

Syntax natürlicher Sprachen

Vorlesung 2: Syntaxanalyse

A. Wisiorek

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

25.10.2022

1. Syntax als linguistische Disziplin

- 1 Syntax als linguistische Disziplin
 - **Syntax- und Grammatikbegriff**
 - **Syntaktische Ausdrucksmittel**
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Syntax (von gr. *syntaksis*, 'Zusammensetzung')

Lexikon der Sprachwissenschaft (Bußmann):

„**Teilbereich der Grammatik natürlicher Sprachen** (auch: Satzlehre).“

„**System von Regeln**, die beschreiben wie **aus einem Inventar von Grundelementen** (Morphemen, Wörtern, Satzgliedern) **durch spezifische syntaktische Mittel** (Morphologische Markierung, Wort- und Satzgliedstellung, Intonation u.a.) **alle wohlgeformten Sätze einer Sprache abgeleitet** werden können.“

<https://mediensprache.net/de/lexikon:>

„**Teilgebiet der Linguistik**, das sich mit der **Kombination von Wörtern zu komplexen Einheiten** (Analyse des Aufbaus von Satzstrukturen und der Zusammenfügung von Wörtern zu größeren Einheiten) beschäftigt, ohne sich für den internen strukturellen Aufbau der Wörter zu interessieren.“

„Der Begriff kann auch benutzt werden, um **den strukturellen Aufbau eines Satzes zu bezeichnen** ('Syntax eines Satzes' und so weiter).“

1.1. Syntax- und Grammatikbegriff

- 1 Syntax als linguistische Disziplin
 - **Syntax- und Grammatikbegriff**
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Allgemeiner Syntaxbegriff der Semiotik

- Syntax als **Struktur einer Zeichenfolge**
→ *Regeln der Kombination elementarer Zeichen zu komplexen Zeichen*

Syntax natürlicher Sprachen

- Syntax als **Struktur von Wortfolgen**
→ *Regeln der Kombination von Wörtern zu größeren Einheiten*
→ *Satz als zentrale sprachliche Einheit: Sprechen in Äußerungen*

Syntax als linguistische Disziplin

- **Satzstruktur-Analyse**
- Teilgebiet der **Grammatik** als Sprachstruktur-Analyse

Grammatik

- (Theorie der) Sprachstruktur

Ebenen des strukturellen Aufbaus von Sprache

- Lautstruktur: **Phonologie**
- Wortstruktur: **Morphologie**
- Satzstruktur: **Syntax**

1 Sprachstruktur

→ *phonologische, morphologische und syntaktische **Regularitäten** einer natürlichen Sprache*

2 Theorie der Sprachstruktur

→ *Sprachwissenschaftliche **Beschreibung der Regularitäten** einer natürlichen Sprache (Modell)*

3 Wissen um Sprachstruktur

→ ***Wissen des Sprechers um diese Regularitäten***

4 Regelbuch

→ ***Lehrwerk, das die Regularitäten einer natürlichen Sprache enthält***

5 Formale Grammatik

→ ***mathematisches Regelsystem** einer formalen Sprache, das zur **Modellierung der grammatikalischen Regelsysteme einer natürlichen Sprache** verwendet werden kann*

Abgrenzung Syntax zu anderen Disziplinen

Abgrenzung zur Morphologie

- Syntax: Analyse des *Strukturaufbaus* sprachlicher Einheiten *oberhalb der Wortebene*

Abgrenzung zu Semantik und Pragmatik

- Syntax: unabhängig von semantischer Interpretation (vgl. Chomsky 1957, 'Syntactic Structures')

Chomsky 1957: *colorless green ideas sleep furiously*

- Satz erfüllt **Wohlgeformtheitsbedingungen**, ist also grammatisch
- hat aber keine sinnvolle semantische Interpretation

Wortartenklassifikation

- **Zusammensetzung** syntaktischer Einheiten
aus **Klassen von Wörtern** (POS = lexikalische Kategorien)
→ *lexikalische Regeln*: ADJ → groß | klein

Flexionsmorphologie

- Analyse von **Wortformen**, insofern sie für die syntaktische Strukturanalyse relevant sind (**Morphosyntax**)
- **Kasus** und **Agreement** als morphologische Ausdrucksmittel syntaktischer Funktionen
- formale Repräsentation als **Merkmalstrukturen**

