NEGRA

SFB 378

Ressourcenadaptive kognitive Prozesse

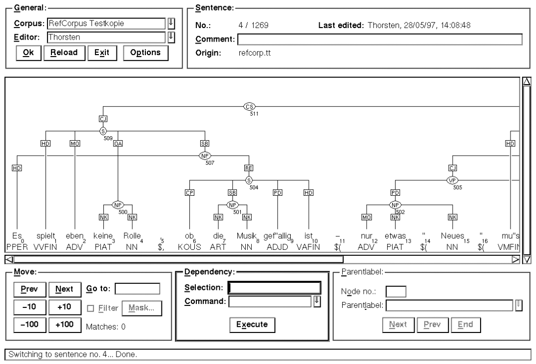
Projekt C3

Nebenläufige grammatische Verarbeitung

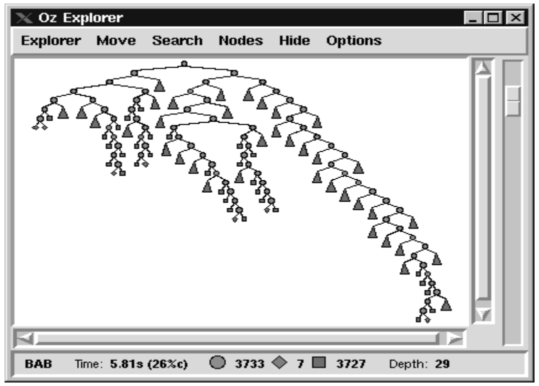
Computerlinguistik	Informatik
Hans Uszkoreit	Gert Smolka
Thorsten Brants	Denys Duchier
Brigitte Krenn	Martin Müller
Stephan Oepen	Joachim Niehren
Wojciech Skut	

