



Computerlinguistische Grundlagen

Jürgen Hermes

Sommersemester 19

Sprachliche Informationsverarbeitung

Institut für Digital Humanities

Universität zu Köln



Parser

- **Begriffe, Anwendungen, Strategien, Überblick**
- **Elementare Parsingalgorithmen**
- **Komplexere Parser**
 - **Chart-Parser**
 - **Statistisches Parsing**



Chart-Parser

- **Aufgabe:** Lösung des Problems der Mehrfachanalysen durch Speicherung der Teilergebnisse in einer Chart.
- **Kanten** (In der Chart gespeicherte Objekte; aktiv vs. passiv):
 - kategorialer Typ der repräsentierten Konstituente
 - Satzabschnitt, über den sich die Kante erstreckt
 - (nur aktive Kanten) Spezifikation des erkannten und des nicht erkannten Teils
- **Operationen:**
 - Eintrag der Kanten in die Chart
 - Erzeugung neuer Kanten in Abhängigkeit von der Chart und vom syntaktischen Wissen.



Earley-Algorithmus (Earley 1970)

- **Datenstruktur:** Chart-Matrix mit Teilungspunkt.
- **Operationen:**
 - prädikative Erweiterung (predictor-Schritt)
 - lexikalische Konsumption (scanner-Schritt)
 - Konstituentenvervollständigung (completer-Schritt)
- **Analyserichtung:**
 - Top-Down (predictor-Schritt)
 - Bottom-Up (completer-Schritt)



Active Chart Parsing (Winograd 1983)

- basierend auf dem Earley-Algorithmus, verwendet allerdings statt geteilter Produktion einen Graphen mit aktiven und passiven Kanten
- passive Kanten: vollständig erkannte Konstituenten; feste Verbindung zweier Chartknoten
- aktive Kanten: partiell erkannte Konstituenten; Quellknoten sucht anhand der gegebenen Grammatik den Zielknoten



Statistisches Parsen

- Ermittlung der wahrscheinlichsten Strukturbeschreibung für Eingabesatz durch Berechnung.
- Basiert auf probabilistischen Grammatiken (syntaktische und lexikalische Regeln sind mit Gewichten versehen; je wahrscheinlicher, desto gewichtiger)



Literatur / Hausaufgabe

Zur Nachbereitung:

Naumann, Langer (1994): Chart-Parsing (S. 102-106)

Strube (1996): „Parsing natürlicher Sprache“ (S. 461 ff)

Jurafsky & Martin (2009): Statistical Parsing (S. 493-500)

Parsen Sie den Satz „Lisa spielt auf dem Saxophon“ in Top-Down und in Bottom-Up-Richtung. Zeichnen Sie dazu zwei Charts (eine in Tabellenform, eine als Graphen), in die jeweils alle passiven Kanten eingetragen werden sollen.

Vervollständigen Sie außerdem die Chart auf S. 500 (Jurafski/Martin)