GERD LAU - HANS DIETER LUTZ

AUTOMATIC ANALYSIS OF THE GERMAN NOUN GROUP AND SOME PROBLEMS

0. Introduction.

Since 1971 the research group " Maschinelle Syntaxanalyse " (MasA) has been working as a part of the project " Linguistische Datenverarbeitung " (LDV) at the " Institut für deutsche Sprache ", supported by the " Bundesminister für Forschung und Technologie " of the Federal Republic of Germany.

The task of the MasA is to elaborate a system for analysing simple sentences of present-day-German. " Simple sentences ", in this case, means any type of sentence with only one finite verb in it.

1. The scope of the analysis.

The team works on the basis of a so-called verb grammar. In this type of grammar the verb represents the nucleus of a sentence, thus being opposed to the subject-predicate-grammar. Proposals for the elaboration of such a grammar have been presented by Helbig-Schenkel,[1] Heringer [2] and Engel.[3]

This type of grammar is based on a verb lexicon, which describes all the obligatory and facultative objects that depend on the verb, e.g.

treffen   with two obligatory objects in
    Ich treffe ihn morgen

---

[1] G. HELBIG, *Der Begriff der Valenz als Mittel der strukturellen Sprachbeschreibung und des Fremdsprachenunterrichts*, in « Deutsch als Fremdsprache », I (1965), pp. 10-23.

[2] H. J. HERINGER, *Theorie der deutschen Syntax*, München 1970.

[3] U. ENGEL, *Bemerkungen zur Dependenzgrammatik*, in « Neue Grammatiktheorien und ihre Anwendung auf das heutige Deutsch. Sprache der Gegenwart », XX (1971), pp. 111-155.
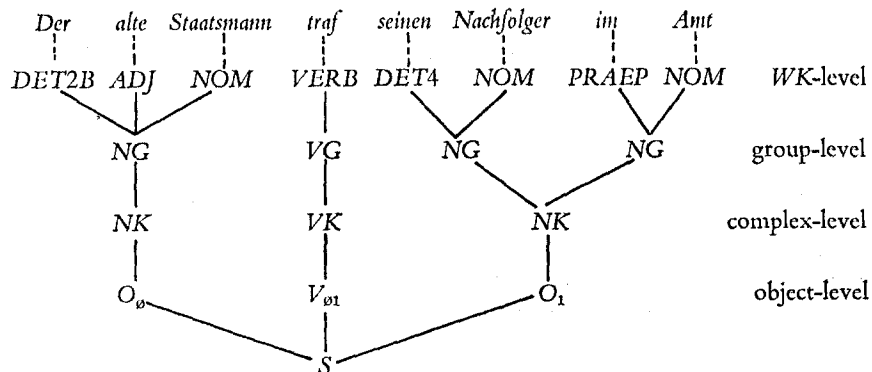
treffen   with one obligatory object and one facultative object in
              Er trifft (den Ball)
treffen auf with two obligatory objects in
              Die Mannschaft traf auf einen schwachen Gegner.

After having detected the number of both the obligatory and facultative objects and specified these objects according to their syntactical functions, one can group the verbs into classes, e.g. the class of all the verbs with one obligatory object (this class contains for example the verb *laufen*), the class of all the verbs with two obligatory objects (containing for example *treffen*), the class of all the verbs with three obligatory objects (containing for example *anbieten*), the class of all the verbs with one obligatory and one facultative object (containing for example *treffen*), and so on. Each of these classes represents a so-called " *Satzbauplan* " (*SBP*).[4]

The systematic levels of the grammar are

a) the word class-level (*WK*-level)
b) the group-level (noun group (*NG*), verbal group (*VG*))
c) the complex-level (noun complex (*NK*), verb complex (*VX*))
d) the object-level (nominative object ($O_\emptyset$), genitive object (O_2), ...).

The following graph can be regarded as an illustration of these four systematic levels:[5]

Der	*alte*	*Staatsmann*	*traf*	*seinen*	*Nachfolger*	*im*	*Amt*	
DET2B	ADJ	NOM	VERB	DET4	NOM	PRAEP	NOM	*WK*-level
	NG		VG	NG		NG		group-level
	NK		VK		NK			complex-level
	$O_\emptyset$		$V_{\emptyset 1}$		O_1			object-level
			S					

---

[4] U. ENGEL, *Die deutschen Satzbaupläne*, in « Wirkendes Wort », XX (1970), pp. 361-392.

