Foundations of Computational Linguistics

man-machine communication in natural language

ROLAND HAUSSER
Computational Linguistics
Universität Erlangen Nürnberg
Germany

CLUE

Part III Morphology and Syntax

13. Words and morphemes 205
13.1 Words and word forms 205
13.2 Segmentation and concatenation 211
13.3 Morphemes and allomorphs 215
13.4 Categorization and lemmatization 217
13.5 Methods of automatic word form recognition 221

14. Word form recognition in LA-Morph 225
14.1 Allo-rules 225
14.2 Phenomena of allomorphy 232
14.3 Left-associative segmentation into allomorphs 238
14.4 Combi-rules 241
14.5 Concatenation patterns 245

15. Corpus analysis 251
15.1 Implementation and application of grammar systems 251
15.2 Subtheoretical variants 254
15.3 Building corpora 258
15.4 Distribution of word forms 260

15.5 Statistical tagging 264

16. Basic concepts of syntax 268
16.1 Delimitation of morphology and syntax 268
16.2 Valency 270
16.3 Agreement 274
16.4 Free word order in German (*LA-D1*) 276
16.5 Fixed word order in English (*LA-E1*) 283

17. LA-syntax for English 286
17.1 Complex fillers in pre- and postverbal position 286
17.2 English field of referents 294
17.3 Complex verb forms 296
17.4 Finite state backbone of LA-syntax (*LA-E2*) 299
17.5 Yes/no-interrogatives (*LA-E3*) and grammatical perplexity 304

18. LA-syntax for German 309
18.1 Standard procedure of syntactic analysis 309
18.2 German field of referents (*LA-D2*) 313
18.3 Verbal positions in English and German 320
18.4 Complex verbs and elementary adverbs (*LA-D3*) 324
18.5 Interrogatives and subordinate clauses (*LA-D4*) 333

Part III
Morphology and Syntax

13. Words and morphemes

13.1 Words and word forms

13.1.1 Different syntactic compatibilities of word forms

*write
*writes
*wrote
John has written *a letter.*
*writing

13.1.2 Francis' & Kučera's 1982 definition of a graphic word

"A word is a string of continuous alphanumeric characters with space on either side; may include hyphens and apostrophes, but no other punctuation marks."

13.1.3 Combination principles of morphology

1. *Inflection* is the systematic variation of a word with which it can perform different syntactic and semantic functions, and adapt to different syntactic environments. Examples are learn, learn/s, learn/ed, and learn/ing.
2. *Derivation* is the combination of a word with an affix. Examples are clear/ness, clear/ly, and un/clear.
3. *Composition* is the combination of two or more words into a new word form. Examples are gas/light, hard/wood, over/indulge, and over-the-counter.

13.1.4 Definition of the notion *word*

Word $=_{def}$ {associated analyzed word forms}

13.1.5 Example of an analyzed word form

[wolves (PN) wolf]

13.1.6 Analysis of an inflecting word

word	*word forms*
wolf $=_{def}$	{[wolf (SN) wolf], [wolf's (GN) wolf], [wolves (PN) wolf], [wolves' (GN) wolf]}

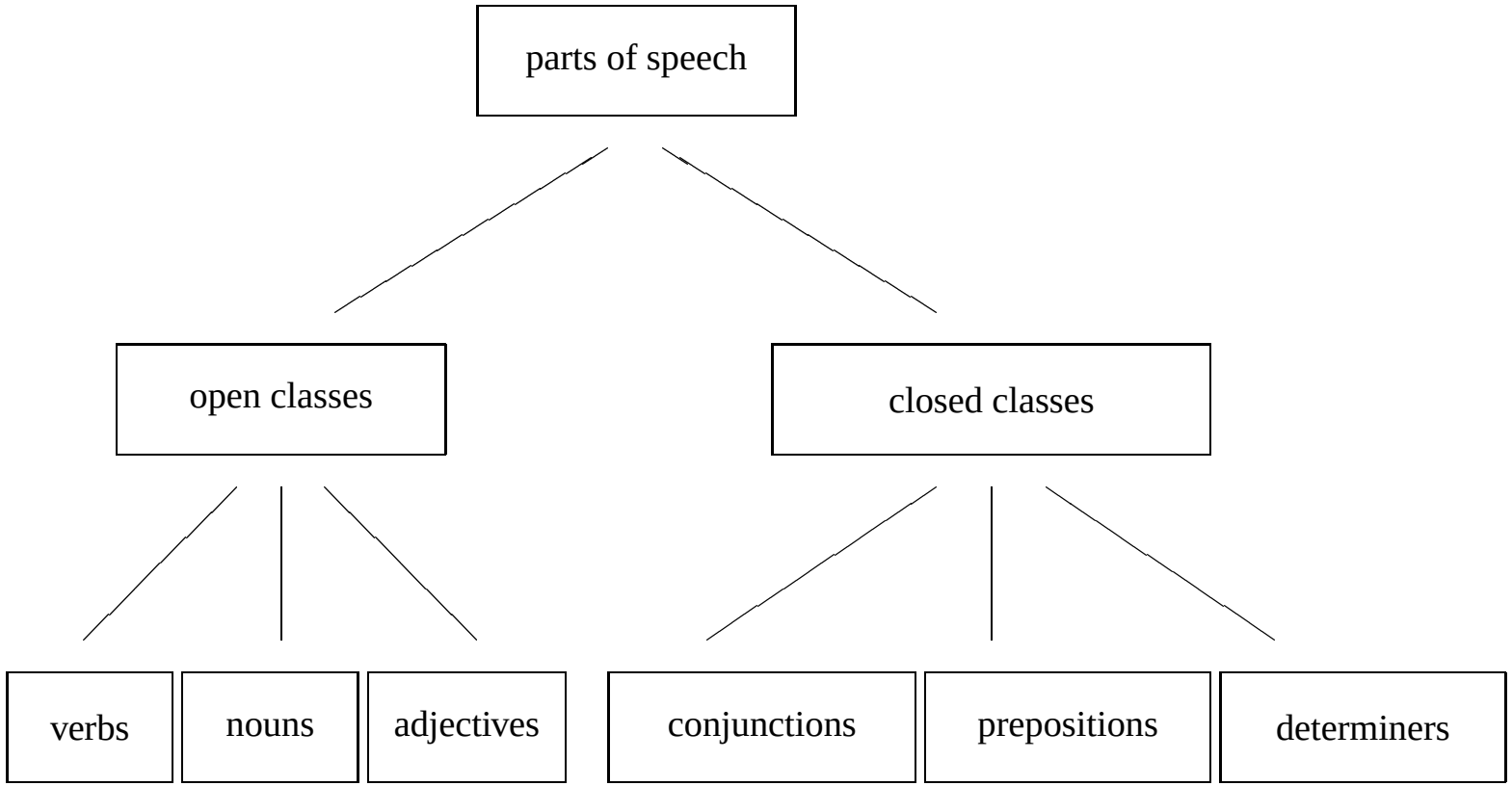
13.1.7 Analysis of a noninflecting word

word	*word forms*
and $=_{def}$	{ [and (cnj) and] }

13.1.8 Parts of speech

- *verbs*, e.g., walk, read, give, help, teach, . . .
- *nouns*, e.g., book, table, woman, messenger, arena, . . .
- *adjective-adverbials*, e.g., quick, good, low, . . .
- *conjunctions*, e.g., and, or, because, after, . . .
- *prepositions*, e.g., in, on, over, under, before, . . .
- *determiners*, e.g., a, the, every, some, all, any, . . .
- *particles*, e.g., only, already, just. . .

13.1.9 Classification of the parts of speech into open and closed classes



13.1.10 Comparison of the open and the closed classes

- The open classes comprise several 10 000 elements, while the closed classes contain only a few hundred words.
- The morphological processes of inflection, derivation, and composition are productive in the open classes, but not in the closed classes.
- In the open classes, the use of words is constantly changing, with new ones entering and obsolete ones leaving the current language, while the closed classes do not show a comparable fluctuation.

13.1.11 Parts of speech and types of signs

The elements of the open classes are also called *content words*, while the elements of the closed classes are also called *function words*. In this distinction, however, the sign type must be taken into consideration besides the category.

This is because only the *symbols* among the nouns, verbs, and adjective-adverbials are content words in the proper sense. *Indices*, on the other hand, e.g. the personal pronouns he, she, it etc., are considered function words even though they are of the category noun. Indexical adverbs like here or now do not even inflect, forming no comparatives and superlatives. The sign type *name* is also a special case among the nouns.

13.2 Segmentation and concatenation

13.2.1 Relation of words and their inflectional forms in German

	base forms	inflectional forms
nouns:	23 000	92 000
verbs:	6 000	144 000
adjective-adverbials:	11 000	198 000
	40 000	434 000

13.2.2 Number of noun-noun compositions

- length two: n^2
  Examples Haus/schuh, Schuh/haus, Jäger/jäger. This means that from 20 000 nouns 400 000 000 possible compounds of length 2 can be derived (base forms).
- length three: n^3
  Examples: Haus/schuh/sohle, Sport/schuh/haus, Jäger/jäger/jäger. This means that an additional 8 000 000 000 000 000 (eight thousand trillion) possible words may be formed.

13.2.3 Possible words, actual words, and neologisms

- Possible words
  Because there is no grammatical limit on the length of noun compounds, the number of possible word forms in German is infinite. These word forms exist potentially because of the inherent productivity of morphology.
- Actual words
  The set of words and word forms used by the language community within a certain interval of time is finite.
- Neologisms
  Neologisms are coined spontaneously by the language users on the basis of known words and the rules of word formation. Neologisms turn possible words into actual words.

13.2.4 Examples of neologisms in English

insurrectionist (inmate)	three-player (set)
copper-jacketed (bullets)	bad-guyness
cyberstalker	trapped-rat (frenzy)
self-tapping (screw)	dismissiveness
migraineur	extraconstitutional (gimmick)

13.2.5 Definition of the notion *morpheme*

morpheme $=_{def}$ {associated analyzed allomorphs}

13.2.6 Formal analysis of the morpheme wolf

morpheme	*allomorphs*
wolf $=_{def}$	{[wolf (SN SR) wolf],
	[wolv (PN SR) wolf]}

13.2.7 Comparing morpheme and word wolf

morpheme	*allomorphs*
wolf $=_{def}$	{wolf,
	wolv}

word	*word forms*
wolf $=_{def}$	{wolf,
	wolf/'s,
	wolv/es,
	wolv/es/'}

13.2.8 Alternative forms of segmentation

allomorphs:	learn/ing
syllables:	lear/ning
phonemes:	l/e/r/n/i/n/g
letters:	l/e/a/r/n/i/n/g

13.3 Morphemes and allomorphs

13.3.1 The regular morpheme learn

morpheme	*allomorphs*
learn $=_{def}$	{[learn (N . . . V) learn]}

13.3.2 The irregular morpheme swim

morpheme	*allomorphs*
swim $=_{def}$	{[swim (N . . . V1) swim], [swimm (. . . B) swim], [swam (N . . . V2) swim], [swum (N . . . V) swim]}

13.3.3 An example of suppletion

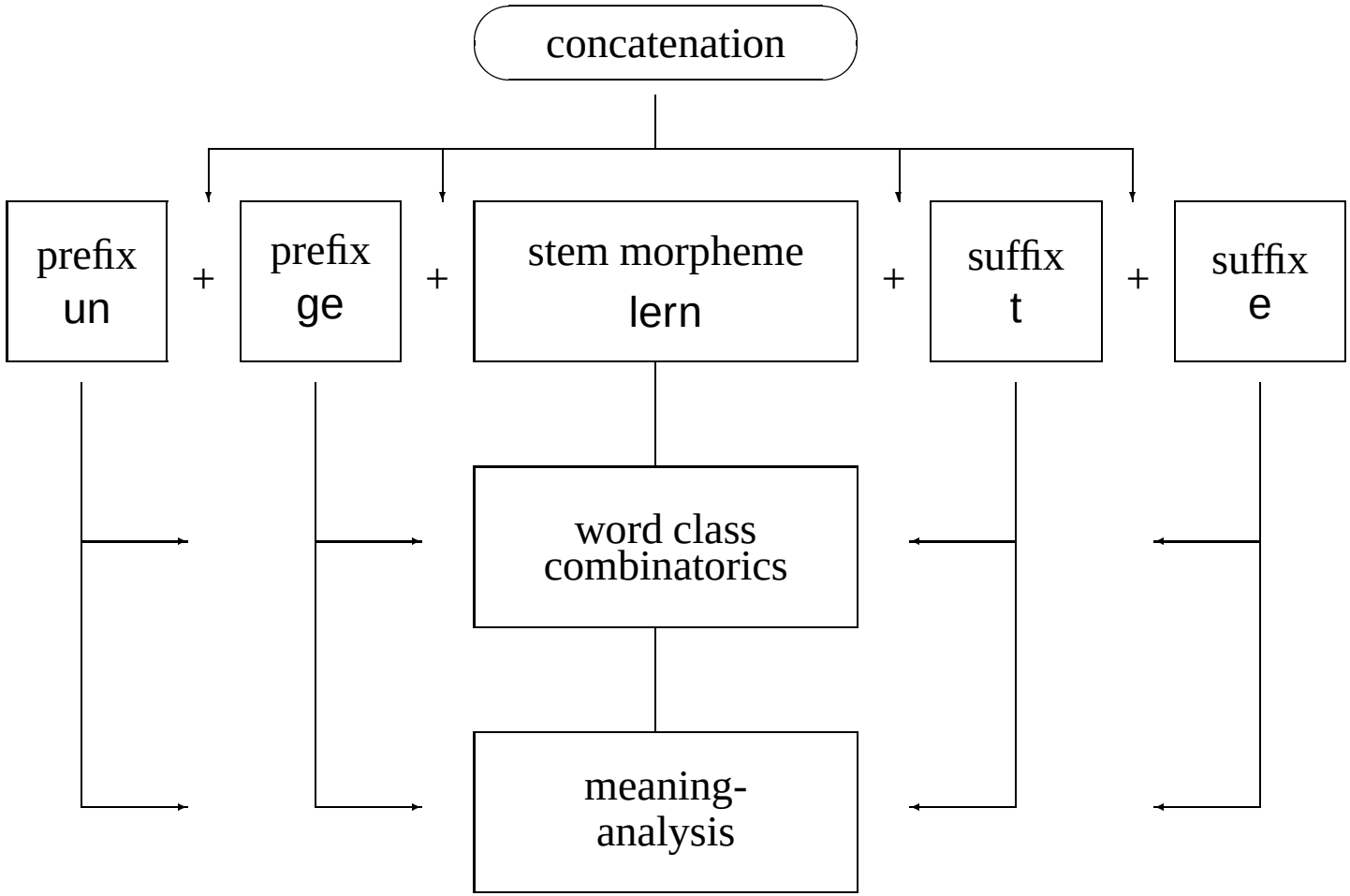
morpheme	*allomorphs*
good $=_{def}$	{[good (ADV IR) good], [bett (CAD IR) good], [b (SAD IR) good]}

13.3.4 Example of a bound morpheme (hypothetical)

morpheme	*allomorphs*
-s $=_{def}$	{[s (PL1) plural], [es (PL2) plural], [en (PL3) plural], [# (PL4) plural]}

13.4 Categorization and lemmatization

13.4.1 Morphological analysis of ungelernte



13.4.2 Schematic derivation in LA-grammar

```
("un" (CAT1) MEAN-a) + ("ge" (CAT2) MEAN-b)
   ("un/ge" (CAT3) MEAN-c) + ("lern" (CAT4) MEAN-d)
      ("un/ge/lern" (CAT5) MEAN-e) + ("t" (CAT6) MEAN-f)
         ("un/ge/lern/t" (CAT7) MEAN-g) + ("e" (CAT8) MEAN-h)
            ("un/ge/lern/t/e" (CAT9) MEAN-i)
```

13.4.3 Components of word form recognition

- *On-line lexicon*
  For each element (e.g. morpheme) of the natural language there must be defined a lexical analysis which is stored electronically.
- *Recognition algorithm*
  Using the on-line lexicon, each unknown word form (e.g. wolves) must be characterized automatically with respect to categorization and lemmatization:
  - *Categorization*
    consists in specifying the part of speech (e.g. noun) and the morphosyntactic properties of the surface (e.g. plural); needed for syntactic analysis.
  - *Lemmatization*
    consists in specifying the correct base form (e.g. wolf); provides access to the corresponding lemma in a semantic lexicon.

13.4.4 Basic structure of a lemma

[surface (lexical description)]

13.4.5 Lemma of a traditional dictionary (*excerpt*)

[1]**wolf** \'wu̇lf\ *n. pl* **wolves** \'wu̇lvz\ *often attributed* [ME, fr. OE *wulf*; akin to OHG *wolf*, L *lupus*, Gk *lykos*] **1** *pl also* **wolf** **a:** any of various large predatory mammals (genus *Canis* and exp. *C. lupus*) that resemble the related dogs, are destructive to game and livestock, and may rarely attack man esp. when in a pack – compare COYOTE, JACKAL **b:** the fur of a wolf ...

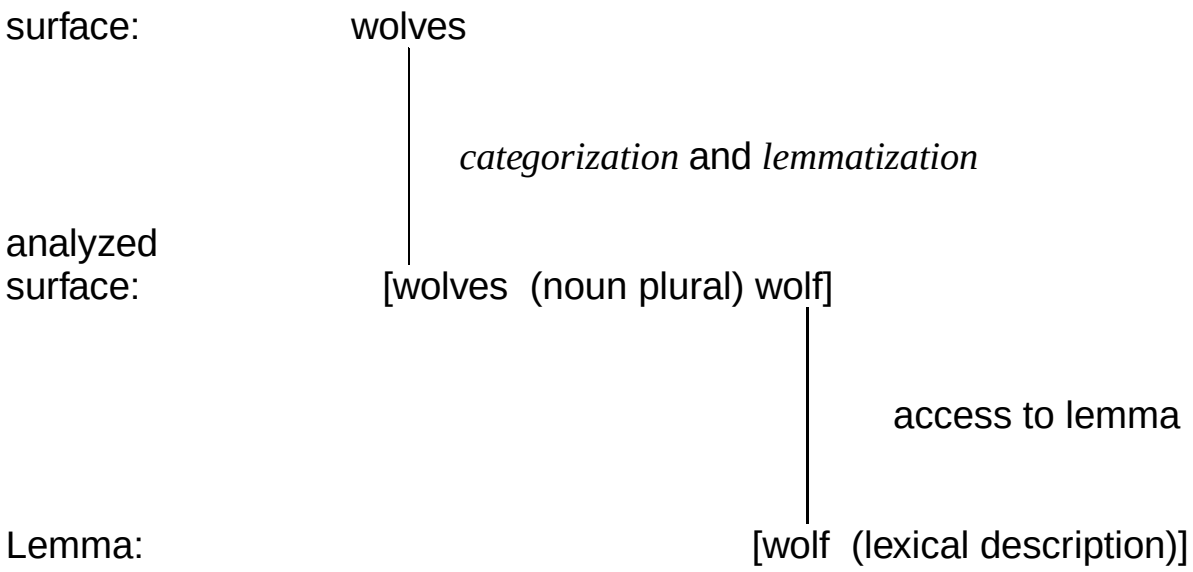
13.4.6 Matching a surface onto a key

word form surface: wolf

matching

lemma: [wolf (lexical description)]

13.4.7 Two-step procedure of word form recognition



13.4.8 Reason for the Two-step procedure

In the natural languages

- the number of word forms is considerably larger than the number of words, at least in inflectional and agglutinating languages, and
- the lexical lemmata normally define words rather than word forms,

13.5 Methods of automatic word form recognition

13.5.1 Word form method

Based on a lexicon of analyzed word forms.

13.5.2 Analyzed word form as lexical lemma

[wolves (part of speech: Subst, num: Pl, case: N,D,A, base form: wolf)]

Categorization and lemmatization are not handled by rules, but solely by the lexical entry.

13.5.3 Advantages and disadvantages of the word form method

- Advantage
  Allows for the simplest recognition algorithm because the surface of the unknown word form, e.g. wolves, is simply matched whole onto the corresponding key in the analysis lexicon.
- Disadvantages
  The production of the analysis lexicon is costly, its size is extremely large, and there is no possibility to recognize neologisms.

13.5.4 Morpheme method

Based on a lexicon of analyzed morphemes.

13.5.5 Schema of the morpheme method

surface:	wolves	
	| |	*segmentation*
allomorphs:	wolv/es	
	⇓ ⇓	*reduction*
morphemes:	wolf+s	*base form lookup* and *concatenation*

(1) segmentation into allomorphs, (2) reduction of allomorphs to the morphemes, (3) recognition of morphemes using an analysis lexicon, and (4) rule-based concatenation of morphemes to derive analyzed word form.

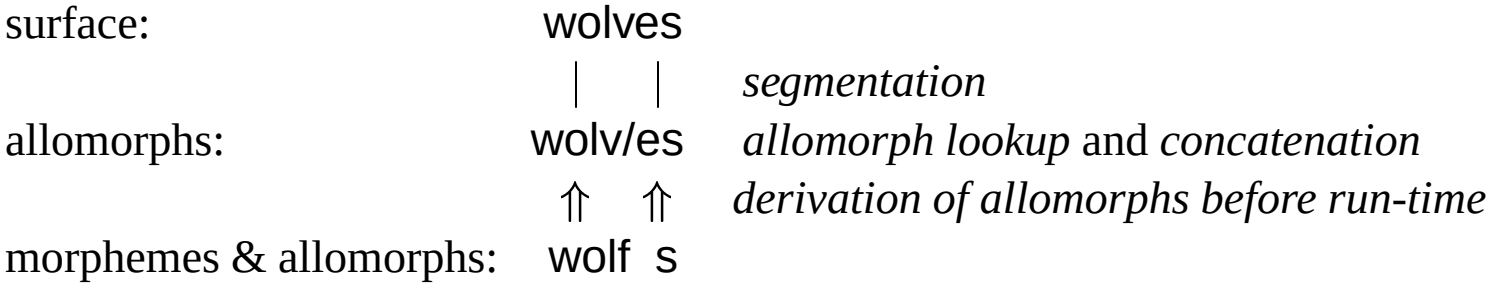
13.5.6 Advantages and disadvantages of the morpheme method

- Advantages
  Uses the smallest analysis lexicon. Neologisms may be analyzed and recognized during run-time using a rule-based segmentation and concatenation of complex word forms into their elements (morphemes).
- Disadvantages
  A maximally complex recognition algorithm ($\mathcal{NP}$ complete).

13.5.7 Allomorph method

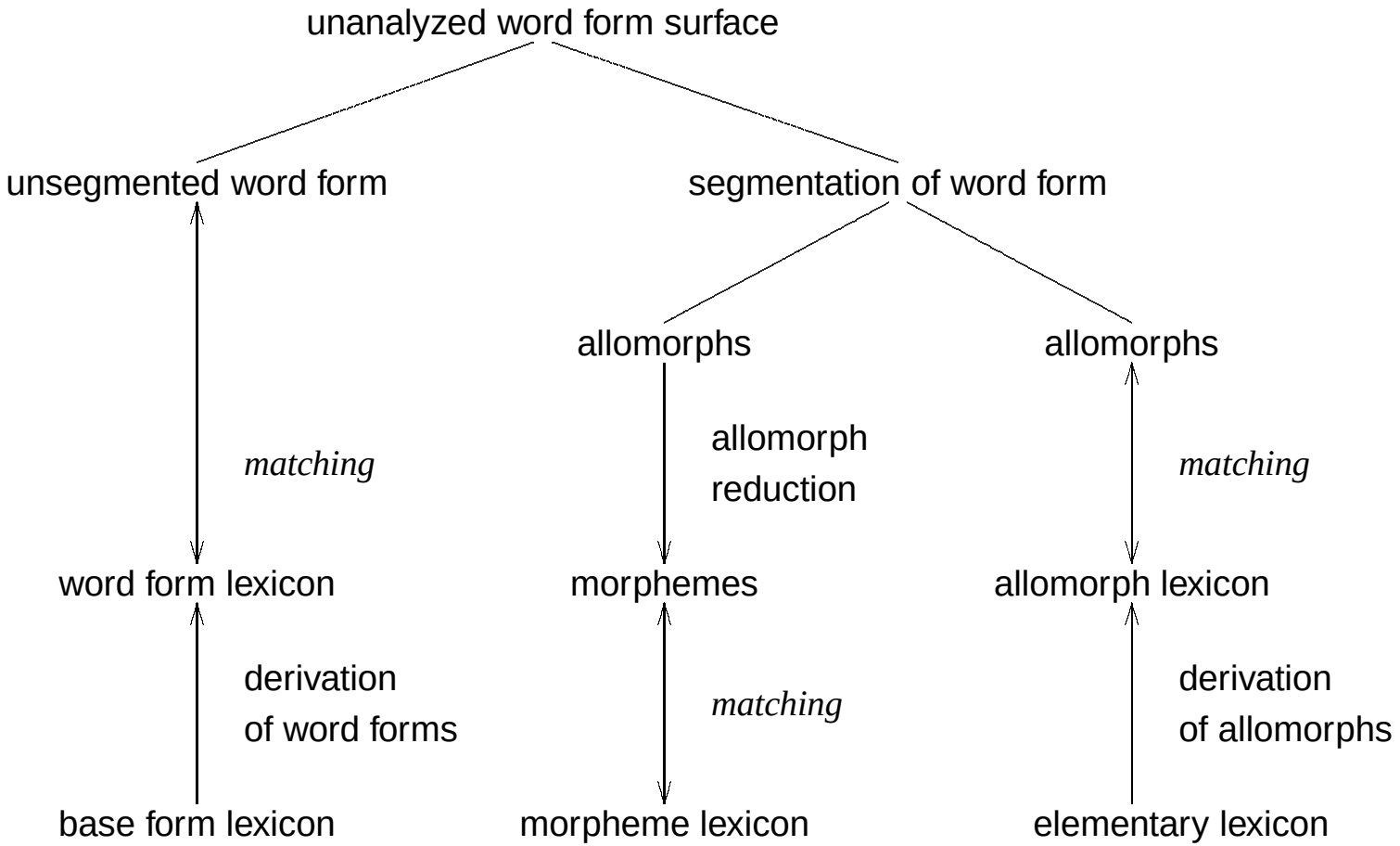
Based on a lexicon of elementary base forms, from which a lexicon of analyzed allomorphs is derived before run by means of allo-rules..

13.5.8 Schema of the allomorph method



During run-time, the allomorphs of the allomorph lexicon are available as precomputed, fully analyzed forms, providing the basis for a maximally simple segmentation: the unknown surface is matched from left to right with suitable allomorphs – without any reduction to morphemes. Concatenation takes place on the level of analyzed allomorphs by means of combi-rules.