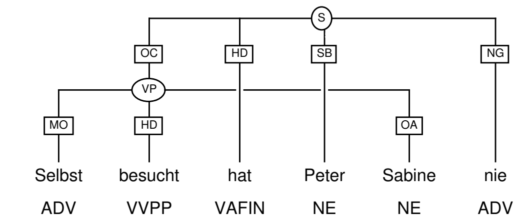
NEGRA

- Development of an Annotation Scheme for Free Word Order Languages

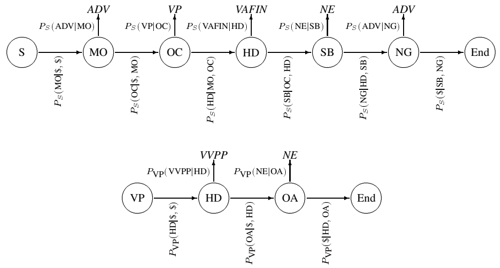


- Statistical and FS Parsing





- Annotation of a German Newspaper Corpus



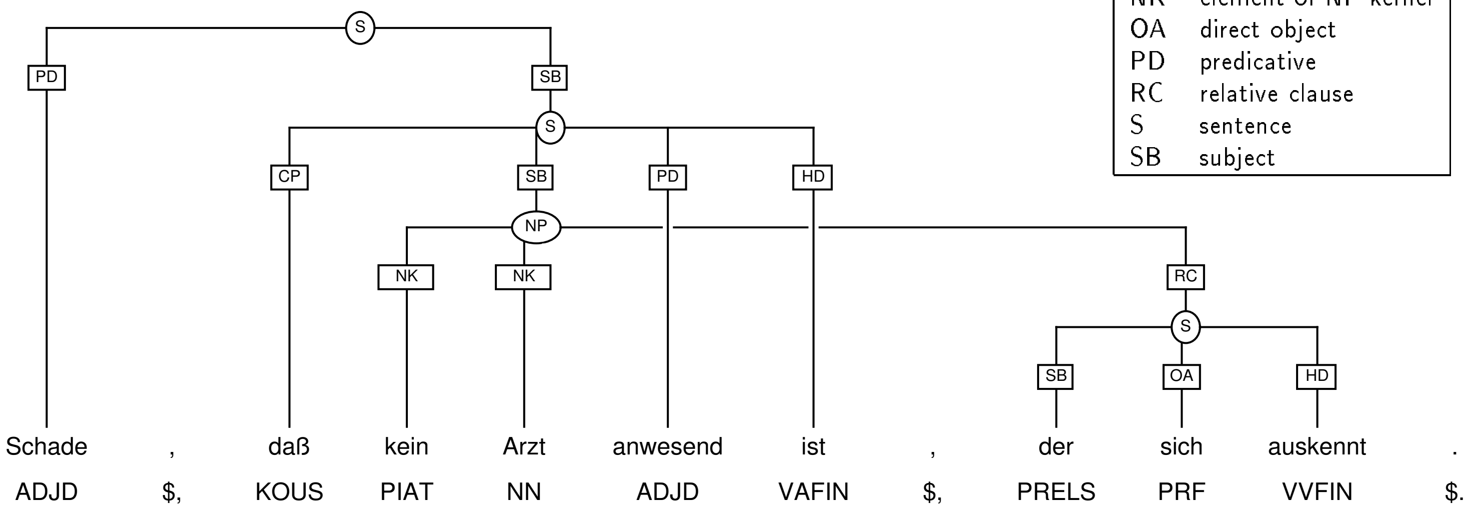
- Parsing in a Concurrent Constraint-Based Framework

- Performance Model of Grammar

Annotation Scheme

- Annotation of predicate-argument-structures
- Four representation levels:
  - argument structure
  - grammatical functions
  - syntactic categories
  - word order

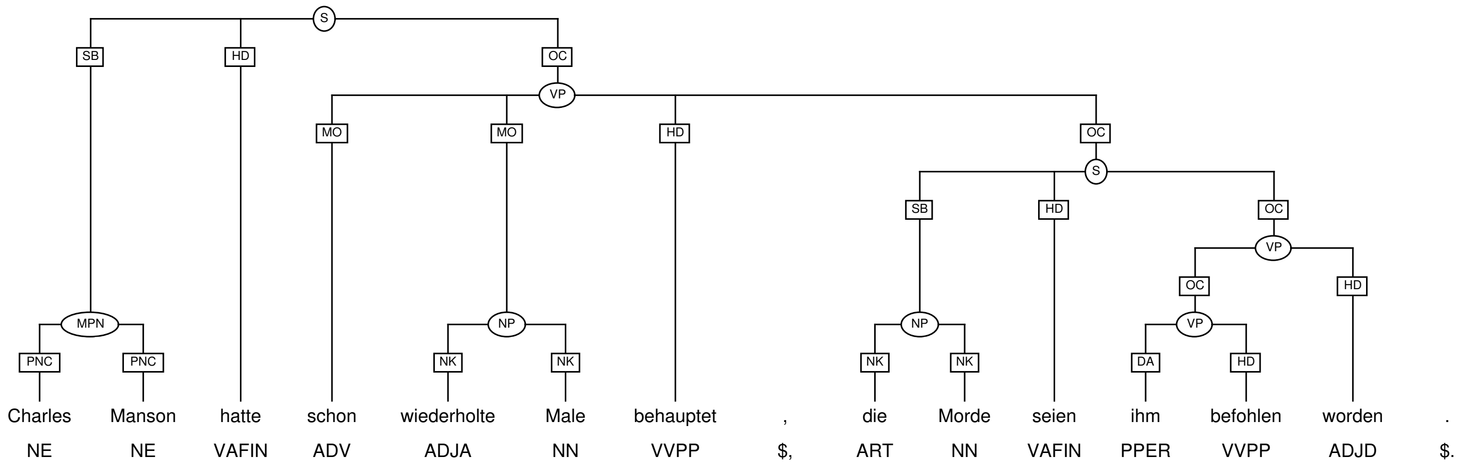
CP	complementizer
HD	head
NK	element of NP kernel
OA	direct object
PD	predicative
RC	relative clause
S	sentence
SB	subject



