

Syntax natürlicher Sprachen

3: X-Bar-Struktur, Rekursion und Ambiguität

A. Wisiorek

Centrum für Informations- und Sprachverarbeitung,
Ludwig-Maximilians-Universität München

29.10.2024

1. Rekursion

1 Rekursion

2 X-Bar-Theorie

- Übergenerierung durch rekursives Phrasenstrukturschema
- X-Bar-Phrasenstrukturschema
- NP-Regeln im X-Bar-Schema
- VP-Regeln im X-Bar-Schema
- Adäquatheit X-Bar als Syntaxmodell

3 Syntaktische Ambiguität

4 Treebanks

Rekursive syntaktische Kategorien

- Erklärung für **Nicht-Endlichkeit** natürlicher Sprachen
- endliche Anzahl an syntaktischen Regeln, aber unbegrenzte Anzahl an bildbaren Sätzen (z. B.: durch Hinzufügung von Adjunkten)
- modellierbar mit **rekursiven Phrasenstrukturregeln**
- Erzeugung **verschachtelter Strukturen**

Beispiele rekursiver Kategorien

- nominale Adjunkte (ADJ- oder PP-Attribute, Genitiv-NP-Attribute)
- verbale Adjunkte (PP-Adverbiale)
- Komplementsätze (z.B. *ich glaube, dass ich glaube, dass ...*)
- Relativsätze, z.B.:
Die Katze, (die der Hund, (der dem Mann entlaufen war), biss), lief weg.

Rekursive Grammatik-Regel

- das Nichtterminalsymbol der linken Regelseite (LHS) ist ableitbar auf das selbe Nichtterminalsymbol
- d.h. (bei einer direkt rekursiven Regel): **das Symbol auf der linken Regelseite (LHS) erscheint auch auf der rechten Regelseite (RHS)**

rekursive Regel

$$X \rightarrow X Y$$

Rekursionstypen

- je nach Position der rekursiven Kategorie auf der rechten Regelseite lassen sich 3 Typen unterscheiden

linksrekursive Regel

$X \rightarrow X Y$ oder $NP \rightarrow NP PP$

rechtsrekursive Regel

$X \rightarrow Y X$ oder $NP \rightarrow ADJ NP$

center-embedding-Rekursion (z.B. Relativsatz)

$X \rightarrow Y X Z$ oder $NP \rightarrow PROPN RELPRO NP V$

Verarbeitung linksrekursiver Regeln

- linksrekursive Regel nicht mit einfachem top-down-Parser verarbeitbar (Recursive-Descent-Parser)
- führt zu Endlosschleife

Verarbeitung center-embedding-Regeln

- center-embedding benötigt mindestens kontextfreie Grammatik
- nicht mit regulärer Grammatik modellierbar

Direkte vs. Indirekte Rekursion

- im Gegensatz zur direkten Rekursion ist bei indirekter die Kategorie der LHS über mehr als eine Regelanwendung auf sich selbst ableitbar

Direkte Rekursion

$X \rightarrow X$ oder $NP \rightarrow DET \ N \ NP$

Indirekte Rekursion

$X \rightarrow Y$ oder $NP \rightarrow DET \ N \ PP$
 $Y \rightarrow X$ $PP \rightarrow P \ NP$

Beispiel direkte Rekursion

- **NP** → DET N **NP**

der Schlüssel des Autos der Firma des Chefs ...

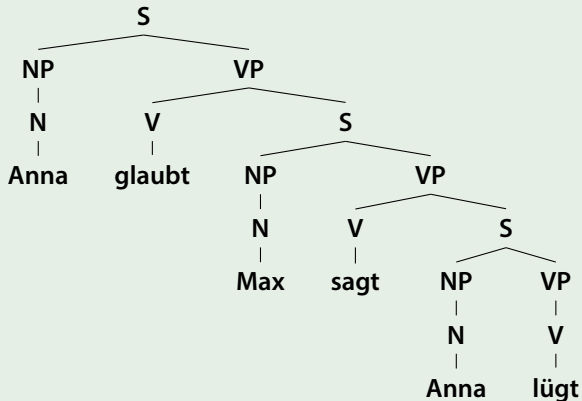
Beispiel indirekte Rekursion

- **NP** → DET N **PP**
- **PP** → P **NP**

in der Frage nach den Problemen in seinen Ausführungen über die Freiheit auf dem Gebiete der Kunst während der 30er Jahre (s. Dürscheid, Kap. 5, nach J. E. Schmidt, 1993:84)

indirekte S-Rekursion

- $S \rightarrow NP VP$
- $VP \rightarrow V S$



2. X-Bar-Theorie

1 Rekursion

2 X-Bar-Theorie

- Übergenerierung durch rekursives Phrasenstrukturschema
- X-Bar-Phrasenstrukturschema
- NP-Regeln im X-Bar-Schema
- VP-Regeln im X-Bar-Schema
- Adäquatheit X-Bar als Syntaxmodell