13.5.9 Schematic comparison of the three basic methods

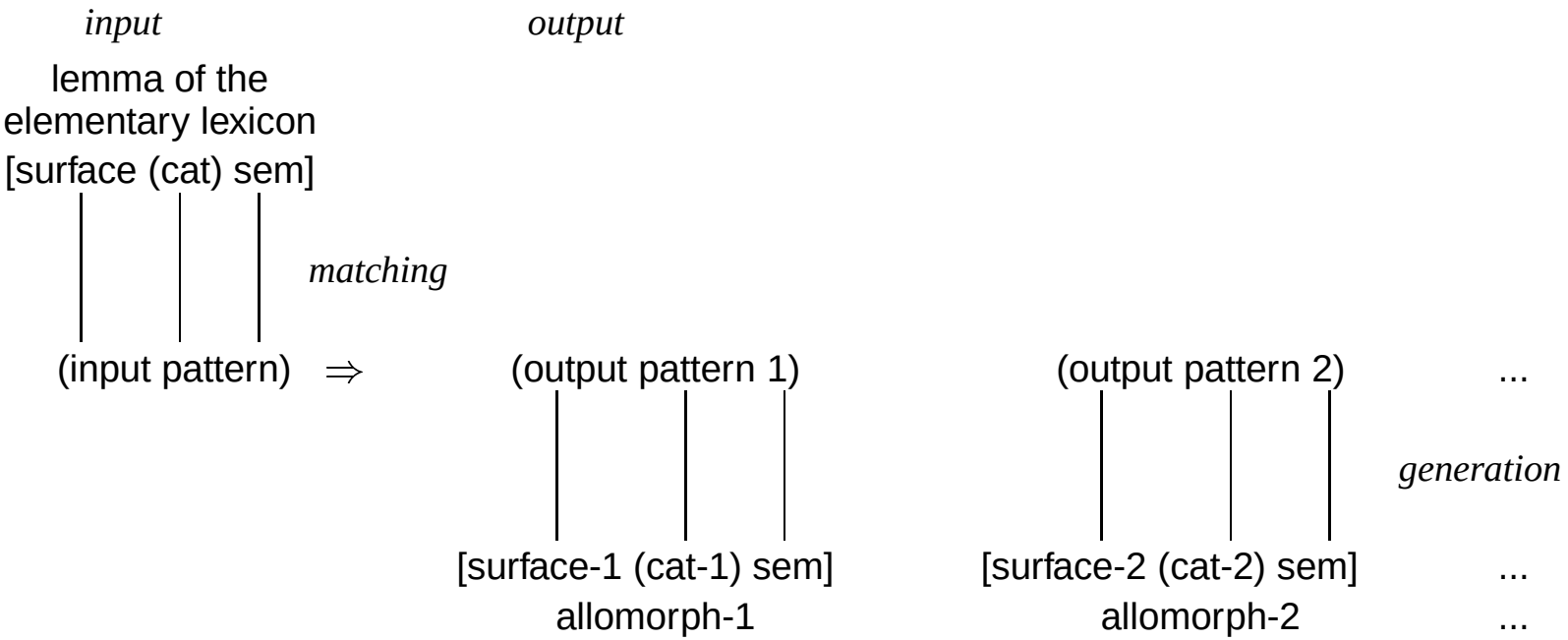


(1) word form method *(2) morpheme method* *(3) allomorph method*

14. Word form recognition in LA-Morph

14.1 Allo-rules

14.1.1 Abstract format of an allo-rule



14.1.2 Example of a base form lemma

```
("derive" (nom a v) derive)
```

14.1.3 Result of applying allo-rules to base form lemma

```
("derive" (sr nom a v) derive)
("deriv" (sr a v) derive)
```

14.1.4 Base form entry of schlafen

```
("schla2fen" (KV VH N GE  {hinueber VS GE } {durch VH A GE }
             {aus VH GE } {ein VS GE }\$ <be VH A GE- >
             <ent VS GE- > <ueber VH A GE- > <ver VH A GE- >)
             schlafen)
```

14.1.5 Output of allo-rules for schlafen

```
("schlaf" (IV V1 VH N GE { hinüber VS GE } { durch VH A GE }
           { aus VH GE } { ein VS GE } $ < be VH A GE- >
           < ent VS GE- > < über VH A GE- > < ver VH A GE- > )
           schlafen)
("schläf" (IV V2 _0 N GE { hinüber VS GE } { durch VH A GE }
           { aus VH GE } { ein VS GE } $ < be VH A GE- >
           < ent VS GE- > < über VH A GE- > < ver VH A GE- > )
           schlafen)
("schlief" (IV V34 _0 N GE { hinüber VS GE } { durch VH A GE }
           { aus VH GE } { ein VS GE } $ < be VH A GE- >
           < ent VS GE- > < über VH A GE- > < ver VH A GE- > )
           schlafen_i)
```

14.1.6 The word forms of schlafen (excerpt)

```
("schlaf/e" (S1 {hinüber}{durch A}{aus}{ein} V) schlafen_p)
("schlaf/e" (S13 {hinüber} {durch A} {aus} {ein} V ) s._k1)
("schlaf/e/n" (P13 {hinüber} {durch A} {aus} {ein} V ) s._pk1)
("schlaf/e/st" (S2 {hinüber} {durch A} {aus} {ein} V ) s._k1)
("schlaf/e/t" (P2 {hinüber} {durch A} {aus} {ein} V ) s._k1)
("schlaf/t" (P2 {hinüber} {durch A} {aus} {ein} V ) s._p)
("schlaf/end" (GER ) schlafen)
("schlaf/end/e" (E ) schlafen)
("schlaf/end/en" (EN ) schlafen)
("schlaf/end/er" (ER ) schlafen)
("schlaf/end/es" (ES ) schlafen)
("schlaf/end/em" (EM ) schlafen)
("schlaf/e/st" (S2 {hinüber} {durch A} {aus} {ein} V ) s._k1)
("schlaf/e/t" (P2 {hinüber} {durch A} {aus} {ein} V ) s._k1)
("schläf/st" (S2 {hinüber} {durch A} {aus} {ein} V ) s._p)
("schläf/t" (S3 {hinüber} {durch A} {aus} {ein} V ) s._p)
("schlief" (S13 {hinüber} {durch A} {aus} {ein} V ) s._i)
("schlief/e" (S13 {hinüber} {durch A} {aus} {ein} V ) s._k2)
("schlief/en" (P13 {hinüber} {durch A} {aus} {ein} V ) s._ik2)
```

```
("schlief/est" (S2 {hinüber} {durch A} {aus} {ein} V ) s._ik2)
("schlief/et" (P2 {hinüber} {durch A} {aus} {ein} V ) s._ik2)
("schlief/st" (S2 {hinüber} {durch A} {aus} {ein} V ) s._ik2)
("schlief/t" (P2 {hinüber} {durch A} {aus} {ein} V ) s._i)
("ge/schlaf/en" (H) schlafen)
("ge/schlaf/en/e" (E) schlafen)
("ge/schlaf/en/en" (EN) schlafen)
("ge/schlaf/en/es" (ES) schlafen)
("ge/schlaf/en/er" (ER) schlafen)
("ge/schlaf/en/em" (EM) schlafen)

("aus/schlaf/e" (S1 V) ausschlafen_pk1)
("aus/schlaf/e" (S13 V ) ausschlafen_k1)
("aus/schlaf/en" (P13 A V ) ausschlafen_pk1)
   ...
("aus/schläf/st" (S2 V) ausschlafen_p)
("aus/schläf/t" (S3 V) ausschlafen_p)
   ...
```

14.1.7 Four degrees of regularity in LA-Morph

- *Regular* inflectional paradigm
  The paradigm is represented by one lemma without any special surface markings, from which one allomorph is derived, e.g. learn ⇒ learn, or book ⇒ book.
- *Semi-regular* inflectional paradigm
  The paradigm is represented by one lemma without any special surface markings, from which more than one allomorph is derived, e.g. derive ⇒ derive, deriv, or wolf ⇒ wolf, wolv.
- *Semi-irregular* inflectional paradigm
  The paradigm is represented by one lemma with a special surface marker, from which more than one allomorph is derived, e.g. swIm ⇒ swim, swimm, swam, swum.
- *Irregular* inflectional paradigm
  The paradigm is represented by several lemmata for suppletive allomorphs which pass through the default rule, e.g. go ⇒ go, went ⇒ went, gone ⇒ gone. The allomorphs serve as input to general combi-rules, as in go/ing.

14.1.8 Tabular presentation of the degrees of regularity

	one lemma per paradigm	lemma without markings	one allomorph per lemma
regular	yes	yes	yes
semi-regular	yes	yes	no
semi-irregular	yes	no	no
irregular	no	no	yes

14.2 Phenomena of allomorphy

14.2.1 Allomorphs of semi-regular nouns

LEX	ALLO1	ALLO2
wolf	wolf	wolv
knife	knife	knive
ability	ability	abiliti
academy	academy	academi
agency	agency	agenci
money	money	moni

14.2.2 Allomorphs of semi-irregular nouns

LEX	ALLO1	ALLO2
analysis	analysis	analyses
larva	larva	larvae
stratum	stratum	strati
matrix	matrix	matrices
thesis	thesis	theses
criterion	criterion	criteria

tempo	tempo	tempi
calculus	calculus	calculi

14.2.3 Allomorphs of semi-regular verbs

LEX	ALLO1	ALLO2
derive	derive	deriv
dangle	dangle	dangl
undulate	undulate	undulat
accompany	accompany	accompani

14.2.4 Allomorphs of semi-irregular verbs

LEX	ALLO1	ALLO2	ALLO3	ALLO4
swIm	swim	swimm	swam	swum
rUN	run	runn	ran	run
bET	bet	bett	bet	bet

14.2.5 Allomorphs of semi-regular adjective-adverbials

LEX	ALLO1	ALLO2
able	able	abl
happy	happy	happi
free	free	fre
true	true	tru

14.2.6 Definition of the allomorph quotient

The allomorph quotient is the percentage of additional allomorphs relative to the number of base form entries.

14.2.7 The allomorph quotient of different languages

Italian: 37%
German: 31%
English: 8,97%

14.2.8 Compounds with ‘pseudo-’ contained in Webster’s New Collegiate Dictionary

pseudoclassic
pseudopregnancy
pseudosalt
pseudoscientific
etc.

14.2.9 Compounds with ‘pseudo-’ not contained in Webster’s New Collegiate Dictionary

pseudogothic
pseudomigrane
pseudoscientist
pseudovegetarian
etc.

14.2.10 Problem for recognition algorithm

In order to recognize the highly productive compositions involving the prefix pseudo, the LA-Morph system must provide a general rule-based analysis. As a consequence, the word forms in 14.2.8, are analyzed as ambiguous whereby the second reading stems from the compositional analysis based on the known forms, e.g. pseudo and classic.

14.2.11 Solution I

Automatic removal of all non-elementary base forms from the on-line lexicon.

14.2.12 Solution II

Leaving the non-elementary base forms like 14.2.8 in the lexicon, but selecting the most likely reading after the word form analysis.

14.2.13 Solution III

Using two lexica. One is an elementary lexicon which does not contain any non-elementary base forms. It is used for the categorization and lemmatization of word forms.

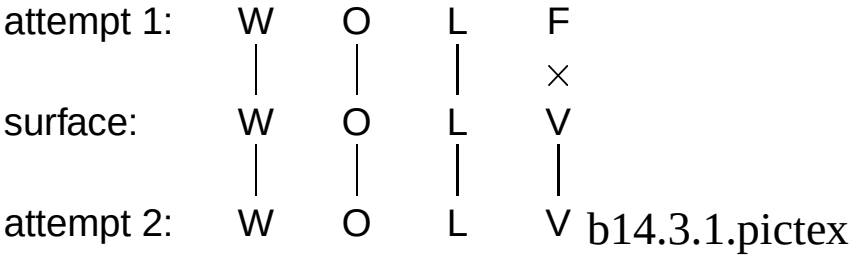
The other is a base form lexicon of content words. It assigns semantic representations to base forms including composita and derivata established in use. During word form analysis the two lexica are related by matching the result of lemmatization onto a corresponding – if present – key word of the base form lexicon (cf. 13.4.7).

14.2.14 Example of solution III

The compositional analysis of kin/ship would be matched onto kinship in the non-elementary base form lexicon, accessing the proper semantic description. In this way, (i) maximal data coverage – including neologisms – is ensured by a rule based analysis, (ii) the possibility of noncompositional meanings is accounted for, and (iii) unnecessary ambiguities are avoided.

14.3 Left-associative segmentation into allomorphs

14.3.1 Left-associative letter by letter matching



14.3.2 Hypothetical examples of English allowing alternative segmentations

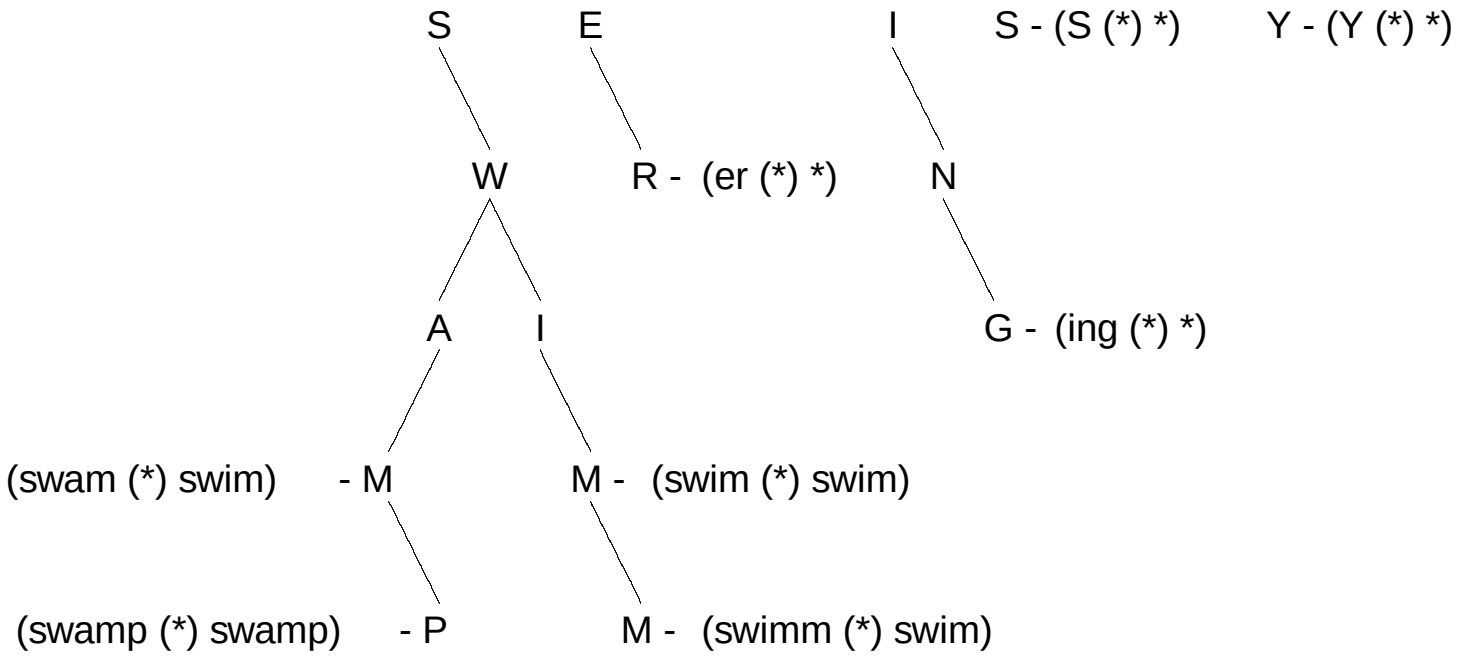
coverage	grandparent	history	lamp/light	land/s/end
cover/age	grandpa/rent	hi/story	lam/plight	land/send
cove/rage		his/tory		

rampage	rampart	scar/face	sing/able	war/plane
ramp/age	ramp/art	scarf/ace	sin/gable	warp/lane
ram/page	ram/part			

14.3.3 Alternative segmentations of a word form in German

surface:	Staubecken	Staubecken
segmentation:	Stau/becken	Staub/ecke/n
translation:	*reservoir*	*dust corners*

14.3.4 Storing allomorphs in a trie structure



14.3.5 Possibilities after finding an entry in the trie structure

- There are no letters left in the surface of the unknown word form, e.g. SWAM. Then the program simply returns the analysis stored at the last letter, here M.
- There are still letters left in the surface of the unknown word form. Then one of the following alternatives applies:
  - The allomorph found so far *is part* of the word form, as swim in SWIMS. Then the program (i) gives the lexical analysis of swim to the combi-rules of the system and (ii) looks for the next allomorph (here s), starting again from the top level of the trie structure.
  - The allomorph found so far *is not part* of the word form, as swam in SWAMPY. In this case the program continues down the trie structure provided there are continuations. In our example, it will find swamp.

  Because it becomes apparent only at the very end of a word form which of these two possibilities applies – or whether they apply simultaneously in the case of an ambiguity – they are pursued simultaneously by the program.

14.4 Combi-rules

14.4.1 Structure of combi-rules

$$\begin{array}{lll} & \textit{input} & \textit{output} \\ \mathrm{r}_n\text{: (pattern of start)} & \text{(pattern of next)} & \Rightarrow \ \mathrm{rp}_n\text{ (pattern of new start)} \end{array}$$

14.4.2 Difference between allo- and combi-rules

Combi-rules differ from allo-rules in that they are defined for different domains and different ranges:

An *allo-rule* takes a lexical entry as input and maps it into one or more allomorphs.
A *combi-rule* takes a word form start and a next allomorph as input and maps it into a new word form start.

14.4.3 Tasks of combi-rules

The combi-rules ensure that

1. the allomorphs found in the surface are not combined into ungrammatical word forms,
   e.g. *swam+ing or *swimm+s (input condition),
2. the surfaces of grammatical allomorph combinations are properly concatenated,
   e.g. swim+s $\Rightarrow$ swims,
3. the categories of the input pair are mapped into the correct result category,
   e.g. (NOM V) + (SX S3) $\Rightarrow$ (S3 V),
4. the correct result is formed on the level of semantic interpretation, and
5. after a successful rule application the correct rule package for the next combination is activated.

14.4.4 Derivation of unduly in LA-Morph

```
 1 +u [NIL . NIL]
 2  +n [NIL . (un (PX PREF) UN)]
 RP:{V-START N-START A-START P-START}; fired: P-START
 3   +d [(un (PX PREF) UN) . (d (GG) NIL)]
     +d [NIL . NIL]
 4    +u [(un (PX PREF) UN) . (du (SR SN) DUE (SR ADJ-V) DUE)]
 RP:{PX+A UN+V}; fired: PX+A
      +u [NIL . NIL]
 5 L [(un+du (SR ADJ) DUE) . (l (GG) NIL (ABBR) LITER)]
 RP:{A+LY}; fired: none
       +l [(un (PX PREF) UN) . NIL]
       +l [NIL . NIL]
 6      +y [(un+du (SR ADJ) DUE) . (ly (SX ADV) LY)]
 RP:{A+LY}; fired: A+LY
("un/du/ly" (ADV) due)
```

14.4.5 Handling of ungrammatical input in LA-Morph

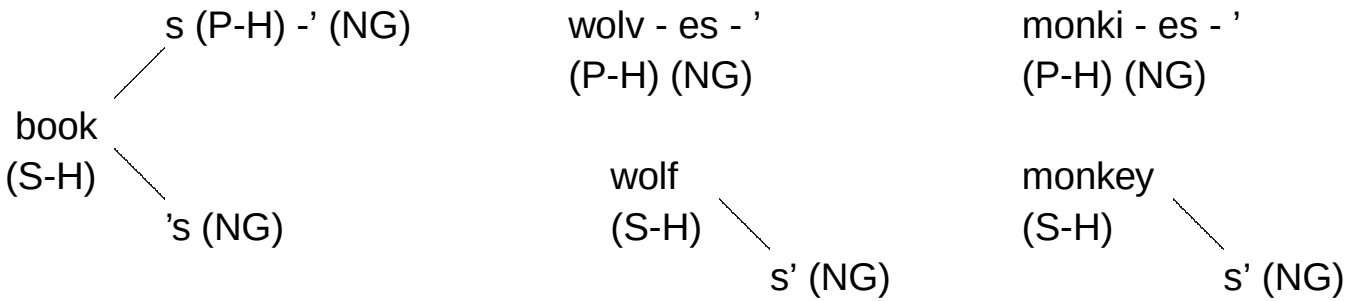
```
 1 +a [NIL . (a (SQ) A)]
 2  +b [NIL . NIL]
 3   +l [NIL . (abl (SR ADJ-A) ABLE)]
 RP:{V-START N-START A-START P-START}; fired: A-START
 4    +e [(abl (SR ADJ) ABLE) . NIL]
      +e [NIL . (able (ADJ) ABLE)]
 RP:{V-START N-START A-START P-START}; fired: none
 5     +l [(abl (SR ADJ) ABLE) . NIL]
ERROR
Unknown word form: "ablely"
NIL
```

14.4.6 Parsing the simplex undulate

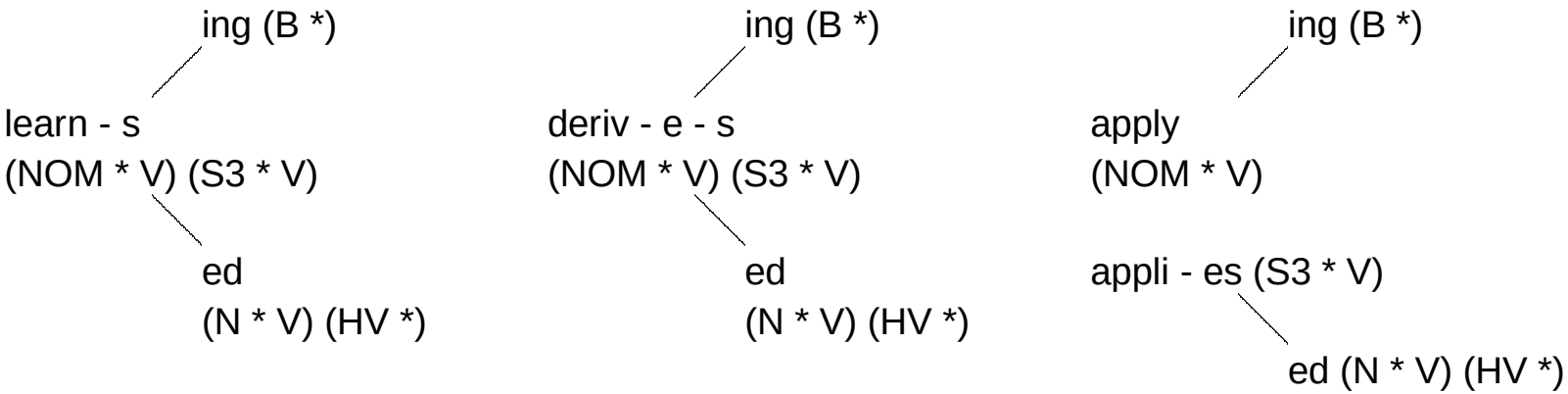
```
 1 +u [NIL . NIL]
 2  +n [NIL . (un (PX PREF) UN)]
 RP:{V-START N-START A-START P-START}; fired: P-START
 3   +d [(un (PX PREF) UN) . (d (GG) NIL)]
     +d [NIL . NIL]
 4    +u [(un (PX PREF) UN) . (du (SR SN) DUE (SR ADJ-V) DUE)]
 RP:{PX+A UN+V}; fired: PX+A
      +u [NIL . NIL]
 5     +l [(un+du (SR ADJ) DUE) . (l (GG) NIL (ABBR) LITER)]
 RP:{A+LY}; fired: none
       +l [(un (PX PREF) UN) . NIL]
       +l [NIL . NIL]
 6      +a [(un+du (SR ADJ) DUE) . NIL]
        +a [NIL . NIL]
 7       +t [(un+du (SR ADJ) DUE) . NIL]
         +t [NIL . (undulat (SR A V) UNDULATE)]
 RP:{V-START N-START A-START P-START}; fired: V-START
 8        +e [(un+du (SR ADJ) DUE) . (late (ADJ-AV) LATE (ADV) LATE)]
 RP:{A+LY}; fired: none
          +e [(undulat (SR A V) UNDULATE) . NIL]
          +e [NIL . (undulate (SR NOM A V) UNDULATE)]
 RP:{V-START N-START A-START P-START}; fired: V-START
("undulate" (NOM A V) UNDULATE)
```

14.5 Concatenation patterns

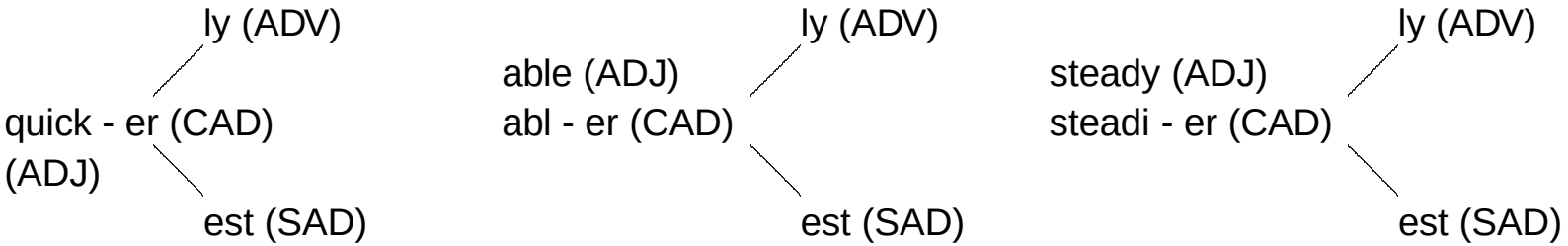
14.5.1 Concatenation patterns of English nouns



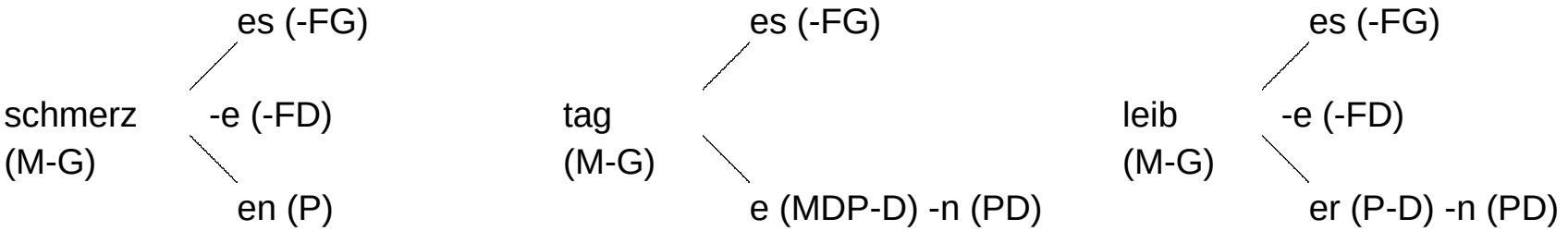
14.5.2 Concatenation patterns of English verbs

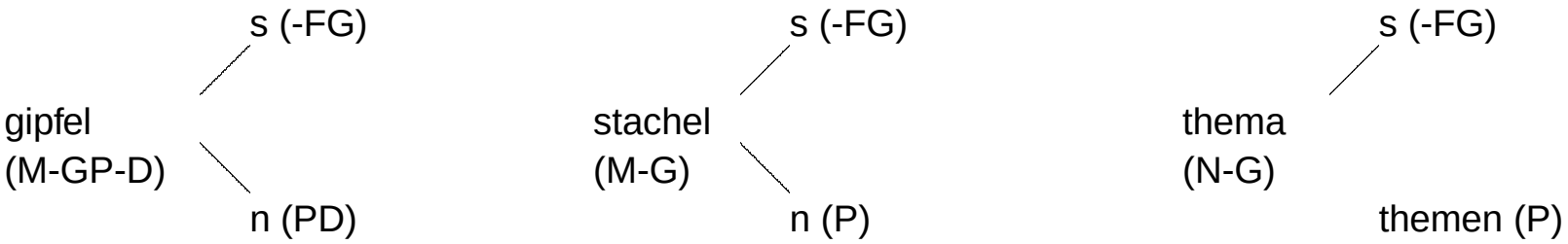


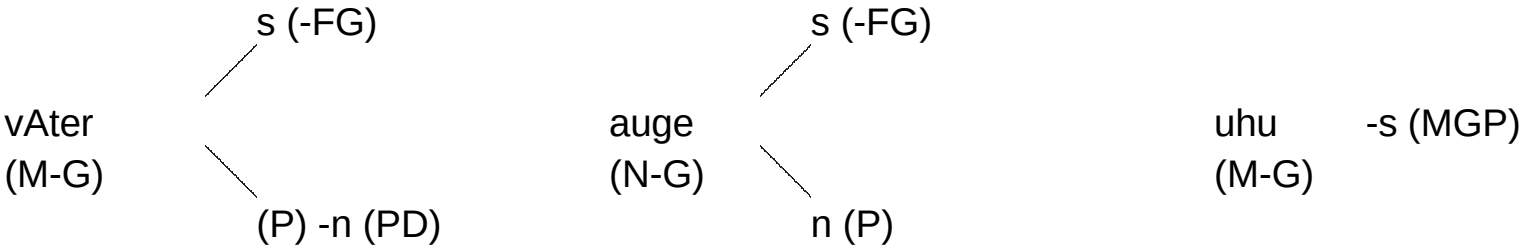
14.5.3 Concatenation patterns of adjective-adverbs



14.5.4 Concatenation patterns of German nouns







braten (M-GP)	-s (-FG)	hAnd (F)	-e (P-D) -n (PD)	frau (F)	-en (P)
drangsal (F)	-e (P-D) -n (PD)	kenntnis (F)	-se (P-D) -n (PD)	mUtter (F)	- (P-D) -n (PD)

14.5.5 Category segments of German noun forms

MN	= Masculinum Nominativ	(Bote)
M-G	= Masculinum no Genitiv	(Tag)
-FG	= no Femininum Genitiv	(Tages, Kindes)
-FD	= no Femininum Dativ	(Schmerze, Kinde)
M-NP	= Masculinum no Nominativ or Plural	(Boten)
M-GP	= Masculinum no Genitiv or Plural	(Braten)
MGP	= Masculinum Genitiv or Plural	(Uhus)
M-GP-D	= Masculinum no Genitiv or Plural no Dativ	(Gipfel)
F	= Femininum	(Frau)
N-G	= Neutrum no Genitiv	(Kind)
NG	= Neutrum Genitiv	(Kindes)
ND	= Neutrum Dativ	(Kinde)
N-GP	= Neutrum no Genitiv or Plural	(Leben)
N-GP-D	= Neutrum no Genitiv or Plural no Dativ	(Wasser)
NDP-D	= Neutrum Dativ or Plural no Dativ	(Schafe)
P	= Plural	(Themen)
P-D	= Plural no Dativ	(Leiber)
PD	= Plural Dativ	(Leibern)

15. Corpus analysis

15.1 Implementation and application of grammar systems

15.1.1 Parts of a grammar system

- Formal algorithm
- Linguistic method

15.1.2 Options for grammar system of word form recognition

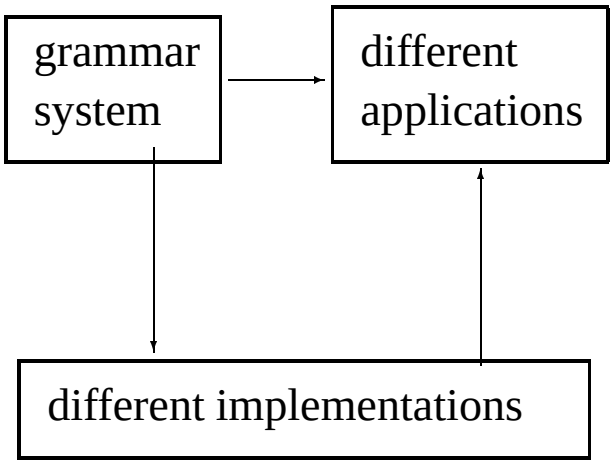
- Formal algorithm:
  C- (Section 7.4), PS- (Section 8.1), or LA-grammar (Section 10.2).
- Linguistic method:
  Word form, morpheme, or allomorph method (cf. Section 13.5).

