

# SATZVERARBEITUNG

Lars Konieczny

*Institut für Informatik und Gesellschaft*

*Abteilung Kognitionswissenschaft*

*Universität Freiburg*

*[lars@cognition.iig.uni-freiburg.de](mailto:lars@cognition.iig.uni-freiburg.de)*

## Fragestellungen

- Welche Rolle spielt Syntax?
  - ▣ Ist Syntax autonom/modular?
  - ▣ Kompetenz vs. Performanz
  - ▣ Parsingalgorithmen
  - ▣ Ambiguitäten
  - ▣ Seriel vs. parallel vs. unterspezifiziert
  - ▣ Komplexität

## Fragestellungen II

- Welche Rolle spielen andere Faktoren?
  - ▣ Diskurskontext / Referenz
  - ▣ Visueller Kontext
  - ▣ Prosodie
  - ▣ Lexikalische Faktoren

Sprache und Kognition: Satzverarbeitung

Lars Konieczny

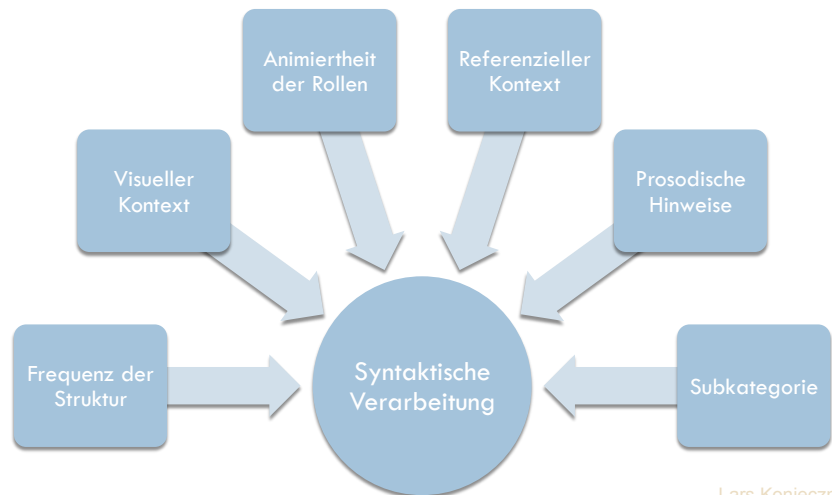
## Ein modularer Ansatz



Sprache und Kognition: Satzverarbeitung

Lars Konieczny

## Ein interaktiver Ansatz



## Kompetenz und Performanz

- Kompetenz: Sprachliches Wissen
  - ▣ Der Hund jagt die Katze
  - ▣ Der Hund, den die Maus sieht, jagt die Katze
  - ▣ Der Hund, den die Maus, die den Vogel erschreckt, jagt die Katze.
  - ▣ ...
- Performanz: Wie wird Sprache verarbeitet?
  - ▣ Kognitive Architektur (modular?)
  - ▣ Beschränkungen des Arbeitsgedächtnis

# Kompetenz und Performanz

- Kompetenz: Sprachliches Wissen
  - Repräsentationen
    - Phrasenstrukturbäume
    - Merkmalstrukturen
    - Beschränkungen (Constraints)
    - Ableitungsregeln
  - Beste Näherung durch linguistische Grammatiktheorie
    - Generativistische Grammatiken
    - Lexikal Functional Grammar
    - Construction Grammar
    - ...

Psycholinguistik

Lars Konieczny

# Kompetenz und Performanz

- Performanz
  - Wie wird das sprachliche Wissen verwendet, um Sprache zu verarbeiten?
  - Welche Verfahren, Algorithmen oder Mechanismen?
    - (Wortweise) *inkrementell*, von links-nach-rechts
    - Auswahl einer Alternative bei Ambiguitäten
  - Architektur
    - Wie arbeiten die verschiedenen Wissens Ebenen zusammen?  
Gibt es Beschränkungen in der Interaktion?
    - Beschränkungen im Arbeitsgedächtnis

Psycholinguistik

Lars Konieczny

## Holzwegsätze (garden-path sent.)

Zum Verständnis der menschlichen Sprachverarbeitung betrachtet man gern Grenzfälle, in denen die korrekte Interpretation nur mit mehr oder weniger Mühe gefunden werden kann.