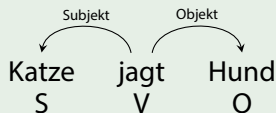
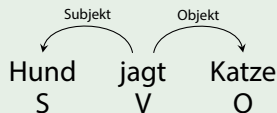
1.2. Syntaktische Ausdrucksmittel

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - **Syntaktische Ausdrucksmittel**
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Dependenzrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Wortstellung

- Markierung syntaktischer Funktion durch *lineare Anordnung*

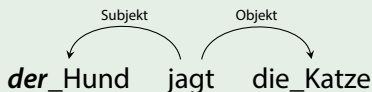
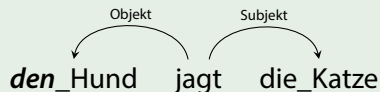
SVO-Wortstellung



Kasus

- *morphologische Markierung* der syntaktischen Funktion am *abhängigen Element*
- Rektion (*dependent-marking*)

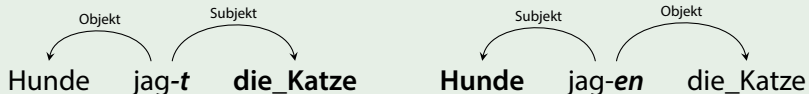
Objekt-Kasus-Markierung



Kongruenz

- Übereinstimmung von Merkmalen zwischen abhängigen Elementen
- *morphologische Markierung* der syntaktischen Funktion des abhängigen Elements am *Kopf* (*head-marking*)

Subjekt-Kongruenz (Numerus und Person)



→ *Verb kongruiert in nominalen Kategorien (Numerus, Person) mit Subjekt-NP*

- (Argument für Regel: $S \rightarrow NP VP$ (Subjekt > Objekt))

2. Syntaktische Struktur

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Dependenzrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

2.1. Satzstruktur

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - **Satzstruktur**
 - Konstituenz- vs. Dependenzrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Linguistisches Wörterbuch (Lewandowski):

„grammatisch, intonatorisch und inhaltlich nach den Regularitäten der jeweiligen Sprache linear und hierarchisch organisierte Einheit als Mittel zu Ausdruck, Darstellung und Appell, zur Kommunikation von Vorstellungen oder Gedanken über Sachverhalte.“

mediensprache.net/de/lexikon:

„kleinste (im Blick auf Inhalt, Struktur und Intonation) selbstständige und vollständige sprachliche Äußerung“

Satz als zentraler Untersuchungsgegenstand der Syntax

- sprachliche Form einer **Äußerung** (einer konkreten Sprachhandlung)
- Beobachtung: lineare Abfolge von Wörtern (**Wortfolge**)

Syntax: Analyse der hierarchischen Struktur von Sätzen

1 Struktur einfacher Sätze (*Clause*):

- **hierarchischer Aufbau** aus Wortgruppen (Konstituenten, insbesondere Phrasen)
- **funktionalen Abhängigkeiten** zwischen diesen syntaktischen Einheiten
→ insbesondere zwischen sog. **Satzgliedern** mit zentralen syntaktischen Funktionen (Subjekt, Objekt, Adverbial, Prädikat)

2 Struktur komplexer Sätze (*Sentence*):

- Aufbau aus einfachen Sätzen
- funktionale Abhängigkeiten zwischen Teilsätzen

Struktur

- **Menge von Relationen**, die zwischen Elementen einer Grundmenge bestehen
- Relation = Menge geordneter Paare

Syntaktische Struktur

- Menge von **Relationen**
 - **zwischen Elementen des Lexikons** einer natürlichen Sprache (Wörtern)
 - und/oder daraus gebildeten **syntaktischen Einheiten**

Beispiel: Dependenzstruktur

- Lexikon: {*der, die, den, Hund, Katze, jagt, sieht*}
- mögliche Dependenz-Satzstruktur:
{(*sieht, Hund*), (*sieht, Katze*), (*Hund, der*), (*Katze, die*)}



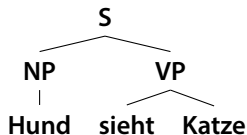
2.2. Konstituenz- vs. Dependenzrelation

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - **Konstituenz- vs. Dependenzrelation**
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Zwei syntaktische Relationstypen

Konstituenz

- **Teil-Ganzes-Beziehung** zwischen Wörtern und aus diesen bestehende syntaktische Einheiten (Konstituenten / Syntagmen)