15.1.3 Minimal standard of well-defined grammar systems

A grammar system is well-defined only if it simultaneously allows

1. different *applications* in a given *implementation*, and
2. different *implementations* in a given *application*.

15.1.4 Modularity of a grammar system



15.1.5 Different implementations of LA-morphology

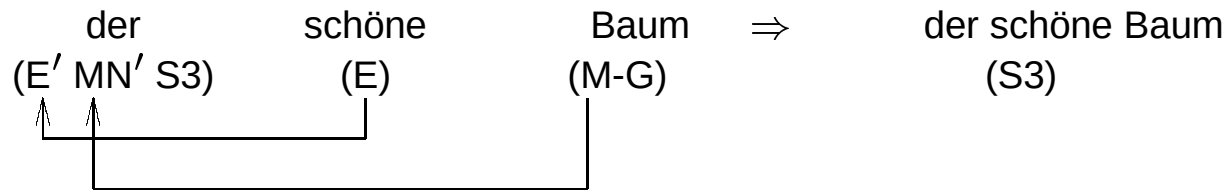
1988 in LISP (Hausser & Todd Kaufmann)
1990 in C (Hausser & Carolyn Ellis)
1992 in C, ‘LAMA’ (Norbert Bröker)
1994 in C, ‘LAP’ (Gerald Schüller)
1995 in C, ‘Malaga’ (Björn Beutel)

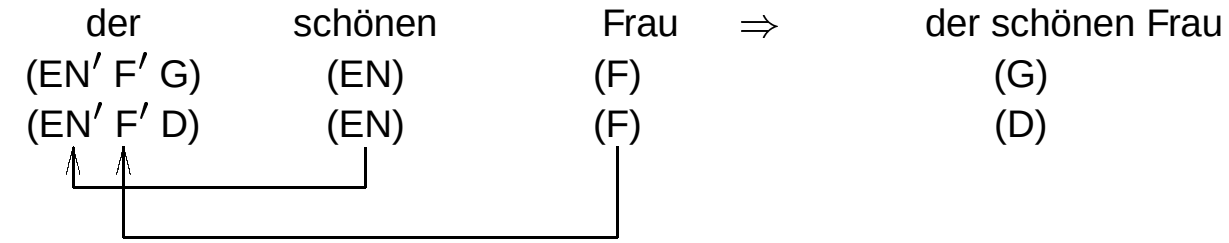
15.1.6 Structural principles common to different LA-Morph implementations

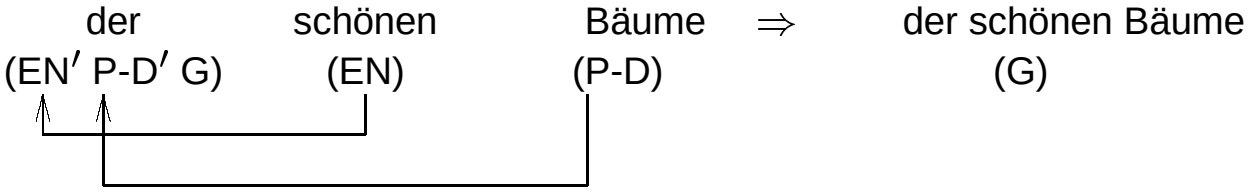
- Specification of the allo- (cf. 14.1.1) and the combi-rules (cf. 14.4.1) on the basis of patterns which are matched onto the input.
- Storage of the analyzed allomorphs in a trie structure and their left-associative lookup with parallel pursuit of alternative hypotheses (cf. Section 14.3).
- Modular separation of motor, rule components, and lexicon, permitting a simple exchange of these parts, for example in the application of the system to new domains or languages.
- Use of the same motor and the same algorithm for the combi-rules of the morphological, syntactic, and semantic components during analysis.
- Use of the same rule components for analysis and generation in morphology, syntax, and semantics.

15.2 Subtheoretical variants

15.2.1 Combinatorics of the German determiner der

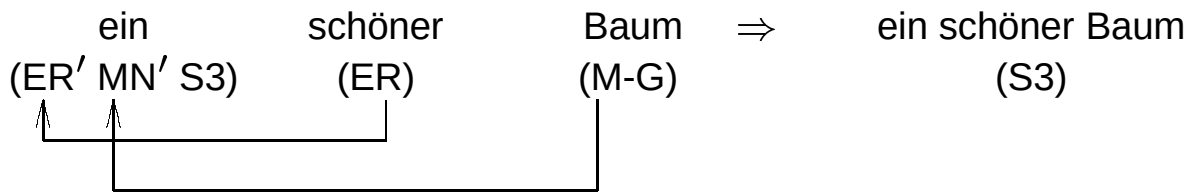




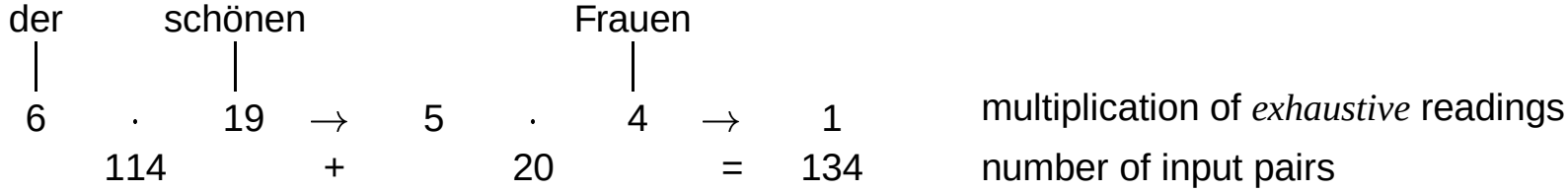


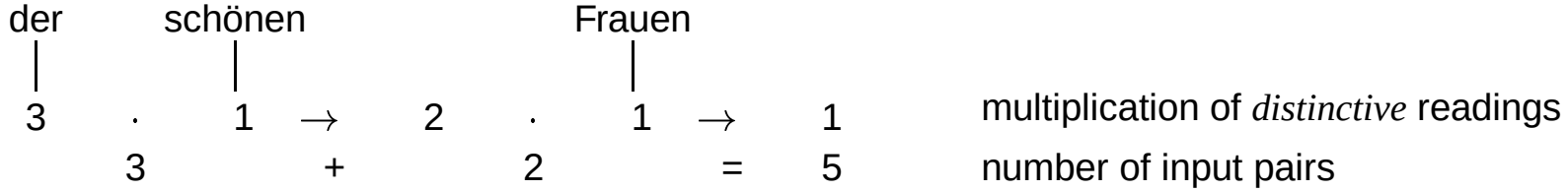
15.2.2 Agreement of adjective-ending with determiner

der schöne Baum (cf. 15.2.1)



15.2.3 Exhaustive versus distinctive categorization in deriving der schönen Frauen





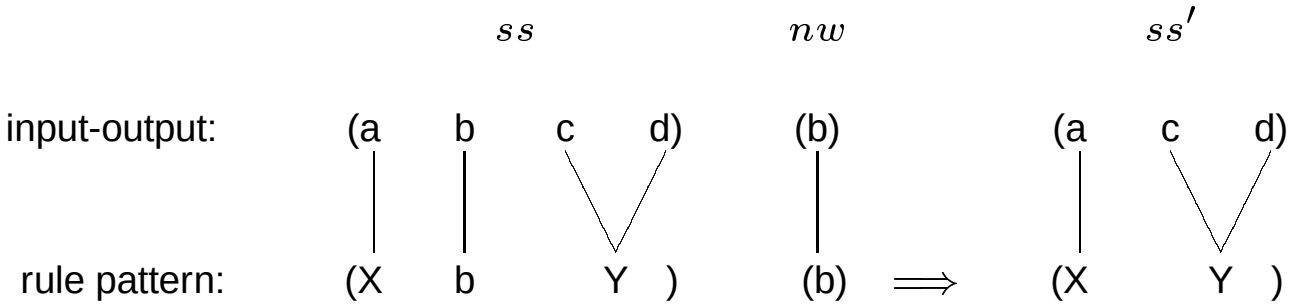
15.2.4 Representing lexical readings via different entries

[der (E′ MN′ S3) DEF-ART]
[der (EN′ F′ G&D) DEF-ART]
[der (EN′ P-D′ G) DEF-ART]

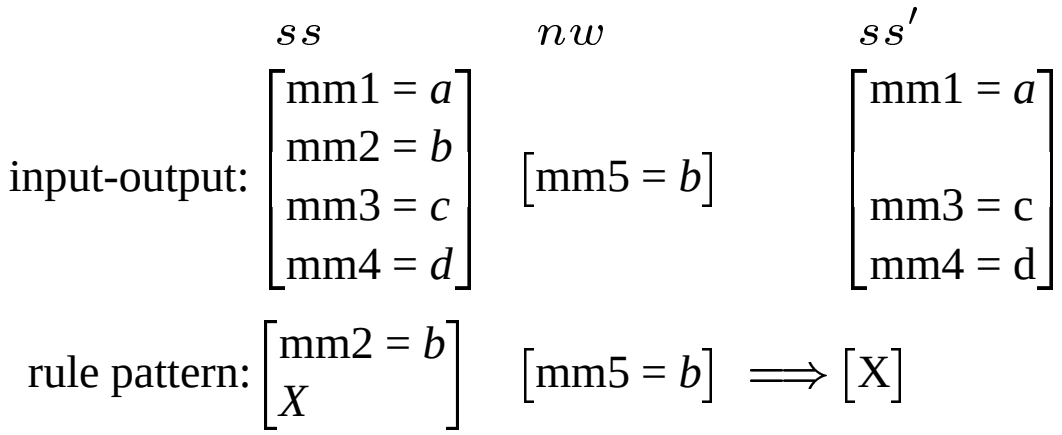
15.2.5 Representing lexical readings via multicats

[der ((E′ MN′ S3) (EN′ F′ G&D) (EN′ P-D′ G)) DEF-ART]

15.2.6 List-based matching (LAP)



15.2.7 Feature-based matching (Malaga)



15.3 Building corpora

15.3.1 Text genres of the Brown and the LOB corpus

	Brown	LOB
A Press: reportage	44	44
B Press: editorial	27	27
C Press: reviews	17	17
D Religion	17	17
E Skills, trade, and hobbies	36	38
F Popular lore	48	44
G Belle lettres, biography, essays	75	77
H Miscellaneous (government documents, foundation records, industry reports, college catalogues, industry house organ)	30	38
J Learned and scientific writing	80	80
K General fiction	29	29
L Mystery and detective fiction	24	24
M Science fiction	6	6
N Adventure and western fiction	29	29
P Romance and love story	29	29
R Humour	9	9
Total	500	500

15.3.2 Kučera & Francis' desiderata for the construction of corpora

1. Definite and specific delimitation of the language texts included, so that scholars using the Corpus may have a precise notion of the composition of the material.
2. Complete synchronicity; texts published in a single calendar year only are included.
3. A predetermined ratio of the various genres represented and a selection of individual samples through a random sampling procedure.
4. Accessibility of the Corpus to automatic retrieval of all information contained in it which can be formally identified.
5. An accurate and complete description of the basic statistical properties of the Corpus and of several subsets of the Corpus with the possibility of expanding such analysis to other sections or properties of the Corpus as may be required.

15.3.3 Difficulties with achieving a representative and balanced corpus

> 'Genre' is not a well-defined concept. Thus genres that have been distinguished so far have been identified on a purely intuitive basis. No empirical evidence has been provided for any of the genre distinctions that have been made.
>
> N. Oostdijk 1988

15.4 Distribution of word forms

15.4.1 Definition of rank

The position of a word form in the frequency list

15.4.2 Definition of frequency class (F-class)

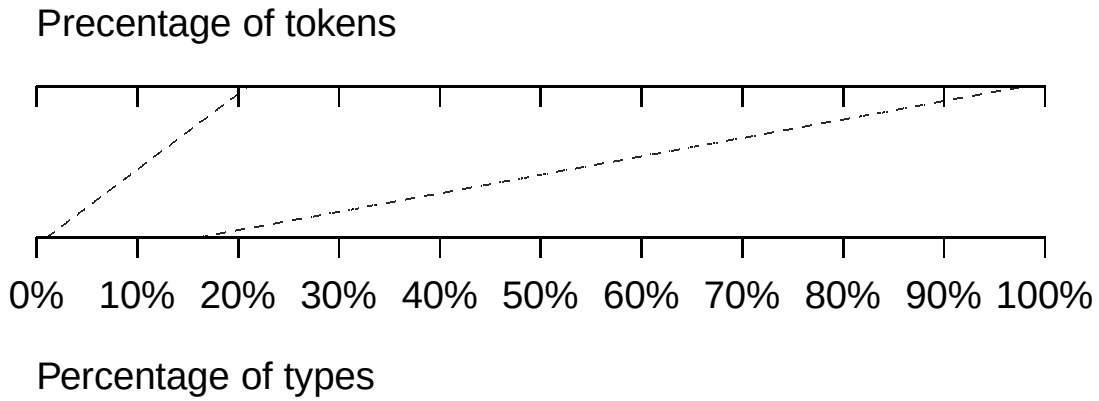
F-class $=_{def}$ [frequency of types # number of types]

There are much fewer F-classes in a corpus than ranks. In the BNC, for example, 655 270 ranks result in 5 301 F-classes. Thus, the number of the F-classes is only 0.8% of the number of ranks. Because of their comparatively small number the F-classes are well suited to bring the type-token correlation into focus.

15.4.3 Type-token distribution in the BNC (*surface-based*)

F-class	start_r	end_r	types	tokens	types-%	tokens-%	
beginning	(the first 9 F-classes)						
1 (the)	1	1	1	5776399	0.000152	6.436776	
2 (of)	2	2	1	2789563	0.000152	3.108475	
3 (and)	3	3	1	2421306	0.000152	2.698118	
4 (to)	4	4	1	2332411	0.000152	2.599060	
5 (a)	5	5	1	1957293	0.000152	2.181057	
6 (in)	6	6	1	1746891	0.000152	1.946601	
7 (is)	7	7	1	893368	0.000152	0.995501	
8 (that)	8	8	1	891498	0.000152	0.993417	
9 (was)	9	9	1	839967	0.000152	0.935995	
sums			9	19 648 696	0.001368 %	21.895 %	
middle	(9 samples)						
1000	1017	1017	1	9608	0.000152	0.010706	
2001	2171	2171	1	4560	0.000152	0.005081	tokens
3000	3591	3591	1	2521	0.000152	0.002809	per
3500	4536	4536	1	1857	0.000152	0.002069	type:
4000	5907	5910	4	5228	0.000607	0.005826	1307
4500	8332	8336	5	4005	0.000758	0.004463	801
4750	10842	10858	17	9367	0.002579	0.010438	551
5000	16012	16049	38	11438	0.005764	0.012746	301
5250	44905	45421	517	26367	0.078420	0.029381	51
end	(the last 9 F-classes)						
5292	108154	114730	6577	59193	0.997620	0.065960	9
5293	114731	122699	7969	63752	1.208763	0.071040	8
5294	122700	132672	9973	69811	1.512736	0.077792	7
5295	132673	145223	12551	75306	1.903775	0.083915	6
5296	145224	161924	16701	83505	2.533260	0.093052	5
5297	161925	186302	24378	97512	3.697732	0.108660	4
5298	186303	225993	39691	119073	6.020456	0.132686	3
5299	225994	311124	85131	170262	12.912938	0.189727	2
5300	311125	659269	348145	348145	52.807732	0.387946	1
sums			551 116	1 086 559	83.595012 %	1.210778 %	

15.4.4 Correlation of type and token frequency



15.4.5 Semantic significance

The higher the freqency, the lower the semantic significance.
Examples: the, of, and, to, a, in, that, was

The lower the freqency, the higher the semantic significance.
Examples: audiophile, butternut, customhouse, dustheap

15.4.6 Hapaxlegomena

Word forms in a corpus which occur only once.

15.4.7 Zipf's law

frequency · rank = constant

15.4.8 Illustration of Zipf's law

word form	rank	·	frequency	=	constant
the	1	·	5 776 399	=	5 776 399
and	2	·	2 789 563	=	5 579 126
...					
was	9	·	839 967	=	7 559 703
...					
holder	3 251	·	2 870	=	9 330 370

15.5 Statistical tagging

15.5.1 Top of Brown corpus frequency list

```
69971-15-500 THE                21341-15-500 IN
36411-15-500 OF                 10595-15-500 THAT
28852-15-500 AND                10099-15-485 IS
26149-15-500 TO                  9816-15-466 WAS
23237-15-500 A                   9543-15-428 HE
```

The entry `9543-15-428 HE`, for example, indicates that the word form `HE` occurs 9 543 times in the Brown corpus, in all 15 genres, and in 428 of the 500 sample texts.

15.5.2 Statistical tagging

is based on categorizing by hand – or half automatically with careful post-editing – a small part of the corpus, called the *core corpus*. The categories used for the classification are called *tags* or *labels*. After hand-tagging the core corpus, the probabilities of the transitions from one word form to the next are computed by means of *Hidden Markov Models* (HMMs).

15.5.3 Subset of the *basic (C5) tagset*

AJ0 Adjective (general or positive) (e.g. good, old, beautiful)
CRD Cardinal number (e.g. one, 3, fifty-five, 3609)
NN0 Common noun, neutral for number (e.g. aircraft, data, committee)
NN1 Singular common noun (e.g. pencil, goose, time, revelation)
NN2 Plural common noun (e.g. pencils, geese, times, revelations)
NP0 Proper noun (e.g. London, Michael, Mars, IBM)
UNC Unclassified items
VVB The finite base form of lexical verbs (e.g. forget, send, live, return)
VVD The past tense form of lexical verbs (e.g. forgot, sent, lived, returned)
VVG The -ing form of lexical verbs (e.g. forgetting, sending, living, returning)
VVI The infinitive form of lexical verbs (e.g. forget, send, live, return)
VVN The past participle form of lexical verbs (e.g. forgotten, sent, lived, returned)
VVZ The -s form of lexical verbs (e.g. forgets, sends, lives, returns)

15.5.4 Sample from the alphabetical word form list of the BNC

```
1 activ nn1-np0 1
1 activ np0 1
2 activa nn1 1
3 activa nn1-np0 1
4 activa np0 2
1 activatd nn1-vvb 1
21 activate np0 4
62 activate vvb 42
219 activate vvi 116
140 activated aj0 48
56 activated aj0-vvd 26
52 activated aj0-vvn 34
5 activated np0 3
85 activated vvd 56
43 activated vvd-vvn 36
312 activated vvn 144
1 activatedness nn1 1
88 activates vvz 60
5 activating aj0 5
8 activating aj0-nn1 6
47 activating aj0-vvg 22
3 activating nn1-vvg 3
14 activating np0 5
371 activating vvg 49
538 activation nn1 93
3 activation nn1-np0 3
2 activation-energy aj0 1
1 activation-inhibition aj0 1
1 activation-synthesis aj0 1
1 activation. nn0 1
1 activation/ unc 1
282 activator nn1 30
6 activator nn1-np0 3
1 activator/ unc 1
1 activator/ unc 1
7 activator/tissue unc 1
61 activators nn2 18
1 activators np0 1
```

Each entry consists (i) of a number detailing the frequency of the tagged word form in the whole corpus, (ii) the surface of the word form, (iii) the label, and (iv) the number of texts in which the word form was found under the assigned label.

15.5.5 Error rates in statistical tagging

The error rate of CLAWS4 is quoted by Leech 1995 at 1.7%, which may seem very good. However, given that the last 1.2% of the low frequency tokens requires 83.6% of the types (cf. 15.4.4), an error rate of 1.7% may also represent a very bad result – namely that about 90% of the types are not analyzed or not analyzed correctly. This conclusion is born out by a closer inspection of sample 15.5.4.

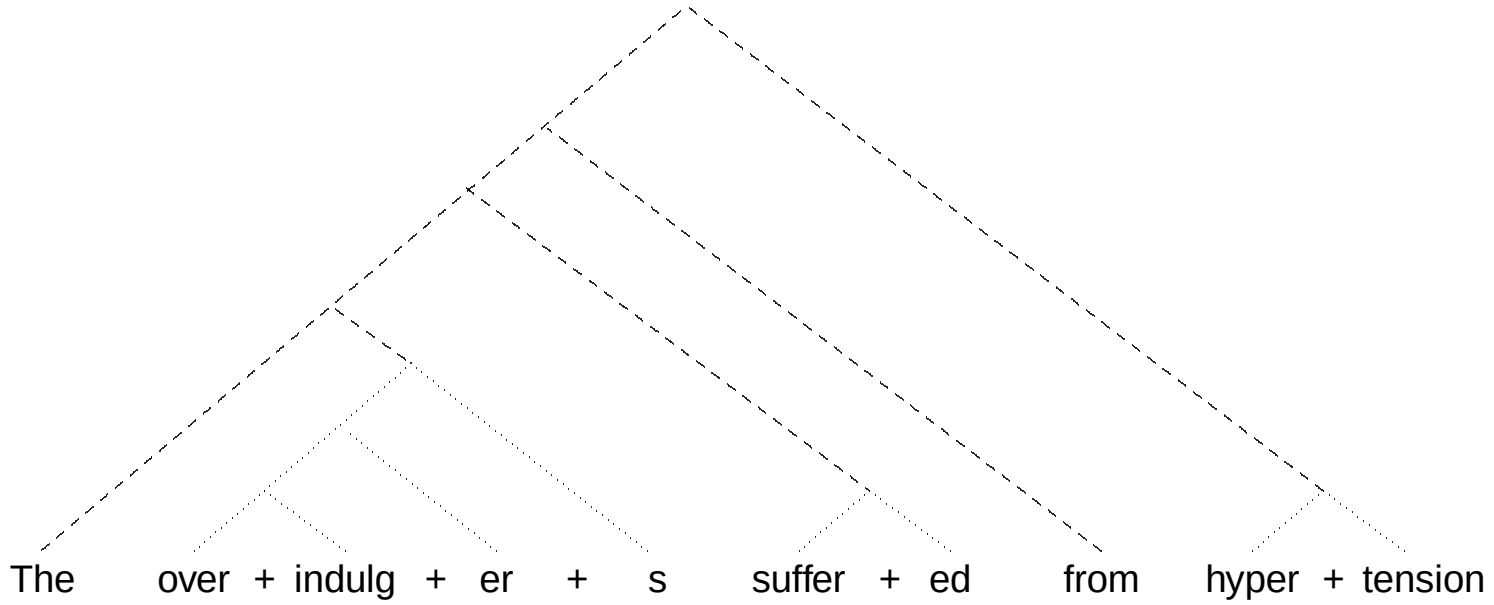
15.5.6 Weaknesses of statistical tagging

1. The categorization is too unreliable to support rule-based syntactic parsing.
2. Word forms can be neither reduced to their base forms (lemmatization) nor segmented into their allomorphs or morphemes.
3. The overall frequency distribution analysis of a corpus is distorted by an artificial inflation of types (e.g., 37.5% in the BNC).
4. Even if the tagger is successfully improved as a whole, its results can never be more than probabilistically-based conjectures.

16. Basic concepts of syntax

16.1 Delimitation of morphology and syntax

16.1.1 Correlation of LA-morphology and LA-syntax



The tree structures of LA-morphology and LA-syntax both satisfy the SLIM-theoretic principles of surface compositionality (S) and time-linear composition (L). However, their respective time-linear compositions occur in different phases.

16.1.2 Treatment of idioms in morphology or syntax?

A syntactic treatment is generally motivated in idioms which (i) retain their compositional meaning as an option, (ii) are subject to normal variations of word order, and (iii) exhibit internal inflectional variation. Otherwise idioms should be handled in the lexicon (e.g. over-the-counter).

16.1.3 Correlation of morphology and syntax in different types of language

Some natural languages compose meaning$_1$ mainly in the syntax (e.g. Chinese) and others mainly in morphology (e.g. Eskimo in which long chains of morphemes are concatenated into a single word form such as [a:wlis-ut-iss?ar-si-niarpu-na] *I am looking for something suitable for a fish-line*). This alternative exists also within a given natural language. For example, in English the complex concept denoted by the word form overindulgers may roughly be expressed analytically as people who eat and drink too much.

16.1.4 Combination principles of syntax

1. *Valency*
2. *Agreement*
3. *Word order*

16.2 Valency

16.2.1 The notions valency carrier and valency filler

go back to the French linguist L. TESNIÈRE 1959, who borrowed them from chemistry. The valency positions of a carrier must be filled, or canceled, by compatible fillers in order for an expression to be syntactically and semantically complete.