Annotation Scheme

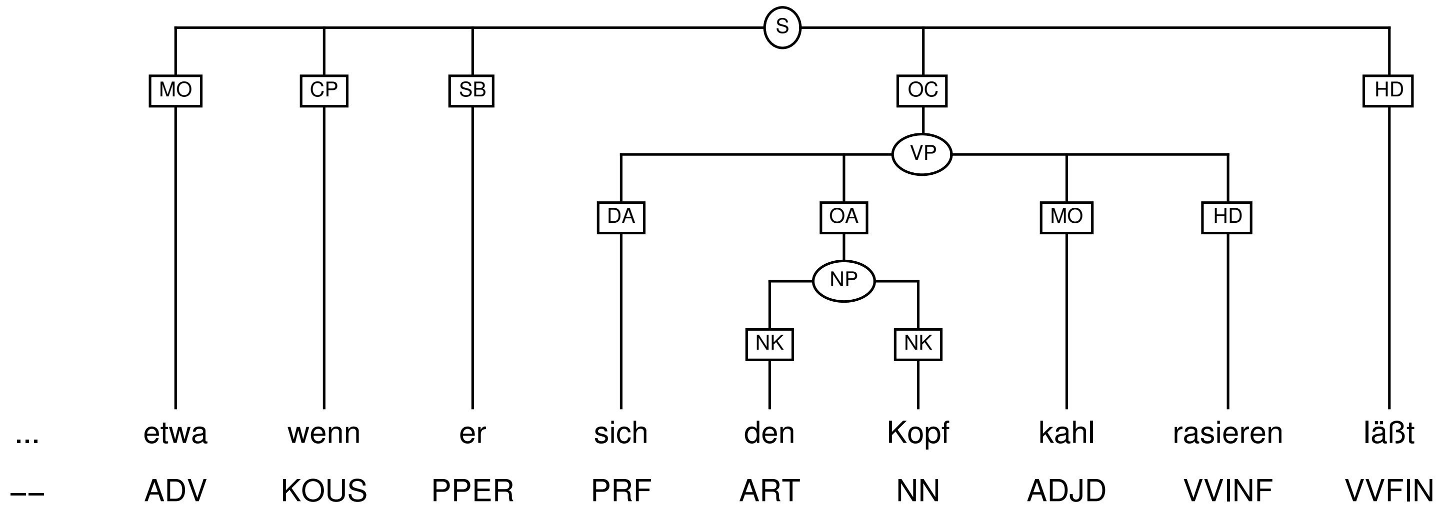
Selected phenomena:

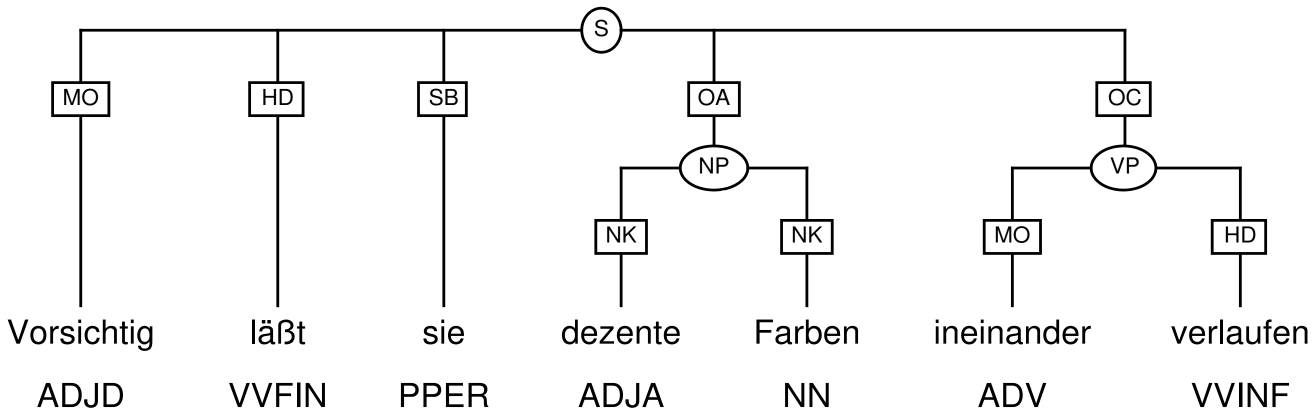
- Noun phrases:
  flat annotation: NP kernel, initial modifiers, pre- and postnominal genitives, appositions, relativ clauses
- Sentences and verb phrases:
  sentences contain finite verbs, VP's contain non-finite verbs as their heads
- Attachment ambiguities:
  Contextually appropriate attachment points are used
  Choose highest attachment point in case of unresolved ambiguities