[5] For the explanation of the word classes, especially *DET2B*, *ADJ*, *NOM*, *DET4*, and *PRAEP* see section 2; $O_\emptyset$ stands for " nominative object ", O_1 for " accusative

The way of finding the word class of a given word form, i.e. the morpho-syntactical description of a word form, is delineated in the contribution of Werner Brecht.[6]

2. *Description of NG-relevant word classes.*

The following word classes are relevant for the constitution of a NG:

a) preposition (postposition) *PRAEP,* e.g. *vor*
b) determinantia *DET,* e.g. *allen*
c) adverbs *ADV,* e.g. *sehr*
d) non-inflected adjectives *ADJU,* e.g. *gut*
e) inflected adjectives *ADJ,* e.g. *sichtbaren*
f) nouns *NOM,* e.g. *Objekten*
g) pronouns *PRON,* e.g. *uns.*

The word classes are defined by the following morphological and/or syntactical properties:

Prepositions are collected in a catalogue and specified by their position (in front of, behind, in front of or behind, in front of and behind the nucleus of a NG), by their government (accusative, genetive, dative) and by the sort of object that is introduced by the preposition (prepositional object, situative object, directional object).

Adverbs are listed, and the list only includes words that meet the following conditions: non-inflexional; no predicative use; certain regularities of position.

Non-inflected adjectives are not listed; they are described by an algorithm, which is to find out the degree of comparison.

Inflected adjectives are not listed, either. They, too, are described

---

object ”; $V_{\emptyset 1}$ stands for the verb class with the “ Satzbauplan ” $\emptyset 1$ (obligatory nominative object plus obligatory accusative object). Note: The example is ambiguous, and only one interpretation is given. The example is ambiguous because of the noun group *im Amt.* This noun group may be an attribute to the noun group *seinen Nachfolger.* It may also be a noun group without any relation to *seinen Nachfolger,* and therefore it must be interpreted as a separate noun complex, which does not depend on *treffen* as a representation of the verb class $V_{\emptyset 1}$. In this case the sequence *im Amt* is regarded as a so-called “ *Angabe* ” (cf. U. ENGEL, Linguistische Studien I *Regeln zur* “ *Satzgliedfolge* ”. *Zur Stellung der Elemente im einfachen Verbalsatz,* in « Sprache der Gegenwart », XIX (1972), pp. 17-75, esp. 24ff).

6 In the first Volume of this book, pp. 9-21.

by an algorithm with respect to case, number, gender, degree of comparison and kind of inflexion (strong, weak).

Nouns are not listed in the present state of our work. There exists an algorithm that provides the description of those words according to case, number, and gender.[7]

In general, the definition of the several word classes is the common one, as found in the traditional grammars. But there are two exceptions: the so-called determinantia and the pronouns.

The so-called determinantia are listed in a catalogue; they are grouped with respect to the possible arrangements of several determinantia within a NG, e.g.

vor allen diesen meinen schönen Büchern.[8]

The results of having found out certain regularities are both a number of distinct subclasses of determinantia and the formulation of context-sensitive rules for analysing arrangements of determinantia.

First, there are 10 subclasses, namely

DET1A	containing	*all, all-*;
DET1B	containing	*manch, solch, welch*;
DET1C	containing	*irgend, irgendwelch*;
DET2A	containing	*dies-, jen-*;
DET2B	containing	*d-, d-jenig*;
DET2C	containing	*ein-*;
DET2D	containing	*kein-, welch-, irgendein-, irgendwelch-, ebend-, ebend-selb-, ebendies-, d-selb*;
DET3	containing	*jed-, jeglich-, jedwed*;
DET4	containing	*mein-, dein-, sein-, uns(e)r-, eu(e)r-, ihr-*;
DET5	containing	*einig-, etlich-, manch-, mehrer-, ein paar.*[9]

Secondly, there are 5 rules for describing the regularities of determinantia arrangements within a NG having a noun as nucleus. For

---

[7] In future it will be possible to incorporate a more effective algorithm which reverts to an almost complete lexicon of German nouns. This lexicon has been placed at our disposal by the research group for " Automatische Lemmatisierung " at the University of Saarbrücken, W-Germany.

[8] The subclassification of the determinantia is a modification of the one given in ENGEL, *Regeln zur Worstellung*, in « Forschungsberichte des Instituts für deutsche Sprache, Mannheim », V (1969), pp. 9-148, esp. 102ff.

[9] The catalogue contains all word forms of the listed stems (indicated by " - ").

the demonstration of these rules we will use a simplified *NG*-structure with an abbreviated notation.

$R1$: $(DET1A)$ $(DET2A)$ $(DET4)$ (ADJ) $NOM \rightarrow NG$
$R2$: $(DET1A)$ $(DET2B)$ (ADJ) $NOM \rightarrow NG$
$R3$: $(DET1B)$ $(DET2C)$ (ADJ) $NOM \rightarrow NG$
$R4$: $(DET1C)/(DET2D)/(DET5)$ (ADJ) $NOM \rightarrow NG$
$R5$: $(DET2C)$ $(DET3)$ (ADJ) $NOM \rightarrow NG$.[10]

It is important that there can be no iteration of a subclass of determinantia.