Peter hat die Ärztin im Krankenhaus gestern sehr geholfen.

Peter<sub>[nom, akk, dat]</sub> Ärztin<sub>[nom, akk]</sub>, Präferenz: Subjekt < Objekt

The boat floated down the river sank.

floated<sub>[main verb past tense, past participle im reduzierten Relativsatz]</sub>  
Präferenz: main verb)

## Syntax

- Das Zusammenfügen von Wörtern zu Sätzen folgt in jeder Sprache syntaktischen Regeln.
  - Z.B. im Deutschen: Artikel < Adjektiv < Nomen
    - der bunte Clown
    - \* bunte der Clown
    - \*Clown bunte der
- Die Struktur von Sätzen ist hierarchisch.
  - Der *Junge*, den die Mitschüler ärgerten, *mochte* den Lehrer.
    - Numerus-Kongruenz zwischen Subjekt und Verb kann über beliebige Distanzen und über beliebig viele andere Verben hinweg hergestellt werden.

## Lexikalisch-syntaktische Merkmale: Argumentstruktur von Verben (1)

- Marion verspricht Lisa, sie zu begleiten.
- Marion versagt Lisa, sie zu begleiten.
  
- Marion verspricht Lisa, sich begleiten zu lassen.
- Marion versagt Lisa, sich begleiten zu lassen.

Psycholinguistik

Lars Konieczny

## Lexikalisch-syntaktische Merkmale: Argumentstruktur von Verben (2)

- **versprechen & versagen**
  - gleiche Komplemente
    - 1. direktes Objekt (NP)
    - 2. Komplementsatz (CP)
  - Bei **versprechen** ist das Subjekt der CP identisch mit dem **AGENS** des Matrixsatzes.
    - Die pronominale Referenz auf das Subjekt eines Satzes im Satz selbst erfolgt mit dem Pronomen **sich**
  - Bei **versagen** oder **erlauben** ist das Subjekt der CP identisch mit dem **PATIENS** des Matrixsatzes.
- Funktionale und thematische Rollen sind Inhalte des mentalen Lexikons.
- LFG (lexical-functional grammar, Bresnan 1982) erfasst beide Aspekte.
- Dass es sich bei **versprechen** und **versagen** um eine Beziehung zwischen den thematischen Rollen des Matrixsatzes und einer funktionalen Rolle (Subjekt) der CP handelt, sieht man im Passiv:
- Lisa wurde von Marion versprochen, sie (nicht sich!) zu begleiten.

Psycholinguistik

## Syntax: Phrasen

- Nominalphrase (NP)  
„Die Vorlesung“
- Verbalphrase (VP)  
„zieht sich ...“
- Präpositionalphrase (PP)  
„in die Länge“
- ...

Psycholinguistik

Lars Konieczny

## Eine einfache Grammatik

### Phrasenstrukturregeln

S → NP VP

NP → pn

NP → det n

NP → det adj n

NP → adj n

NP → NP PP

PP → p NP

VP → v

VP → v NP

VP → v NP PP

VP → v S

### Lexikon (deutsch)

det = {der, die, den, dem, ..., ein, eine, ... jede, ...}

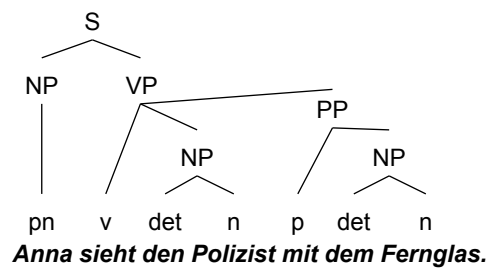
n = {Frau, Mann, Polizist, Freude, Fernglas, ...}

