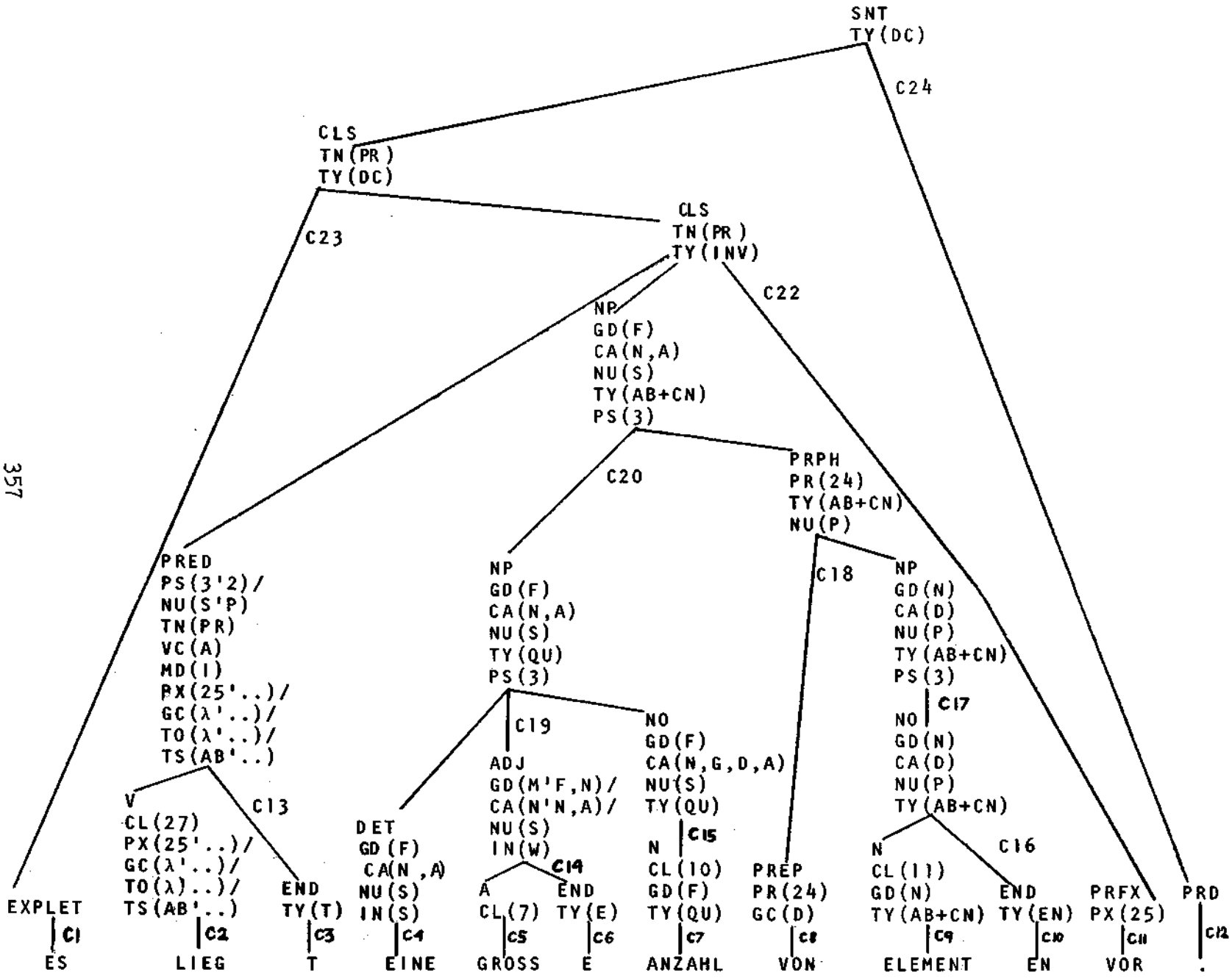
APPENDIX

Analysis of *Es liegt eine grosse Anzahl von Elementen vor.*

by Annette Stachowitz

Linguistics Research Center

The University of Texas at Austin


SNT
TY(DC)
C24
CLS
TN(PR)
TY(DC)
C23
CLS
TN(PR)
TY(INV)
C22
NP
GD(F)
CA(N,A)
NU(S)
TY(AB+CN)
PS(3)
C20
PRPH
PR(24)
TY(AB+CN)
NU(P)
PRED
PS(3'2)/
NU(S'P)
TN(PR)
VC(A)
MD(I)
PX(25'..)/
GC(λ'..)/
TO(λ'..)/
TS(AB'..)
NP
GD(F)
CA(N,A)
NU(S)
TY(QU)
PS(3)
C18
NP
GD(N)
CA(D)
NU(P)
TY(AB+CN)
PS(3)
C17
V
CL(27)
PX(25'..)/
GC(λ'..)/
TO(λ)..)/
TS(AB'..)
C13
END
TY(T)
C19
ADJ
GD(M'F,N)/
CA(N'N,A)/
NU(S)
IN(W)
C14
NO
GD(F)
CA(N,G,D,A)
NU(S)
TY(QU)
N
C15
NO
GD(N)
CA(D)
NU(P)
TY(AB+CN)
N
C16
DET
GD(F)
CA(N,A)
NU(S)
IN(S)
A
END
CL(7)
TY(E)
CL(10)
GD(F)
TY(QU)
PREP
PR(24)
GC(D)
CL(11)
GD(N)
TY(AB+CN)
END
TY(EN)
PRFX
PX(25)
PRD
EXPLET
C1
ES
C2
LIEG
C3
T
C4
EINE
C5
GROSS
C6
E
C7
ANZAHL
C8
VON
C9
ELEMENT
C10
EN
C11
VOR
C12
.

Rules:

```
C1      V EXPLET                        =       * ES

C2      V V                             =       * LIEG
        + CL(27)
        + PX(..'25'..)/
        + GC(..'λ'..)/
        + TØ(..'λ'..)/
        + TS(..'AB'..)

C2      V V                             =       * LAG
        + CL(9)
        + PX(..'25'..)/
        + GC(..'λ'..)/
        + TØ(..'λ'..)/
        + TS(..'AB'..)
```

(analogous C2 rules for the stems laeg, leg)

```
C3      V END                           =       * T
        + TY(T)

C4      V DET                           =       * EINE
        + GD(F)/
        + CA(N,A)/
        + NU(S)
        + IN(S)

C5      V A                             =       * GROSS
        + CL(7)

C6      V END                           =       * E
        + TY(E)

C7      V N                             =       * ANZAHL
        + CL(10)
        + GD(F)
        + TY(QU)

C8      V PREP                          =       * VON
        + PR(24)
        + GC(D)
```

```
C9    V N                 =   * ELEMENT
      + CL(11)
      + GD(N)
      + TY(AB+CN)

C10   V END               = * EN