The rules mentioned above can be illustrated by the diagram on the next page

The pronouns are considered as pro-forms of a *NG* with a nominal nucleus. A part of the pronouns can be grouped in several subclasses according to three significant features: possibility of using a preposition in front of the pronoun, possibility of putting determinantia in front of the pronoun, possibility of using a genitive or prepositional attribute in the post-nuclear field.

There are 9 subclasses, namely

PRON1   containing *man*;
PRON2   containing *du, er, es, ich, ihr, sie, wir*;
PRON3   containing *dein(er), dich, dir, einander, euch, euer, ihr(er), ihm, ihn(en), mein(er), mich, mir, sein(er), sich, sie, uns(er)*;
PRON4   containing *all-, ebend-, ebend-selb, ebendies-, einig-, ein paar, etlich-, etwas, irgendein-, irgendetwas, irgendw-, irgendwelch-, jedermann-, jemand-, kein-, manch-, mehrer-, niemand-, nichts, w-, welch-*;
PRON5A containing *d-, d-selb-, d-jenig-, dies-, jen-*;
PRON5B containing *ein-*;
PRON5C containing *jed-, jedwed-, jeglich-*;
PRON5D containing *dein-, eu(e)r-, ihr-, mein-, sein-, uns(e)r-*;
PRON6   containing *selber, selbst*.

---

[10] The brackets stand for " facultative "; the slash stands for " exclusive o r ". A full description of the structure of a noun group with a nominal nucleus will be given in section 5.

class \ rule	DET1A	DET1B	DET1C	DET2A	DET2B	DET2C	DET2D	DET3	DET4	DET5	ADJ	NOM
R1:	alle			diese							treuen	Hunde
R2:	all	manch			die				meine		treuen	Hunde
R3:						ein					treuer	Hund
R4:			irgendwelch								treue	Hunde
							ebendiese				treuen	Hunde
										einige	treue	Hunde
R5:						ein		jeder			treue	Hund
rule \ class	DET1A	DET1B	DET1C	DET2A	DET2B	DET2C	DET2D	DET3	DET4	DET5	ADJ	NOM

But, there is another group of pronouns without the characteristics mentioned above, i.e. without pre-nuclear field, without a preposition in front of them, without post-nuclear field. These pronouns do not differ at the NG- or NK-level. But they differ at the O-level, when you try to detect if such a pronoun is a substitute for an object (depending on the verb) or a substitute for a so-called " *Angabe* " (which does not depend on a verb).[11] Furthermore, pronouns that can be substitutes for an object, can be differentiated according to the kind of these objects.

So, there are six more subclasses, namely

> *PRON7A* containing *wobei, wofür, womit, wonach, woraus, worum, wovon*;
>
> *PRON7B* containing *wo, wohinter, woneben*;
>
> *PRON7C* containing *woher, woherauf, woheraus, woherein, woher- über, wohin, wohinaus, wohinein, wohinüber*;
>
> *PRON7D* containing *woran, worauf, worin, worüber, worunter*;
>
> *PRON7E* containing *wogegen*;
>
> *PRON8*   containing *inwiefern, inwieweit, wann, warum, weshalb, weswegen, wie, wieso, wodurch.*

Pronouns of *PRON7A* are substitutes for prepositional objects (O_4); pronouns of *PRON7B* are substitutes for situative objects (O_5); pronouns of *PRON7C* are substitutes for directional objects (O_6); pronouns of *PRON7D* are substitutes for O_4 or O_5; pronouns of *PRON7E* are substitutes for O_4 or O_6; pronouns of *PRON8* are substitutes for so-called " Angaben ".

The final division of pronouns in 15 subclasses makes it possible to set up rules for detecting and describing NGs with a pronominal nucleus. For the demonstration of these rules we will use a simplified and abbreviated notation.

> R6: $PRON1 \mid PRON2 \rightarrow NG_\emptyset$
>
> R7: $PRON6 \rightarrow NG_? \rightarrow NG_{Attr.}$
>
> R8: $(PRAEP) \{PRON3 \mid PRON4\} \rightarrow NG_?$
>
> R9: $(PRAEP) \{\{(DET1A)\ PRON5A\} \mid \{(DET1B)\ PRON5B\}$
> $\mid \{(DET1C)\ PRON5C\} \mid \{(DET1A)\ (DET2A/$
> $DET2B)\ PRON5D\}\} \rightarrow NG_?$

---

[11] The reason for these pronouns not being included in *PRON1* (which meets the same conditions at the first glance) is that *man* of *PRON1* covers always the position of a nominative object ($0_\emptyset$), in contrast to the pronouns of *PRON7A* to *PRON8*.