3 Syntaktische Ambiguität

4 Treebanks

2.1. Übergenerierung durch rekursives Phrasenstrukturschema

- 1 Rekursion
- 2 X-Bar-Theorie
 - **Übergenerierung durch rekursives Phrasenstrukturschema**
 - X-Bar-Phrasenstrukturschema
 - NP-Regeln im X-Bar-Schema
 - VP-Regeln im X-Bar-Schema
 - Adäquatheit X-Bar als Syntaxmodell
- 3 Syntaktische Ambiguität
- 4 Treebanks

vorläufige NP-Strukturregeln (rekursiv)

Phrasenschema NP (vereinfacht)

DET (ADJP)* N (PP)*

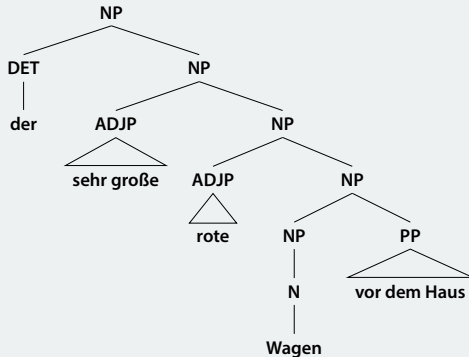
NP-Produktionsregeln mit Rekursion (vorläufig)

$NP \rightarrow DET\ NP$ (?)

$NP \rightarrow ADJP\ NP$

$NP \rightarrow NP\ PP$

$NP \rightarrow N$



NLTK: Übergenerierung vorläufiger NP-Grammatik

```
1 grammar = nltk.CFG.fromstring("""
2 NP -> DET NP
3 NP -> ADJP NP
4 NP -> N
5 DET -> 'das'
6 ADJP -> 'rote'
7 N -> 'Auto'
8 """)
9
10 for sentence in generate(grammar, depth=9):
11     print(' '.join(sentence))
12
13 #...
14 #das das rote das rote Auto
15 #...
```

Problem mit direkt rekursiver DET-Regel

rekursive DET-Regel: $NP \rightarrow DET\ NP$

Mit der Regel können zwar NPs gemäß des Phrasenschemas abgeleitet werden

Problem: Übergenerierung

- die Grammatik ist aber stark **übergenerierend** und damit kein adäquates Modell des NP-internen Strukturaufbaus
- **Übergenerierung 1**: DET rekursiv wiederholbar an erster Position (richtige Strukturposition aber falsche Anzahl)
- **Übergenerierung 2**: DET wiederholbar an falscher Strukturposition (z. B.: ADJP DET N)

Lösung: X-Bar-Struktur

- **Strukturbegrenzung notwendig!**
- möglich durch **X-Bar-Theorie** (Chomsky)

2.2. X-Bar-Phrasenstrukturschema

- 1 Rekursion
- 2 X-Bar-Theorie
 - Übergenerierung durch rekursives Phrasenstrukturschema
 - **X-Bar-Phrasenstrukturschema**
 - NP-Regeln im X-Bar-Schema
 - VP-Regeln im X-Bar-Schema
 - Adäquatheit X-Bar als Syntaxmodell
- 3 Syntaktische Ambiguität
- 4 Treebanks

- Entwicklung durch Chomsky im Rahmen der *Government & Binding*-Theorie
- ursprüngliches Symbol: Balken (*bar*): \bar{X} , meist als X' geschrieben