pn = {Peter, Anna, ...}

p = {auf, unter, neben, in, mit, ...}

v = {liest, arbeitet, schreibt, sieht, ...}

adj = {große, großer, großen, kleine, rote, ...}



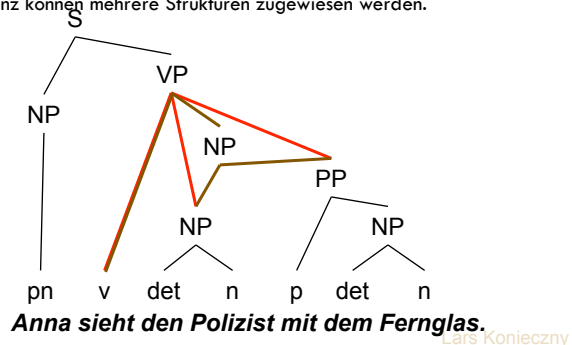
Psycholinguistik

# Syntaktische Ambiguität

- Lexikalisch: Ein Wort kann mehreren Kategorien angehören.
  - n → {train, hammer, book ...}
  - v → {read, train, witnessed, ...}
  - *The old train the young*
  - *The old train was broken*
- Strukturell: Einer Wortsequenz können mehrere Strukturen zugewiesen werden.

## PS-Regeln

NP	→	NP PP
VP	→	v NP
VP	→	v NP PP



Psycholinguistik

Lars Konieczny

# Lexikalische und strukturelle Ambiguität

## *Time flies like an arrow.*

- TIME
  - Nomen: die Zeit
  - Nominalkompositum: Zeit-
  - Verb: Zeit nehmen (mit Stoppuhr)
- FLIES
  - Nomen: die Fliege (pl.)
  - Nomen: der Hosenschlitz (pl.)
  - Verb: fliegen
- LIKE
  - Adverb: so ... wie ...
  - Verb: mögen, angenehm finden

Psycholinguistik

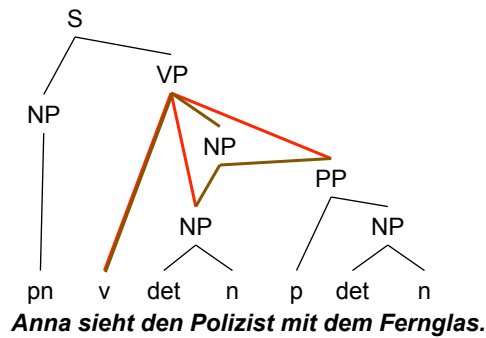


## Dimensionen der Satzverarbeitung

- Horizontal
  - ▣ links-nach-rechts
  - ▣ rechts-nach-links
  - ▣ Inselparsing
    - head-driven
    - key-driven

- Vertikal
  - ▣ bottom-up
  - ▣ top-down
  - ▣ mixed-mode

- Verarbeitung von Ambiguitäten
  - ▣ Seriell
  - ▣ parallel
  - ▣ deterministisch



Präferenzen?  
Auswege aus Holzwegen

Psycholinguistik

Lars Konieczny

## Inkrementelle Verarbeitung I

- (1) ..., daß **sie** nach dem Ergebnis zu fragen **erlaubt hat**.  
 ..., that **she** for the result to ask permitted has.  
 „... that **she** has given permission to ask for the result.“
- (2) ..., daß **sie** nach dem Ergebnis zu fragen **erlaubt war**.  
 ..., that **her** for the result to ask permitted was.  
 „... that it was permitted to ask **her** for the result.“
- ❖ Kamide & Mitchell (1999)
- (3) Kyooju-ga gakusee-ni shokansisho-ga **kasita** mezurasii komonjo-o **miseta**.  
 Prof.-nom student-dat librarian-nom **lent** ... manuscript-acc **showed**  
 a. “The professor showed the ... manuscript which the librarian had lent the student” (low)  
 b. “The professor showed the student the ... manuscript which the librarian had lent” (high)

Psycholinguistik

Lars Konieczny

## Inkrementelle Verarbeitung II

- Inkrementelle Präferenzen bei der Ambiguitätsauflösung sprechen für Strukturaufbau schon vor dem Kopf der Phrase.
- Aber auch bei einfachen eindeutigen Strukturen lassen sich Hinweise auf einen frühen Strukturaufbau finden:
  - Hemforth et al. (1993)
    - Der hungrige Fuchs **bemerkte** die fette Henne.
    - Den hungrigen **Fuchs** **bemerkte** die fette Henne.
  - Konieczny (1996)
    - Der hungrige Fuchs **bemerkte** die fette Henne.
    - Den **hungrigen Fuchs** **bemerkte** die fette Henne.

