

Syntax natürlicher Sprachen

Einführung

A. Wisiorek

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

14.10.2025

1. Kursübersicht

- 1 Kursübersicht
- 2 Syntaxbegriff
- 3 Syntax als linguistische Disziplin
- 4 Syntaktische Ausdrucksmittel
- 5 Konstituenz vs. Dependenz
- 6 Repräsentation syntaktischer Strukturen

- Die Vorlesung behandelt **Grundbegriffe der Grammatik** und wesentliche syntaktische Konstruktionen *des Deutschen* im Hinblick auf eine **Verwendung in der maschinellen Sprachverarbeitung**.
- Dazu werden die in neueren Grammatiktheorien verwendeten **Klassifizierungen von Phrasen**, deren **innere Struktur** sowie deren **relationale Abhängigkeiten** erklärt.
- Außerdem werden in der Computerlinguistik übliche **Grammatikformalismen** (wie Kontextfreie Grammatiken, Unifikationsgrammatiken, PCFGs, Datenbasierte Abhängigkeitsgrammatiken, Partielle Parsingmodelle) besprochen.
- ebenso wie **syntaktische Annotationsstandards** (z.B. Penn Treebank, Universal Dependencies) vorgestellt und verwendet, um typische oder schwierige syntaktische Konstruktionen genau zu beschreiben.

- Kenntnis und Anwendung **funktionaler und struktureller Begriffe der grammatischen Beschreibung**:
 - Kongruenz
 - Rektion
 - Subkategorisierung
 - Valenz
- Kenntnis **Grammatikformalismus** und Anwendung für Analysen natürlichsprachlicher Sätze:
 - CFG (Kontextfreie Grammatik: *Konstituentenstruktur*)
 - FCFG (Feature-based CFG: *Modellierung grammatischer Merkmale*)
 - PCFG (Probabilistische CFG: *Gewichtung von CFG-Regeln*)
 - DG (Dependency Grammar: *Dependenzstruktur*)
 - **Chunk-Parser** (u.a. RegExpParser: *partiell parsing, 'Parsing as Tagging'*)
- Kenntnis und Bedienung eines **Programmsystems**, das einen Grammatikformalismus verwendet
→ *NLTK zum Schreiben von CFGs, FCFGs, PCFGs und DGs*

NLTK: Anwendung regelbasierter Parsing-Grundlagen

- Schreiben von Grammatikregeln und Anwendung in Parsingalgorithmen
- konstituentenbasierte sowie dependenzbasierte Grammatiken
- 'Toy Grammar' für einzelne Beispielsätze

Informationen zum NLTK (Natural Language Toolkit)

- Bündel von Python-Bibliotheken und Programmen für computerlinguistische Anwendungen
- quelloffen, für Lehre entwickelt
- bietet auch Interfaces, z.B. für *Stanford Parser*
- Lehrbuch: <http://www.nltk.org/book>
- Dokumentation: <http://www.nltk.org/howto>
- Daten (Korpora, Grammatiken): <http://www.nltk.org/data.html>

2. Syntaxbegriff

- 1 Kursübersicht
- 2 Syntaxbegriff**
- 3 Syntax als linguistische Disziplin
- 4 Syntaktische Ausdrucksmittel
- 5 Konstituenz vs. Dependenz
- 6 Repräsentation syntaktischer Strukturen

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).“

mediensprache.net/de/lexikon:

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

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.“

3. Syntax als linguistische Disziplin

- 1 Kursübersicht
- 2 Syntaxbegriff
- 3 Syntax als linguistische Disziplin**
- 4 Syntaktische Ausdrucksmittel
- 5 Konstituenz vs. Dependenz
- 6 Repräsentation syntaktischer Strukturen

Syntax als Satzstruktur-Analyse

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 wie Sätzen*
→ *Satz als zentrale sprachliche Einheit*
→ *kleinste abgeschlossene Äußerung im Sprachgebrauch*

Syntax als linguistische Disziplin

- Teilgebiet der **Grammatik** (Phonologie, Morphologie, Syntax)

Grammatik (von gr. (*technē*) *grammatikē*, 'Schreibkunst')

- (Theorie/Analyse der) **Sprachstruktur**

Abgrenzung Syntax zu anderen Disziplinen

Abgrenzung zur Phonologie und Morphologie

- **Syntax:** Analyse des *Strukturaufbaus* sprachlicher Einheiten *oberhalb der Wortebene*
 - **Phonologie:** Lautstruktur = kleinste bedeutungsunterscheidende Elemente
 - **Morphologie:** Wortstruktur = kleinste bedeutungstragende Elemente

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**

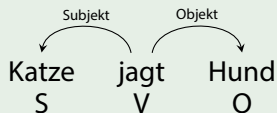
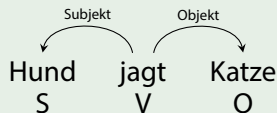
4. Syntaktische Ausdrucksmittel

- 1 Kursübersicht
- 2 Syntaxbegriff
- 3 Syntax als linguistische Disziplin
- 4 Syntaktische Ausdrucksmittel**
- 5 Konstituenz vs. Dependenz
- 6 Repräsentation syntaktischer Strukturen