Beschränkung der Struktur von Phrasen

- 1 **binäre Verzweigung** ($A \rightarrow B C$)
- 2 Einführung **phrasaler Zwischenebene** X' (im NLTK NOM bzw. VERBAL)
- 3 **gleicher Strukturaufbau für alle Phrasenarten** (X als Wortart-Variable)

bisher: Anzahl und Art unmittelbarer Konstituenten einer Phrase nicht beschränkt

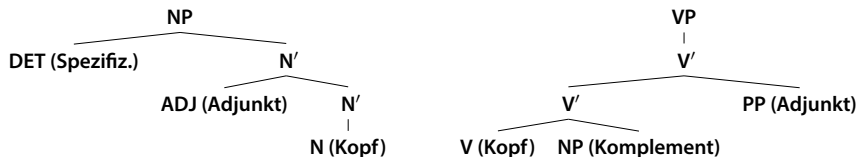
- Mischung aus nicht festgelegter Anzahl aus lexikalischen und phrasalen Kategorien
- keine festgelegte Ordnung zwischen Kopf und Erweiterungen

Problem: Übergenerierung

führt zu Problemen bei rekursiver Strukturdefinition

Lösung: phrasale Zwischenebene: X'

- Einführung **phrasaler Zwischenebene** (X') zwischen **Gesamtphrase** (XP, **maximale Projektion**) und **Kopf** (X)
- Erlaubt die Differenzierung verschiedener Arten von Dependents in Phrase durch Strukturposition

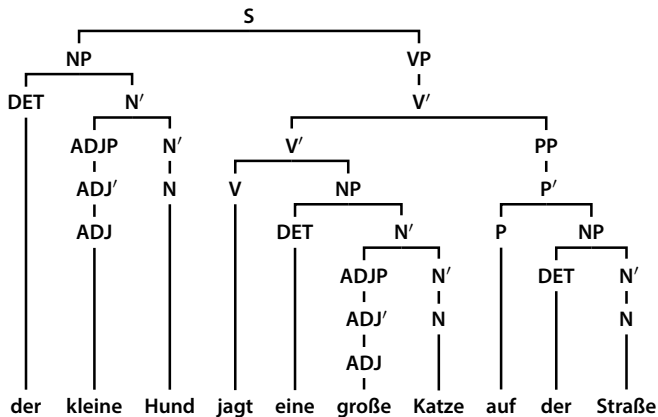


Spezifizierer

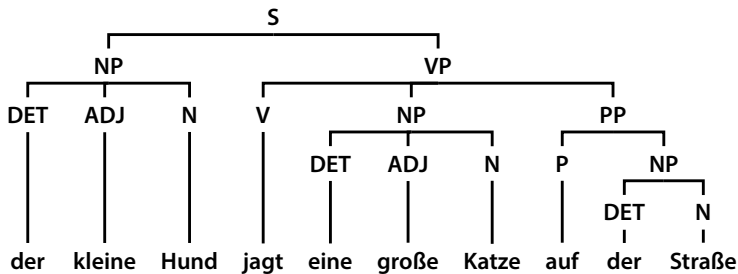
phrasenabschließende nicht-obligatorische Erweiterung, Anzahl begrenzt (nur ein Element)

- Verbindung mit X' , bilden zusammen den Abschluß der XP-Phrase
- Deutsche NP: Determinierer, Possessivpronomen, Quantifizierer, sächsischer Genitiv (*seines Glückes Schmied*)
- verallgemeinert: als Strukturposition ($XP \rightarrow SPEC X'$) im Schema für Elemente mit entsprechenden Eigenschaften
- z. B. AUX als VP-Spezifizierer (s. unten)

Beispiel: vollständige X-Bar-Analyse



zum Vergleich: flache Phrasenstrukturanalyse



Verallgemeinerung verbaler Komplement-Adjunkt-Distinktion

Komplement

obligatorische (valenzgeforderte) **Erweiterung**, Anzahl abhängig von Valenz/Subkategorisierungsrahmen

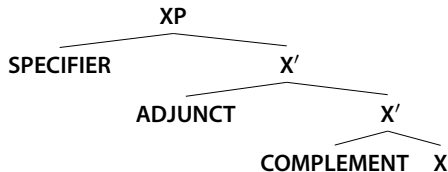
- *direkte Verbindung mit Phrasenkopf X, bilden zusammen X'-Phrase*
- *Modellierung der engen Verbindung Komplement mit Kopf*
- *Deutsche NP: Genitiv-NP-Erweiterung oder von-PP*

Adjunkt

nicht-obligatorische Erweiterung, Anzahl nicht begrenzt

- *Verbindung mit X'-Konstituente, bilden zusammen wieder X'*
- *Deutsche NP: Adjektiv-Attribut, PP-Attribut*

Allgemeines X-Bar-Schema (wortstellungsunabhängig)