[DA=dative; HD=head; MO=modifier; NK=NP kernel; OA=direct object; OC=clausal object; PNC=part of proper noun; SB=subject]

Complements of 'Lassen'

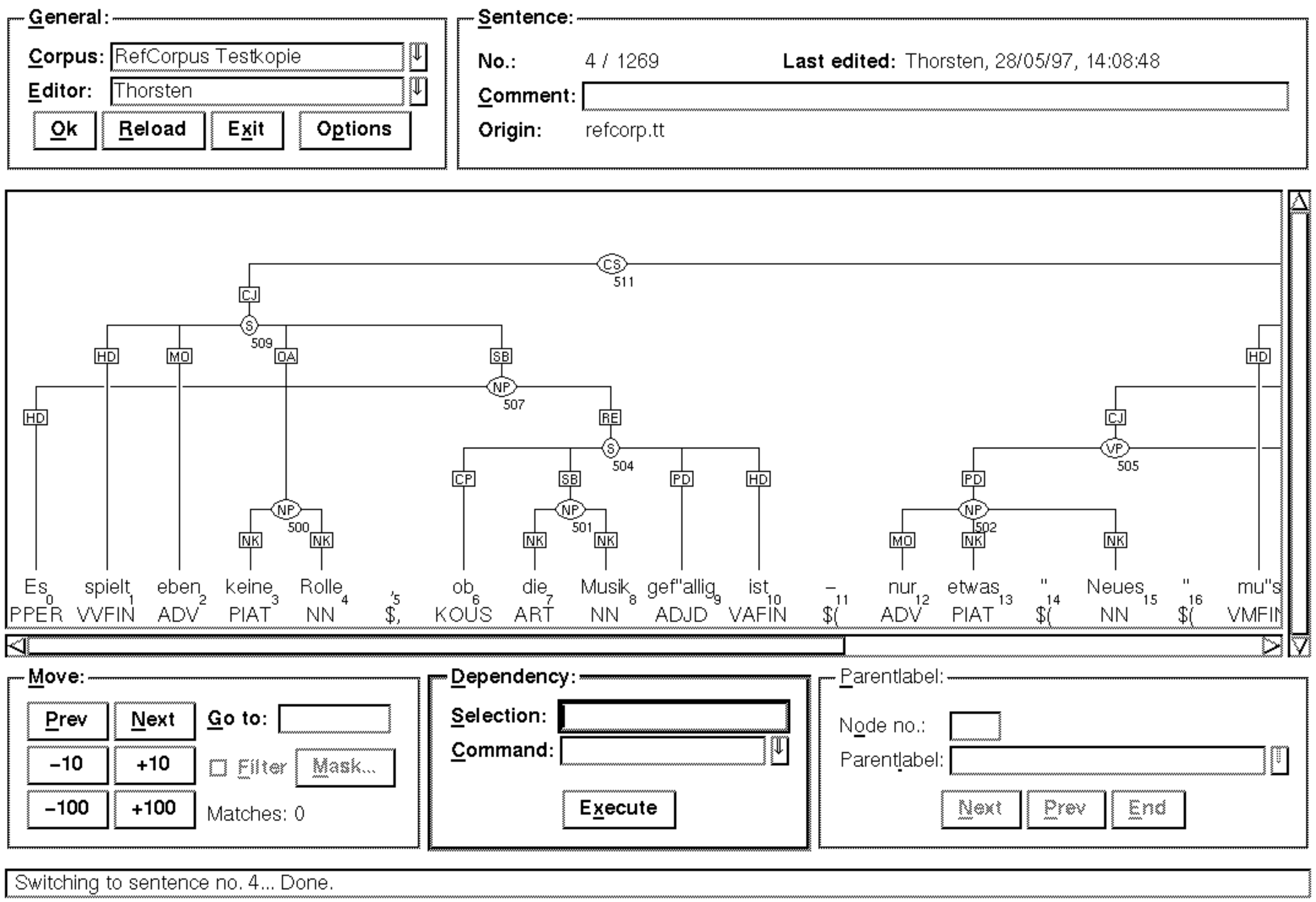




[CP=complementizer; DA=dative; HD=head; MO=modifier; NK=NP kernel; OA=direct object; OC=clausal object; SB=subject]

Annotation Tool

- Immediate graphical visualization
- Plausibility checks
- Interaction of human annotator and parser



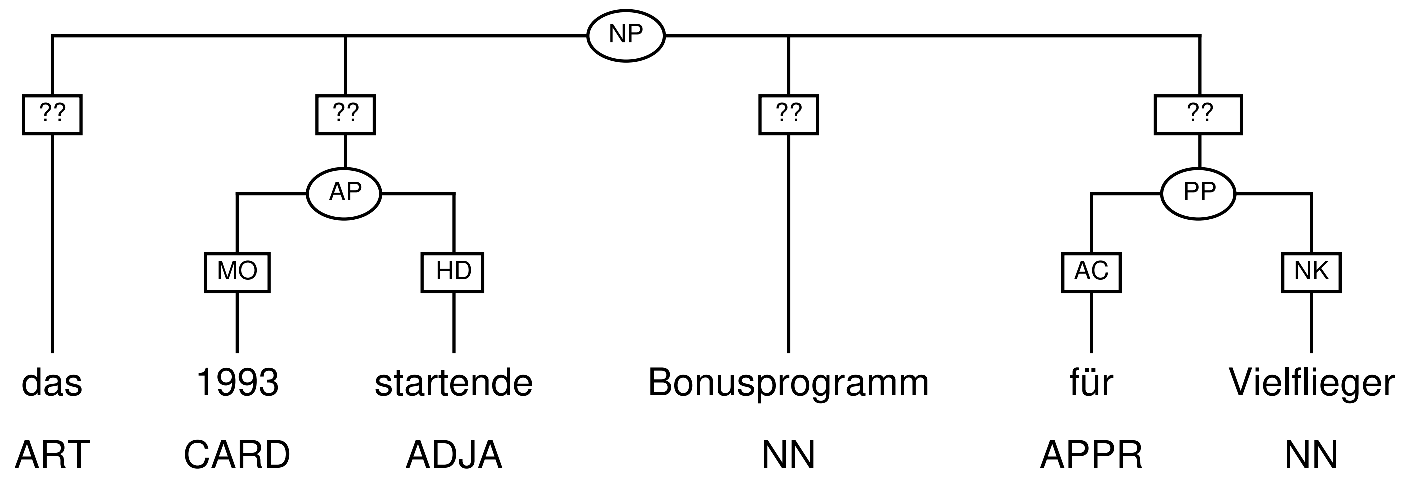
Automation

Levels of Automation:

0) Pure manual annotation
1) User: constituents and phrase categories
   Tool: grammatical functions
2) User: constituents
   Tool: phrase categories and grammatical functions
3) User: constituents without explicit marking of sub-phrases
   Tool: structures of fixed (small) depths, phrase categories, and gramm. functions
4) User: complex phrases
   Tool: simple phrases, phrase categories, and grammatical functions
5) User: revision
   Tool: complete structures