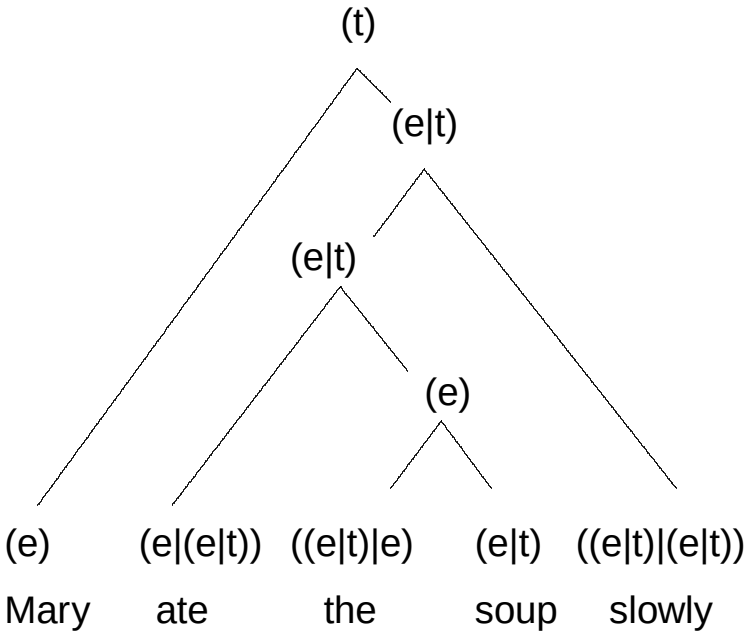
16.2.2 Coding the structure of valency carriers in LA-grammar

Composite syntactic categories are defined as lists of category segments. For example, the English verb form ate is analyzed as follows.

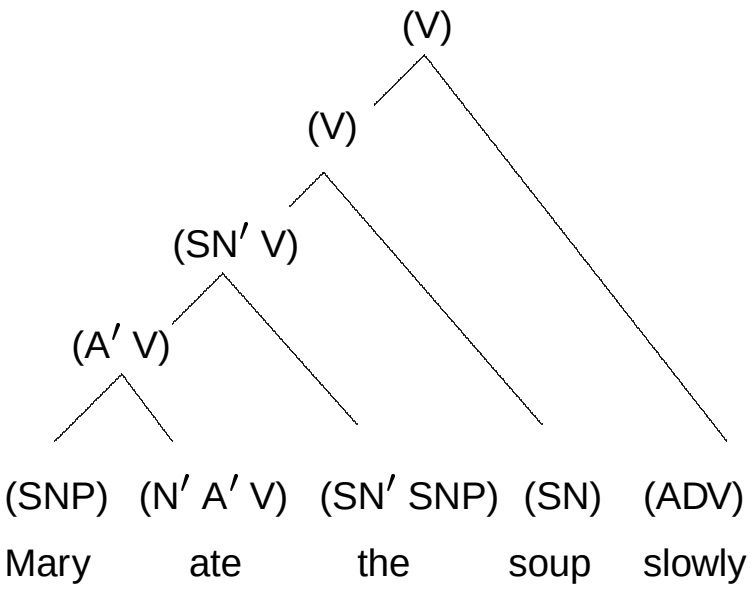
[ate (N′ A′ V) eat]

16.2.3 Carriers, fillers, and modifiers in CG and LAG

C-grammar analysis



LA-grammar analysis



16.2.4 Examples of different valency carriers in German

- the one-place verb form schläfst (*sleep*):

  [schläfst (S2′ V) schlafen]

  result segment

  valency position

- the two-place verb form liest (*read*):

  [liest (S23′ A′ V) lesen]

  result segment

  valency positions

- the two-place verb form hilft (*help*):

  [hilft (S3′ D′ V) helfen]

  result segment

  valency positions

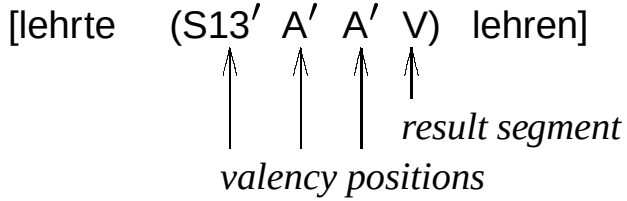
- The three-place verb form gebt (*give*):

  [gebt (P2′ D′ A′ V) geben]

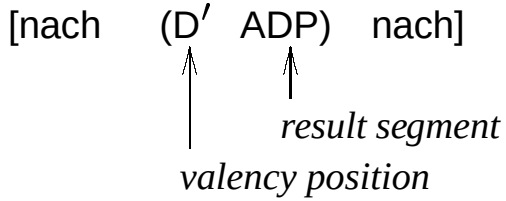
  result segment

  valency positions

- The three-place verb form lehrte (*taught*):



- The one-place preposition nach (*after*):



16.2.5 Category structure of valency fillers and modifiers

[Bücher (P-D) buch]	(*books*)
[ihm (D) er]	(*him*)
[gestern (ADV) gestern]	(*yesterday*)

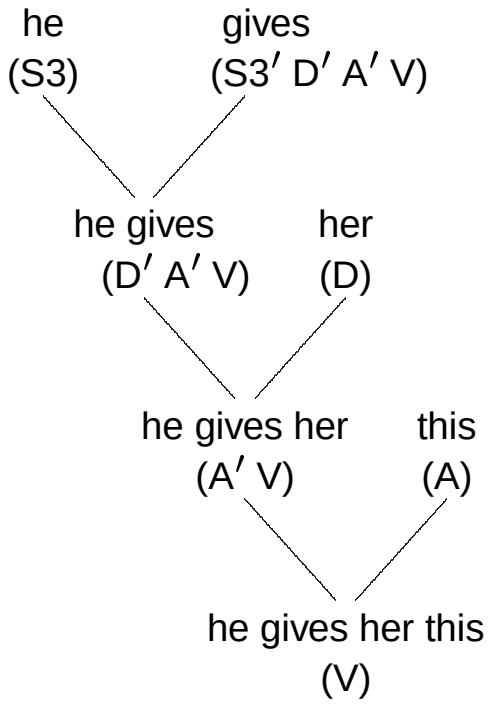
Valency carriers may also function as valency fillers using their result segment, e.g V, as the filler segment. In this case, the segments representing valency positions are attached at the beginning of the category resulting from the composition.

16.3 Agreement

16.3.1 Agreement violation in English

*Every girls need a mother.

16.3.2 Identity-based agreement in a simple LA-syntactic analysis



16.3.3 An LA-grammar for 6.3.2 (*LA-plaster*)

LX $=_{def}$ { [he (S3) *], [her (D) *], [this (A) *], [gives (S3′ D′ A′ V) *]}

$ST_S =_{def}$ { [(S3) {MAIN+FV}] }

MAIN+FV:	(S3) (S3′ D′ A′ V)	⇒ (D′ A′ V)	{FV+MAIN1}
FV+MAIN1:	(D′ A′ V) (D)	⇒ (A′ V)	{FV+MAIN2}
FV+MAIN2:	(A′ V) (A)	⇒ (V)	{ }

$ST_F =_{def}$ { [(V) $rp_{FV+MAIN2}$] }

16.3.4 Example of an error in identity-based agreement

I	+	gives	⇒	Error: ungrammatical continuation
(S1)		(S3′ D′ A′ V)		

16.4 Free word order in German (*LA-D1*)

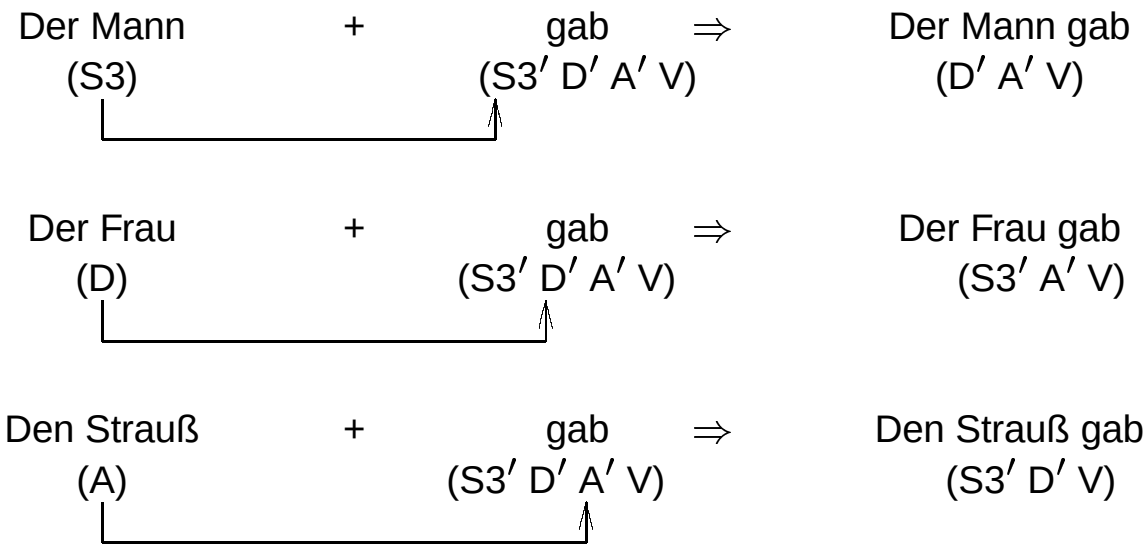
16.4.1 Word order variations in a declarative main clause of German

Der Mann gab der Frau den Strauß.
(the man gave the woman the bouquet.)
Der Mann gab den Strauß der Frau.
(the man gave the bouquet the woman.)
Der Frau gab der Mann den Strauß.
(the woman gave the man the bouquet.)
Der Frau gab den Strauß der Mann.
(the woman gave the bouquet the man.)
Den Strauß gab der Mann der Frau.
(the bouquet gave the man the woman.)
Den Strauß gab der Frau der Mann.
(the bouquet gave the woman the man.)

16.4.2 Word order violation in German

*Der Mann der Frau gab einen Strauß.
(the man the woman gave the bouquet.)

16.4.3 Free canceling of valency positions in a carrier of German

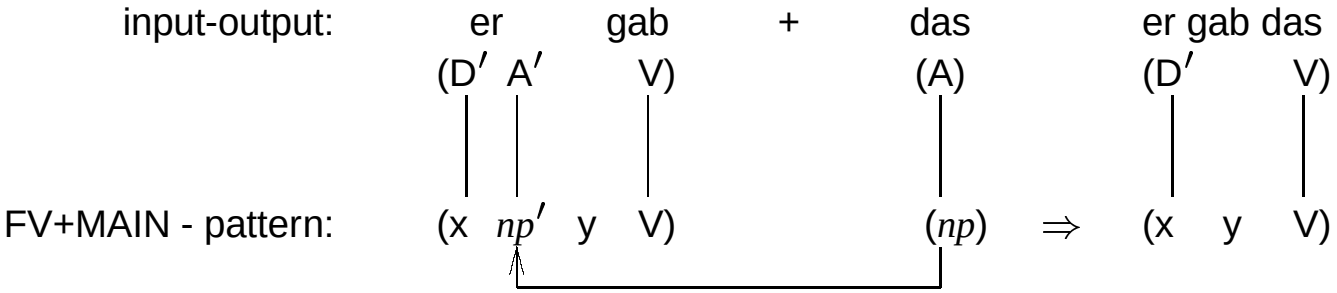


16.4.4 German LA-grammar with partial free word order

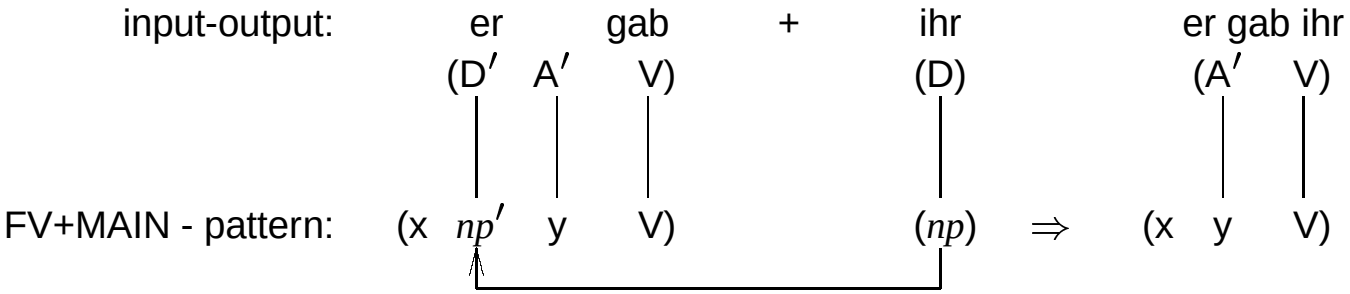
LX $=_{def}$ { [er (S3) *], [ihr (D) *], [das (A) *], [gab (S3′ D′ A′ V) *]}
Variable definition: *np* ε {D, A}, with *np*′ correspondingly D′ or A′
x, y = .?.?.?.? (i.e. an arbitrary sequence up to length 4)

$ST_S =_{def}$ { [(S3) {MAIN+FV}] }
MAIN+FV: (S3) (S3′ D′ A′ V) ⇒ (D A V) {FV+MAIN}
FV+MAIN: (x *np*′ y V) (*np*) ⇒ (x y V) {FV+MAIN}
$ST_F =_{def}$ { [(V) $rp_{FV+MAIN}$] }

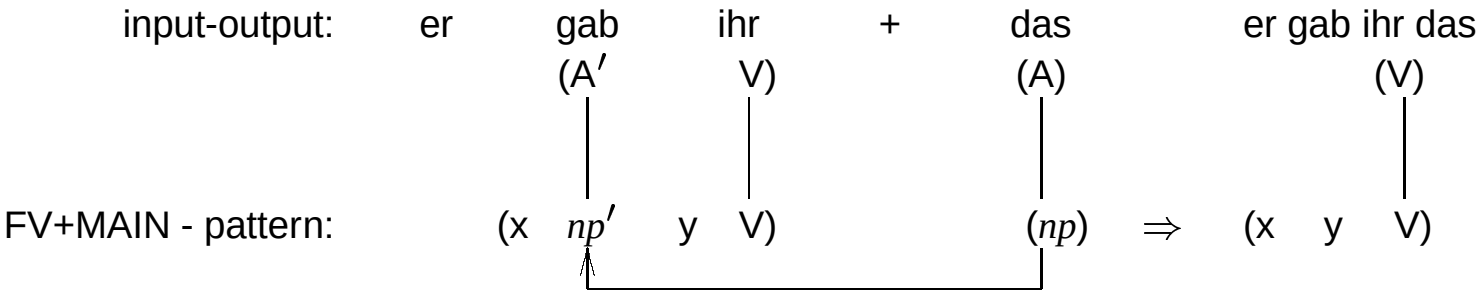
16.4.5 FV+MAIN matching a next word accusative



16.4.6 FV+MAIN matching a next word dative



16.4.7 Reapplication of FV+MAIN



16.4.8 German LA-grammar with free word order (*LA-D1*)

LX $=_{def}$ { [er (S3) *], [ihr (D) *], [das (A) *], [gab (S3′ D′ A′ V) *]}

Variable definition: *np* ε {S3, D, A}, with *np*′ correspondingly S3′, D′ or A′
x, y = .?.?.?.? (i.e. an arbitrary sequence up to length 4)
$ST_S =_{def}$ { [(*np*) {MAIN+FV}] }
MAIN+FV: (*np*) (x *np*′ y V) $\Rightarrow$ (x y V) {FV+MAIN}
FV+MAIN: (x *np*′ y V) (*np*) $\Rightarrow$ (x y V) {FV+MAIN}
$ST_F =_{def}$ { [(V) $rp_{FV+MAIN}$] }

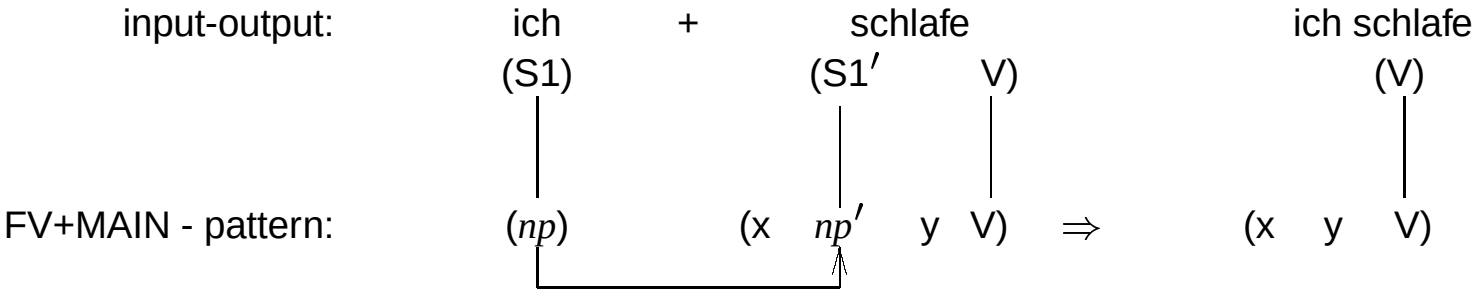
16.4.9 Word order variants of *LA-D1*

er gab ihr das	das gab er ihr	ihr gab er das
er gab das ihr	das gab ihr er	ihr gab das er

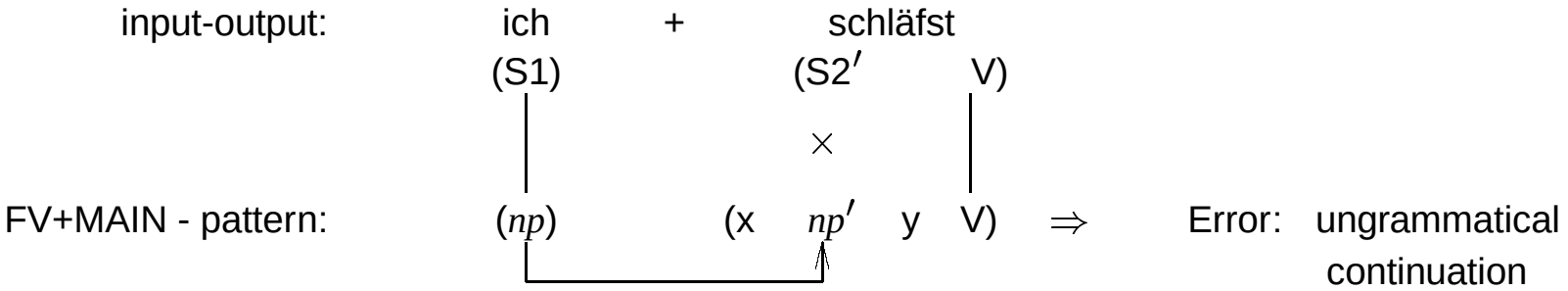
16.4.10 Extending the lexion of *LA-D1*

[ich (S1) *], [du (S2) *], [wir (P1) *], [schlafe (S1′ V) *], [schläfst (S2′ V) *], [schläft (S3′ V) *], [schlafen (P1′ V) *], [lese (S1′ A′ V) *], [liest (S2′ A′ V) *], [las (S3′ A′ V) *], [helfe (S1′ D′ V) *], [hilfst (S2′ D′ V) *], [half (S3′ D′ V) *], [lehre (S1′ A′ A′ V) *], [lehrst (S2′ A′ A′ V) *], [lehrt (S3′ A′ A′ V) *], [gebe (S1′ D′ A′ V) *], [gibst (S2′ D′ A′ V) *].

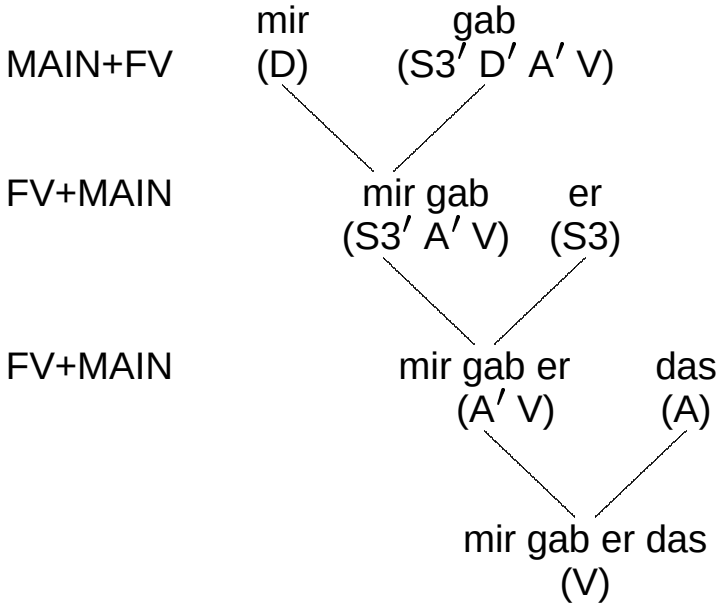
16.4.11 Identity-based subject-verb agreement in German



16.4.12 Agreement violation in German

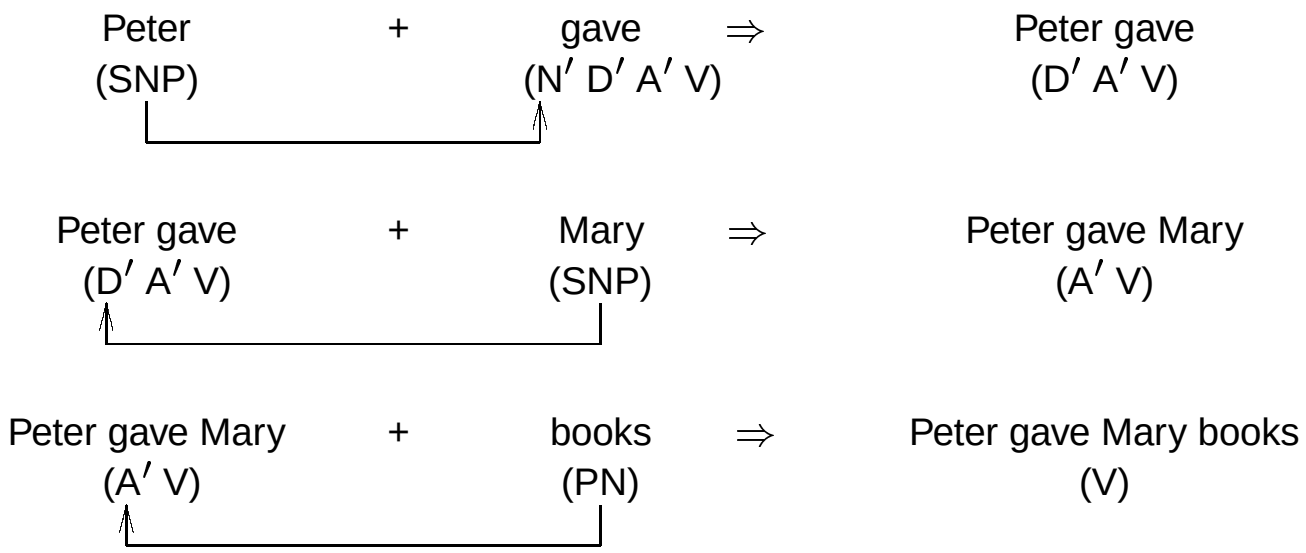


16.4.13 Derivation in *LA-D1* (identity-based agreement)



16.5 Fixed word order in English (*LA-E1*)

16.5.1 Fixed canceling of valency positions in a carrier of English

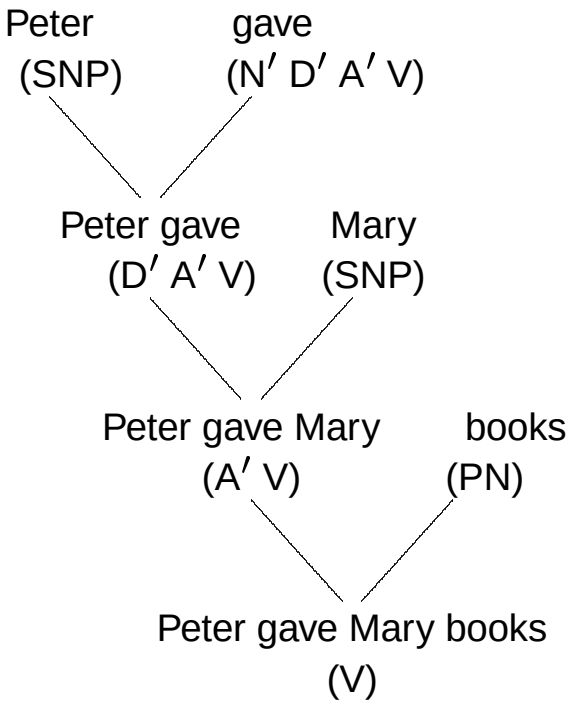


16.5.2 English LA-grammar with fixed word order (*LA-E1*)

LX $=_{def}$ { [Peter (SNP) *], [Mary (SNP) *], [books (PN) *],
[gave (N′ D′ A′ V) *]}
Variable definition: *np* ε {SNP, PN}, *np′* ε {N′, D′, A′},
x = .?.?.?.? (i.e. an arbitrary sequence up to length 4)

$ST_S =_{def}$ { [(x) {NOM+FV}] }
NOM+FV: (*np*) (*np′* x V) $\Rightarrow$ (y V) {FV+MAIN}
FV+MAIN: (*np′* x V) (*np*) $\Rightarrow$ (y V) {FV+MAIN}
$ST_F =_{def}$ { [(V) $rp_{FV+MAIN}$] }

16.5.3 Derivation in *LA-E1* (definition-based agreement)



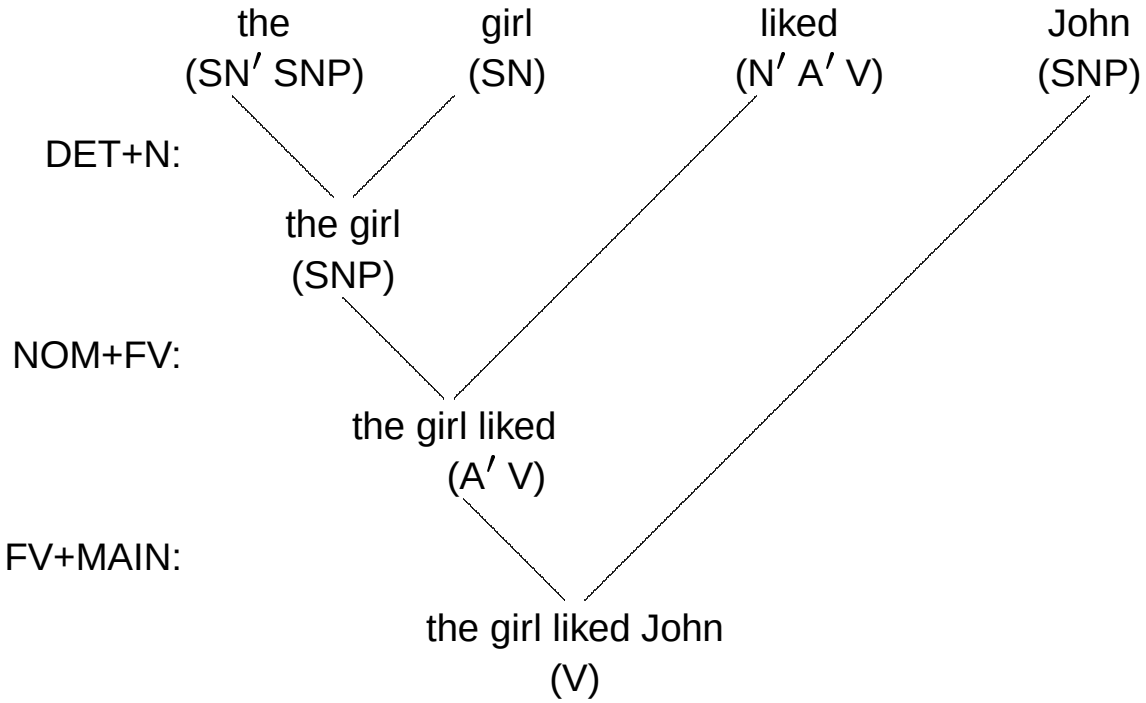
17. LA-syntax for English

17.1 Complex fillers in pre- and postverbal position

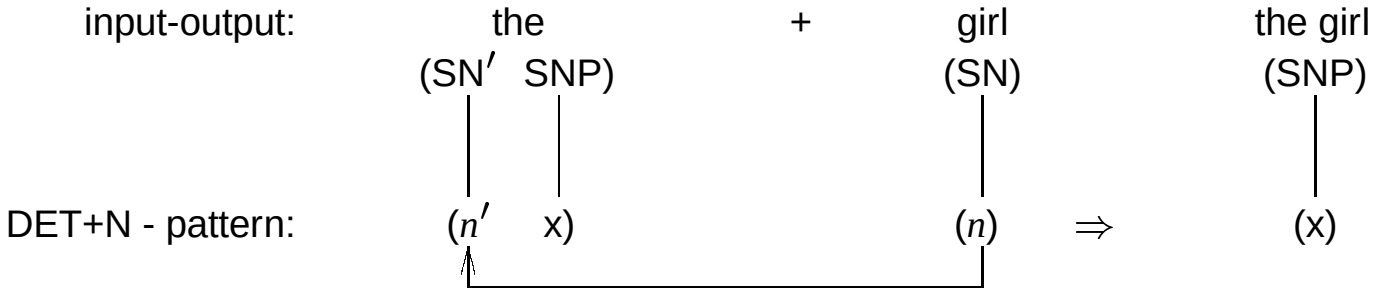
17.1.1 Determiner and noun categories of English

categories	*surfaces*	*examples of lemmata*
singular and plural determiners:		
(SN′ SNP)	a, an, every, the	[a (SN′ SNP) *]
(PN′ PNP)	all, several, the	[all (PN′ PNP) *]
singular and plural nouns:		
(SN)	man, woman, book, car	[woman (SN) *]
(PN)	men, women, books, cars	[men (PN) *]