Spezifizierer-Regel (Phrasenabschluss)

$XP \rightarrow SPEC, X'$

rekursive Adjunkt-Regel (rechtsrekursiv)

$X' \rightarrow ADJUNCT, X'$

Komplement-Regel

$X' \rightarrow COMPL, X$

Links- vs. Rechtsverzweigung

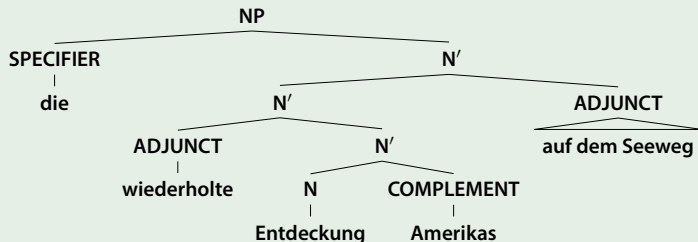
- X-Bar-Schema: **ohne implizierte lineare Struktur** (Wortstellung)
- Einzelsprachliche Regeln!
- **Deutsche NP: links- und rechtsverzweigend**

left- vs. rightbranching

- Linksverzweigung: **head-final** (OV-Sprachen)
- Rechtsverzweigung: **head-initial** (VO-Sprachen)

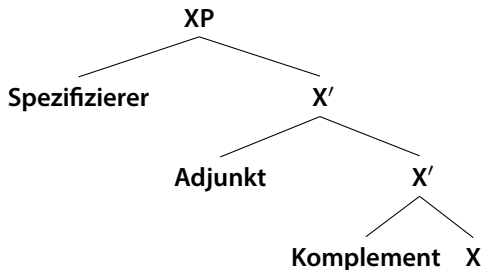
Beispiel für Links- und Rechtsverzweigung im Deutschen

die wiederholte Entdeckung Amerikas auf dem Seeweg



Strukturpositionen im X-Bar-PSG-Schema

	X (Kopf)	X'	XP
Komplement	Schwester	Tochter	
Adjunkt		Schwester und Tochter	
Spezifizierer		Schwester	Tochter



2.3. NP-Regeln im X-Bar-Schema

- 1 Rekursion
- 2 X-Bar-Theorie
 - Übergenerierung durch rekursives Phrasenstrukturschema
 - X-Bar-Phrasenstrukturschema
 - **NP-Regeln im X-Bar-Schema**
 - VP-Regeln im X-Bar-Schema
 - Adäquatheit X-Bar als Syntaxmodell
- 3 Syntaktische Ambiguität
- 4 Treebanks

X-Bar-NP-Regeln

Spezifizierer-Regel (Phrasenabschluss)

$NP \rightarrow DET N'$

Adjektiv-Adjunkt-Regel (rechtsrekursiv)

$N' \rightarrow ADJP N'$

PP-Adjunkt-Regel (linksrekursiv)

$N' \rightarrow N' PP$

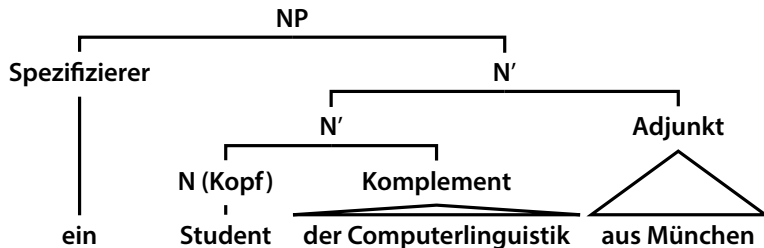
Genitiv-NP-Komplement-Regel (Ko-Konstituente/Schwester Kopf)