## Inkrementelle Verarbeitung III

→ *Jedes neue Word wird sofort im gegebenen Kontext interpretiert (wenn möglich)*

*„Immediacy hypothesis“*

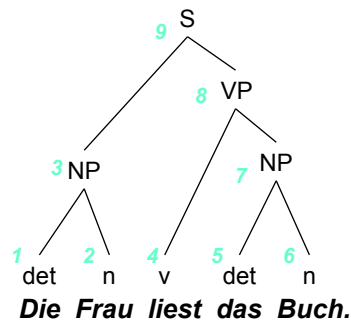
# Bottom-up Parsing

## Phrasenstrukturregeln

S	→	NP VP
NP	→	det n
VP	→	v NP

## Lexikon (deutsch)

det = {der, die, das, ...}  
 n = {Frau, Mann, Polizist, Buch, ...}  
 v = {liest, arbeitet, schreibt, sieht, ...}



**Inkrementell?**

**Nein!**

Psycholinguistik

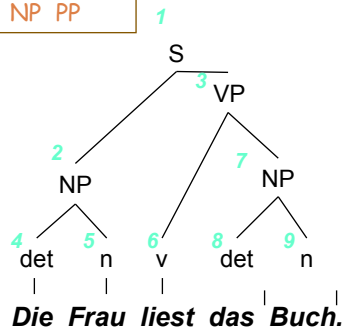
# Top-down Parsing

## Phrasenstrukturregeln

S	→	NP VP
NP	→	det n
VP	→	v NP
NP	→	NP PP

## Lexikon (deutsch)

det = {der, die, das, ...}  
 n = {Frau, Mann, Polizist, Buch, ...}  
 v = {liest, arbeitet, schreibt, sieht, ...}



**Inkrementell?**

**Ja!**

**Aber:**

**arbeitet unabhängig vom Input!**

**Links-Rekursion!**

Psycholinguistik

## Left-corner Parsing (“arc eager”)

### Phrasenstrukturregeln

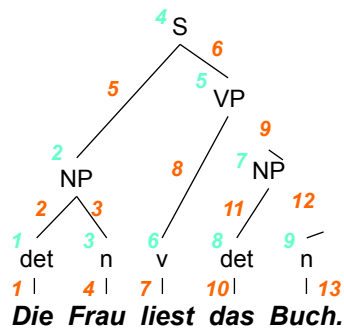
S	→	NP VP
NP	→	det n
VP	→	v NP

### Lexikon (deutsch)

det= {der, die, das, ...}

n = {Frau, Mann, Polizist, Buch, ...}

v = {liest, arbeitet, schreibt, sieht, ...}



Psycholinguistik

Inkrementell?

Fast.

Input-gesteuert?

Ja!

## Left-corner Parsing (“arc hyper-eager”)

### Phrasenstrukturregeln

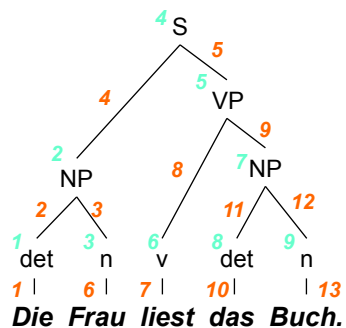
S	→	NP VP
NP	→	det n
VP	→	v NP

### Lexikon (deutsch)

det= {der, die, das, ...}

n = {Frau, Mann, Polizist, Buch, ...}

v = {liest, arbeitet, schreibt, sieht, ...}



Psycholinguistik

Inkrementell?

Ja!

Input-gesteuert?

Ja!

## Parsing und Arbeitsgedächtnis

### PS-Regeln

S → NP VP  
 NP → det n  
 NP → NP RS  
 VP → v  
 RS → relPro NP v  
 ...