Dependenz

- **Abhängigkeitsbeziehungen** zwischen Wörtern
- Kopf (*Head*) bestimmt Vorkommen und/oder Form von Dependent



Konstituenten-Struktur

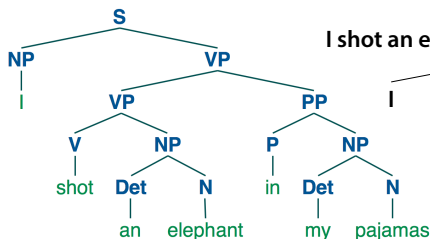
- Aus welchen **syntaktischen Einheiten** besteht ein Satz?
- Wie läßt sich eine Satz **hierarchisch zerlegen**?
- *Hintergrund: Aristotelische Logik (binäre Subjekt-Prädikat-Struktur)*

Syntagmatische Satzanalyse

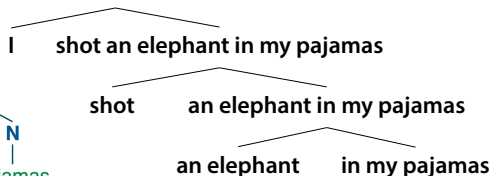
- Feststellung von syntaktischen Einheiten oberhalb Wortebene durch **syntaktische Tests**
- hierarchische Strukturregeln zur Erklärung des **Aufbaus von Sätzen** als lineare Wortfolgen
- Einfacher Satz (*Clause*) besteht aus:
 - **NP**
 - **VP**

I shot an elephant in my pajamas

↓ *Analysen der Konstituentenstruktur* ↓



I shot an elephant in my pajamas



Dependenz-Struktur

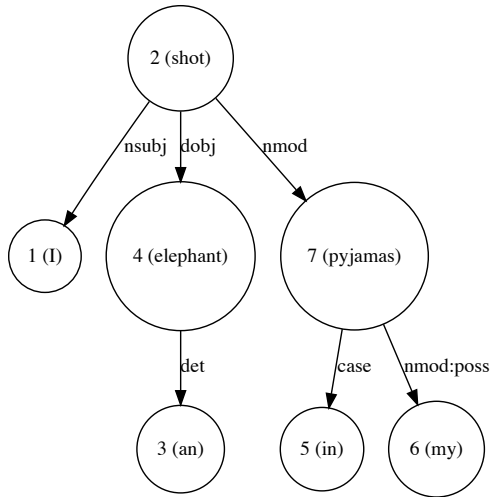
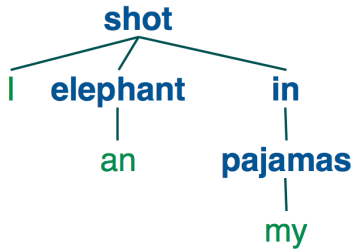
- In welcher **syntaktische Beziehung** stehen Wörter?
- Welche **Funktion** haben sie im Satz?
- *Hintergrund: Frege/Prädikatenlogik (mehrstellige Prädikate)*

Funktionale Satzanalyse

- notwendige und nicht-notwendige Einheiten im Satz
- **Abhängigkeitsverhältnisse** zwischen Wörtern
- Einfacher Satz (*Clause*) benötigt:
 - **Prädikat**
 - **Komplemente** (notwendige Erweiterungen)
 - **Adjunkte** (nicht notwendige Erweiterungen)

I shot an elephant in my pajamas

↓ *Analysen der Dependenzstruktur* ↓



2.3. Repräsentation syntaktischer Strukturen

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - **Repräsentation syntaktischer Strukturen**
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Syntaxbaum (auch: Parsebaum, Ableitungsbaum)

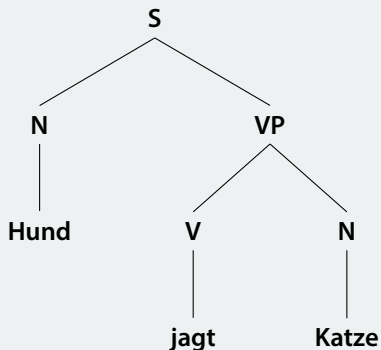
Baumstruktur = gerichteter Graph

- mathematische Repräsentation hierarchischer Struktur

Komponenten einer Baumstruktur

- **Knoten** = Elemente der Struktur
- **Kanten** = geordnete Paare von Knoten (ggf. gelabelt)
 - *Repräsentation der Relation zwischen zwei Knoten*
 - *bei Konstituentenstruktur: Teil-Ganzes-Relation*