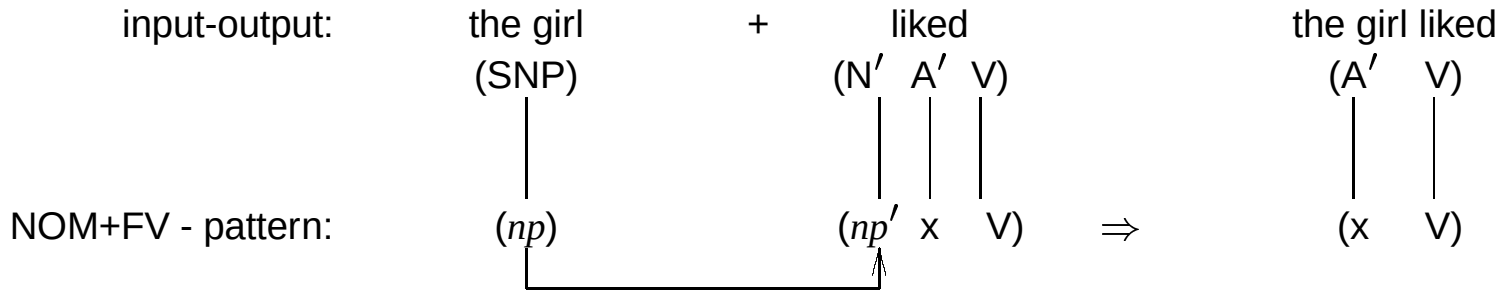
17.1.2 Complex noun phrase before the valency carrier



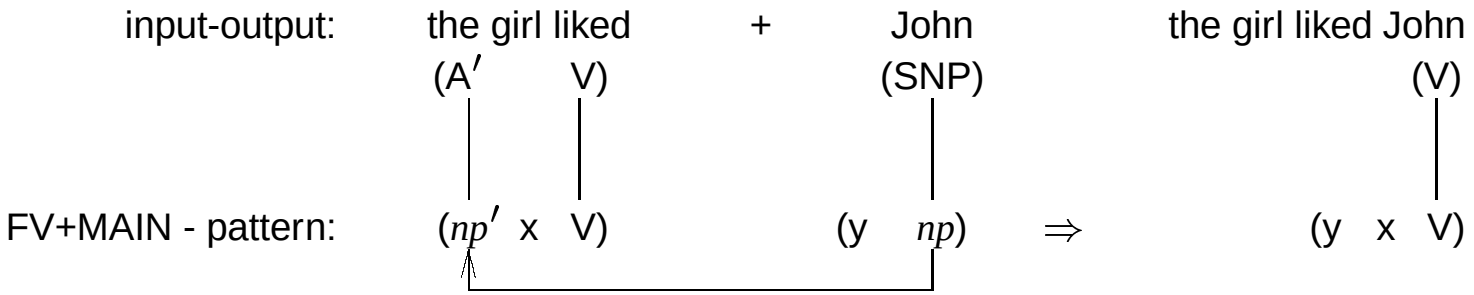
17.1.3 Preverbal application of Det+N



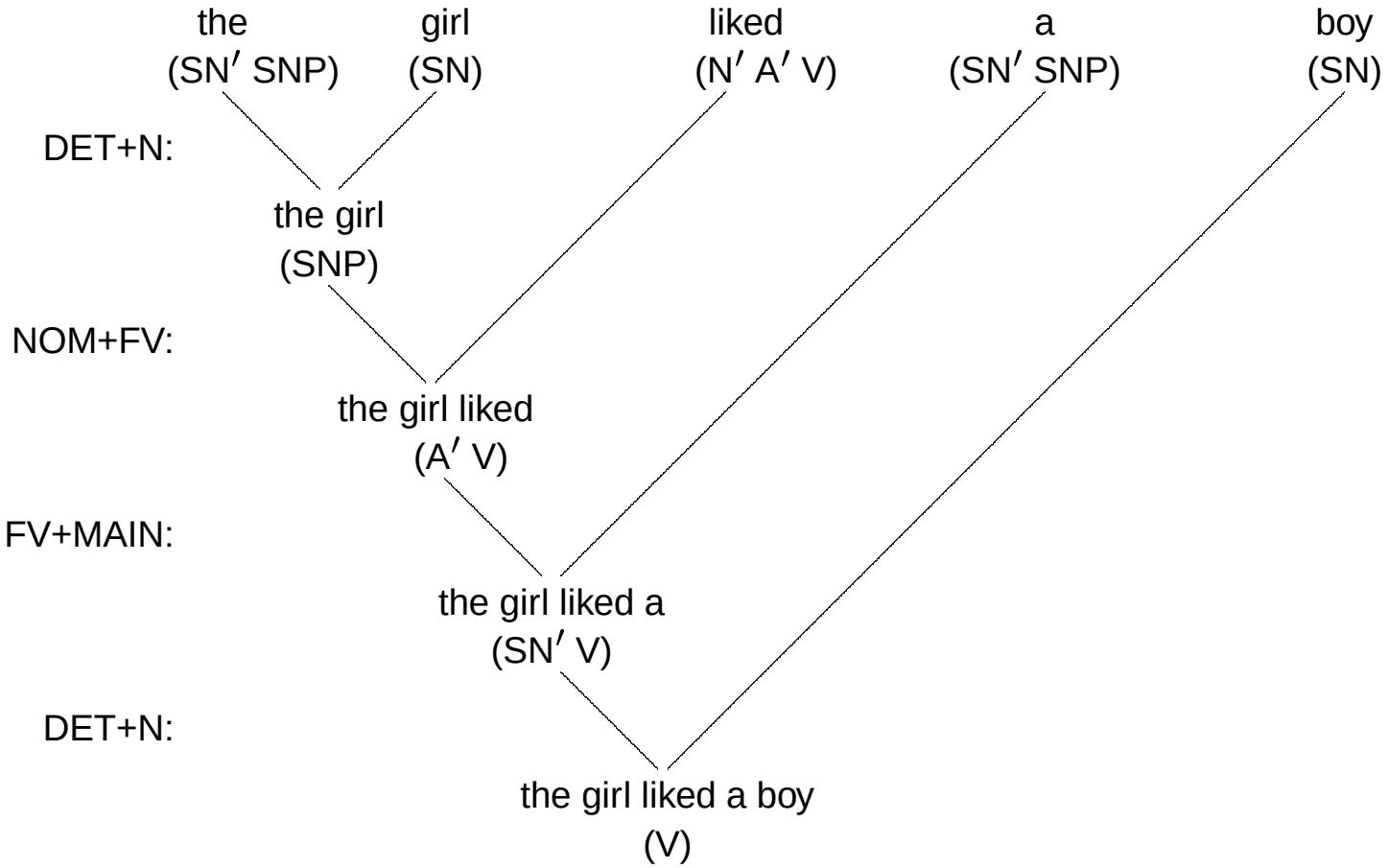
17.1.4 Application of NOM+FV to complex nominative NP



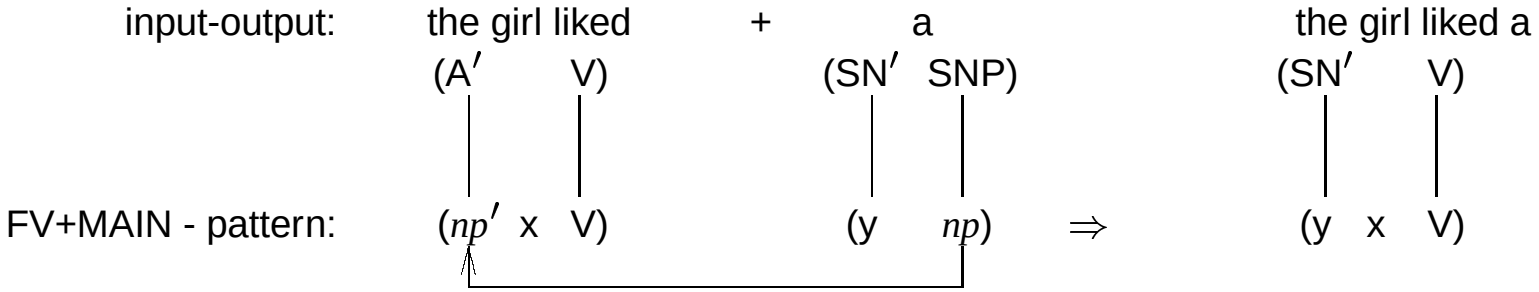
17.1.5 FV+MAIN adding elementary object NP



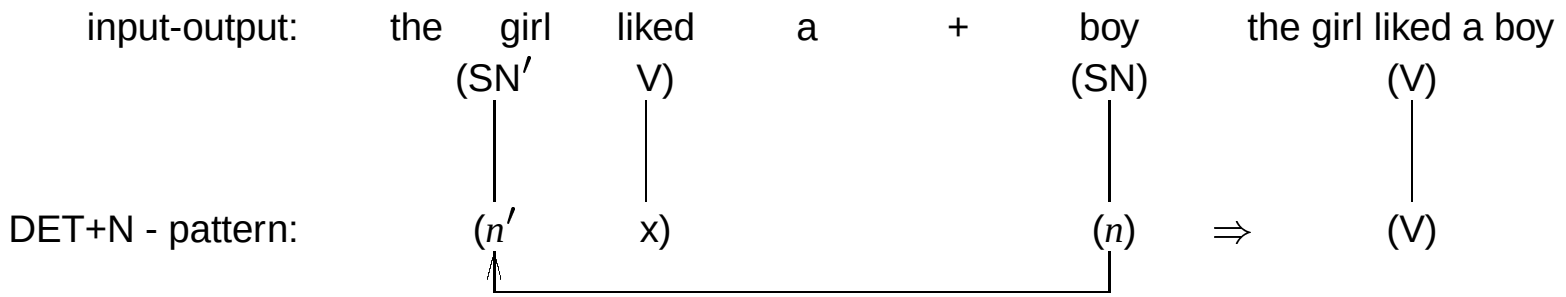
17.1.6 Complex noun phrase after valency carrier



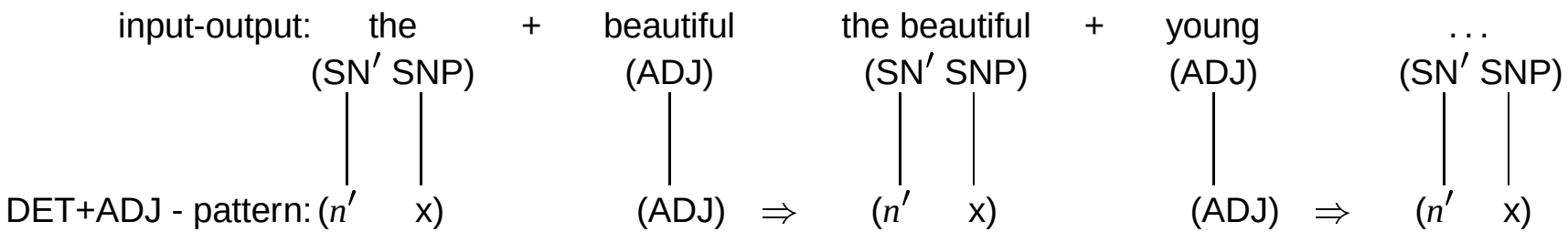
17.1.7 FV+Main adding beginning of complex object NP



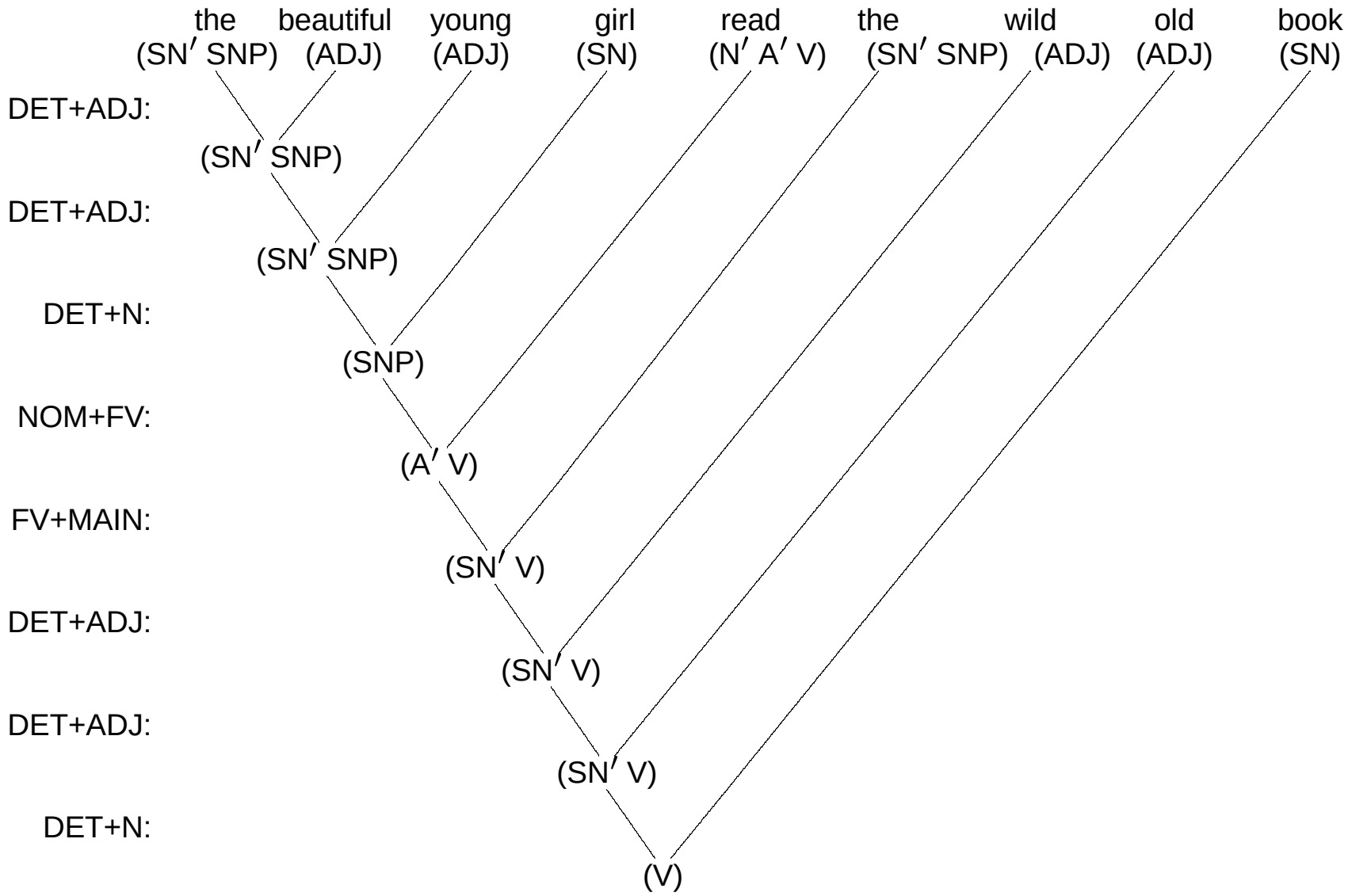
17.1.8 Postverbal application of Det+N



17.1.9 DET+ADJ recursively adding adjectives

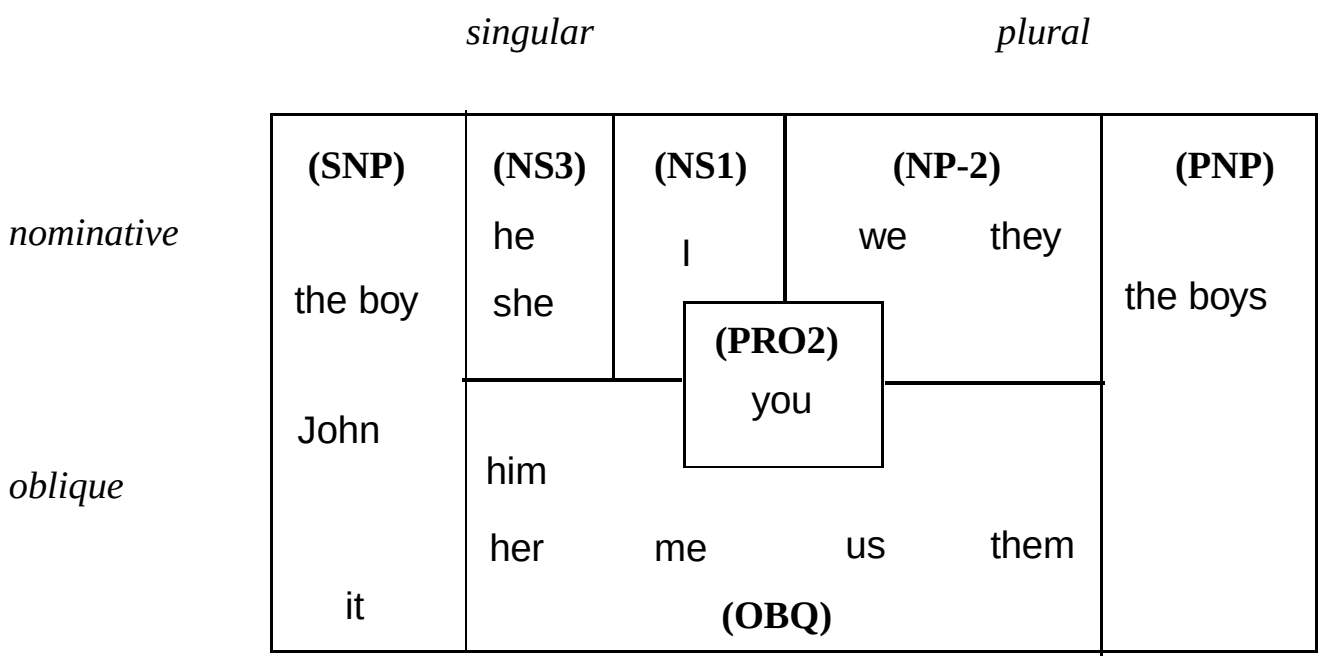


17.1.10 Complex noun phrases with adjectives

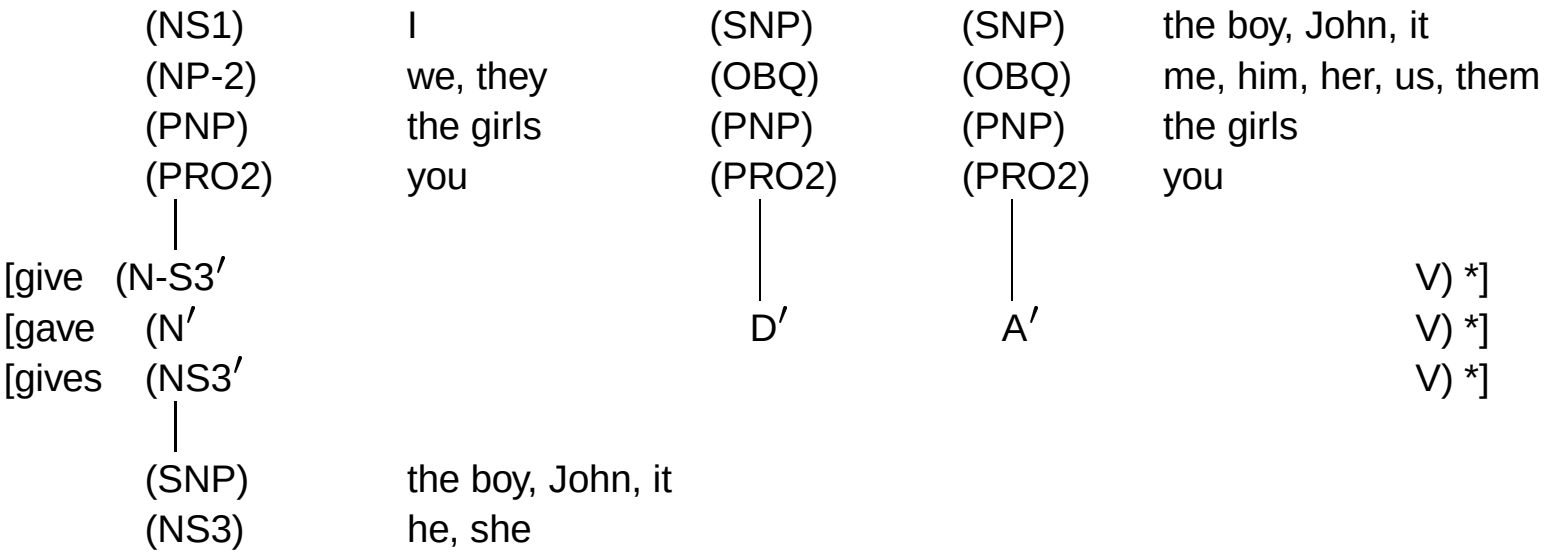


17.2 English field of referents

17.2.1 Categories of nominal valency fillers in English

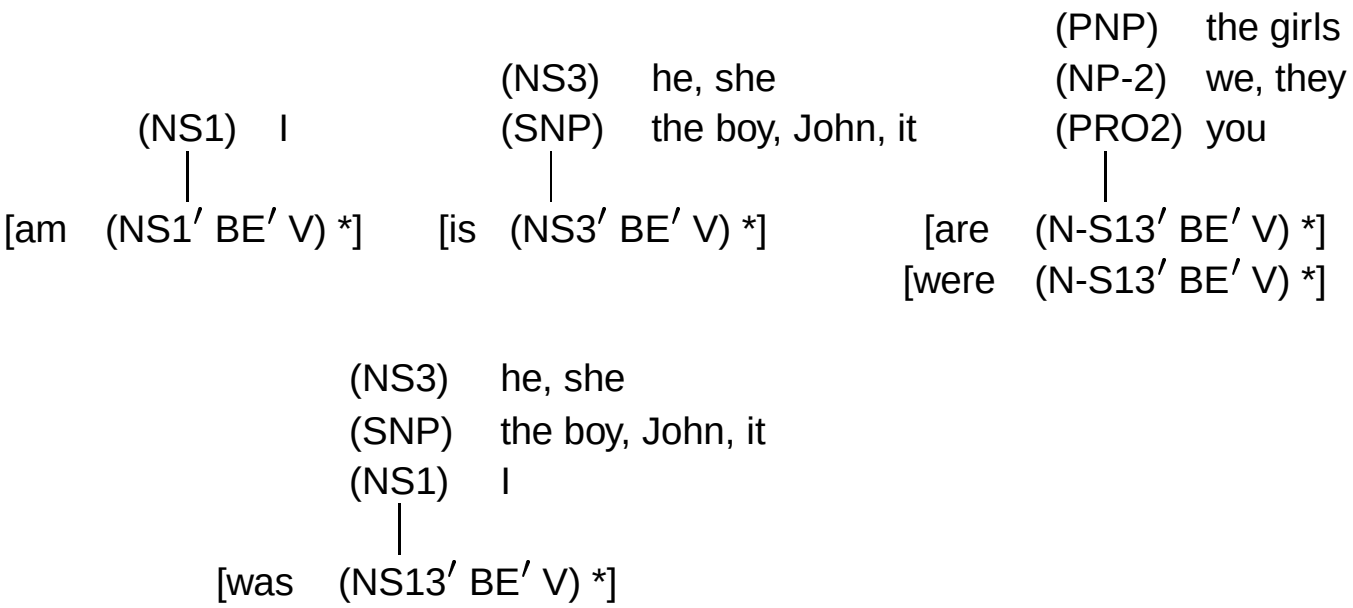


17.2.2 Agreement of fillers and valency in main verbs

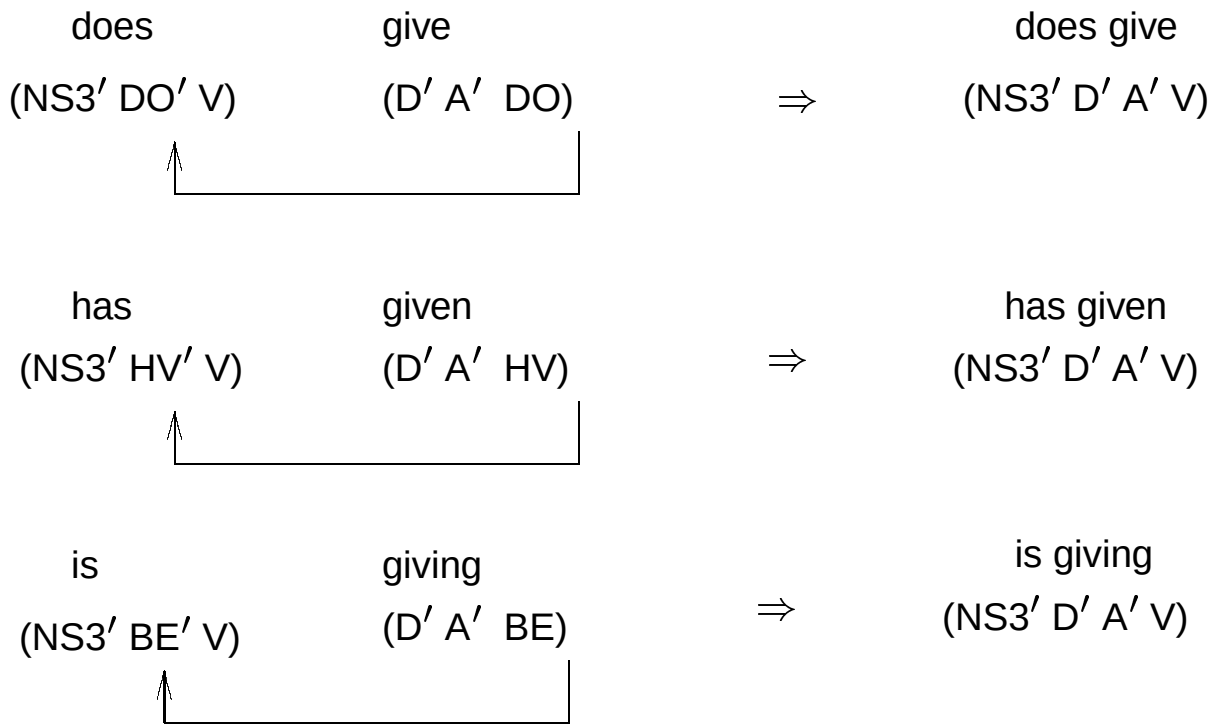


17.3 Complex verb forms

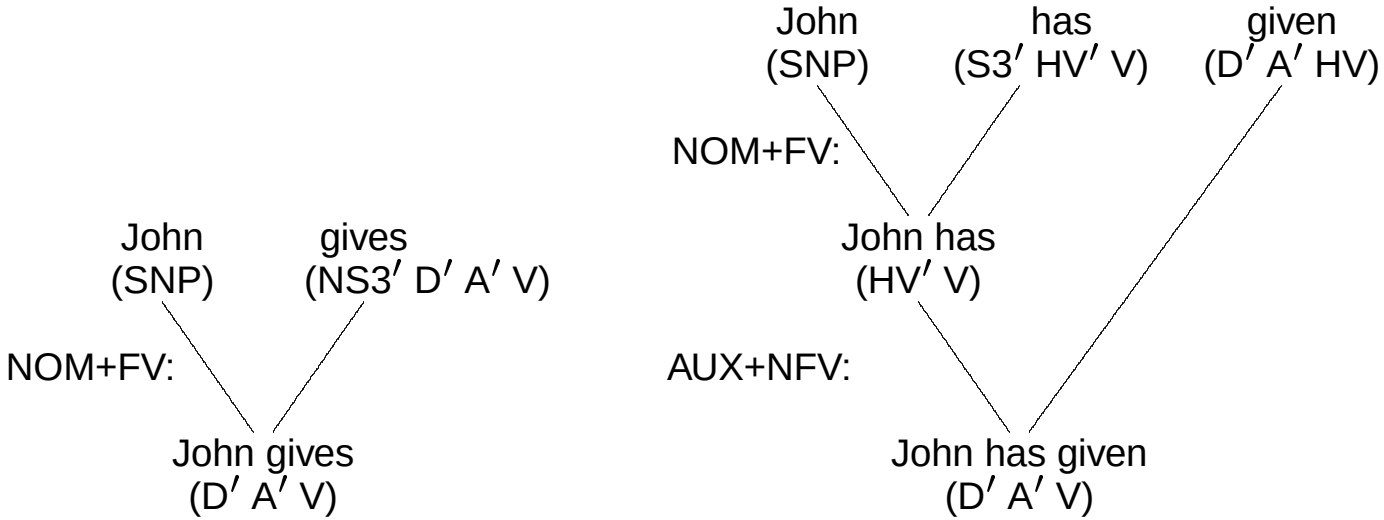
17.3.1 Nominative agreement of the auxiliary be