      + TY(EN)

C11   V PRFX              = * VOR
      + PX(25)

C12   V PRD               = * .

C13   V PRED              = V V                V END
      + PS(3'2)/            $ CL(..,27)        $ TY(T)
      + NU(S'P)                                B
      + TN(PR)
      + VC(A)
      + MD(I)
      ^ 2

C14   V ADJ               = V A                V END
      + GD(M'F,N)/          $ CL(..,7)         $ TY(E)
      + CA(N'N,A)/                             B
      + NU(S)
      + IN(W)

C15   V NO                = V N
      + CA(N,G,D,A)/        $ CL(..,10)
      + NU(S)
      ^ 2

C16   V NO                = V N                V END
      + CA(D)               $ CL(..,11)        $ TY(EN)
      + NU(P)                                  B
      ^ 2

C17   V NP                = V NO
      + PS(3)               $ NU(P)
      $ 2.1NU
      ^ 2

C18   V PRPH              = V PREP             V NP
      ^ 2,3                 . 3.1GC            $ CA
                                               $ GD

C19   V NP                = V DET              V ADJ        V NØ
      + PS(3)               $ GD/              . 4.1GD/     $ GD
      $*2.4GD/              $ NU/              . 4.2NU/     $ NU/
      $*2.5NU/              $ CA               . 4.3CA      $ CA
      $*2.6CA               . 3.1,W2.1/        $ IN
      ^ 4                   . 3.2,W2.2/
                            . 3.3,W2.3
                            .*3.4IN
```

```
C20  V NP              =  V NP          V PRPH
     ^ 2,3                $ TY(QU)      $ PR(54)
                                        $ NU(P)

C22  V CLS             =  V PRED        V NP        V PRFX
     + TY(INV)            $ PX          . 2.2PS/    . 2.1PX
     $ 3.4                $ PS/         . 2.3NU
     $ 2.5                $ NU          . 2.4TY
                          $ TS          ? PRN
                          $ TN          $ CA(N)
                          $ VC
                          $ MD
                          $ GC(λ)
```