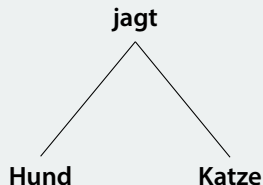
Baumdarstellung:



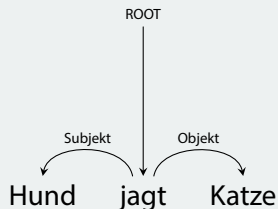
Klammerausdruck:

[S [N Hund] [VP [V jagt] [N Katze]]]

Baumdarstellung (Stemma):



mit gelabelten Kanten:



Klammerausdruck:

[jagt [Hund] [Katze]]

Notation als Tripel:

(jagt, Subjekt, Hund), (jagt, Objekt, Katze)

3. Automatische Syntaxanalyse

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 **Automatische Syntaxanalyse**
 - **Formale Grammatiken als Syntaxmodelle**
 - **Parsing als automatische Syntaxanalyse**
 - **Besondere syntaktische Strukturen**
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - **Parsing in NLP**

3.1. Formale Grammatiken als Syntaxmodelle

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - **Formale Grammatiken als Syntaxmodelle**
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

1 Beschreibung des Sprachsystems

→ *traditionelle Buch-Grammatik; nicht-computational*

2 Aufzählung aller grammatischen Sätze

→ *Problem 1: natürliche Sprachen sind unendlich*

→ *Problem 2: Struktur nicht repräsentiert*

3 Beschreibung durch *formale Grammatik*

→ *mathematisches Modell des syntaktischen Regelsystems
(computational)*

→ *ermöglicht die Analyse der Struktur einer unendlichen Menge an Sätzen mit endlichen Mitteln*

- **Formales Regelsystem** zur eindeutigen **Beschreibung** und **Erzeugung** einer *formalen (!) Sprache*
 - *Generierung aller wohlgeformten Sätze (= Sprache)*
 - generative Grammatik
- **Produktionsregeln**
 - *ausgehend von Startsymbol*
 - *Ersetzung linker mit rechter Regelseite*
- kann auch als Modell zur **Erkennung** und **Wiedergabe** der syntaktischen Struktur *natürlicher Sprachen* verwendet werden

- **Formale Sprache = Menge aller** aus Grundsymbolen (z. B. $\{a, b, c\}$) mit den Grammatikregeln **ableitbaren formalsprachlichen Wörter** (z. B. $\{a, aa, aba, \dots\}$)
- in Analyse Syntax **natürlicher Sprache**:
 - *Grundsymbole sind Wörter des Lexikons:*
 $\{die, der, den, Hund, Katze, jagt\}$
 - die aus der entsprechenden formalen Grammatik als Syntaxmodell ableitbaren *formalsprachlichen Wörter* sind *natürlichsprachliche Sätze*:
 $\{der - Hund - jagt - die - Katze, die - Katze - jagt - den - Hund, \dots\}$

Syntaktische Regeln

- 1 $S \rightarrow NP \ VP$
- 2 $PP \rightarrow P \ NP$
- 3 $NP \rightarrow DET \ N \mid Det \ N \ PP$
- 4 $VP \rightarrow V \ NP \mid VP \ PP$

Lexikalische Regeln

- 1 $DET \rightarrow 'an' \mid 'my'$
- 2 $N \rightarrow 'elephant' \mid 'pajamas'$
- 3 $V \rightarrow 'shot'$
- 4 $P \rightarrow 'in'$

Startsymbol: erstes Symbol LHS (*left-hand-side*)

1 | S → ...