### □ “Zentraleinbettung”:

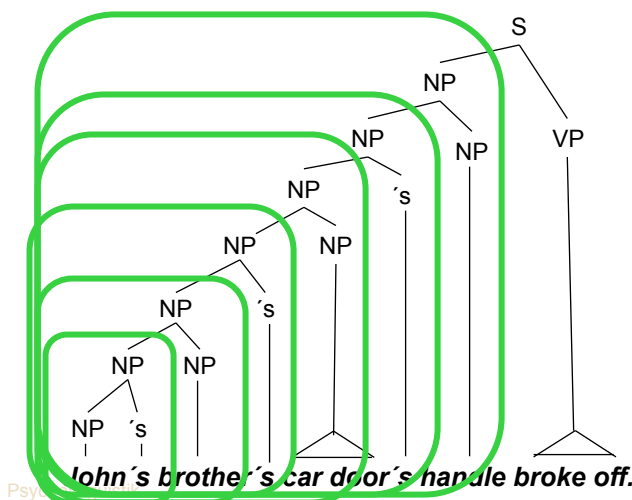
- “Die Maus starb”
- “Die Maus, die die Katze jagte, starb”
- “Die Maus, die die Katze, die der Hund biss, jagte, starb”
- ...

- Keine Ambiguität
- Grammatikalisch
- Warum schwierig, wenn mehr als ein Satz zentral eingebettet ist?

Lars Konieczny

## Left embedding: bottom-up

- [[[John’s brother]’s car door]’s handle] broke off.

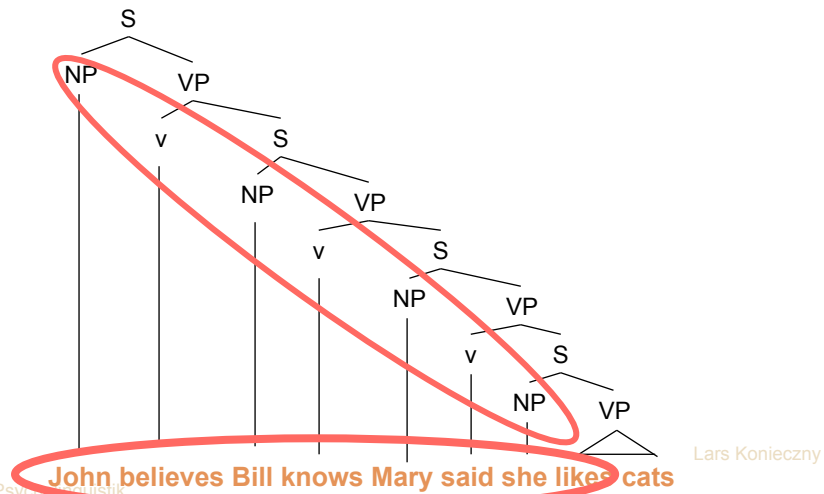


Lars Konieczny



## Right embedding: bottom-up

- John believes [Bill knows [Mary said [she likes cats]]]



## Parsing und Gedächtnisbelastung

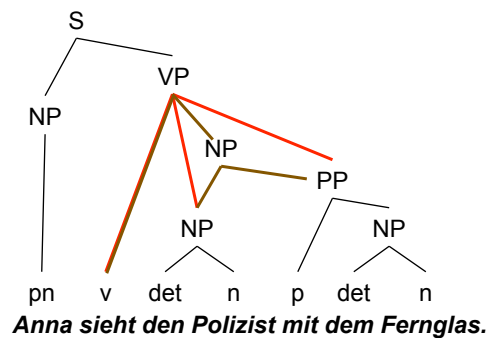
- Links-Einbettung (LE) ist "leicht":
  - [[[John's brother]'s car door]'s handle] broke off.
- Ebenso rechts-Einbettung (RE):
  - John believes [Bill knows [Mary said [she likes cats]]]
- Aber: Zentraleinbettung (ZE) ist hart:
  - [The mouse [the cat [the dog bit] chased] died]
- Top-down: LE: **hart** ZE: **hart** RE: leicht
- Bottom-up: LE: leicht ZE: **hart** RE: **hart**
- Left-corner: LE: leicht ZE: **hart** RE: leicht
- Lit: Abney & Johnson (1991)

## Strukturelle Ambiguität

- Einer Wortsequenz können mehrere Strukturen zugewiesen werden.