The subscript PRN in the NP constituent is added to the clause label only if NP dominates a pronoun:

```
V NP      =     V PRN
+ PRN           $ TY(PS)
^ 2
```

```
C23  V CLS             =  V EXPLET      V CLS
     ^ 3                                $ TY(INV)
                                        * PRN
```

(This rule specifies that a clause with inverted word order may only be preceded by an expletive es if its subject is not a personal pronoun: Es kommen drei Personen in Frage. But: * Es kommen sie in Frage.)

```
C24  V SNT             =  V CLS         V PRD
     $ 2.1                $ TY(DC)      B
```

This analysis may show the difficulties that have to be accounted for in the analysis of surface strings with context-free phrase structure rules. Apart from the problems of discontinuity of elements in the surface structure and of phrasal dictionary elements, the amount of information in lexical elements which is relevant for correct analysis and translation is extremely large. Almost every verb can have different readings (and translations) depending on which one of a (sometimes very large) number of selection

restrictions or feature packets it is associated with. (Feature packets may include separable prefixes, case government including prepositional objects governed, types of objects and subjects required, etc.). For example, the German verb liegen may be associated with 30 different feature packets, resulting in 30 different readings of which a few are shown here (these translations, with a few exceptions, are taken from Wildhagen and Héraucourt, German-English / English-German Dictionary, Vol. II German-English, Brandstetter Verlag, Wiesbaden, 1957):

1. liegen, intransitive, requiring a physical object as subject, with a locative adverb: to lie, to rest, to be located or situated;
2. liegen, governing a dative object which must be human and with a subject which must be abstract: to suit a.p., to appeal to sb.;
3. liegen, associated with the separable prefix an, with an inanimate concrete subject, governing a dative object or a prepositional object with the preposition an and an NP which must be concrete and inanimate: border on, be adjacent to;
4. liegen, with the separable prefix an, with a human subject and a human dative object: to entreat a.p.;
5. liegen, with the separable prefix bei, intransitive, with a concrete inanimate subject, with the auxiliary sein if used in the perfect tense: to be enclosed;
6. liegen, with the separable prefix danieder, intransitive, and with a human subject: to be lying ill;
7. liegen, with the separable prefix vor, intransitive, and with an abstract subject: exist.

The subscript format, in which the rules for this analysis are written, makes surface analysis possible

because of the following two characteristics:
a) Rule constituents are only subconfigurations of work space configurations, i.e. only the features relevant in a particular rule are mentioned in that rule while all others are disregarded. For example, rule C13 (p. 3) only states the condition that a verb stem must be classified as belonging to the paradigmatic class 27 in order to be concatenable with the verb ending -t, thus forming a predicate with the indicated features. The remaining properties of the verb (prefix, case government, type of object and subject required) are irrelevant in this concatenation rule and are merely "carried up the structural tree" by means of the operation specified by the symbols ∧2 on the left side of that rule.
b) Agreement and government are specified as set theoretical operations between the values of rule constituents. For example, rule C19 (p. 3) very generally states that in a German sentence the sequence determiner-adjective-nominal should be analyzed as a noun phrase provided that they agree in gender, number and case, and that the adjective and the determiner must not agree in type of inflection (weak or strong). These conditions are expressed by the operations specified in the second and following lines of each constituent of this rule. (All other features of the nominal head are not specifically mentioned in the rule and are simply carried up the tree.) Thus, very large numbers of rules can be represented by one rule in this subscript format. This makes it possible to incorporate and refer to the large amount of information necessary for analysis and translation in the dictionary and syntax of a surface grammar. Access to this information available in the surface string would be practically impossible with a context-free phrase structure grammar with simple symbols because of the unmanageable number of lexical classes and morphological

and syntactic rules building on these classes.

In spite of the greater economy of subscript rules, however, problems resulting from permutations of elements of phrasal and idiomatic expressions cannot be easily solved in surface analysis. For this reason, the analysis of sentences containing such elements is, in practice, performed in two steps at the LRC: surface analysis and standard analysis. In standard analysis the elements of phrasal and idiomatic expressions are re-ordered to a pre-determined standard order and are then treated as one single dictionary item, possibly with internal variable slots. A detailed description of standard analysis may be found in Research in German-English Machine Translation on Syntactic Level, Final Technical Report, RADC-TR-69-368, Volume II, August 1970.