Level 1

User: constituents and phrase categories
Tool: grammatical functions

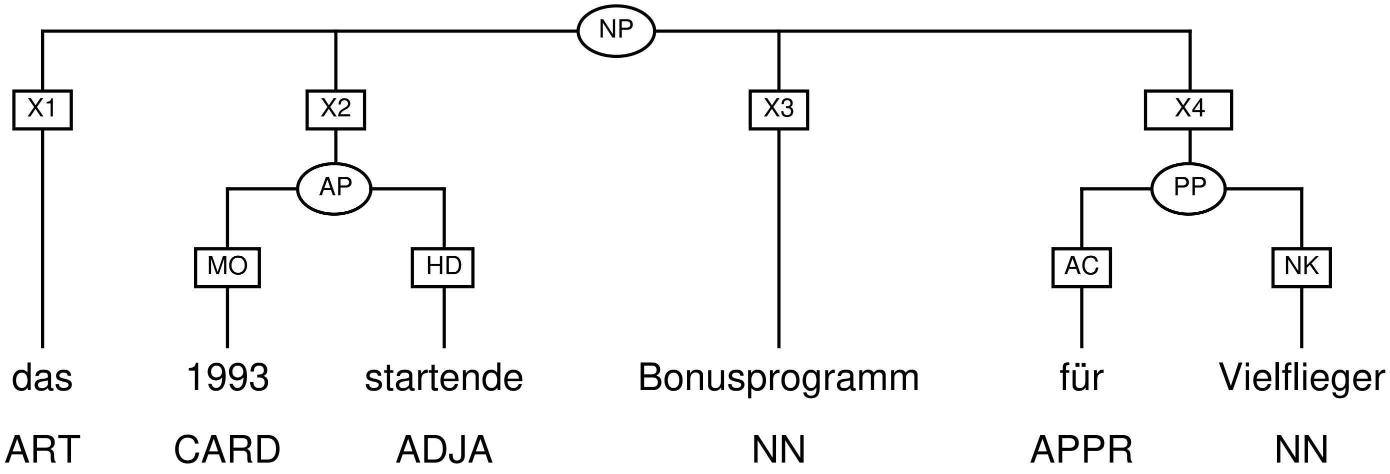


AP = adjective phrase, NP = noun phrase, PP = prepositional phrase

AC = adpositional case marker, HD = head, MO = modifier, NK = noun kernel element

ADJA = attribute adjective, APPR = proposition, ART = article, CARD = cardinal, NN = common noun,

