls' + 'ship'
16 - 19	'birdies'
20 - 21	'feeding'

This compositional division will obviously meet the net of connections of the communication units of the primary text. The orphan composition units /3, 4, 8, 15, 22/ include the orphan communication units.

The composition units of the second degree of the unit linked by the 'sea' form two composition units of the

- 38 -

third degree:

/5-7/ + /9-11/   : Person 1   /my ship', 'I remained',
                              'I sent', 'my desire'/

/This includes the composition unit 8
as well./

/12-14/ + /16-19/:   'bird' /either as a primary lexical
                            unit or as a semantic and-
                            or thesauristic element/

/This includes the composition unit 15
as well./

The compositional structure of the whole text can be
summarized the following way:

Birds above the sea

0   1   2   3

I. Gulls above the Danube   1- 2
II. Birds above the sea   3-22
    A. But I have seen gulls above the sea, too 3
    B. Then the weather had already turned to
       winter-like   4
    C. 1. My ship passed from a Finnish
          harbour towards Stockholm   5- 7
       2. The wind was wuthering on the deck   8
       3. I sent farewell to the homely
          Finnish coasts   9-11
    D. 1. Gulls joined the track of the ship   12-14
       2. Does the sea belong only to the
          gulls?   15
       3. A group of the crying birdies was
          flying over the ship.   16-19
       4. The sailors feed the gulls, but
          who feeds the crying birdies?   20-21
    E. This recklessness that shoulders the
       journey above the sea is encouraging   22

In the hierarchical structure we cannot establish
levels equally valid to every text, but we can define the
'depth-level' of every composition unit as the
compositional distance measured from the 'whole of the
work of art' as the broadest composition unit. /The upper

- 39 -

line of numbers marks the depth-levels./ We can observe
that we can meet eg. a composition unit of the first
degree on the most different depth-levels.

4.
CONCLUSIONS

My paper dealt with the questions of the co-
textual analysis of a special text-type. I tried to
deliberate on the complete text-analysing process
algorithmically. This process has parts than can be done
automatically already now, and theoretically -- I
suppose-- all its parts can be automatized. But the
practical realisation expects the solution of different
problems that are among the key-problems of the
linguistics and the documentation. My intention was first
of all to show how these are connected even at the
analysis of the simplest text. Finally I should like to
point to some problems of basic importance again.

4.1   Before the compilement of the thesaurus we have
        to select the so-called 'notion-words'. Mostly
these 'notion-words' take place in the sector of the
thesauristic definitions.

We also have to elaborate the 'semantic basic
language' by which the words of the given language can
be semantically defined and we have to define the system
of the relations to be used in the semantic definitions.
/The 'notion-words' are words, the elements of the
'semantic basic language' are of feature character./

The thesaurus has to provide the interrelations
among the different 'notion-words' and the different
'lexical units'. These interrelations are of one-many

- 40 -

character:

$$TDEF_i \begin{cases} LDEF_{i1} \\ LDEF_{i2} \\ \vdots \\ LDEF_{in} \end{cases}$$
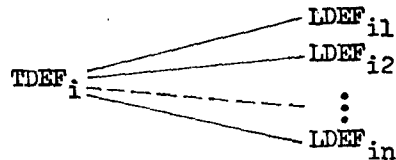
That is: more lexical units /LDEF/ may belong to one 'notion word' /TDEF/.

4.2 The relation 'logical semantic representation' — 'linguistic semantic representation' is an analogous equivalent of the relation 'notion word'—'lexical unit'.

The 'logical semantic representation' /LOSR/ represents 'logical connections' among 'notion words', while the 'linguistic semantic representation' /LISR/ represents 'verbal connections' among 'lexical units'. /In fact, LISR is a deep structure that demonstrates the semantic character of the constituents, too./

In connection with these we have to define the systems of both the logical and linguistic connections, the way of their representation, the correspondence of these two systems of connections, and the way of passing over from one to the other.

The interrelations are here interrelations of sets and they are also of one-many character. /That is: to a given set of elementary logical relations more different text-structures can belong. These text structures one by one may contain only as many independent deep structures as the number of the elementary logical relations./

The efficacity of the analysing system first of all depends on the determinedness of the elements and relations of the logical and linguistic system and of the rules of

the matching between the two systems.

4.3   We have to mention that the relation of the
       phonetic and phonological representation is also
analogous to the above told one, though this analogy
contains an oblique symmetry. This is a natural
consequence of the continuous and one-way one-many relation
passing from the logical representation to the phonetic
one.

4.4   From the point of view of the relation of the
       linguistic deep and surface structure the structure
of the syntactic informations and the condition
informations of the LDEF is of primary importance. It is
significant to approach it from the transformations.
/That is, we have to examine to what extent the concrete
lexical units occuring in a given deep structure determine
the character of the transformations to be used./

4.5   It is only a solution of the above sketched
       problems happening on a theoretical basis forming
one coherent system that makes possible the solution of
such questions as
       - the automatic realization of the transitions
            from the surface structure to the incomplete
               deep structure,
            from the incomplete to the complete deep
               structure,
            from the deep structure to the logical
               representation,
       - the automatic discovery of the thematic connections
            built on the TDEF-s,
       - the automatic establishment of the --in the given
            connection-- irrelevant elements of the
            deep structure,
       - the automatic realization of the reduction
            building on the irrelevant elements,

finally
- the automatic dispatching of the generating of
G' character necessary during the process.

4.6 The demonstrating material in the chapters 2 and 3
of my paper was a so-called 'simple narrative text'.
When listing the problems we must not ignore the fact
that the different kinds of texts will enlarge the above
enumerated list of the basic problems with their special
problems as well.

Notes

1. We can find an excellent bibliography on the later investigations in generative grammar in the issue 14 of the Langages /Tendances nouvelles en syntaxe generative, par N.Ruwet/, 1969 juin.

2. See for example : A.K.Zholkovski, I.A.Mel'chuk, O semanticheskom sinteze. Problemy kibernetiki 19./1967/ 177-238.

3. One that gives a survey of the works dealing with the problems of literary text analysis: J.Ihwe, Linguistik und literaturwissenschaft /Bibliographie/. Replik 2./1968/ 28-42.

4. See: J.S.Petőfi, On the structural linguistic analysis of poetik works of art. Computational Linguistics /Budapest/ VI./1967/ 53-82.

5. See:
   D.Soergel, Klassifikationssysteme und thesauri. Frankfurt am Main, 1969.
   J.S.Petőfi, A tezaurusz-kérdés jelenlegi helyzete /The present state of the thesaurus problem/ Budapest, OMKDK 1969.

   The abbreviations below are taken from Soergel's work.

6. Our intention was only to summarize the types of informations. In connection with this see the works of M.Bierwisch, Ch.Fillmore and M.Gross.

7. The Hungarian language is very rich in suffixes. See to this:
   A magyar nyelv szóvégmutató szótára /Reverse-alphabetized dictionary of the Hungarian language/ Compiled by F. Pap. Budapest, 1969.
   Verzeichnis der ungarischen Suffixe und Suffixkombinationen. Zusammengestellt von W.Veenker, Hamburg, 1968.

8. Áprily Lajos was a Hungarian poet and translator. His analysed short-novel is taken from his volume "Fecskék, őzek, farkasok" /Swallows, deer, wolves/ Budapest, 1965.

9. See: J.S.Petőfi, Notes on semantic interpretation of verbal works of art. Computational Linguistics /Budapest/ VII./1968/ 79-105