$R10$: $PRON7A \rightarrow \ldots O_4$
$R11$: $PRON7B \rightarrow \ldots O_5$
$R12$: $PRON7C \rightarrow \ldots O_6$
$R13$: $PRON7D \rightarrow \ldots O_4 \mid O_5$
$R14$: $PRON7E \rightarrow \ldots O_4 \mid O_6$
$R15$: $PRON8 \quad \rightarrow \ldots A_?$.[12]

This collection of rules shows the difference between $PRON1$ ... $PRON6$ and $PRON7A$ ... $PRON8$. A comparison with the rules $R1$ to $R5$ illustrates that there are less possibilities to combine determinantia in front of the nucleus of a pronominal NG. Other differences will be discussed in section 3.

3. *Two types of NGs.*

It is no new observation that there are at least two different types of NGs in German.

The first type covers all NGs having a noun as nucleus of the NG, e.g. *die Mitbestimmung,*

the second type covers all NGs with a pronoun as nucleus of the NG, e.g. *sie, die unsere, woran.*

The structure of the NG depends on what kind of nucleus appears.

In a NG with a pronominal nucleus, there can be no adverbs, no non-inflected adjectives, and no adjective – in opposition to a NG with a noun as a nucleus.

In a pronominal NG there are less possibilities of arranging determinantia than in a nominal NG (see rules $R1$-$R15$). The reason is quite evident: by suppressing the noun (nucleus) in a nominal NG, the last determinans turns into a pronoun, i.e. it represents the nucleus in a pronominal NG. Compare the following two examples:

$$
\begin{array}{ccc}
alle\ diese\ B\ddot{u}cher & \text{vs.} & alle\ diese \\
DET1A\ DET2A\ NOM & & DET1A\ PRON5A \\
\uparrow & & \uparrow \\
\text{nucleus} & & \text{nucleus}
\end{array}
$$

---

[12] The " ? " indicates that the case of the NG is not predictable in general. " A " is an abbreviation of " *Angabe* ". The braces are to denote larger units belonging to each other. For the explanation of other signs see footnote 10.

There is another difference beween a nominal *NG* and a pronominal one: in nominal *NG*s one can find other *NG*s embedded:

> *die Ansprachen haltenden Parlamentarier,* or (rather complex)
> *die nach Marx alle menschlichen Handlungen wesentlich*
> *und zu aller Zeit zum Nachteil der Mehrheit der Bevölkerung bestim-*
> *menden Kapitalgesetze.*

In both cases the first *die* belongs to the last word, to *Parlamentarier* and *Kapitalgesetze; die Parlamentarier* and *die Kapitalgesetze* form the hierarchically highest *NG*. If you transform the determinans of this highest *NG* into a pronoun such as *sie,* there is no possibility of embedding other *NG*s in this pronominal *NG*. This explains the difference between a nominal *NG* and a pronominal one, too.

4. *Criteria for the analysis.*

The different elements of a German *NG* underlie different kinds of syntagmatic relations. The number of these relations varies corresponding to the two types of *NG*s.

Four kinds of relations are relevant for the nominal *NG*:

> congruity between adjectives and nouns (concerning case, number, and gender)
>> between determinantia and nouns (concerning case, number, and gender)
>> between prepositions and nouns (concerning the case of a noun and the possible government of a preposition)
>> between determinantia and adjectives (concerning case, number, and gender);
> inflectional relation between determinantia and adjectives (concerning the inflexion of a determinans and the inflexion of a following adjective;
> degree of comparison of the adjective of the nominal *NG* (providing the connection with further parts of the syntactical analysis);
> compatibility of subclasses of determinantia according to the rules $R1$ to $R5$; the rule of compatibility of nominal *NG* has the form:

17

$$((\mathit{det} = 1) \Rightarrow ((M(WK, \ I, \ n\text{-}i) = DET1A \vee (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET2A \vee DET2B \vee DET4)) \vee$$
$$\vee ((M(WK, \ I, \ n\text{-}i) = DET2A \wedge (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET4))) \vee$$
$$\vee ((M(WK, \ I, \ n\text{-}i) = DET1B \wedge (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET2C))) \vee$$
$$\vee ((M(WK, \ I, \ n\text{-}i) = DET2C \wedge (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET3)))) \vee$$
$$\vee ((\mathit{det} = 2) \Rightarrow (M(WK, \ I, \ n\text{-}i) = DET1A \wedge (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET2A \wedge (M(WK, \ I, \ n\text{-}i + 2) = DET4))).[13]$$