Terminale: RHS lexikalischer Regeln (*right-hand-side*)

- natürlichsprachliche Wörter (Blätter im Syntaxbaum)

1 | ... → 'an' | 'my'

2 | ... → 'elephant' | 'pajamas'

Nicht-Terminal: RHS und LHS syntaktischer Regeln

- insbesondere: **syntaktische Kategorien** (Phrasentypen)
- **CFG-Beschränkung**: nur 1 Nicht-Terminal auf LHS

```
1 | S → NP VP
2 | PP → P NP
```

Präterminale: LHS lexikalischer Regeln

- **Wortarten** (lexikalische Kategorien)
- gehören formal zu den **Nicht-Terminal-Symbolen**

```
1 | DET → ...
2 | N → ...
```

- **modellierte Relation**

→ *Konstituentengrammatik : Dependenzgrammatik*

- **Kategorien**

→ *atomare Kategorien (CFGs) : komplexe Merkmalstrukturen (feature-based CFGs)*

- **Komplexität der Grammatik (Chomsky-Hierarchie)**

→ *regulär : kontextfrei : kontext-sensitiv : rekursiv aufzählbar*

- **Analysetiefe der Grammatik (Rekursion?)**

→ *flach : verschachtelt*

- unendliche Menge an Sätzen **mit endlichen Mitteln beschreibbar**
- **rechnergestützt verarbeitbar** durch Parsingalgorithmen
- Beantwortung Fragen zur **Komplexität** natürlicher Sprache (ist jede natürliche Sprache kontextfrei?)
- psycholinguistische Anwendung: Parser als **Modell menschlicher Sprachverarbeitung**

- keine **Disambiguierung** bei struktureller Ambiguität
- **Übergenerierung** durch neuen Grammatikregeln
- **vollständige Beschreibung** der Grammatik natürlicher Sprache schwierig
→ *handgeschriebenen Grammatiken modellieren immer nur Ausschnitt*

3.2. Parsing als automatische Syntaxanalyse

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - **Parsing als automatische Syntaxanalyse**
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Verarbeitung formaler Grammatiken

- Formale Grammatik = Syntaktisches Strukturmodell
→ *dessen Regeln sind aber nicht mehr als **Sammlung von Strings***
- **Verfahren** notwendig, um zu entscheiden, ob eine Eingabe gemäß einer gegebenen formalen Grammatik **wohlgeformt** ist

Parsing-Algorithmen

- Verfahren zur **Verarbeitung** von formalen Grammatiken zur **Strukturerkennung und -Analyse** der **Eingabe** (Satz als Tokensequenz)

Strukturerkennung

- **Überprüfung** der grammatischen Struktur einer Eingabe als Suche einer Ableitung aus den Regeln einer formalen Grammatik (ob Satz in formaler Sprache enthalten ist)

Strukturzuweisung

- gleichzeitig **Wiedergabe** der in der Suche aufgebauten grammatischen Struktur der Eingabe (Syntaxbaum)

- **CFG-Parsing**

- *top-down vs. bottom-up vs. dynamische Programmierung*

- **Unifikationsparsing**

- *Verarbeitung von Merkmalstrukturen*

- **statistische Parsingalgorithmen**

- *Viterbi-Algorithmus* (effizientes Auffinden der wahrscheinlichsten Ableitung; Wahrscheinlichkeiten aus syntaktisch annotiertem Korpus gelernt = Treebank)

- **Dependency Parsing**

- *u.a. transition-based* (basierend auf aus Dependency Treebank gelerntem Modell)

- **Partielles Parsing / Chunk-Parsing**

- *Parsing as Tagging* (mit regulärer Grammatik oder Training eines Klassifikators)

3.3. Besondere syntaktische Strukturen

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Dependenzrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - **Besondere syntaktische Strukturen**
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP

Problem

- mehr als eine Strukturanalyse möglich

Lösung

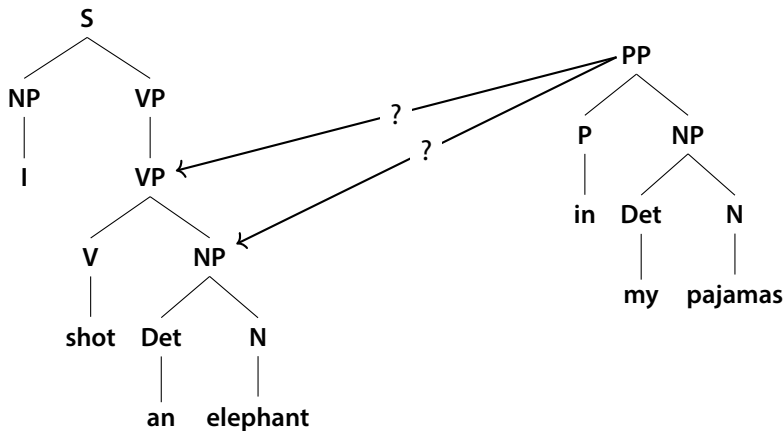
- Disambiguierung struktureller Ambiguität durch **PCFGs** (mit statistischen Informationen zu Regelwahrscheinlichkeiten angereicherte CFGs)
- weitere Disambiguierung durch **Lexikalisierung** (erfasst z.B. die Präferenz für PP-Attachment an VP bei *setzen/ stellen/ legen*-Verben)