### PS-Regeln

NP	→	NP PP
VP	→	v NP (PP)



Psycholinguistik

Lars Konieczny

## Strukturelle Ambiguitäten und Parsing

- Beim inkrementellen Parsing (von links nach rechts) müssen Regeln ausgewählt werden, mit denen jedes neue Wort in die bestehende Struktur integriert werden kann.
- Manchmal können mehrere Regeln (Regelmengen) ausgewählt werden, die jeweils eine wohlgeformte Struktur erzeugen



**Lokale (temporäre) Ambiguität:** Nur eine Analyse stellt sich später als richtig heraus

The boat **floated down the river sank.**



**Globale Ambiguität:** Alle Analysen bleiben bis zum Satzende korrekt.

Anna **sieht den Polizist mit dem Fernglas.**

Psycholinguistik

Lars Konieczny



## Was geschieht bei einer Ambiguität?

- Parallele Verarbeitung?
  - ▣ Mehrere Alternativen werden parallel weiterverfolgt
  - ▣ Wenn eine Alternative später scheitert, wird sie gelöscht.
  - ▣ Problem: Multiple lokale Ambiguitäten lassen den Suchraum anwachsen!
  - ▣ Unbegrenzte Parallelverarbeitung empirisch widerlegt: "Garden-path"-Sätze
  - ▣ Begrenzte Parallelverarbeitung: Nur einige (die besten) Alternativen werden weiterverfolgt.
- Serielle Verarbeitung?
  - ▣ Nur eine Alternative wird weiterverfolgt. Welche? (→ MA, LC)
  - ▣ Was geschieht, wenn diese scheitert? (→ „backtracking“, selektive Reanalyse)
- Unterspezifikation?
  - ▣ Verzögere die Entscheidung
  - ▣ Unterspezifizierte Repräsentationen

Psycholinguistik

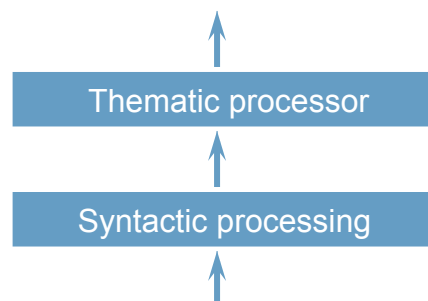
Lars Konieczny

## Ein serielles Modell: "Garden Path Theory"

(Frazier & Rayner, 1982)

### 1. Modular

- ▣ Parser/Syntactic processor arbeitet **autonom**.  
Entscheidungen werden ausschl. auf der Basis struktureller Ökonomie gefällt.
- ▣ Der „thematische Prozessor“ kann allerdings eine Reanalyse einleiten.



Psycholinguistik

Lars Konieczny

## Ein serielles Modell: “Garden Path Theory”

(Frazier & Rayner, 1982)

### 2. Inkrementell

- Binde jedes neue Wort unmittelbar an die bisher aufgebaute Struktur ein (*immediacy hypothesis*)  
→ Left-corner parsing

Psycholinguistik

Lars Konieczny

## Ein serielles Modell: “Garden Path Theory”

(Frazier & Rayner, 1982)

### 3. Ökonomische Auflösung von strukturellen Ambiguitäten

- **Minimal Attachment (MA)**  
*Postuliere keine phrasalen Knoten, die möglicherweise unnötig sind.*
- **Late Closure (LC)**  
*Binde neues Material an die zuletzt verarbeitete Phrase an.*

Psycholinguistik

Lars Konieczny

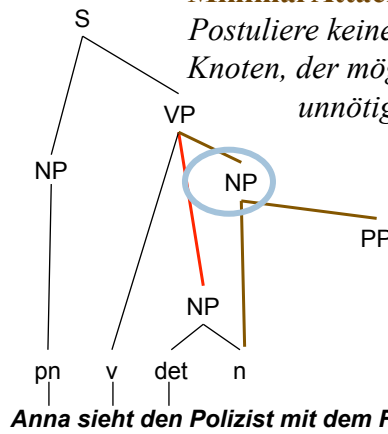
# Minimal Attachment (MA)

## Phrasenstrukturregeln

S	→	NP VP
NP	→	pn
NP	→	det n
NP	→	adj n
NP	→	NP PP
PP	→	p NP
VP	→	v (NP) (PP)
VP	→	v S

## Minimal Attachment (MA)

Postuliere keinen phrasalen Knoten, der möglicherweise unnötig ist!