Only two kinds of relations are relevant for the pronominal NG:
congruity between determinantia and pronouns (concerning case, number, and gender)
between prepositions and pronouns (concerning the case of a pronoun and the possible government of a preposition);
compatibility of subclasses of determinantia according to the rules R6 to R9; the rule of compatibility of a pronominal NG has the form:

$$((\mathit{det} = 1) \Rightarrow ((M(WK, \ I, \ n\text{-}i) = DET1A) \wedge (M(WK, \ I, \ n\text{-}i + 1) =$$
$$= DET2A \vee DET2B))).$$

Summing up, there are only two kinds of relations of the pronominal NG, but four kinds of relations of the nominal NG; the set of the various congruity relations noticeable at the nominal NG differs from the one noticeable at the pronominal NG; and the compatibility

---

[13] This notation is a conglomerate of matrix-notation, logical signs, and of programmer's conventions. It can be read: "If the value of the counter "det" equals 1, then the value of the place " WK " of the matrix " M " at the level " I " at the position " n-i " must have been " DET1A " and the value of the place " WK " of the matrix " M " at the level " I " at the position " n — + 1 " must have been DET2A or DET2B or DET4, or the value of... " and so on.

M is a three-dimensional matrix which gets built up during the morphological analysis of an input sentence, and which contains all computed information of all the words of this sentence. " I " stands for " interpretation " and is a variable, running from 1 to 50. At the moment, " WK " stands for " word class " with 40 different specifications. " n " indicates the position of the noun in a given sentence, and " i " is a position counter; so, if $n = 3$ and $i = 1$, then $n — i = 2$, i.e. the position in question is the next to the noun on the left side.

rule for the nominal *NG* si much more complex than that for the pronominal *NG*.

5. *An algorithm as representation of a hyperrule.*

The preceding outlines displayed the complexity of contextual restrictions concerning German noun groups. It sounds extremely clumsy to talk about these restrictions in a natural language. We therefore developed a formal notation for the communication between linguists and programmers of our team. It consists of a matrix-notation, sentential connectives, functions and flow chart conventions. The compatibility rules were a small cut-out. For more details see our forthcoming publication.[14]

In the following, we will discuss the nominal *NG* only, because its structure and its difficulties are more interesting than the structure of the pronominal *NG*.

At the present state of our algorithm the following constructions have been excluded: coordination of adjectives, fusion of prepositions and certain determinantia (i.e. *am* being *an + dem*), postponed prepositions, prepositions like *um ... willen*, and embedded participle constructions.

The reason for excluding embedded participle constructions is that, in our opinion, complex constructions like these can be analyzed after the successful interpretation of simple sentences only. The other restrictions mentioned above are not so important, i.e. the solution of these problems are, more or less, not so difficult, and, in fact, one of our collegues has made a proposal for the treatment of all kinds of prepositions.

At the moment, we will consider the following hyperrule and we will interpret it as a statement about the word order only:

(1)   *(PRAEP) (DET) (DET) (DET) (((ADV*) ADJU*)*
        ADJ)* NOM → *NG.*[15]

---

[14] ARBEITSGRUPPE MASA, *Zur maschinellen Syntaxanalyse I. Morphosyntaktische Voraussetzungen für eine maschinelle Sprachanalyse des Deutschen*, in «Forschungsberichte des Instituts für deutsche Sprache, Mannheim», XVIII, 1, XVIII, 2, 1974.

[15] The asterisk indicates the possibility of iterating. For the explanation of the other signs see footnote 10.

The program that is intended to analyze nominal NGs first looks for a *NOM*, proceeding from the left to the right. After having found a *NOM*, the whole algorithm will be called up in order to search an *ADJ* to the left of the *NOM*. *ADV** and *ADJU** can appear only in case an *ADJ* has been found. Otherwise the algorithm turns to *DET*, which tests the word order and compatibility of the *DETs* if there exists a *DET*. There are no more than three *DETs* admitted. *PRAEP* is the word class which stands at the extreme left in German NGs (according to our restrictions); it is the last one to be looked for.

In many cases the output of this analysis depends on homography. An example may prove this as well as the fact that there remain lots of ambiguities at this level of analysis:

(2)   *nahe   so   sehr   gründlich   vergifteten   Wäldern*

  |*PRAEP   ADV   ADV   ADJU   ADJ   NOM*| description 1

  VERB |*ADV   ADV   ADJU   ADJ   NOM*| description 2

  VERB   ADV|*ADV   ADJU   ADJ   NOM*| description 3

  VERB   ADV   ADV |*ADJU   ADJ   NOM*| description 4

  VERB   ADV   ADV   ADJU |*ADJ   NOM*| description 5

  VERB   ADV   ADV   ADJU   VERB |*NOM*| description 6

(3)   *vor   so   sehr   gründlich   vergifteten   Wäldern*

  |*PRAEP   ADV   ADV   ADJU   ADJ   NOM*| description 1

If *PRAEP* is a homograph, *ADV** and *ADJU** may or may not be constituents of the nominal NG (see (2)). If *PRAEP* is no homograph and if there is no *DET*, the whole string must be a NG (see (3)).