Beispielsatz für PP-Attachment-Ambiguität

- *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

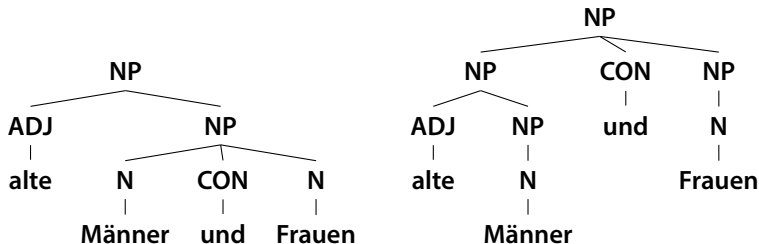
- Konstituente kann im Parsebaum an mehr als einer Stelle angebunden werden



Beispiel Koordinierungsambiguität

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

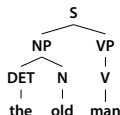
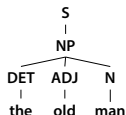
- Skopus des Adjektivs unklar:



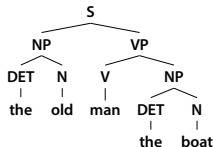
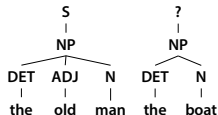
Beispiel Garden-Path-Satz

- *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!**

Rekursive Grammatik-Regel

- Symbol auf left-hand-side (LHS) erscheint auch auf right-hand-side (RHS)

linksrekursive Regel

1 | NP \rightarrow NP PP

- linksrekursive Regel nicht mit Recursive-Descent-Parser verarbeitbar
- führt zu Endlosschleife

rechtsrekursive Regel

1 | NP \rightarrow ADJ NP

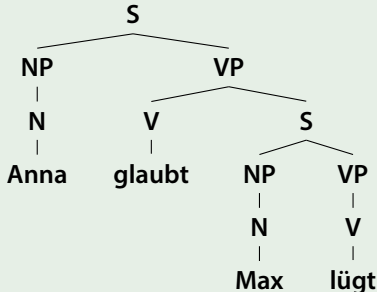
center-embedding-Rekursion

1 | $NP \rightarrow N \text{ RELPRO } NP \text{ V}$

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

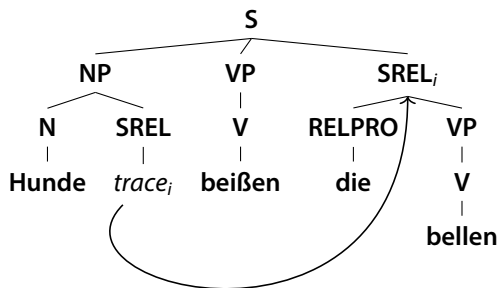
indirekte Rekursion

1 | $S \rightarrow NP \text{ VP}$
2 | $VP \rightarrow V \text{ S}$



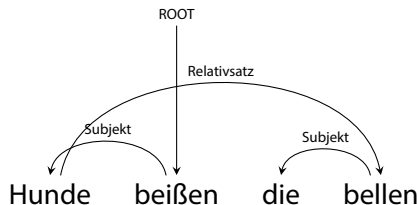
Nicht-projektive Strukturen: CFG

- auch: diskontinuierliche Strukturen, *long distance dependencies*
 - entstehen durch Herausbewegung von Phrasen/Konstituenten
- mit CFGs: nicht direkt modellierbar (überkreuzende Kanten)
 - nur mit *traces* (Leerstelle + Movement/Transformationsregeln)



- Dependenzgrammatik: unabhängig von linearer Anordnung (Abhängigkeitsrelationen)
- aber: bestimmte Dependency Parsingalgorithmen können nicht-projektive Strukturen nicht verarbeiten

beißen
|
Hunde
|
bellen
|
die



3.4. Parsing in NLP

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - **Parsing in NLP**

Voraussetzungsschritte für automatische Syntaxanalyse

- Sentence Segmentation
- Tokenisierung
- Part-of-Speech-Tagging
- Stemming
- morphologisches Parsing (Kasus, Agreement)