$N' \rightarrow N NP$

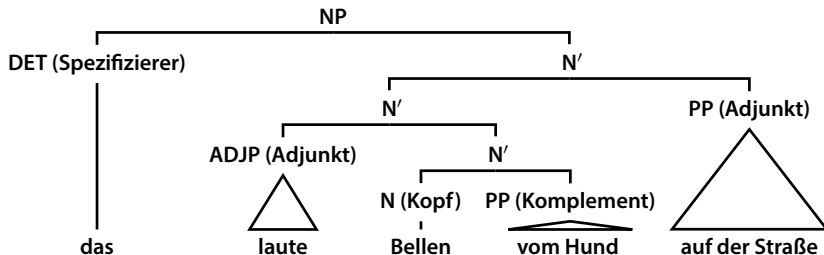
Kopf-Regel

$N' \rightarrow N$

Beispiel 1: X-Bar-Analyse NP



Beispiel 2: X-Bar-Analyse NP



2.4. VP-Regeln im X-Bar-Schema

- 1 Rekursion
- 2 X-Bar-Theorie
 - Übergenerierung durch rekursives Phrasenstrukturschema
 - X-Bar-Phrasenstrukturschema
 - NP-Regeln im X-Bar-Schema
 - **VP-Regeln im X-Bar-Schema**
 - Adäquatheit X-Bar als Syntaxmodell
- 3 Syntaktische Ambiguität
- 4 Treebanks

Phrasenabschluss (ohne Spezifizierer)

$$VP \rightarrow V'$$

PP-Adjunkt-Regel (ebenso: ADVP, ADJP)

$$V' \rightarrow V' PP$$

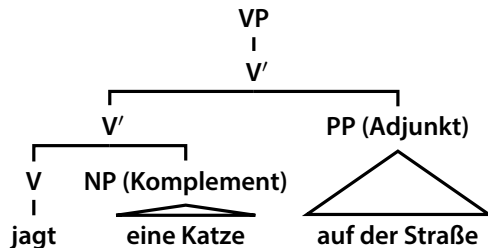
Komplement-Regel (di)transitives Verb

$$V' \rightarrow V NP (NP)$$

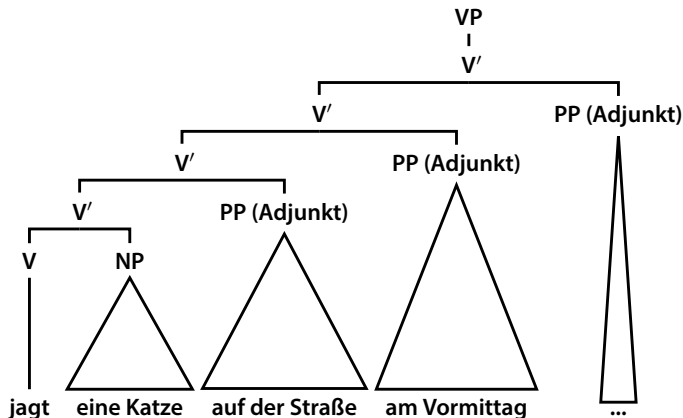
Komplement-Regel intransitives Verb (kein Komplement)

$$V' \rightarrow V$$

Beispiel 1: X-Bar-Analyse VP



Beispiel 2: X-Bar-Analyse VP (rekursive Adjunktion)



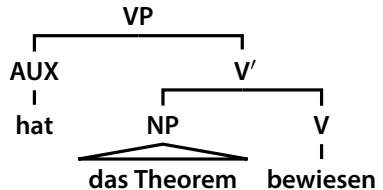
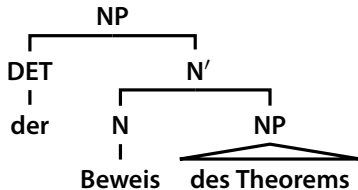
Auxiliare

- Hilfsverben
- begleiten Verb (Träger lexikalischer Bedeutung)
- Ausdruck von **grammatischen Merkmalen** des Verbs
 - Tempus
 - Modus
 - Diathese
 - Flexionsmerkmale

verschiedene Analysen

- Analyse ist stark **theorieabhängig** (welche Strukturposition)
- u. a. als Spezifizierer (s. o.)
- ebenso: Analyse als Verbgruppe (Verb + Auxiliare)
- ebenso: Eintrag in Subkategorisierungslexikon

Parallele X-Bar-Strukturanalyse für NP und VP: Auxiliar als Spezifizierer



2.5. Adäquatheit X-Bar als Syntaxmodell

- 1 Rekursion
- 2 X-Bar-Theorie
 - Übergenerierung durch rekursives Phrasenstrukturschema
 - X-Bar-Phrasenstrukturschema
 - NP-Regeln im X-Bar-Schema
 - VP-Regeln im X-Bar-Schema
 - **Adäquatheit X-Bar als Syntaxmodell**
- 3 Syntaktische Ambiguität
- 4 Treebanks

X-Bar-Phrasenstrukturgrammatiken

- schränken zwar durch ihre Strukturrestriktionen die Übergenerierung ein

Problem

- bestimmte Formen der Übergenerierung bleiben aber weiter bestehen

Gründe für Übergenerierung von CFGs (auch X-Bar)