As homographs may exist, the algorithm is constructed in the following way: in a first step the longest nominal NG possible is built up, and in a second step this nominal NG is minimized in all possible interpretations (in the case of (2) there are five interpretations which cover shorter nominal NGs than the first one).

6. *Problems concerning the identification of a " nominal complex " (consisting of a NG or of a number of NGs).*

A nominal complex *NK* is a noun group or a string of noun groups which are part of a sentence. Yet, this gives no information about the role the *NK* plays in the sentence: a *NK* can be either an object which depends on the finite verb, or an " *Angabe* ". *NKs* in the nominative case can serve as objects only, i.e. as subjects of sentences. The following schedule displays that any *NK* with an *O*-value between 1 and 6 can be either an object or an " *Angabe* ":

O	Object	" *Angabe* "
1	*Er wiegt z w e i K i l o*	*Er läuft z w e i M e t e r*
2	*Er gedenkt d e s M o r g e n s*	*Er läuft d e s M o r g e n s*
3	*Er gibt m i r Wein*	*Er steigt m i r auf den Fuß* *Er trägt m i r den Koffer*
4	*Er rechnet m i t d e m V o r -* *t r a g des Bruders*	*Er verdient m i t d e m V o r -* *t r a g des Bruders*
5	*Er wohnt i m e r s t e n S t o c k*	*Er schläft i m e r s t e n S t o c k*
6	*Er geht i n s K i n o hinein*	*Er schläft i n d e n T a g hinein*

Thus we cannot derive any knowledge about a *NK*'s function from its grammatical description only. But semantical criteria can provide some solutions at this level of analysis:

(4)   *Er rechnet mit dem Vortrag seines Bruders.*
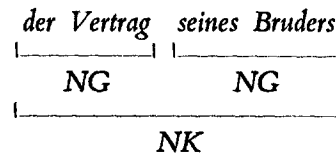
(4) contains the verb *rechnen mit* + 1 with the " *Satzbauplan* " *SBP* Ø4. But (4) contains also the verb *rechnen* with *SBP* Ø. As the semantic features of *Vortrag* tell us that *Vortrag* is no means of calculating, the *SBP* Ø will be excluded as an incorrect derivation.

(5)   *Er rechnet mit der Rechenmaschine.*

(5) contains the same verbs as (4). As *Rechenmaschine* is a means of calculating, the NK is also an " *Angabe* ". Therefore both *SBP* Ø and *SBP* Ø4 are possible analyses.

But we do not yet follow semantics. We try to analyze with syntactic means as long as possible.

If we combine two NGs in order to build a NK, the right NG serves as an attribute:

$$\text{der Vertrag} \quad \text{seines Bruders}$$

```
der Vertrag   seines Bruders
|_____|  |_____|
    NG            NG
|_____|
            NK
```

An attribute is a NG with one of the following O-values:

$O = 2$ .... genitive phrase: *der Vortrag s e i n e s B r u d e r s*
$O = 4$ .... prepositional phrase: *der Mann m i t d e m K r o p f*
$O = 5$ .... situative phrase: *der Baum i m G a r t e n*
$O = 6$ .... directional phrase: *die Fahrt i n d i e S t a d t*

The rules $RR1$ and $RR2$ define an algorithm which is to build NKs of NG-combinations. Further on, $RR3$ and $RR5$ change a NK into an object, and $RR4$ an NK into an " *Angabe* ".

L_i ... leftmost position of a NG_i or a NK_j
R_i ... rightmost position of a NG_i or a NK_j
O_i ... value of the " object class " of a NG_i or NK_j.

$RR1$: $NG_i \ (L_i, \ R_i, \ O_i) = NK_j \ (L_i, \ R_i, \ O_i)$

$RR2$: $NG_i \ (L_i, \ R_i, \ O_i) + NK_j \ (L_j, \ R_j, \ O_j) = NK_k \ (L_i, \ R_j, \ O_i)$
on condition that 1) $R_i + 1 = L_j$;
                             2) $O_j =$ either 2 or 4 or 5 or 6;
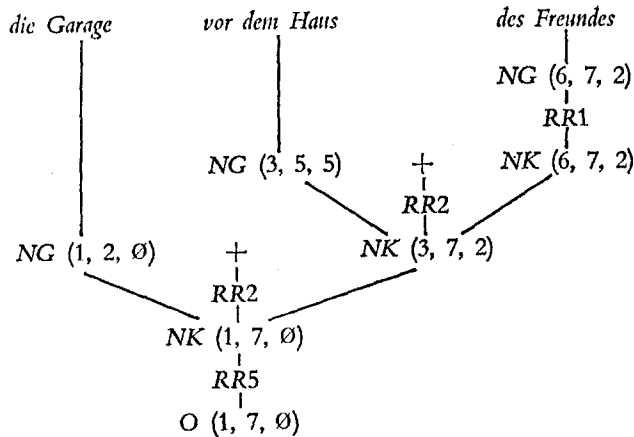   3) NG_i does not contain a preposition behind the noun.