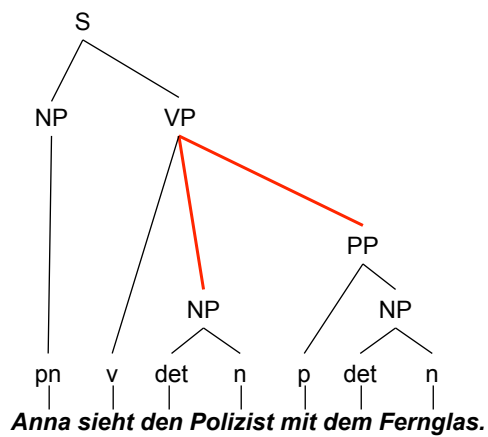


Psycholinguistik

# Minimal Attachment (MA)

## Phrasenstrukturregeln

S	→	NP VP
NP	→	pn
NP	→	det n
NP	→	adj n
NP	→	NP PP
PP	→	p NP
VP	→	v (NP) (PP)
VP	→	v S

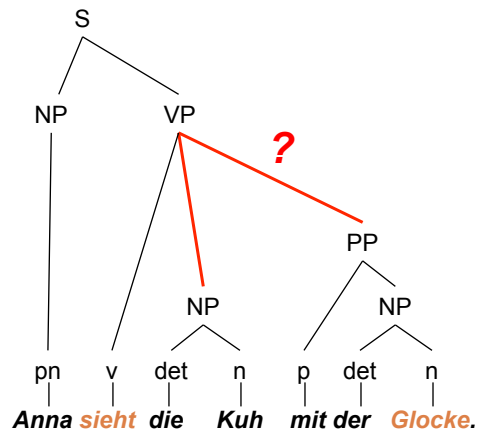


Psycholinguistik

# Minimal Attachment (MA)

## Phrasenstrukturregeln

- S → NP VP
- NP → pn
- NP → det n
- NP → adj n
- NP → NP PP
- PP → p NP
- VP → v (NP) (PP)
- VP → v S

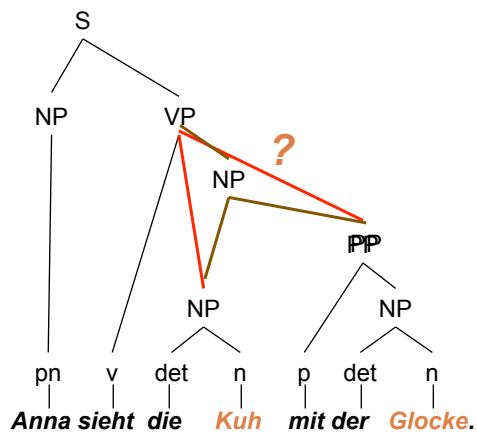


Psycholinguistik

# Minimal Attachment (MA)

## Phrasenstrukturregeln

- S → NP VP
- NP → pn
- NP → det n
- NP → adj n
- NP → NP PP
- PP → p NP
- VP → v (NP) (PP)
- VP → v S



Psycholinguistik

# Late Closure (LC)

## PS-Regeln

S → NP VP (AP)