Nichtberücksichtigung von Morphologie

- Rektion (Kasus):

**der Mann sieht des Kindes*

- Kongruenz (Agreement in Merkmalen):

**das Kinder*

Nichtberücksichtigung von Subkategorisierung

- Art und/oder Anzahl von Komplementen

**der Hund geht die Katze*

Lösung

Modellierung von grammat. Merkmalen durch Feature-based grammars (FCFGs)

3. Syntaktische Ambiguität

1 Rekursion

2 X-Bar-Theorie

- Übergenerierung durch rekursives Phrasenstrukturschema
- X-Bar-Phrasenstrukturschema
- NP-Regeln im X-Bar-Schema
- VP-Regeln im X-Bar-Schema
- Adäquatheit X-Bar als Syntaxmodell

3 Syntaktische Ambiguität

4 Treebanks

Problem

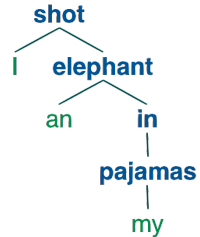
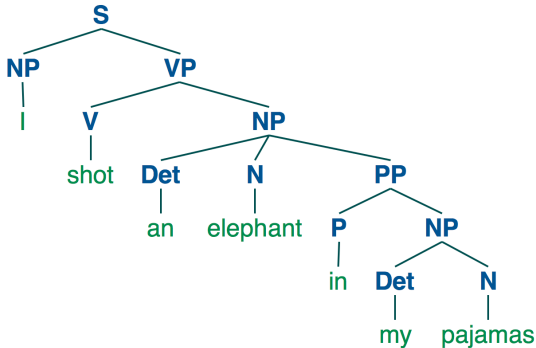
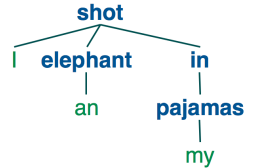
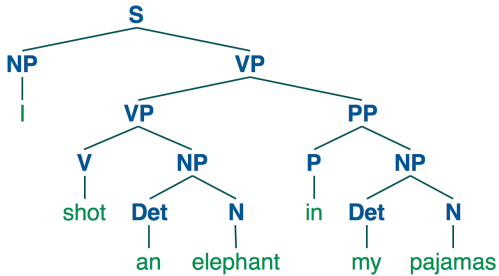
- mehr als eine Position im Syntaxbaum, d.h. mehr als eine Ableitung möglich (verschiedene syntaktische Interpretationen)
- im Gegensatz zur Übergenerierung ist hier die Modellierung mit CFG-Regeln grundsätzlich korrekt, aber eine **Disambiguierung** ist notwendig

Lösung

- in Parsingsystemen: Disambiguierung z.B. über probabil. lexikalisierte CFGs (**PCFGs**), d.h. Bestimmung der wahrsch. Ableitung im Kontext

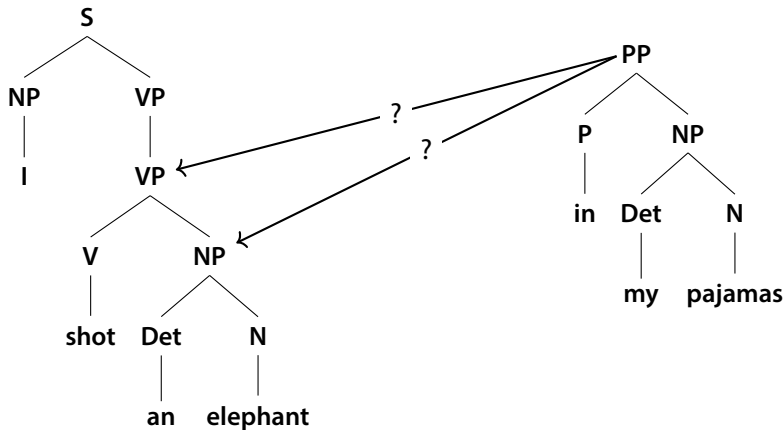
Beispielsatz für PP-Attachment

- *One morning I shot an elephant in my pajamas.
How he got into my pajamas I don't know.*
(Groucho Marx, Animal Crackers, 1930)