The following is an explanation of the symbols used in the structural tree. The symbols are defined going from left to right in the sentence and from the bottom to the top of the tree.

Lexical level:

EXPLET = Expletive es; not a pronoun but rather a syntactically empty placeholder for the subject of the sentence.

V
CL(27)
PX(25'...)/
GC(λ'...)/
TO(λ'...)/
TS(AB'...) = This verb of paradigmatic class 27 may be used with any of a number of specified separable prefixes, among them prefix 25, which is the German prefix vor. If it is used in conjunction with this particular prefix, it is intransitive (governs case λ; semantic type of object λ) and takes a subject of the semantic class type abstract.

END TY(T)	=	Ending of type -t
DET GD(F) CA(N,A) NU(S) IN(S)	=	Determiner, gender feminine, ambiguous with respect to case, i.e. it may be considered nominative or accusative, number singular, strongly inflected.
A CL(7)	=	Adjective of paradigmatic class 7.
END TY(E)	=	Ending of type -e
N CL(10) GD(F) TY(QU)	=	Noun of paradigmatic class 10, gender feminine, type quantifier, i.e. a quantifying noun which may be followed by a von PRPH and then constitutes a modifier of the head noun in that PRPH.
PREP PR(24) GC(D)	=	The preposition is identified as preposition number 24 (von) and has the feature "governs case dative".
N CL(11) GD(N) TY(AB+CN)	=	A noun of the paradigmatic class 11, gender neuter, and semantic type abstract and countable.
END TY(EN)	=	Ending of the type -en.
PRFX PX(25)	=	This prefix is identified as prefix number 25 (vor).
PRD	=	The period is marked as being a marginal symbol, i.e. it constitutes the boundary of a word and of a sentence.

Morphological level:

PRED PS(3'2)/ NU(S'P) TN(PR) VC(A) MD(I) PX(25'...)/ GC(λ'...)/ TO(λ'...)/ TS(AB'...)	=	The predicate (finite verb) has all the features of the underlying verb stem: prefix, case government, type of object and subject required. (The feature CL

- paradigmatic class - is dropped because it is no longer relevant.) In addition, it has the features person and number which mark it as either 3rd person singular or 2nd person plural. (The apostrophe and slash establish this relation between the individual features.) It is also marked as: tense present, voice active, and mood indicative.

ADJ
GD(M'F,N)/
CA(N'N,A)
NU(S)
IN(W)

= With respect to gender and case, the inflected adjective is characterized as masculine nominative; or feminine or neuter nominative or accusative. In number it is singular; the inflection is weak.

NO
GD(F)
CA(N,G,D,A)
NU(S)
TY(QU)

= The inflected nominal has the same gender and type information as the dictionary noun entry and in addition has the tags number singular, case 4-way ambiguous, i.e. it is either nominative, genitive, dative, or accusative, depending on its environment.

NO
GD(N)
CA(D)
NU(P)
TY(AB+CN)

= Inflected nominal with the gender and type of the underlying noun stem, case dative, number plural.

Phrase level:

NP
GD(F)
CA(N,A)
NU(S)
TY(QU)
PS(3)

= The noun phrase has the gender, case, and number characteristics in which the underlying determiner, adjective and noun agree, namely feminine nominative or accusative singular; the type is that of the head noun; the NP is marked as 3rd person.

NP
GD(N)
CA(D)
NU(P)
TY(AB+CN)
PS(3)

= Noun phrase with all syntactic and semantic features of the underlying nominal, identified as 3rd person.

PRPH
PR(24)
TY(AB+CN)
NU(P)

= This prepositional phrase is identified as dominating preposition 24, i.e. von, and an NP with a head noun of type abstract and countable, number plural.

NP
GD(F)
CA(N,A)
NU(S)
TY(AB+CN)
PS(3)

= This noun phrase, which dominates an NP followed by a von PRPH, has the syntactic features of the dominated NP: gender feminine, case nominative or accusative, number singular, and the semantic features of the head noun of the dominated PRPH: type abstract and countable. It is also marked as an NP in the 3rd person.

Clause and sentence level:

CLS
TY(INV)
TN(PR)

= This clause is of the type with inverted word order; it may be followed by a "?" to form a question or, as in this sentence, it may be preceded by an expletive es to form a declarative sentence; its tense is present.

CLS
TY(DC)
TN(PR)

= A clause of type declarative, tense present.

SNT
TY(DC)

= A sentence of type declarative.