$RR3$: $NK_j \ (L_j, \ R_j, \ O_j) = O_j \ (L_j, \ R_j, \ O_j)$ (i.e. a derivation of
             on condition that $O_j \neq$ Ø.   an object.)

$RR4$: $NK_j\ (L_j,\ R_j,\ O_j) = A_j\ (L_j,\ R_j,\ O_j)$ (i.e. a derivation of
on condition that $O_j \neq \emptyset$.   an " $Angabe$ ".)

$RR5$: $NK_j\ (L_j,\ R_j,\ \emptyset\ ) = O_j\ (L_j,\ R_j,\ \emptyset\ )$ (i.e. the derivation of
the subject.)

Consider an example of a possible application of $RR1$-$RR5$ in order to combine three NGs and thus make a single NK:

Exceptional case I: Prepositions in front of, and behind, a NG or NK.

(6) *um des Freundes willen*

(7) *um des Freundes in Paris willen*

Evidently (7) is *no* string of two NGs. Therefore, $RR1$ and $RR2$ are not applicable to this type of NK, and we must construct a special algorithm for the analysis of it.

Exceptional case II: Preposition behind a NG or a NK.

In German there are some prepositions which can appear in front of, or behind, a NG, for instance *nach, wegen,* ...

Er zittert wegen des Kampfes or *Er zittert des Kampfes wegen.*

When we consider (8), (9) and (10) we see that a postposition in a NG at the same time is a postposition of a NK.

(8)   *des Kampfes wegen*
(9)   *des Kampfes in der Sporthalle wegen*
(10)   *des Kampfes in der Sporthalle in Bonn wegen*

We need a special algorithm for the analysis of NKs of the type (8)–
(10). It must provide, for instance, the following analyses of (11):

(11)   *Er entsinnt sich des Beifalls wegen des Gesanges.*
        |_____| |_____|
        analysis *a*) ... NG (4, 5, 2)   NG (6, 8, 4)
        |_____| |_____|
        analysis *b*) ... NG (4, 6, 4)   NG (7, 8, 2)

It is possible to unite both NGs of *a*) into one NK by application of
RR1 and RR2. But in *b*) this is impossible; condition 3. of RR2 is not
met. Therefore both NGs in *b*) must be changed into separate NKs
by RR1.
    Another quite complicated example:

(12)   *Er entledigt sich des Freundes des Boxens wegen des Messers*
        |_____| |_____| |_____|
        NG (4, 5, 2) NG (6, 7, 2)   NG (8, 10, 4)
        |_____| |_____|
                              NK (6, 10, 2)
        |_____|
        analysis *a*) ...   NK (4, 10, 2)

(12)   *Er entledigt sich des Freundes des Boxens wegen des Messers*
        |_____| |_____| |_____|
        NG (4, 5, 2)   NG (6, 8, 4)   NG (9, 10, 2)
        |_____| |_____|
        analysis *b*) ...   NK (4, 8, 2)   NK (9, 10, 2)

RR1-RR5 and the special algorithms for exceptional case I and
II are the theoretical basis to find all possible combinations of NGs
and to derive objects and "*Angaben*" from NKs.
    Next we check the syntax for useful hints so as to achieve a correct
analysis of NKs of a given sentence, and hints so as to decide, which
NKs are objects and which are "Angaben".

In (13) and (14) we find NGs of the same structure. But the correct analysis of (13) differs completely from that of (14).

(13)   weil der Mann mit der Verletzung der Elle im linken Arm schläft
      |_____| |_____| |_____| |_____|
                                  :
     |_____|
                   NK $(2, 11, \varnothing)$
                        |
                   O $(2, 11, \varnothing)$

(14)   weil der Mann mit der Frau des Morgens im alten Cafe tanzt
      |_____| |_____| |_____| |_____|
    NK $(2, 3, \varnothing)$ NK $(4, 6, 4)$ NK $(7, 8, 2)$ NK $(9, 11, 5)$
        | | | |
    O $(2, 3, \varnothing)$ A $(4, 6, 4)$ A $(7, 8, 2)$ A $(9, 11, 5)$

As long as we do not find any hints, the rules must be applied in a systematic order, so as to have *all* possible analyses as a result. A semantic component then can choose the subset of reasonable analyses out of the whole set. Thus the analysis of (13) would have to be an analysis of (14), and vice versa.