Wortstellung

- Markierung syntaktischer Funktion durch *lineare Anordnung*

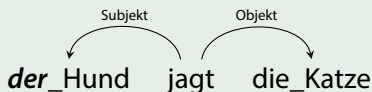
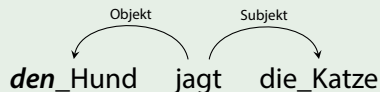
SVO-Wortstellung



Kasus

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

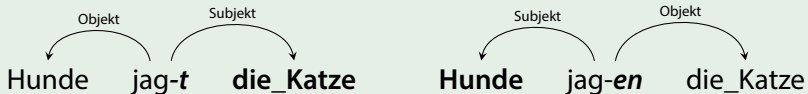
Objekt-Kasus-Markierung



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*

5. Konstituenz vs. Dependenz

- 1 Kursübersicht
- 2 Syntaxbegriff
- 3 Syntax als linguistische Disziplin
- 4 Syntaktische Ausdrucksmittel
- 5 Konstituenz vs. Dependenz**
- 6 Repräsentation syntaktischer Strukturen

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

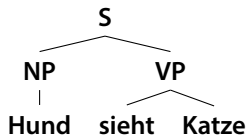
Syntaktische Struktur

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

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

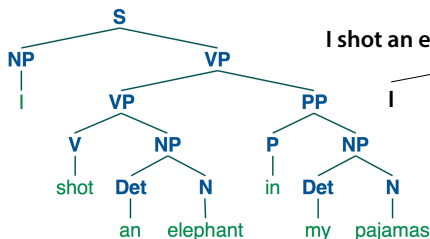
- Aus welchen **syntaktischen Einheiten** besteht ein Satz?
- Wie läßt sich eine Satz **hierarchisch zerlegen**?
- *Hintergrund: Aristotelische Logik (binäre Subjekt-Prädikat-Struktur)*

Syntagmatische Satzanalyse

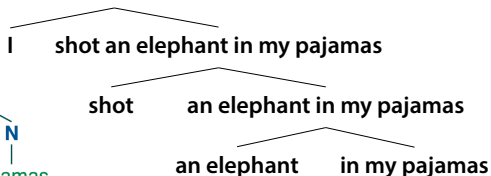
- Feststellung von syntaktischen Einheiten oberhalb Wortebene durch **syntaktische Tests**
- hierarchische Strukturregeln zur Erklärung des **Aufbaus von Sätzen** als lineare Wortfolgen
- Einfacher Satz (*Clause*) besteht aus:
 - **NP**
 - **VP**

I shot an elephant in my pajamas

⇓ *Analysen der Konstituentenstruktur* ⇓



I shot an elephant in my pajamas



Dependenz-Struktur

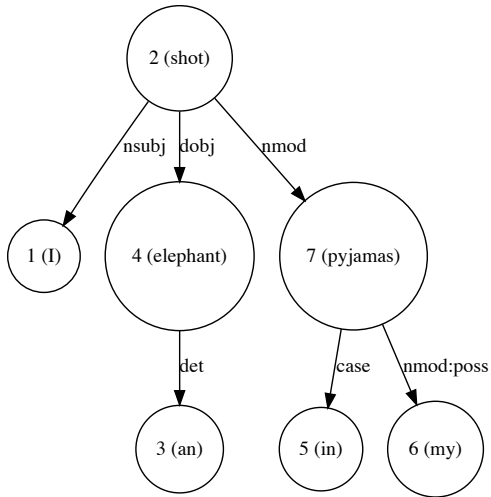
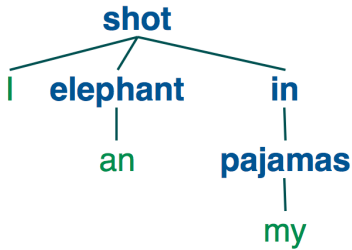
- In welcher **syntaktische Beziehung** stehen Wörter?
- Welche **Funktion** haben sie im Satz?
- *Hintergrund: Frege/Prädikatenlogik (mehrstellige Prädikate)*

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* ↓



6. Repräsentation syntaktischer Strukturen

- 1 Kursübersicht
- 2 Syntaxbegriff
- 3 Syntax als linguistische Disziplin
- 4 Syntaktische Ausdrucksmittel
- 5 Konstituenz vs. Dependenz
- 6 Repräsentation syntaktischer Strukturen**

Syntaxbaum (auch: Parsebaum, Ableitungsbaum)

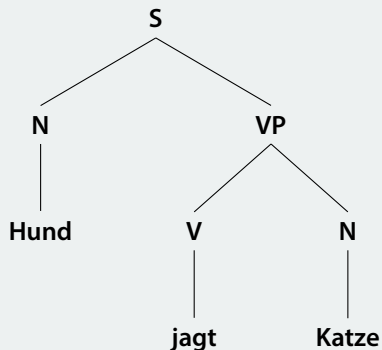
Baumstruktur = gerichteter Graph

- 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*

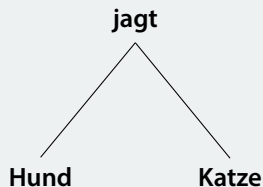
Baumdarstellung:



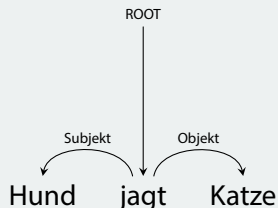
Klammerausdruck:

[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)