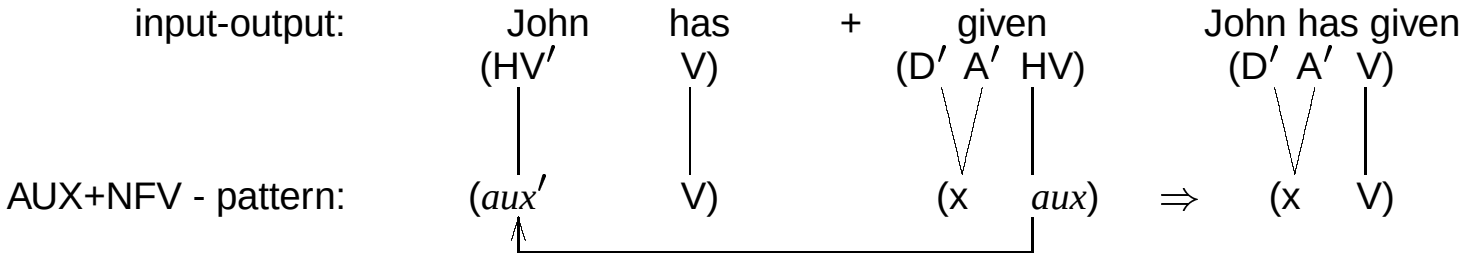
17.3.2 Complex verb forms of English



17.3.3 Comparing basic and complex verb forms of English



17.3.4 AUX+NFV adding a nonfinite verb



17.4 Finite state backbone of LA-syntax (*LA-E2*)

17.4.1 *LA-E2*: an English LA-syntax with complex NPs

LX $=_{def}$ {[Julia (SNP) *], [John (SNP) *], [Suzy (SNP) *], [it (SNP) *],
[boy (SN) *], [boys (PN) *], [girl (SN) *], [girls (PN) *], [book (SN) *],
[books (PN) *], [a (SN′ SNP) *], [every (SN′ SNP) *], [the (SN′ SNP) *],
[all (PN′ PNP) *], [several (PN′ PNP) *], [the (PN′ PNP) *]
[I (NS1) *], [you (PRO2), [he (NS3) *], [she (NS3) *], [it (SNP) *],
[we (NP-2) *], [they (NP-2) *], [me (OBQ) *], [him (OBQ) *],
[her (OBQ) *], [us (OBQ) *], [them (OBQ) *]
[am (NS1′ BE′ V) *], [is (NS3′ BE′ V) *], [are (N-S13′ BE′ V) *]
[was (NS13′ BE′ V) *], [were (N-S13′ BE′ V) *]
[have (N-S3′ HV′ V) *], [has (NS3′ HV′ V) *], [had (N′ HV′ V) *]
[do (N-S3′ DO′ V) *], [does (NS3′ DO′ V) *], [did (N′ DO′ V) *]
[give (N-S3′ D′ A′ V) *], [gives (NS3′ D′ A′ V), [gave (N′ D′ A′ V) *],
[give (D′ A′ DO) *], [given (D′ A′ HV) *], [giving (D A BE) *]
[like (N-S3′ A′ V) *], [likes (NS3′ A′ V), [liked (N′ A′ V) *]
[like (A′ DO) *], [liked (A′ HV) *], [liking (A′ BE) *]
[sleep (N-S3′ V) *], [sleeps (NS3′ V) *], [slept (N′ V) *]
[sleep (DO) *], [slept (HV) *], [sleeping (BE) *]}

Variable definition:

np' ε {N', N-S3', NS1', NS3', NS13', N-S13', D', A'}, (valency positions)

np ε {PRO2, NS1, NS3, NP-2, SNP, PNP, PN, OBQ} (valency fillers), and

if *np* = PRO2, then $np' \epsilon$ {N′, N-S3′, N-S13′, D′, A′},

if *np* = NS1, then $np' \epsilon$ {N′, N-S3′, NS1′, NS13′},

if *np* = NS3, then $np' \epsilon$ {NS3′, NS13′},

if *np* = NP-2, then $np' \epsilon$ { N′, N-S3′},

if *np* = SNP, then $np' \epsilon$ { N′, NS3′, NS13′, D′, A′},

if *np* = PNP, then $np' \epsilon$ {N′, N-S3′, N-S13′, D′, A′},

if *np* = OBQ, then $np' \epsilon$ {D′, A′},

n ϵ {SN, PN} and n' correspondingly SN′ or PN′,

aux ϵ {DO, HV, BE} and aux' correspondingly DO′, HV′ or BE′

x, y = .?.?.?.? (arbitrary sequence up to length 4)

$ST_S =_{def}$ { [(x) {1 DET+ADJ, 2 DET+N, 3 NOM+FV}] }

DET+ADJ: (n' x) (ADJ) $\Rightarrow$ (*n* x) {4 DET+ADJ, 5 DET+N}

DET+N: (n' x) (*n*) $\Rightarrow$ (x) {6 NOM+FV, 7 FV+MAIN}

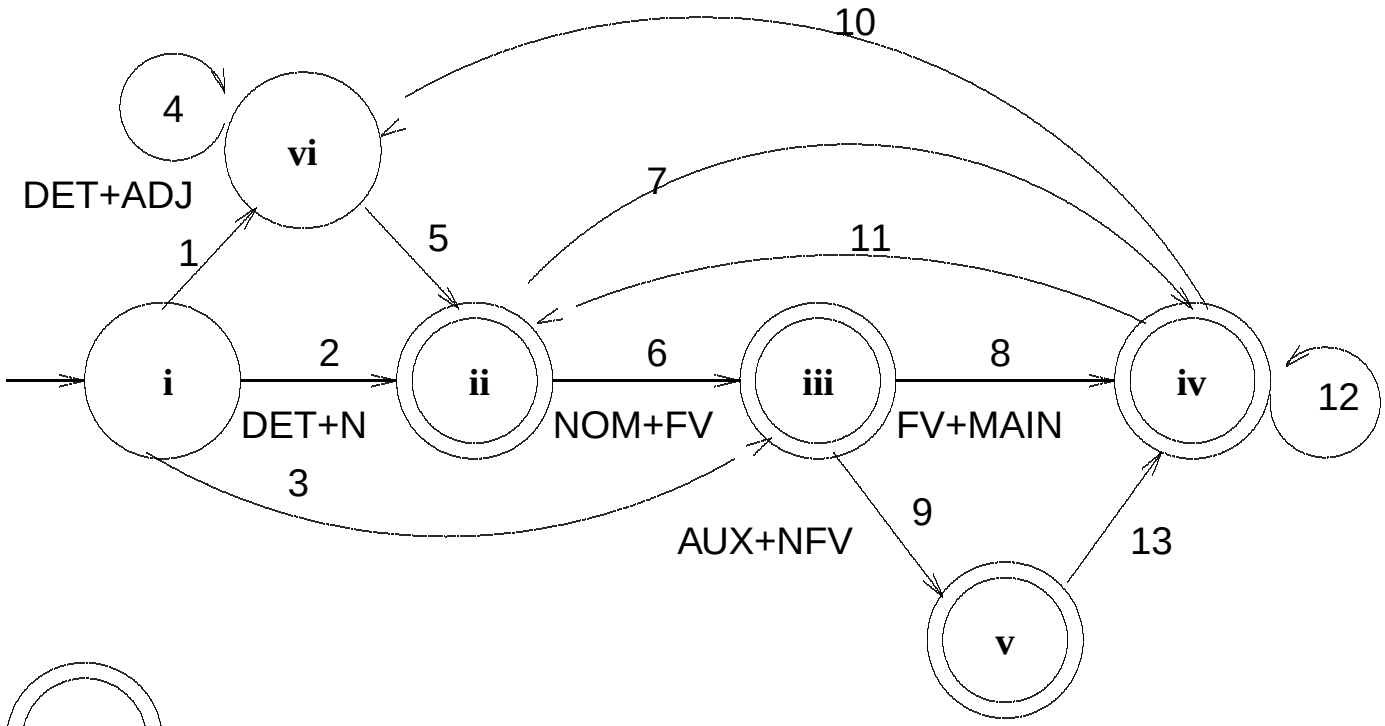
NOM+FV: (*np*) (np' x V) $\Rightarrow$ (x V) {8 FV+MAIN, 9 AUX+NFV}

FV+MAIN: (np' x V) (y *np*) $\Rightarrow$ (y x V) {10 DET+ADJ, 11 DET+N, 12 FV+MAIN}

AUX+NFV: (aux' V) (x *aux*) $\Rightarrow$ (x V) {13 FV+MAIN}

$ST_F =_{def}$ { [(V) rp_{nom+fv}], [(V) $rp_{aux+nfv}$], [(V) $rp_{fv+main}$], [(V) rp_{det+n}]}

17.4.2 The finite state backbone of *LA-E2*



= possible final state

= not a possible final state

(ii)	2, 5, 11	DET+N
(iii)	3, 6	NOM+FV
(iv)	7, 8, 12, 13	FV+MAIN
(v)	9	AUX+NFV
(vi)	1, 4, 10	DET+ADJ

17.4.3 Specifying the transition numbers in the input

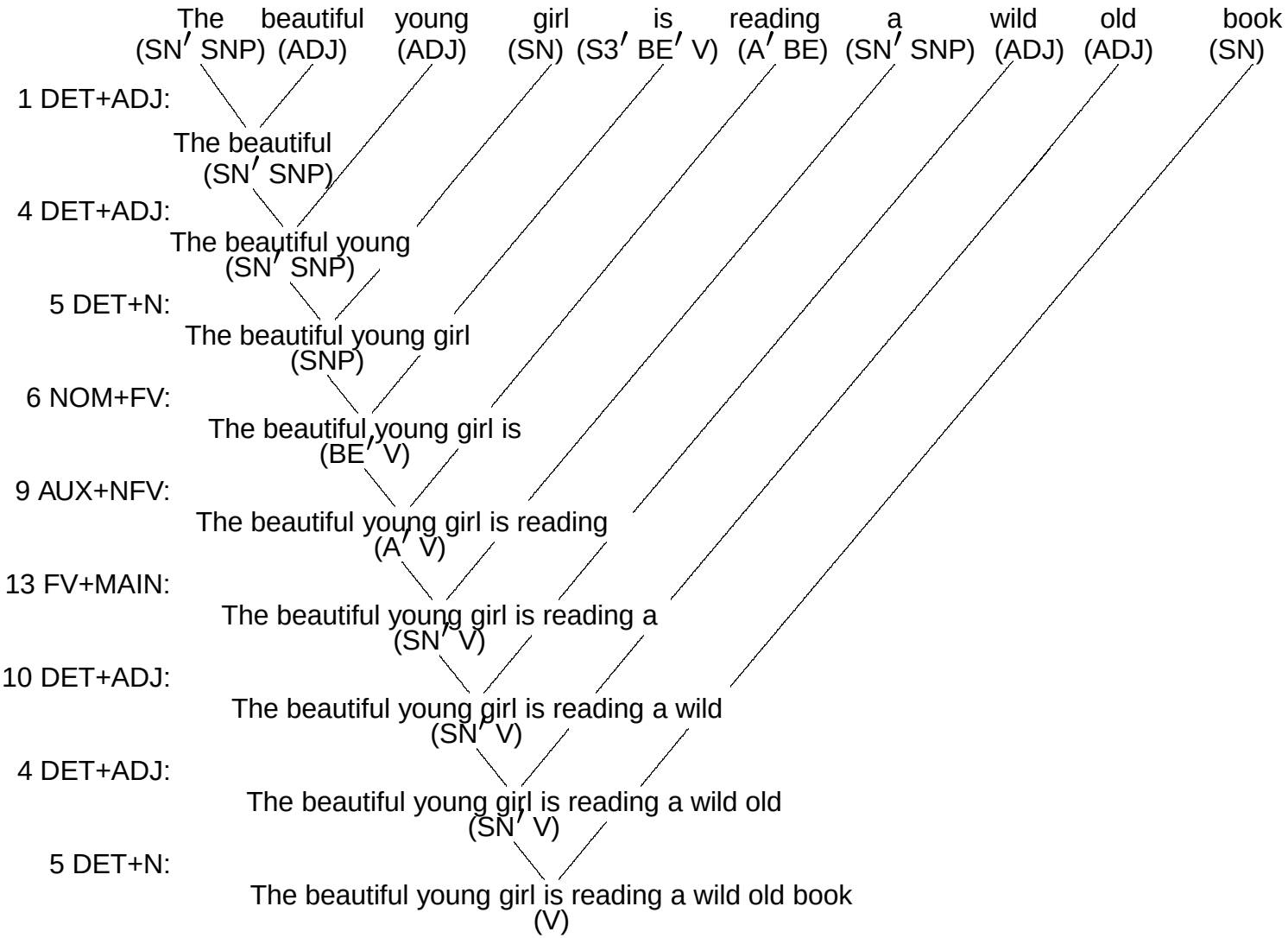
Peter 3 gave 8 Mary 12 a 11 book

the 1 beautiful 4 young 5 girl 6 is 9 reading 13 a 10 wild 4 old 5 book

the 2 boy 6 gave 8 the 11 girl 7 a 11 book

Peter 3 gave 8 Mary 12 Suzy

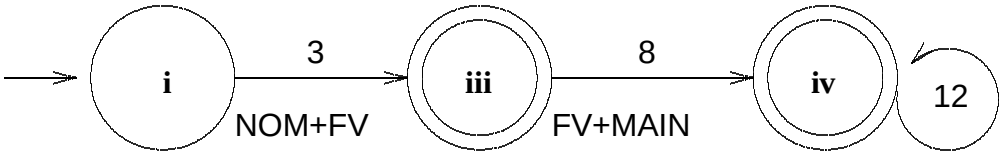
17.4.4 Syntactic analysis with transition numbers



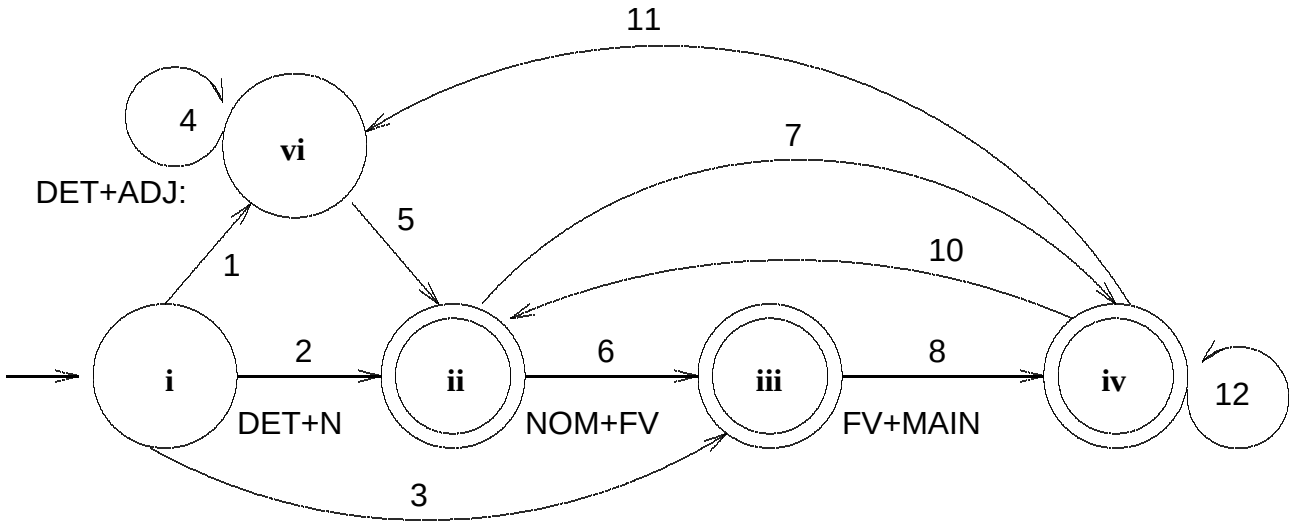
17.5 Yes/no-interrogatives (*LA-E3*) and grammatical perplexity

17.5.1 Expanding *LA-E1* to *LA-E1.5* handling complex NPs

LA-E1



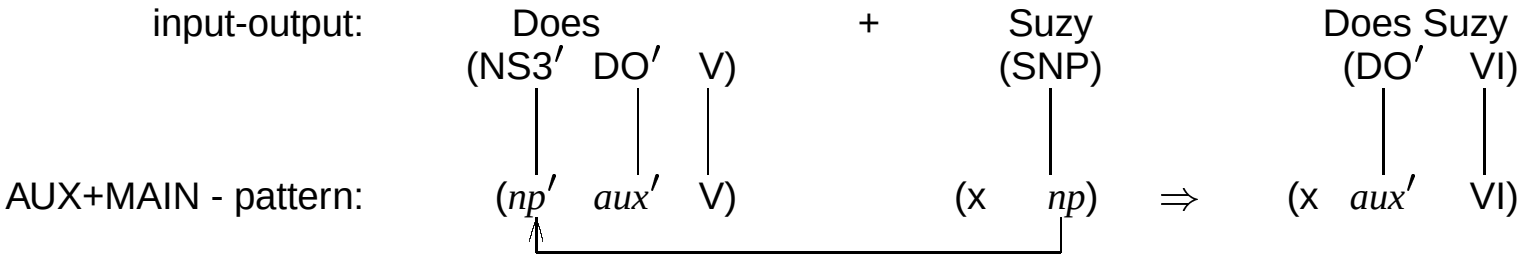
LA-E1.5



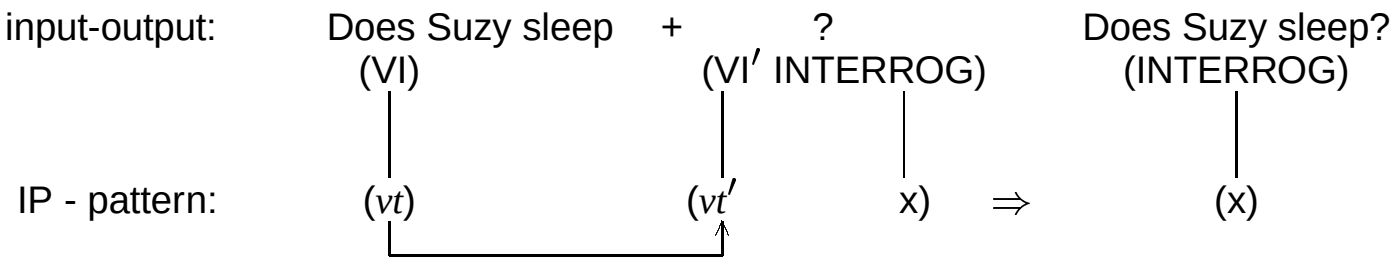
17.5.2 Comparing declaratives and Yes/No-Interrogatives

Suzy does like the book.	Does Suzy like the book?
Suzy has liked the book.	Has Suzy liked the book?
Suzy is liking the book.	Is Suzy liking the book?

17.5.3 Categorial operation of AUX+MAIN



17.5.4 Categorial operation of IP

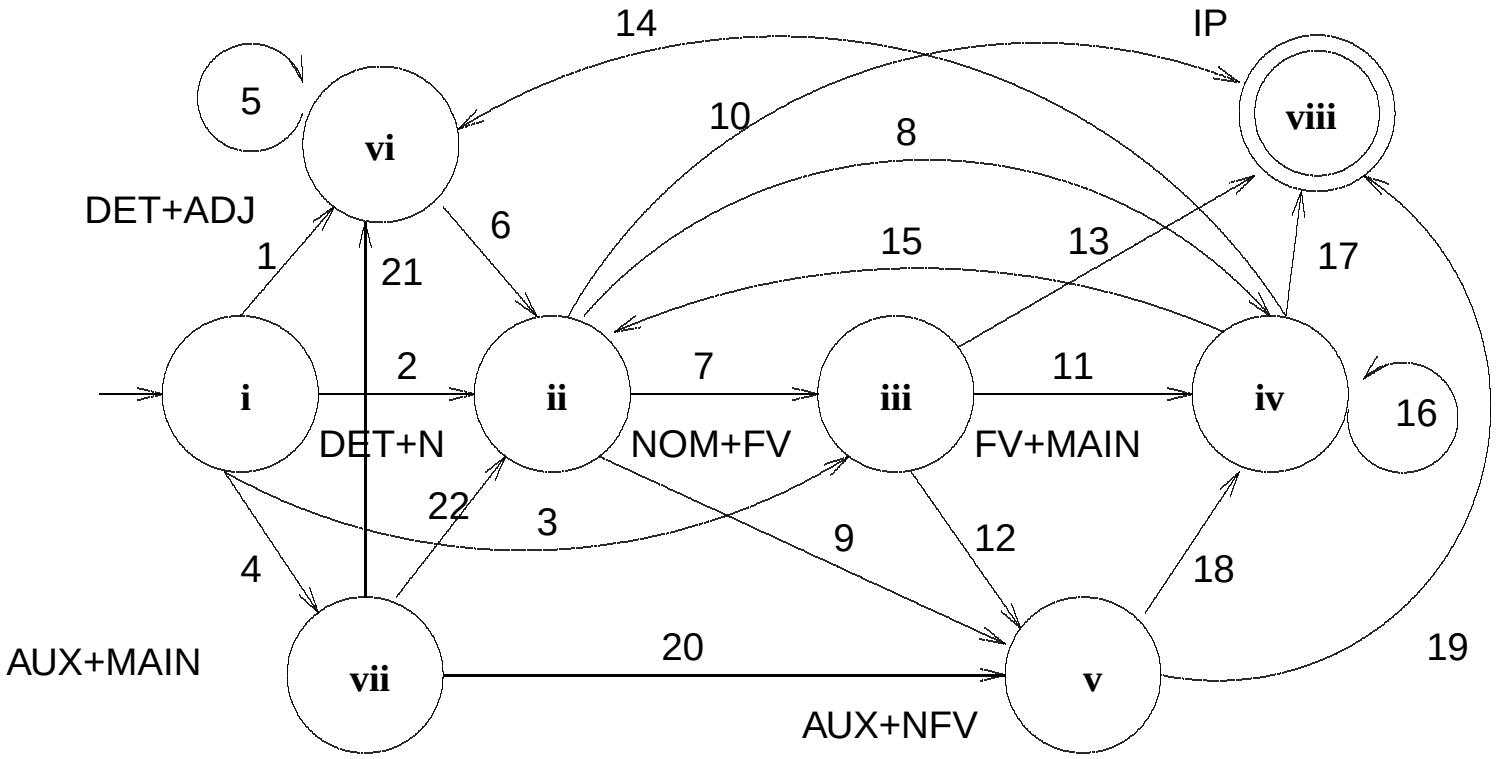


17.5.5 *LA-E3* for English yes/no-interrogatives

LX = LX of *LA-E2* plus {[. (V′ decl) *], [? (V′ interrog) *], [? (VI′ interrog)*]}
Variable definitions = that of *LA-E2* plus $vt \in$ {V, VI},

$ST_S =_{def}$ { [(x) {1 DET+ADJ, 2 DET+N, 3 NOM+FV, 4 AUX+MAIN}] }
DET+ADJ: (*n*′ x) (ADJ) ⇒ (*n*′ x) {5 DET+ADJ, 6 DET+N}
DET+N: (*n*′ x) (*n*) ⇒ (x) {7 NOM+FV, 8 FV+MAIN, 9 AUX+NFV, 10 IP}
NOM+FV: (*np*) (*np*′ x V) ⇒ (x V) {11 FV+MAIN, 12 AUX+NFV, 13 IP}
FV+MAIN: (*np*′ x V) (y *np*) ⇒ (y x V) {14 DET+ADJ, 15 DET+N, 16 FV+MAIN, 17 IP}
AUX+NFV: (*aux*′ V) (x *aux*) ⇒ (x V) {18 FV+MAIN, 19 IP}
AUX+MAIN:(*np*′ *aux*′ V) (x *np*) ⇒ (x *aux*′ VI) {20 AUX+NFV, 21 DET+ADJ, 22 DET+N}
IP: (*vt*) (*vt*′ x) ⇒ (x) {}
$ST_F =_{def}$ { [(decl) rp_{ip}], [(interrog) rp_{ip}]}

17.5.6 The finite state backbone of *LA-E3*



ii	2, 6, 15, 22	DET+N	v	9, 12, 20	AUX+NFV
iii	3, 7	NOM+FV	vi	1, 5, 14, 21	DET+ADJ
iv	8, 11, 16, 18	FV+MAIN	vii	4	AUX+MAIN
			viii	10, 13, 17, 19	IP

17.5.7 Perplexity

> Perplexity is, crudely speaking, a measure of the size of the set of words from which the next word is chosen given that we observe the history of the spoken words.
>
> S. Roukos 1995

18. LA-syntax for German

18.1 Standard procedure of syntactic analysis

18.1.1 Differences between natural languages

Natural languages are all based on the same time-linear derivation order. They differ ony in their language specific handling of

- *agreement*
- *word order*
- *valency structure* (lexicalization)

18.1.2 Phase I of standard procedure

1. Formal treatment of declarative main clauses with elementary finite verbs and elementary nominal fillers determines the basic typological properties of the natural language,
2. Extension to complex nominal fillers requires treatment of the internal and the external agreement restrictions of derived noun phrases, and the time-linear derivation of complex fillers in pre- and postverbal position.
3. The extension to complex verb phrases treats complex tenses and modalities.

18.1.3 What has to be done before Phase II

A theoretically well-founded semantic and pragmatic interpretation for the syntactic analysis developed so far.

18.1.4 Phase II of the standard procedure

The syntactic analyses of the second phase should be developed directly out of the semantic and pragmatic interpretation, and be provided for both, the speaker and the hearer mode.

Topics: (i) addition of basic and derived modifiers ranging from adverbs over prepositional phrases to subordinate clauses, (ii) treatment of sentential subjects and objects including infinitive constructions, (iii) handling of different syntactic moods like interrogative and different verbal moods like passive, and (iv) treatment of conjunctions including gapping constructions.