It would need too much space to print the set of theoretically possible analyses of (15) here. The reader easily sees, in how many different ways RR1-RR5 can be applied. The whole set of possible analyses is greater than 100! We just show a single analysis. We note the rule being applied in the leftmost column. Now we try to use the " *Satzbauplan* " as a means of improving the exhaustive analysis-procedure. In (15) *wartete* is the single finite verb. It has the SBP $\varnothing$ 4 *auf* + 1. That means, it takes both an object in the nominative case and a prepositional object. Having found a certain SBP, our algorithm runs as follows:

    1. Look for a NG (, , $\varnothing$) and make it an object. (According to RR1, RR5.)

    2. Look for a NG (, , 4) which contains the preposition *auf*, and make it an object. (According to RR1, RR3.)

    3. Apply RR1, RR2 and RR4 to the NGs which are not objects.

    4. Try to analyze one of the NGs which are not objects as an attribute of an object. If this is possible, restart at 3.; if not, stop!

(15)

	weil der Mann	mit der Frau	des Morgens	in der Küche	auf den Bediensteten	wartete
RR1(NG₁)	NG₁(2,3,∅)	NG₂(4,6,4)	NG₃(7,8,2)	NG₄(9,11,5) NG₅(9,11,4)	NG₆(12,14,4) NG₇(12,14,6)	
RR1(NG₂)	NK₁(2,3,∅)					
RR1(NG₃)		NK₂(4,6,4)				
RR1(NG₄)			NK₃(7,8,2)	NK₄(9,11,5)		
RR1(NG₆)					NK₆(12,13,4)	
RR5(NK₁)	O(2,3,∅)					
RR4(NK₂)		A(4,6,4)				
RR4(NK₃)			A(7,8,2)			
RR4(NK₄)				A(9,11,5)		
RR3(NK₅)					O(12,14,4)	

(15) weil der Mann   mit der Frau   des Morgens   in der Küche   auf den Bediensteten   wartete

O(2,3,∅)   A(4,6,4)   A(7,8,2)   A(9,11,5)   O(12,14,4)

mit dem Kropf

O(2,6,∅)   mit dem Hut   A(7,8,2)   A(9,11,5)   O(12,14,4)

meines Vaters

O(2,8,∅)   A(9,11,5)   O(12,14,4)

meines Bruders   in der Schweiz

O(2,11,∅)   O(12,14,4)

mit dem Besitzer   des Hauses

O(2,3,∅)   A(4,8,4)   A(9,11,5)   O(12,14,4)

O(2,3,∅)   A(4,11,4)   O(12,14,4)

O(2,3,∅)   A(4,6,4)   A(7,11,2)   O(12,14,4)

O(2,6,∅)   A(7,11,2)   O(12,14,4)

(The last two analyses are not fine semantically, but maybe somebody else can find examples, for which these analyses are correct.)

In the case of (15), this algorithm produces no more than 8 analyses. That means that the *unshortenend* result of more than 100 analyses can be reduced to 8 by using an improved version of analysis. This shows that the *SBP* gives us the syntactical information which is necessary to make the correct analysis.

On the opposite page we show that all 8 analyses are possible with the *SBP Ø 4 auf + 1* and the structure of *NGs* as in (15).

Sometimes one can find more than one finite verb in a sentence:

(16)   *Er liebe   bestimmte nahe   Wälder*

$$\text{verb}_1 \quad \text{verb}_2 \qquad \text{verb}_3$$
$$\text{adj}_1 \quad \text{adj}_2 \qquad \text{adj}_3$$
$$\text{prep}_1$$
$$\text{adju}_1$$

In these cases one has to find out which of the verbs is the actual finite verb. In German the finite verb can appear at the first position in the sentence, at the second position, and at the end. It is easy to fill up the following matrix according to the verbs in (16):

verb	1st position	2nd posizion	end
liebe	—	+	—
bestimmte	—	+	—
nahe	—	+	—

Neither verb_1, verb_2 or verb_3 is standing at the first position or at the end of the sentence. From that we can conclude that each of them *must* be at the second position of the sentence. We still know another criterion to find the right verb: The string in front of it, at position 1, *must* be an *NK*. In (16) only verb_1 meets this condition. Therefore *liebe* is the finite verb.

Only if one knows a lot of syntactical heuristics like these, one is able to decide at which level of sentence-analysis semantics should be employed. We think there will be need for an interaction between a syntactical and semantical component.