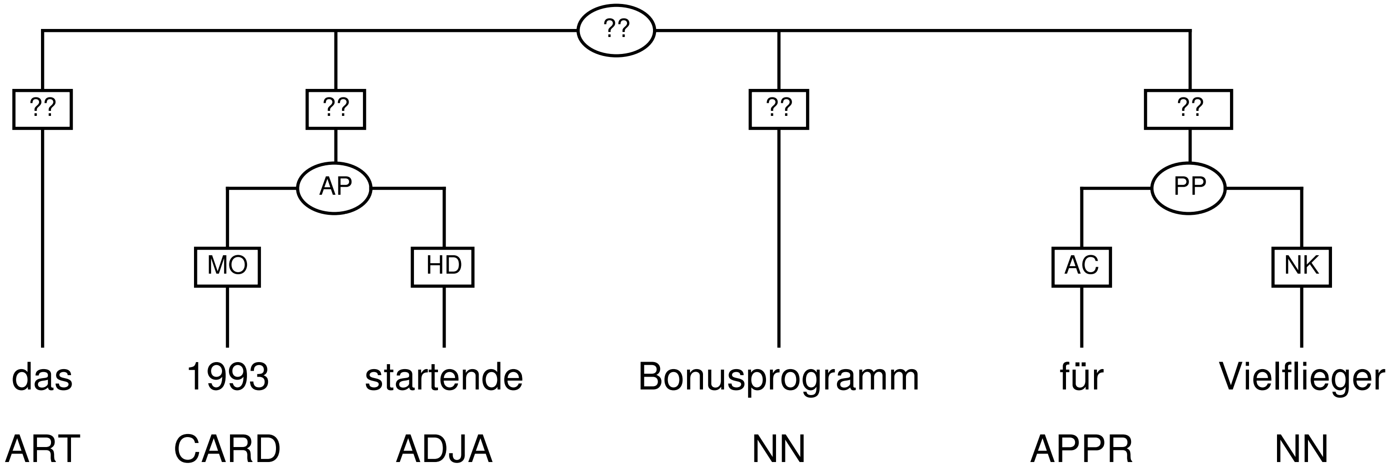
Assignment of Grammatical Functions



$$
\begin{aligned}
\mathrm{argmax}_{X_1,X_2,X_3,X_4} \;& P_{NP}(X_1|\$) \cdot P_{NP}(\mathsf{ART}|X1) \\
& \cdot P_{NP}(X_2|\$, X_1) \cdot P_{NP}(\mathsf{AP}|X_2) \\
& \cdot P_{NP}(X_3|X_1, X_2) \cdot P_{NP}(\mathsf{NN}|X_3) \\
& \cdot P_{NP}(X_4|X_2, X_3) \cdot P_{NP}(\mathsf{PP}|X_3) \\
& \cdot P_{NP}(\$|X_3, X_4)
\end{aligned}
$$

Level 2

User: constituents
Tool: phrase categories and grammatical functions



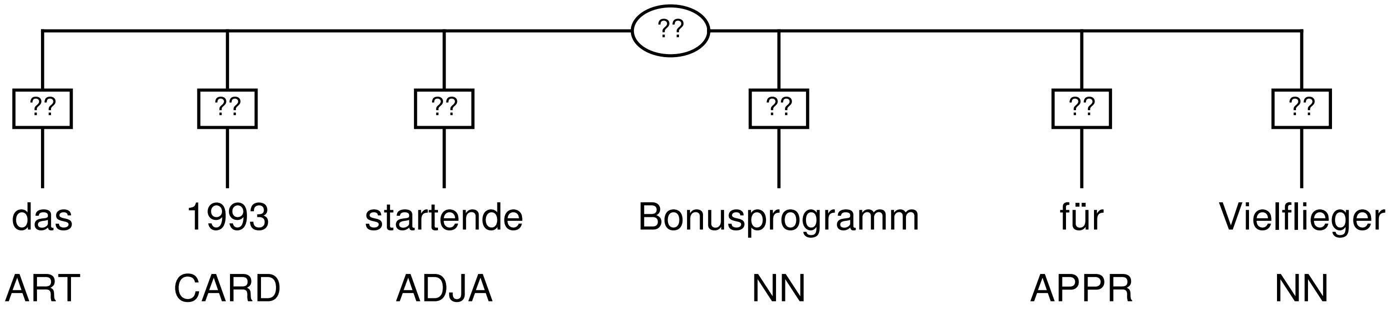
AP = adjective phrase, NP = noun phrase, PP = prepositional phrase

AC = adpositional case marker, HD = head, MO = modifier, NK = noun kernel element

ADJA = attribute adjective, APPR = proposition, ART = article, CARD = cardinal, NN = common noun,

Level 3

User: constituents without explicit marking of sub-phrases
Tool: structures of fixed (small) depths, phrase categories, and gramm. functions



AP = adjective phrase, NP = noun phrase, PP = prepositional phrase

AC = adpositional case marker, HD = head, MO = modifier, NK = noun kernel element

ADJA = attribute adjective, APPR = proposition, ART = article, CARD = cardinal, NN = common noun,

Cascaded Finte State Parsing

- Grammar divided in strata/levels (chunks & clauses)
- Pipeline of finite-state recognizers/transducers