18.1.5 Distinctive categorization of determiners

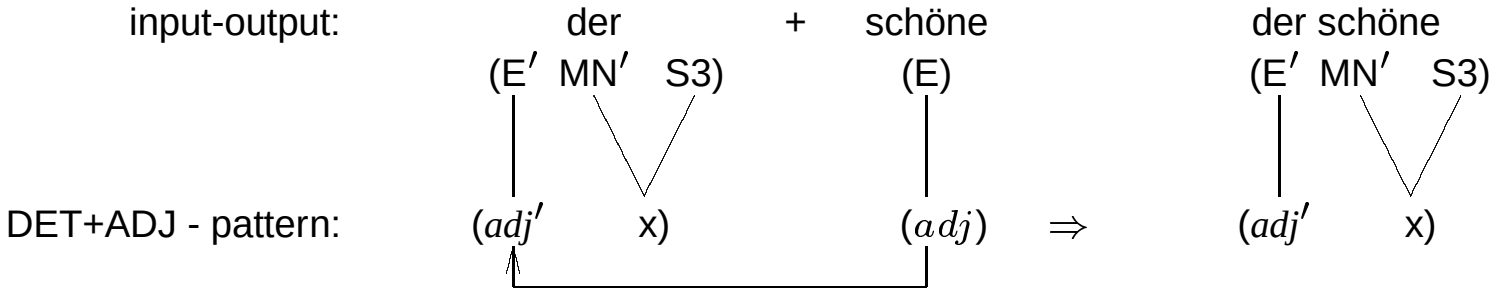
definite article

[der	(E′ MN′ S3)	
	(EN′ F′ G&D)	
	(EN′ P-D′ G)	DEF-ART]
[des	(EN′ -FG′ G)	DEF-ART]
[dem	(EN′ -FD′ D)	DEF-ART]
[den	(EN′ M-N′ A)	
	(EN′ PD′ D)	DEF-ART]
[das	(E′ N-G′ S3&A)	DEF-ART]
[die	(E′ F′ S3&A)	
	(EN′ P-D′ P3&A)	DEF-ART]

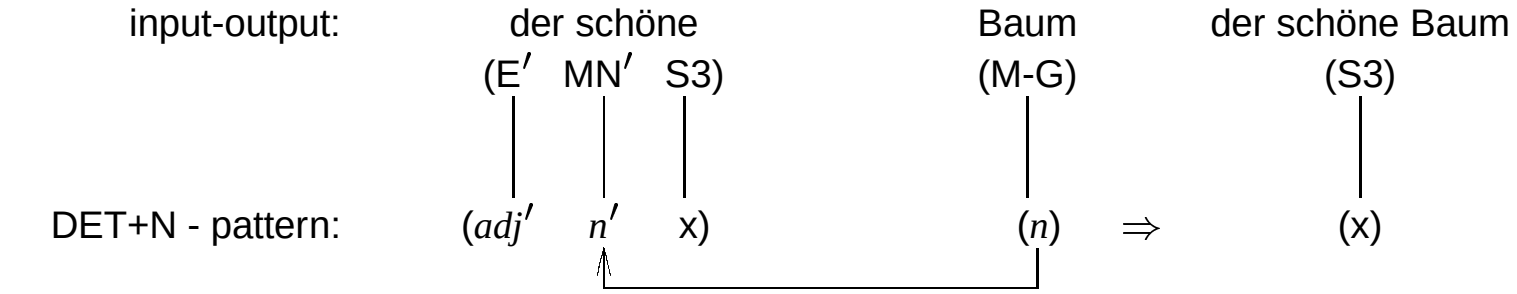
indefinite article

[ein	(ER′ MN′ S3)	
	(ES′ N-G′ S3&A)	INDEF-ART]
[eines	(EN′ -FG′ G)	INDEF-ART]
[einem	(EN′ -FD′ D)	INDEF-ART]
[einen	(EN′ M-N′ A)	INDEF-ART]
[eine	(E′ F′ S3&A)	INDEF-ART]
[einer	(EN′ F′ G&D)	INDEF-ART]

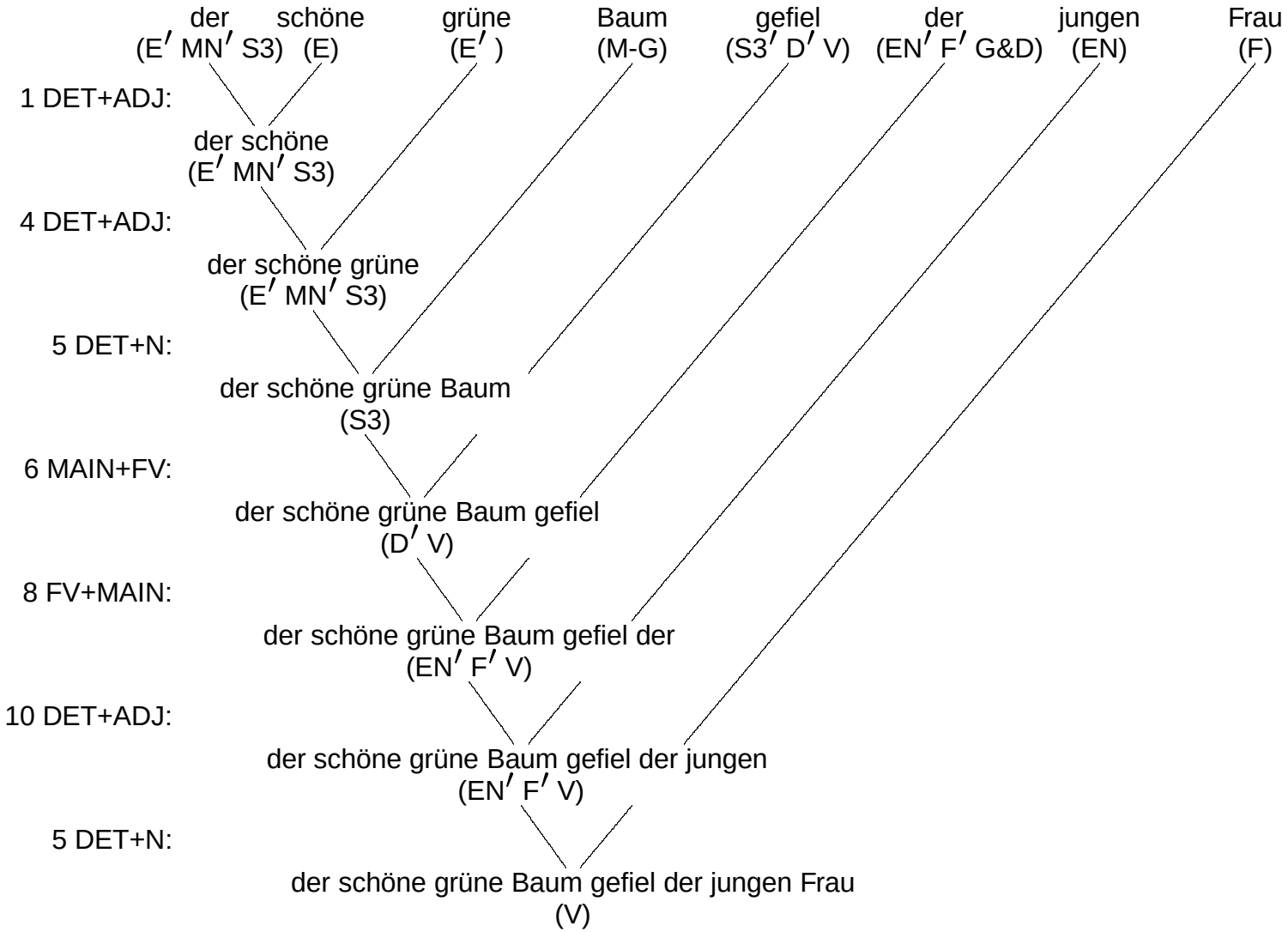
18.1.6 Categorial operation of DET+ADJ



18.1.7 Categorial operation of DET+N



18.1.8 Pre- and postverbal derivation of noun phrases

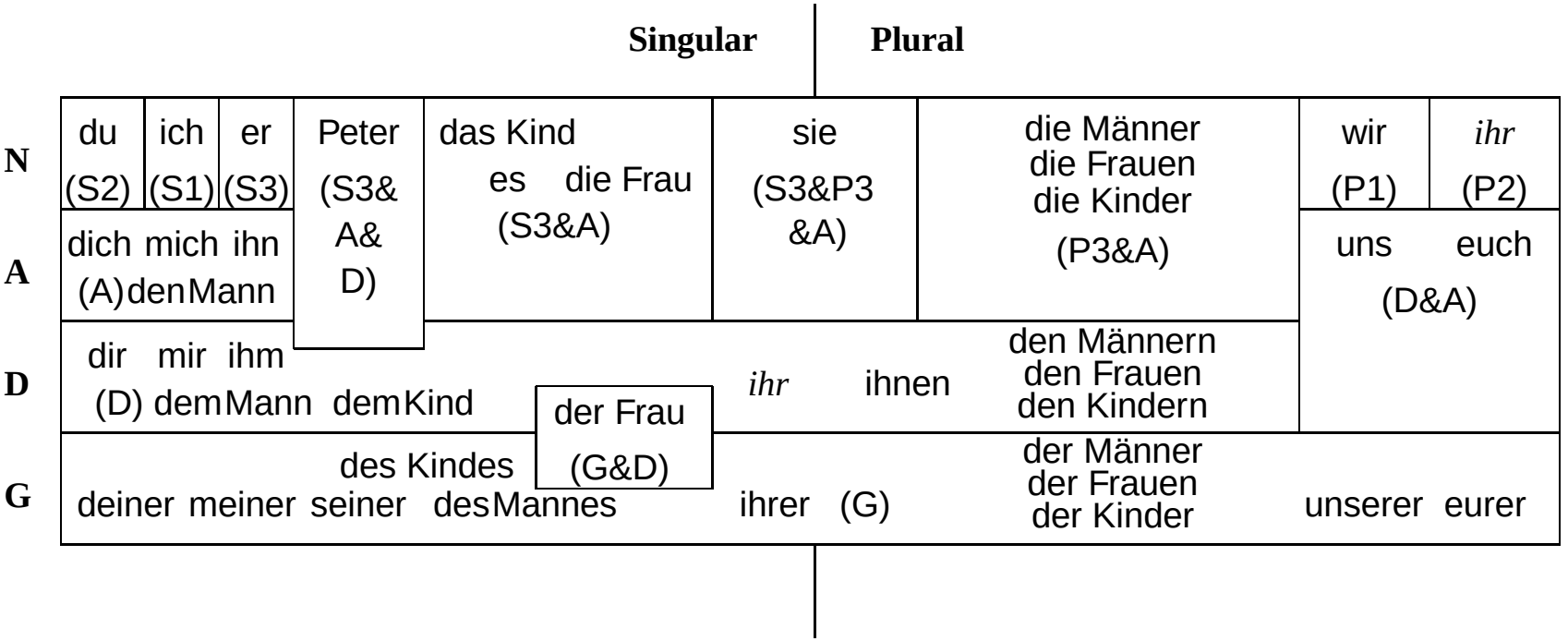


18.2 German field of referents (*LA-D2*)

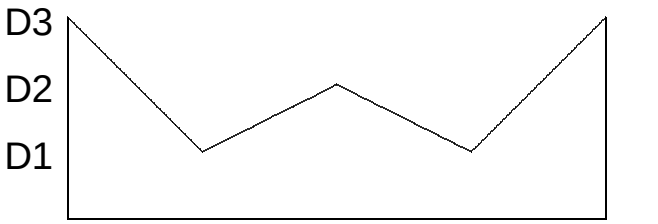
18.2.1 Traditional paradigms of German noun phrases

	Masculinum	*Femininum*	*Neutrum*	*Plural*
Nominative	der Mann	die Frau	das Kind	die Männer, etc.
Genitive	des Mannes	der Frau	des Kindes	der Männer, etc.
Dative	dem Mann	der Frau	dem Kind	den Männern, etc.
Accusative	den Mann	die Frau	das Kind	die Männer, etc.

18.2.2 Distinctive categories of nominal fillers (German)



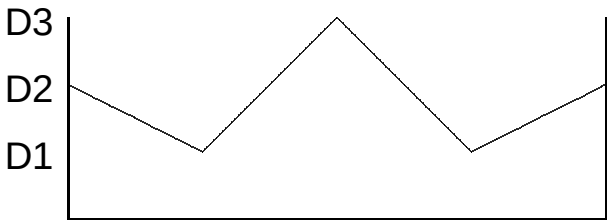
18.2.3 Centering and distance in fields of reference



he I you we they

S3 S1 SP2 P1 P3

English

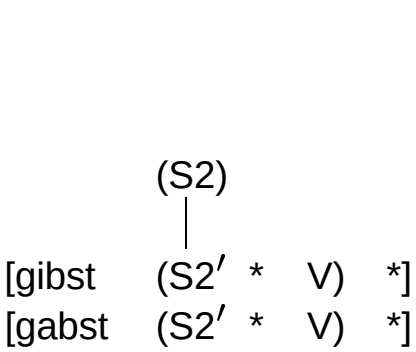


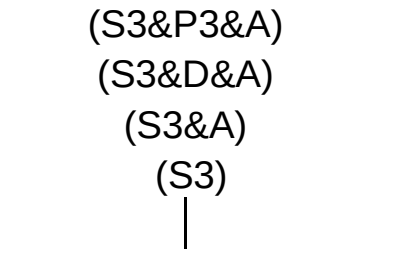
du ich sie wir ihr

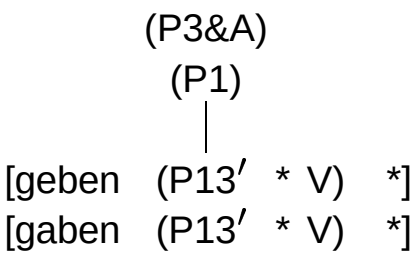
S2 S1 SP3 P1 P2

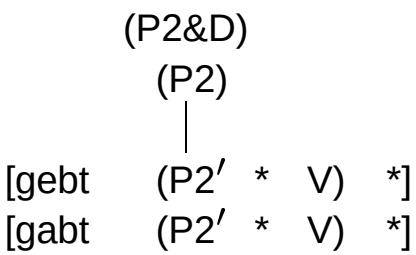
German

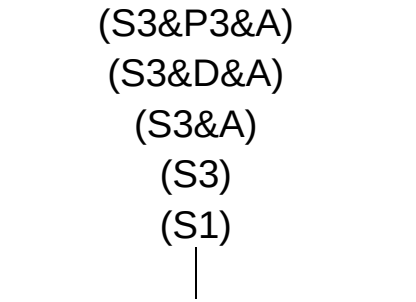
18.2.4 Agreement of nominal fillers and verbal valencies











18.2.5 German LA-grammar handling complex fillers (*LA-D2*)

LX = LX of LA-D1 plus the determiners defined in 18.1.5, the nouns defined in 14.5.1, 14.5.2, and the following pronouns

[ich (S1) *], [du (S2) *], [er (S3) *], [es (S3&A) *], [wir (P1) *],
[ihr (P2&D) *], [sie (S3&P3&A) *], [deiner (G) *], [uns (D&A) *],
[euch (D&A) *], [mir (D) *], [dir (D) *], [ihm (D) *], [mich (A) *],
[dich (A) *], [ihn (A) *]

plus adjectives with comparation

[schöne (E) *]	[schönere (E) *]	[schönste (E) *]
[schönen (EN) *]	[schöneren (EN) *]	[schönsten (EN) *]
[schöner (ER) *]	[schönerer (ER) *]	[schönster (ER) *]
[schönes (ES) *]	[schöneres (ES) *]	[schönstes (ES) *]

plus finite main verb forms of differing valency structures

[gebe (S1′ D′ A′ V) *]	[lese (S1′ A′ V) *]	[schlafe (S1′ V) *]
[gibst (S2′ D′ A′ V) *]	[liest (S23′ A′ V) *]	[schläfst (S2′ V) *]
[gibt (S3′ D′ A′ V) *]	[lesen (P13′ A′ V) *]	[schläft (S3′ V) *]
[geben (P13′ D′ A′ V) *]	[lest (P2′ A′ V) *]	[schlafen (P13′ V) *]
[gebt (P2′ D′ A′ V) *]	[las (S13′ A′ V) *]	[schlaft (P2′ V) *]
[gab (S13′ D′ A′ V) *]	[last (S2P2′ A′ V) *]	[schlief (S13′ V) *]
[gabst (S2′ D′ A′ V) *]	[lasen (P13′ A′ V) *]	[schliefst (S2′ V) *]

[gaben (P13′ D′ A′ V) *] [schliefen (P13′ V) *]
[gabt (P2′ D′ A′ V) *] [schlieft (P2′ V) *]

variable definition
np ∈ {S1, S2, S3, P1, P2, P2&D, G, G&D, D, A, S3&A, S3&D&A, D&A,
P3&A, S3&P3&A}
np′ ∈ {S1′, S13′, S2′, S23′, S2P2′, S3′, P13′, P2′, G′, D′, A′}
and if *np* ∈ {G, D, A}, then *np*′ is correspondingly G′, D′, or A′
if *np* = P1, then *np*′ = P13′
if *np* = S1, then *np*′ ∈ {S1′, S13′}
if *np* = S2, then *np*′ ∈ {S2′, S23′}
if *np* = S3, then *np*′ ∈ {S3′, S23′}
if *np* = P3&A, then *np*′ ∈ {P13′, A′}
if *np* = P2&D, then *np*′ ∈ {P2′, D′}
if *np* = G&D, then *np*′ ∈ {G′, D′}
if *np* = D&A, then *np*′ ∈ {D′, A′}
if *np* = S3&A, then *np*′ ∈ {S3′, S23′, A′}
if *np* = S3&D&A, then *np*′ ∈ {S3′, S23′, D′, A′}
if *np* = S3&P3&A, then *np*′ ∈ {S3′, S23′, P13′, A′}
n ∈ {MN, M-G, M-NP, M-GP, MGP, M-GP-D, F, N-G, -FG, -FD, N-GP, N-GP-D,

NDP-D, P, P-D, PD},

n' ϵ {MN′, M-N′, F′, N-G′, -FG′, -FD′, P-D′, PD′}, and

if $n \epsilon$ {MN, -FG, -FD, F, P-D, PD}, then n' is corresponding

if n = M-G, then $n' \epsilon$ {MN′, M-N′}

if n = M-NP, then $n' \epsilon$ {-FG′, -FD′, P-D′, PD′ }

if n = M-GP, then $n' \epsilon$ {MN′, -FD′, M-N′, P-D′, PD′}

if n = MGP, then $n' \epsilon$ {-FG′, P-D′, PD′}

if n = M-GP-D, then $n' \epsilon$ {MN′, -FD′, M-N′, P-D′}

if n = N-G, then $n' \epsilon$ {N-G′, -FG′, -FD′}

if n = N-GP, then $n' \epsilon$ {N-G′, -FG′, -FD′, P-D′, PD′}

if n = N-GP-D, then $n' \epsilon$ {N-G′, -FG′, -FD′, P-D′}

if n = NDP-D, then $n' \epsilon$ {-FD′, P-D′}

if n = P, then $n' \epsilon$ {P-D′, PD′}

adj ϵ {e, en, es, er} and *adj′* is corresponding

$ST_S =_{def}$ { [(x) {1 DET+ADJ, 2 DET+N, 3 MAIN+FV}] }

DET+ADJ: (*adj′* x) (*adj*) $\Rightarrow$ (*adj′* x) {4 DET+ADJ, 5 DET+N}

DET+N: (*adj′ n′* x) (*n*) $\Rightarrow$ (x) {6 MAIN+FV, 7 FV+MAIN}

MAIN+FV: (*np*) (x *np′* y V) $\Rightarrow$ (x y V) {8 FV+MAIN}

FV+MAIN: (x *np′* y V) (z *np*) $\Rightarrow$ (z x y V) {9 FV+MAIN, 10 DET+ADJ, 11 DET+N}

$ST_F =_{def}$ { [(V) $rp_{MAIN+FV}$], [(V) $rp_{FV+MAIN}$], [(V) rp_{DET+N}] }

18.3 Verbal positions in English and German

18.3.1 Finite verb position in declarative main clauses

	English: post-nominative	*German*: verb-second
1.	Julia *read* a book	Julia *las* ein Buch
2.	*a book *read* Julia	Ein Buch *las* Julia
3.	Yesterday Julia *read* a book	*Gestern Julia *las* ein Buch
4.	*Yesterday *read* Julia a book	Gestern *las* Julia ein Buch
5.	Julia yesterday *read* a book	*Julia gestern *las* ein Buch
6.	*While Mary slept, *read* Julia a book	Als Maria schlief, *las* Julia ein Buch
7.	While Mary slept, Julia *read* a book	*Als Maria schlief, Julia *las* ein Buch

18.3.2 Nonfinite main verb position in declarative main clauses

English: contact position	*German*: distance position
1. Julia *has slept*	Julia *hat geschlafen*
2. Julia *has read* a book	*Julia *hat gelesen* ein Buch
3. *Julia *has* a book *read*	Julia *hat* ein Buch *gelesen*
4. Yesterday Julia *has read* a book	*Gestern Julia *hat gelesen* ein Buch
5. *Yesterday *has* Julia a book *read*	Gestern *hat* Julia ein Buch *gelesen*
6. Julia *has given* M. a book yesterday	*Julia *hat gegeben* M. ein Buch gestern
7. *Julia *has* M. yesterday a book *given*	Julia *hat* M. gestern ein Buch *gegeben*

18.3.3 Satzklammer in German

Julia has the offer of the opposing party yesterday afternoon

Julia hat das Angebot der Gegenseite gestern nachmittag *abgelehnt.*	declined
verworfen.	refused
kritisiert.	criticized
zurückgewiesen.	rejected

18.3.4 Verb position in subordinate clauses

English: post-nominative	*German*: clause final
1. before Julia *slept*	bevor Julia *schlief*
2. before Julia *had slept*	*bevor Julia *hatte geschlafen*
3. *before Julia *slept had*	bevor Julia *geschlafen hatte*
4. before Julia *bought* the book	*bevor Julia *kaufte* das Buch
5. *before Julia the book *bought*	bevor Julia das Buch *kaufte*
6. before Julia *had bought* the book	*bevor Julia *hatte gekauft* das Buch
7. *before the book a man *bought*	bevor das Buch ein Mann *kaufte*

18.4 Complex verbs and elementary adverbs (*LA-D3*)

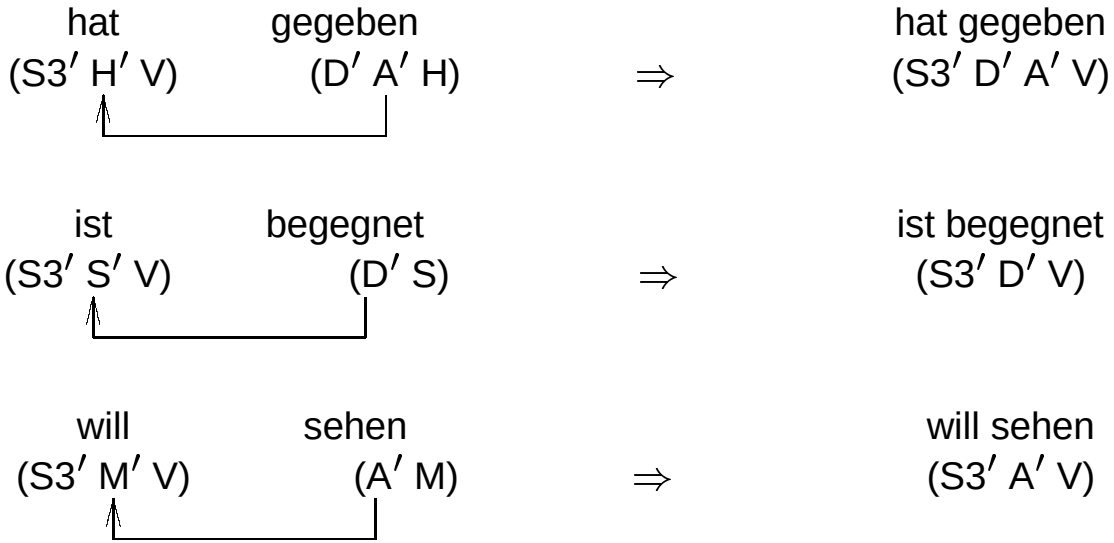
18.4.1 LA-paradigms of German auxiliaries and modals

[bin (S1′ S′ V) *]
[bist (S2′ S′ V) *]
[ist (S3′ S′ V) *]
[sind (P13′ S′ V) *]
[seid (P2′ S′ V) *]
[war (S13′ S′ V) *]
[warst (S2′ S′ V) *]
[waren (P13′ S′ V) *]
[wart (P2′ S′ V) *]

[habe (S1′ H′ V) *]
[hast (S2′ H′ V) *]
[hat (S3′ H′ V) *]
[haben (P13′ H′ V) *]
[habt (P2′ H′ V) *]
[hatte (S13′ H′ V) *]
[hattest (S2′ H′ V) *]
[hatten (P13′ H′ V) *]
[hattet (P2′ H′ V) *]

[kann (S13′ M′ V) *]
[kannst (S2′ M′ V) *]
[können (P13′ M′ V) *]
[könnt (P2′ M′ V) *]
[konnte (S13′ M′ V) *]
[konntest (S2′ M′ V) *]
[konnten (P13′ M′ V) *]
[konntet (P2′ M′ V) *]

18.4.2 Complex verb forms of German



18.4.3 Declarative main clauses with a finite main verb

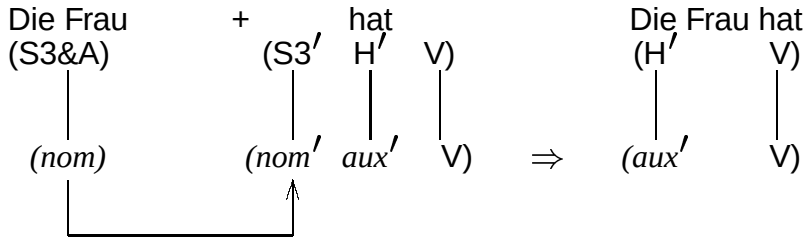
Die Frau *gab* dem Kind den Apfel
Dem Kind *gab* die Frau den Apfel

18.4.4 Declarative main clauses with an auxiliary construction

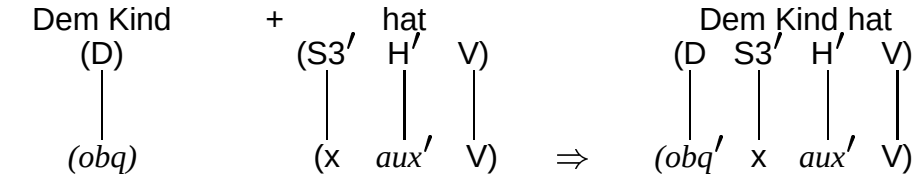
Die Frau *hat* dem Kind den Apfel *gegeben*
Dem Kind *hat* die Frau den Apfel *gegeben*

18.4.5 +FV alternatives of adding the auxiliary

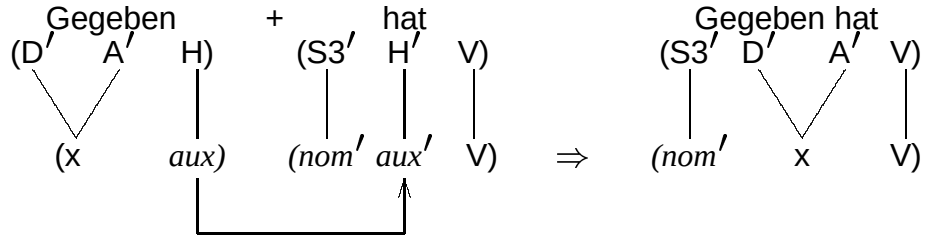
1. input-output: Die Frau + hat → Die Frau hat



