



Computerlinguistische Grundlagen

Jürgen Hermes

Sommersemester 20

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.



Chart-Parser

① Das ② Buch ③ lag ④ auf ⑤ dem ⑥ Tisch ⑦ gestern ⑧

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---------|----------|---------|---------|---------|-----------|---------------|
| 0 | | D → Das | NP → D N | | | | | S → NP VP |
| 1 | | | N → Buch | | | | | |
| 2 | | | | V → lag | | | | VP → V PP Adv |
| 3 | | | | | P → auf | | PP → P NP | |
| 4 | | | | | | D → dem | NP → D N | |
| 5 | | | | | | | N → Tisch | |
| 6 | | | | | | | | Adv → gestern |



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)



Earley-Algorithmus (Earley 1970)

①Karl②liebt③die④warmen⑤Sommermonate⑥

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|-----------------------------------------------------|------------------------------|------------------------------|
| 0 | $S \rightarrow \cdot NP VP$ $NP \rightarrow \cdot N$ $NP \rightarrow \cdot D A N$ | $S \rightarrow NP \cdot VP$ $NP \rightarrow N \cdot$ $N \rightarrow Karl$ | | | | $S \rightarrow NP VP \cdot$ |
| 1 | | $VP \rightarrow \cdot V NP$ $VP \rightarrow \cdot V PP$ | $VP \rightarrow V \cdot NP$ $VP \rightarrow V \cdot PP$ $V \rightarrow liebt$ | | | $VP \rightarrow V NP \cdot$ |
| 2 | | | $NP \rightarrow \cdot N$ $NP \rightarrow \cdot D A N$ $PP \rightarrow \cdot P NP$ | $NP \rightarrow D \cdot A N$ $D \rightarrow die$ | $NP \rightarrow D A \cdot N$ | $NP \rightarrow D A N \cdot$ |
| 3 | | | | | $A \rightarrow warmen$ | |
| 4 | | | | | | $N \rightarrow Sommermonate$ |
| 5 | | | | | | |

Syntax

$S \rightarrow NP VP$
 $NP \rightarrow N$
 $NP \rightarrow D A N$
 $VP \rightarrow V NP$
 $VP \rightarrow V PP$
 $PP \rightarrow P NP$

Lexikon

$A \rightarrow warmen$
 $D \rightarrow die$
 $N \rightarrow Karl$
 $N \rightarrow Sommermonate$
 $V \rightarrow liebt$

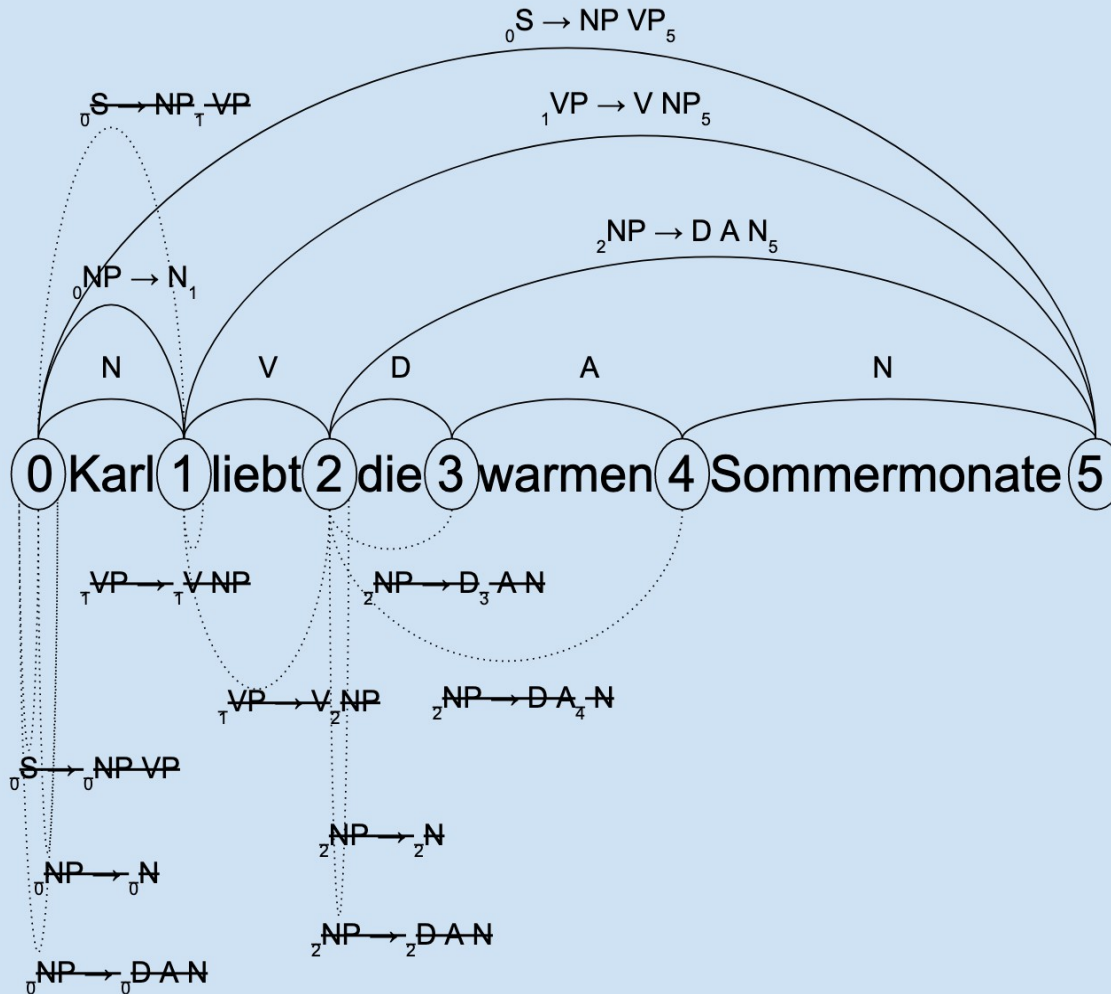


Active Chart Parsing (Winograd 1983)

- 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



Active Chart Parsing (Winograd 1983)



Syntax

$S \rightarrow NP VP$
 $NP \rightarrow N$
 $NP \rightarrow D A N$
 $VP \rightarrow V NP$

Lexikon

$A \rightarrow \text{warmen}$
 $D \rightarrow \text{die}$
 $N \rightarrow \text{Karl}$
 $N \rightarrow \text{Sommermonate}$
 $V \rightarrow \text{liebt}$



Literatur / Hausaufgabe

Zur Nachbereitung:

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

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

Parsen Sie den Satz „Lisa spielt auf dem Saxophon“ mithilfe des Earley-Algorithmus und zeichnen Sie außerdem die Chart als Graphen. Die genauere Aufgabenstellung findet sich bei ILIAS.