Mögliche Folgeanwendungen

- **Entity Extraction**
- **Relation Extraction**
- **Semantic Parsing** (Natural Language Understanding)

Parsing in NLP-Pipeline

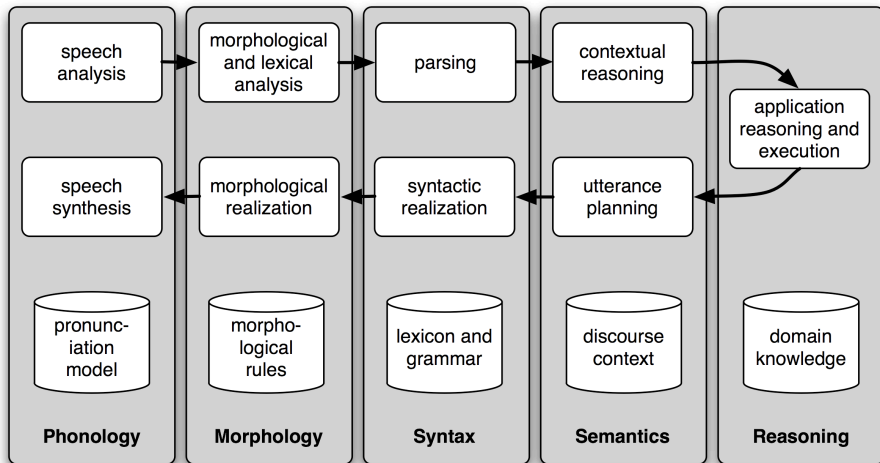
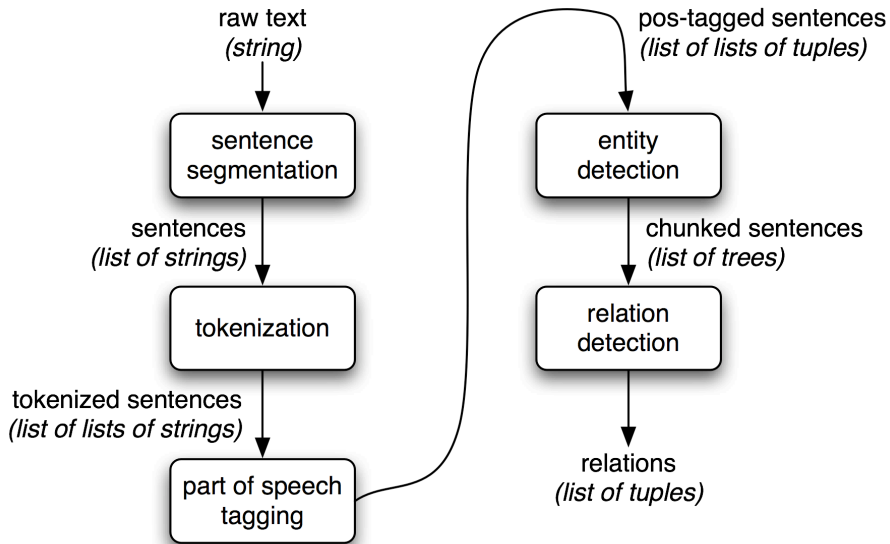


Abbildung: Parsing in NLP-Pipeline (*Spoken Dialogue System*),

<http://www.nltk.org/book/ch01.html#fig-sds>

Parsing in IE-Pipeline



Rückblick auf heutige Themen

- 1 Syntax als linguistische Disziplin
 - Syntax- und Grammatikbegriff
 - Syntaktische Ausdrucksmittel
- 2 Syntaktische Struktur
 - Satzstruktur
 - Konstituenz- vs. Abhängigkeitsrelation
 - Repräsentation syntaktischer Strukturen
- 3 Automatische Syntaxanalyse
 - Formale Grammatiken als Syntaxmodelle
 - Parsing als automatische Syntaxanalyse
 - Besondere syntaktische Strukturen
 - Syntaktische Ambiguität (Disambiguierung)
 - Temporale Ambiguität (Backtracking)
 - Rekursive Regeln
 - Nicht-projektive Strukturen
 - Parsing in NLP