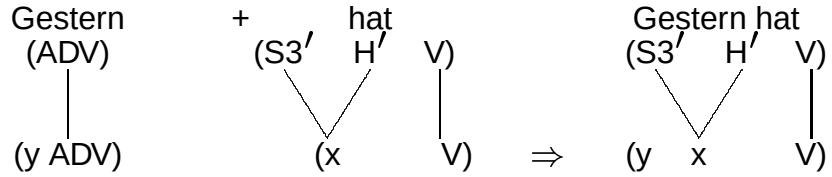
2. input-output: Dem Kind + hat → Dem Kind hat



3. input-output: Gegeben + hat → Gegeben hat



5. input-output: Gestern + hat → Gestern hat



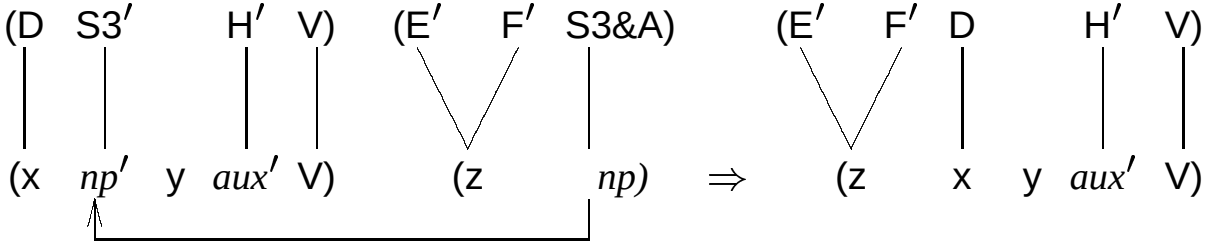
18.4.6 Extending MAIN+FV into +FV using clauses

+FV:

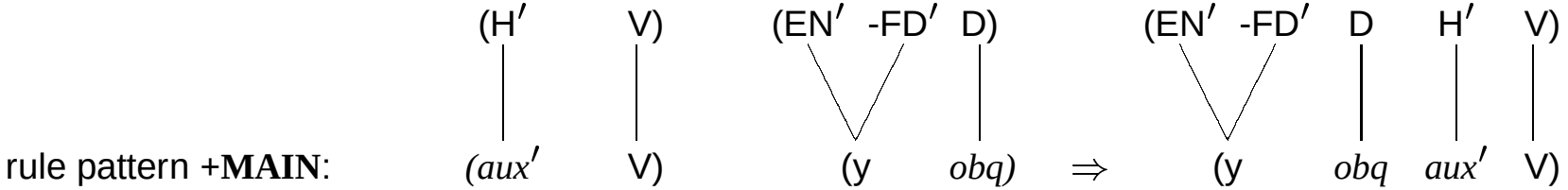
1. (*nom*)(*nom′ aux′* V) ⇒ (*aux′* V)
2. (*obq*)(x *aux′* V) ⇒ (*obq* x *aux′* V)
3. (x *aux*)(*nom′ aux′* V) ⇒ (*nom′* x V)
4. (*np*)(x *np′* y V) ⇒ (x y V)
5. (y ADV)(x V) ⇒ (y x V) {+MAIN, +NFV, +FV, +IP}

18.4.7 +MAIN Alternatives after the auxiliary

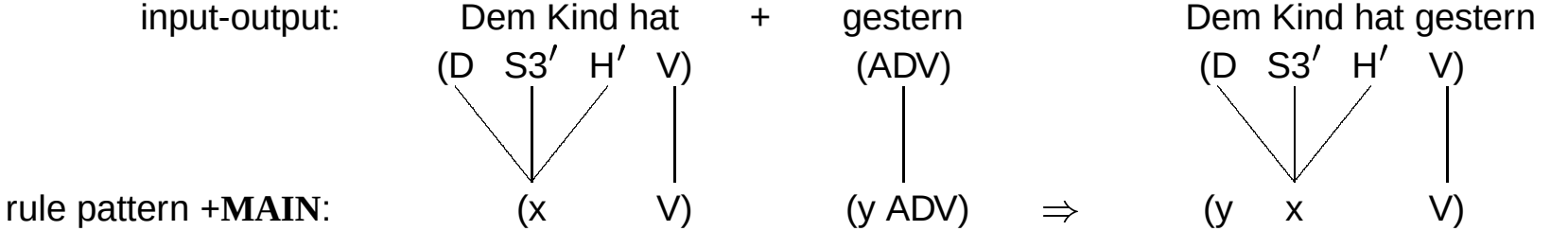
1. input-output: Dem Kind hat + die → Dem Kind hat die



2. input-output: Die Frau hat + dem → Die Frau hat dem



4. input-output: Dem Kind hat + gestern → Dem Kind hat gestern

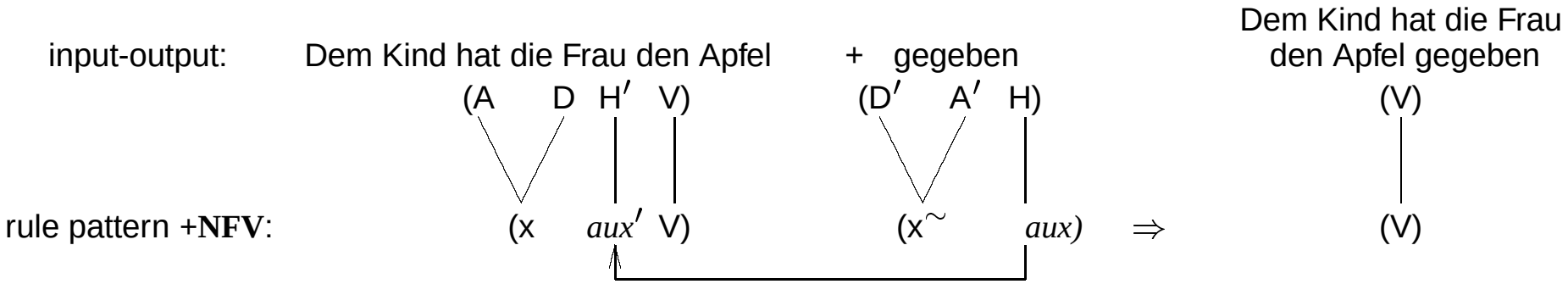


18.4.8 Extending FV+MAIN into +MAIN using clauses

+MAIN:
1. (x *nom′* y *aux′* V)(z *nom*) ⇒ (z x y *aux′* V)
2. (x *aux′* V)(y *obq*) ⇒ (y *obq* x *aux′* V)
3. (x *np′* y V)(z *np*) ⇒ (z x y V)
4. (x V)(y ADV) ⇒ (y x V)

{+ADJ, +N, +MAIN, +NFV, +FV, +IP}

18.4.9 Categorial operation of +NFV



18.4.10 German grammar handling complex verb forms (*LA-D3*)

LX = LX of *LA-D2* plus auxiliaries defined in 18.4.1, plus
nonfinite main verb form of 18.4.2, plus adverbials
[gestern (ADV) *], [hier (ADV) *], [jetzt (ADV) *], plus punctuation signs
[. (V′ DECL) *], [? (VI′ INTERROG) *], [? (V′ INTERROG) *]
variable definition = variable definition of *LA-D2* plus

$nom \in np \setminus \{D, A, D\&A\}$	nominative filler[1]
$nom' \in np \setminus \{D, A\}$	nominative valency positions
$obq \in \{D, A, D\&A\}$	oblique filler
$aux \in \{H, B, M\}$,	auxiliaries and modals
$vt \in \{V, VI\}$,	mood marker
$sm \in \{DECL, INTERROG\}$,	sentence mood

$ST_S =_{def}$ { [(x) {1 +ADJ, 2 +N, 3 +FV, 4 +NFV}] }

+ADJ: (*adj′* x) (*adj*) ⇒ (*adj* x) {5 +ADJ, 6 +N}

+N: (*adj′ n′* x) (*n*) ⇒ (x) {7 +FV, 8 +MAIN, 9 +NFV, 10 +IP}

+FV: (*nom*)(*nom′ aux′* V) ⇒ (*aux′* V)
(*obq*)(x *aux′* V) ⇒ (*obq* x *aux′* V)
(x *aux*)(*nom′ aux′* V) ⇒ (*nom′* x V)
(*np*)(x *np′* y V) ⇒ (x y V)
(ADV)(x V) ⇒ (x V) {11 +MAIN, 12 +NFV, 13 +IP}

+MAIN: (x *nom′* y *aux′* V)(z *nom*) ⇒ (z x y *aux′* V)

(x *aux′* V)(y *obq*)⇒ (y *obq* x *aux′* V)

(x *np′* y V)(z *np*) ⇒ (z x y V)

(x V)(y ADV) ⇒ (y x V) {14 +ADJ, 15 +N, 16 +MAIN, 17 +NFV, 18 +FV, 19 +IP}

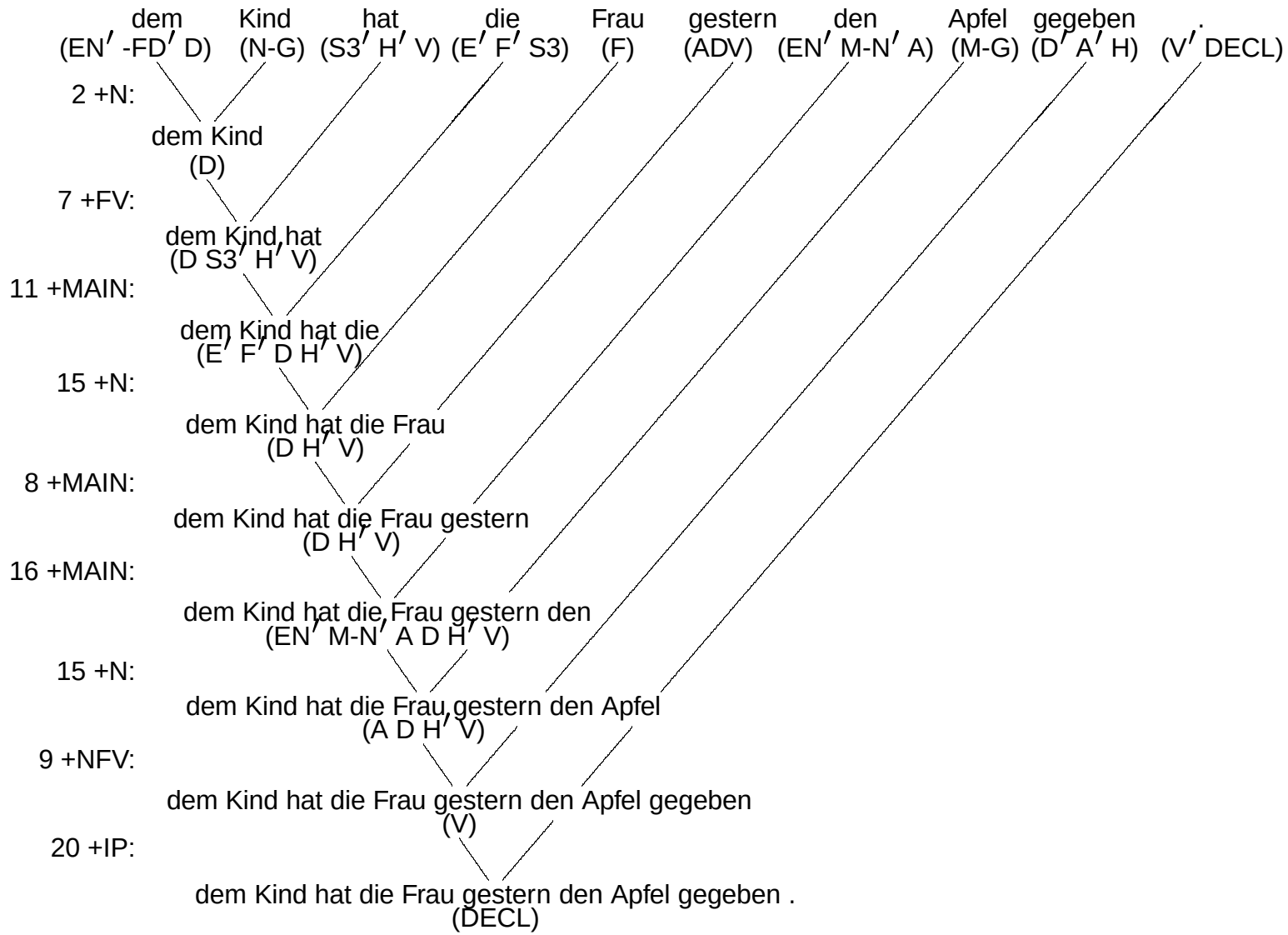
+NFV: (x *aux′* V)(x$^{\sim}$ *aux*)

(x = x$^{\sim}$) ⇒ (V) {20 +IP}

+IP: (*vt*) (*vt′ sm*) ⇒ (*sm*) {}

$ST_F =_{def}$ { [(*sm*) rp_{+ipt}]}

18.4.11 Declarative with dative preceding auxiliary



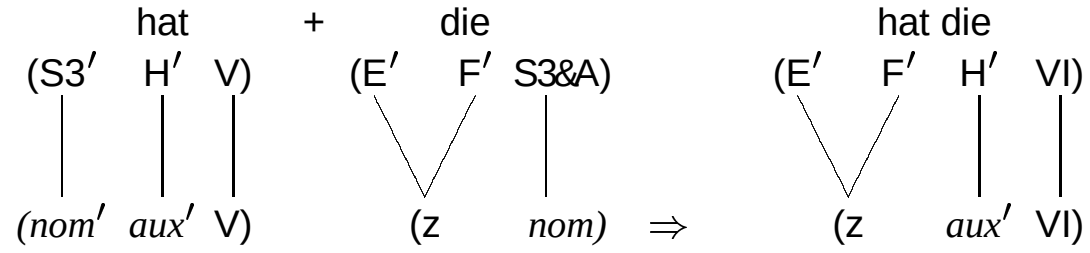
18.5 Interrogatives and subordinate clauses (*LA-D4*)

18.5.1 Interrogative with and without auxiliary

1. *Hat* die Frau dem Kind gestern den Apfel *gegeben* ?
   (*Has the woman the child yesterday the apple given* ?)
2. *Hat* dem Kind gestern die Frau den Apfel *gegeben*?
3. *Hat* gestern die Frau dem Kind den Apfel *gegeben*?
4. *Gab* die Frau dem Kind gestern den Apfel ?
   (*Gave the woman the child yesterday the apple* ?)
5. *Gab* gestern die Frau dem Kind den Apfel?

18.5.2 ?+MAIN starting an interrogative main clause

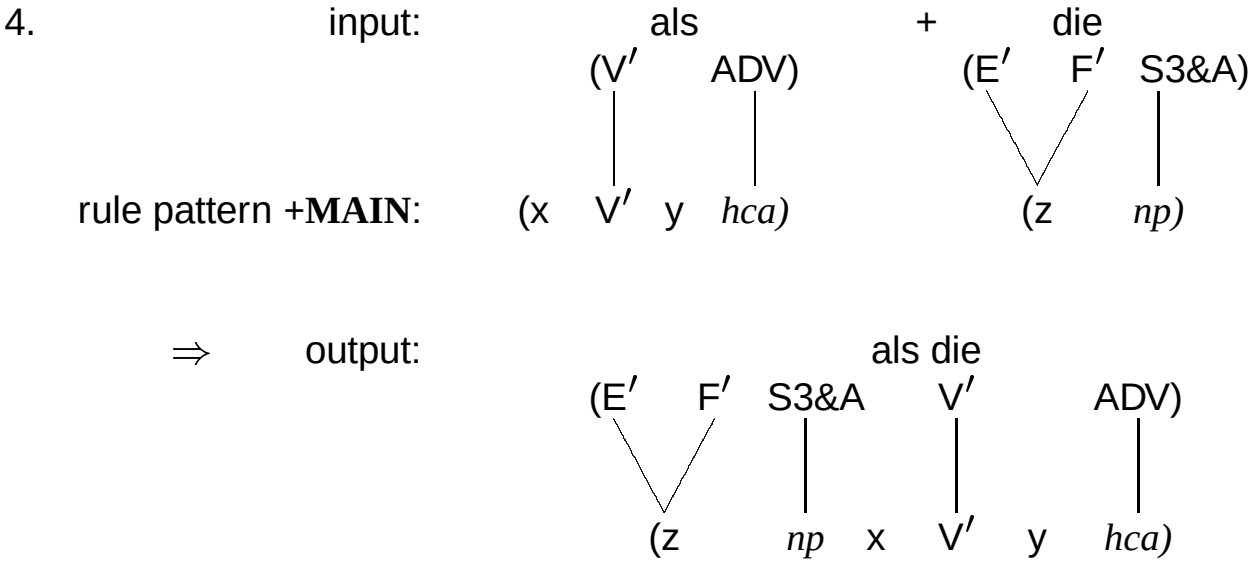
1. input-output:



18.5.3 Subordinate clauses with and without auxiliary

1. *Als* die Frau dem Kind gestern den Apfel *gegeben hat*
   (*When the woman the child yesterday the apple given has*)
2. *Als* dem Kind gestern die Frau den Apfel *gegeben hat*
3. *Als* gestern die Frau dem Kind den Apfel *gegeben hat*
4. *Als* die Frau dem Kind gestern den Apfel *gab*
   (*When the woman the child yesterday the apple gave*)
5. *Als* gestern die Frau dem Kind den Apfel *gab*

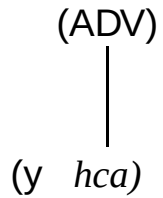
18.5.4 +MAIN starting an adverbial subclause



18.5.5 +FV concluding subclause with finite main verb

2. input: als die Frau dem Kind den Apfel + gab

⇒ output: als die Frau dem Kind den Apfel gab



18.5.6 Beginning of an adverbial subclause in postverbal position

Julia las,	+	als		Julia las, als	+	Maria	Julia las, als Maria
(A′ V)		(V′ ADV)	⇒	(V′ A′ V)		(S3&D&A)	⇒ (S3&D&A V′ A′ V)

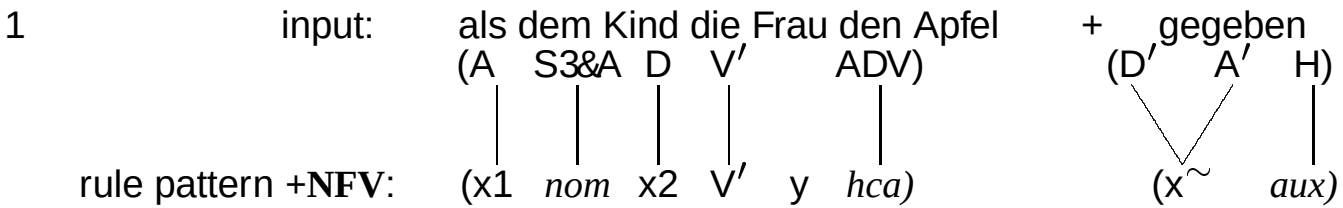
18.5.7 Completion of an adverbial subclause in postverbal position

Julia las, als Maria	+	schlief		Julia las, als Maria schlief,
(S3&D&A V′ A′ V)		(S3′ V)	⇒	(A′ V)

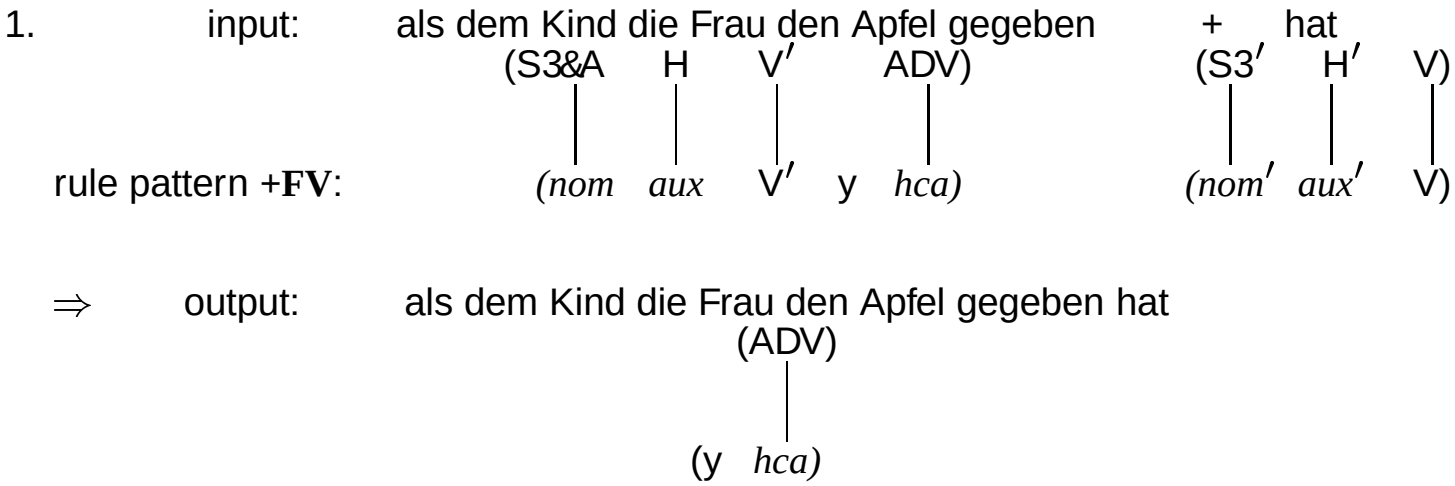
18.5.8 Nesting of adverbial subclauses in preverbal position

Als Maria, obwohl Julia die Zeitung	+	las	Als Maria, obwohl Julia die Zeitung las,	
(A S3&D&A V′ S3&D&A V′ ADV)		(S3′ A′ V)	⇒	(S3&D&A V′ ADV)

18.5.9 +NFV adds nonfinite main verb to subclause



18.5.10 +FV concludes subclause with finite auxiliary



18.5.11 LAG handling interrogative and adverbial clauses (*LA-D4*)

LX = LX of *LA-D3* plus subordinating conjunctions
[als (V′ ADV) *], [nachdem (V′ ADV) *], [obwohl (V′ ADV) *]
variable definition = variable definition of *LA-D3* plus *hca* ϵ {V, VI, ADV}
$ST_S =_{def}$ { [(x) {1 +ADJ, 2 +N, 3 +FV, 4 +MAIN, 5 ?+MAIN}] }

+N: (*adj′ n′* x) (*n*) ⇒ (x) {6 +FV, 7 +MAIN, 8 +NFV, 9 +IP}

+ADJ: (*adj′* x) (*adj*) ⇒ (*adj′* x) {10 +ADJ, 11 +N}

?+MAIN: (*nom′ aux′* V)(z *nom*) ⇒ (z *aux′* VI)
(*nom′ aux′* V)(y *obq*) ⇒ (y *obq nom′ aux′* VI)
(x *np′* y V)(z *np*) ⇒ (z x y VI)
(x V)(y ADV) ⇒ (y x VI) {12 +ADJ, 13 +N, 14 +MAIN, 15 +NFV, 16 +IP}

+FV: (*nom aux* V′ y *hca*) (*nom′ aux′* V) ⇒ (y *hca*)
(x V′ y *hca*)(x~ V)
[x = x~] ⇒ (y *hca*)
(*nom*)(*nom′ aux′* V) ⇒ (*aux′* V)
(*obq*)(x *aux′* V) ⇒ (*obq* x *aux′* V)
(x *aux*)(*np′ aux′* V) ⇒ (x *np′* V)
(*np*)(x *np′* y V) ⇒ (x y V)
(ADV)(x V) ⇒ (x V) {17 +MAIN, 18 +NFV, 19 +FV, 20 +IP}

+MAIN: (x *nom′* y *aux′* V)(z *nom*) ⇒ (z x y *aux′* V)

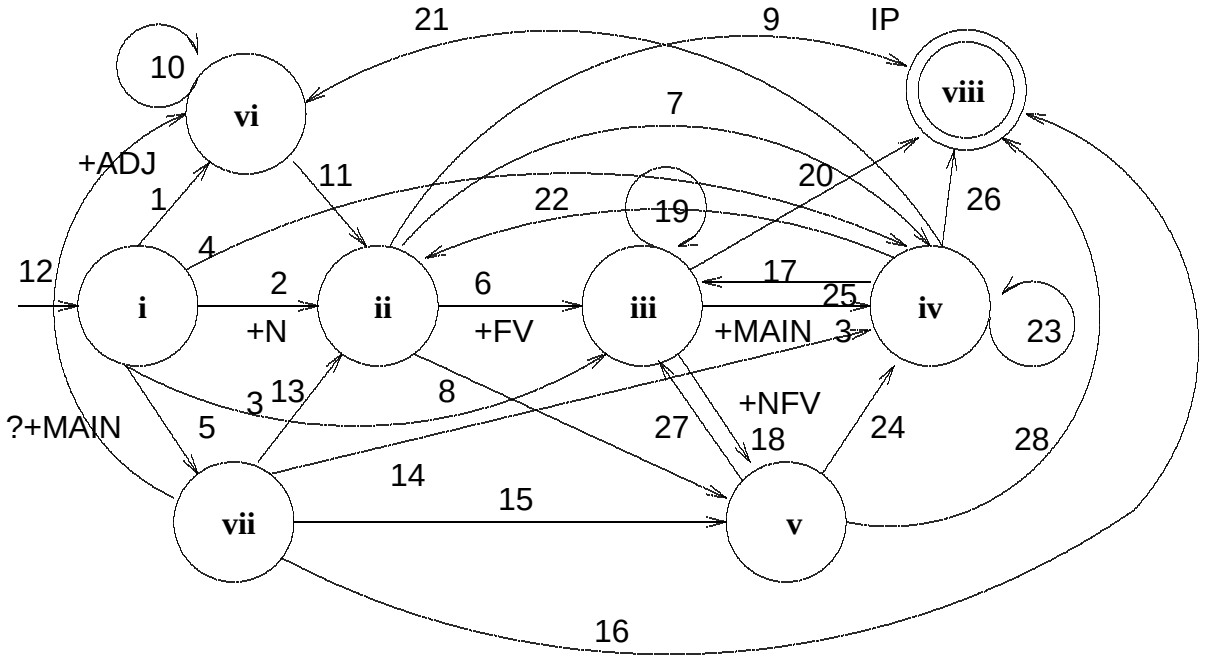
(x *aux*′ V)(y *obq*)⇒ (y *obq* x *aux*′ V)
(x *np*′ y V)(z *np*) ⇒ (z x y V)
(x V′ y *hca*)(z *np*) ⇒ (z *np* x V′ y *hca*)
(x V)(y ADV) ⇒ (y x V) {21 +ADJ, 22 +N, 23 +MAIN, 24 +NFV, 25 +FV, 26 +IP}

+NFV: (x1 *nom* x2 V′ y *hca*)(x$^{\sim}$ *aux*)
[(x1 ∘ x2) = x$^{\sim}$] ⇒ (*nom aux* V′ y *hca*)
(x *aux*′ V)(x$^{\sim}$ *aux*) ⇒(V) {27 +FV, 28 +IP}

+IP: (*vt*) (*vt*′ *sm*) ⇒ (*sm*) {}

$ST_F =_{def}$ { [(V) rp_{+ipt}], [(VI) rp_{+ipt}]}

18.5.12 The finite state backbone of *LA-D4*



ii	2, 11, 13,19, 22,	+N
iii	3, 6, 8,17, 19, 27	+FV
iv	4, 14, 7, 23, 24, 25	+MAIN
v.	8, 15, 18,	+NFV
vi.	1, 10, 12, 21,	+ADJ
vii:	5,	+MAIN
viii:	9, 16, 20, 26,	+IP

18.5.13 Verification of grammars

1. *Syntactic verification*

   The formal grammars for English and German developed so far should be implemented as parsers and tested automatically on increasing sets of positive and negative test sentences.

2. *Morphological and lexical verification*

   The word form recognition of these grammars should be changed from the preliminary full form lexica LX to suitable applications of LA-Morph and be tested on corpus-based word lists in order to provide extensions with sufficient data coverage of the lexicon and the morphology.

3. *Functional verification in communication*

   The formal grammars and parsers for natural languages should be supplemented with an automatic semantic and pragmatic interpretation that is (i) in line with the basis assumptions of the SLIM theory of language and (ii) demonstrated to be functional in automatic applications.