Attachment-Ambiguität

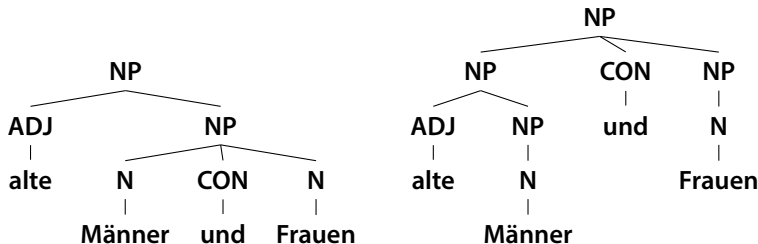
- Konstituente kann im Parsebaum an mehr als einer Stelle angebunden werden



Beispiel Koordinierungsambiguität

- *[alte [Männer und Frauen]]*
- *[alte Männer] und [Frauen]*

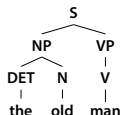
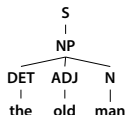
- Skopus des Adjektivs unklar:



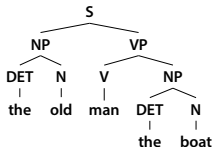
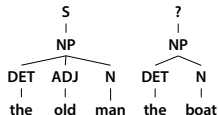
Beispiel Garden-Path-Satz

- *The old man the boat.*

- Teilfolge im Satz mit zwei Strukturanalysen:



- aber nur eine Analyse ermöglicht vollständige Analyse (Parse/Ableitung) für den gesamten Satz:



Parsingalgorithmus benötigt **Backtracking!**

4. Treebanks

- 1 Rekursion
- 2 X-Bar-Theorie
 - Übergenerierung durch rekursives Phrasenstrukturschema
 - X-Bar-Phrasenstrukturschema
 - NP-Regeln im X-Bar-Schema
 - VP-Regeln im X-Bar-Schema
 - Adäquatheit X-Bar als Syntaxmodell
- 3 Syntaktische Ambiguität
- 4 Treebanks

Treebank

Sammlung von per Hand annotierten Syntaxbäumen in bestimmtem Annotationsschema

grammar induction zur Erstellung von Syntaxmodellen

- Treebanks als **implizite Grammatik**
- CFG-Regeln können aus Treebank-Korpus gewonnen werden

- englisches Zeitungskorpus (bekannteste Treebank)
- relativ flache Struktur (z.B. NP → DT JJ NN)
- Schema verwendet im englischen Stanford-Parser-Modell

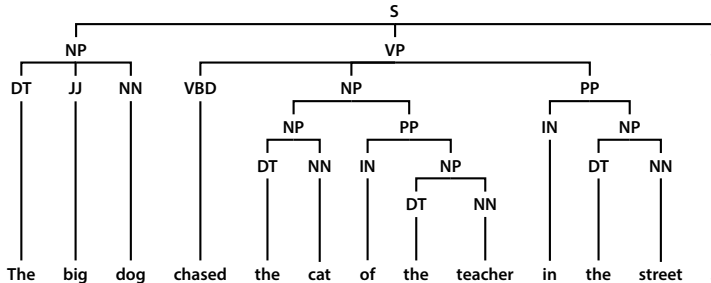


Abbildung: Beispiel-Parse Stanford-Parser en. (trainiert auf Penn-Treebank)

- wichtigste deutsche Treebank
- noch flacher als Penn-Treebank (z.B PP → APPR ART NN)
- Grundlage deutsches Stanford-Parser-Modell

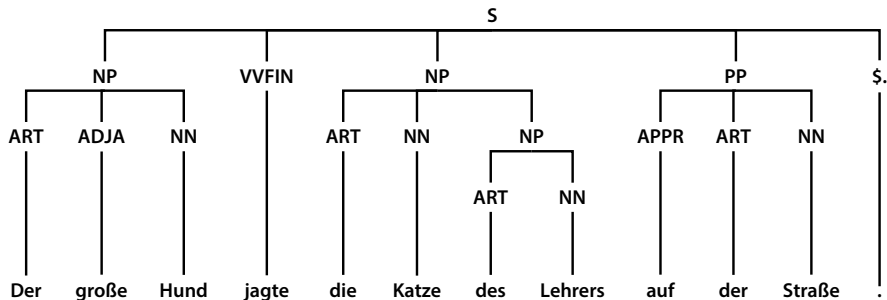


Abbildung: Beispiel-Parse Stanford-Parser dt. (trainiert auf TIGER-Korpus)