S' → comp S

NP → pn

NP → det n

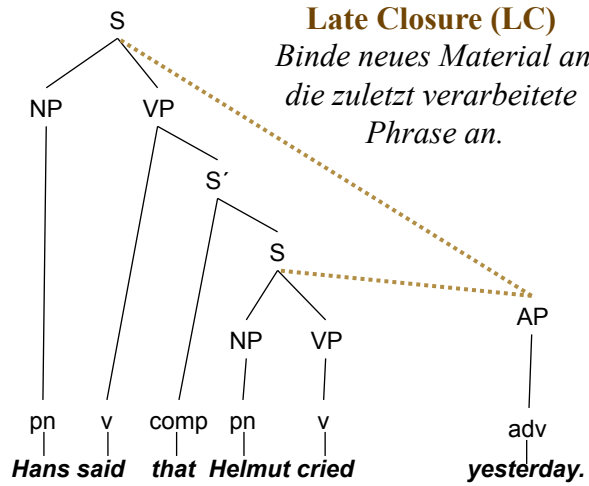
NP → NP PP

PP → p NP

VP → v (NP) (PP)

VP → v S'

AP → adv



Psycholinguistik

# Late Closure (LC)

## PS-Regeln

S → NP VP (AP)

S' → comp S

NP → pn

NP → det n

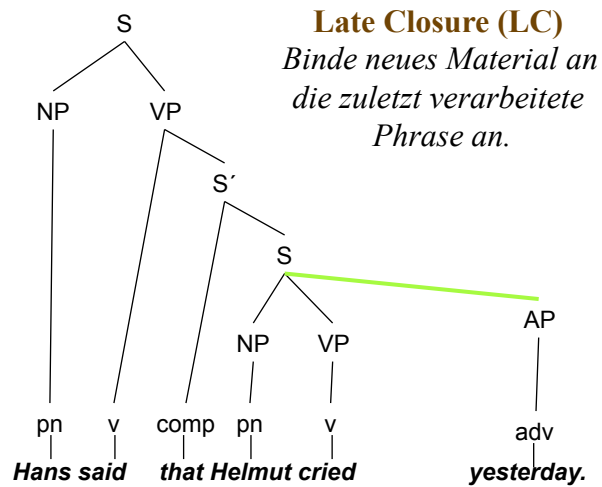
NP → NP PP

PP → p NP

VP → v (NP) (PP)

VP → v S'

AP → adv



Psycholinguistik

## Garden-Path Theory: Evidenz

- VP-Attachment:
  - ▣ “The cop saw the burglar **with the gun**”
- NP/S Attachment:
  - ▣ “The athlete realised **his aims** were too high”
- Object-Attachment:
  - ▣ “After the man left **the shop** closed”
  - ▣ “After the child sneezed **the doctor** arrived”
- Active over Reduced Relative:
  - ▣ “The horse **raced** past the barn fell”
  - ▣ “The man **delivered** the junkmail threw it away”
- Complement vs Relative Clause:
  - ▣ “The doctor told the woman **that** he was in love with to leave”

Psycholinguistik

Lars Konieczny

## Weak vs. strong garden paths

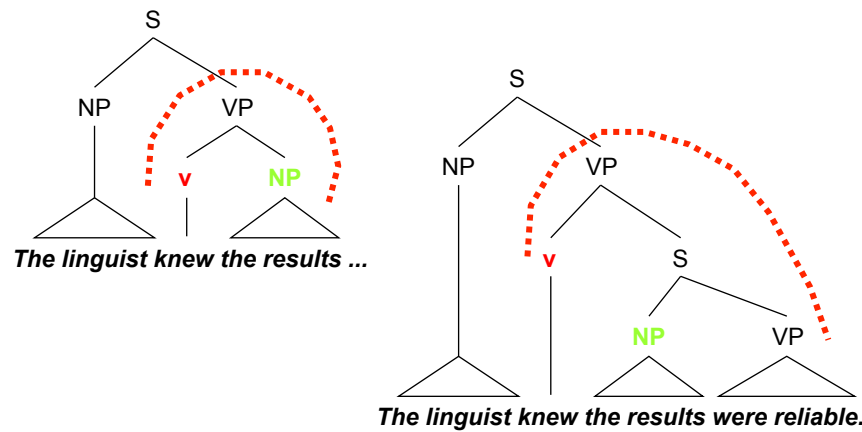
- Observation:
  - ▣ Some garden-paths are easy to recover from, others are hard
    - The linguist knew **the results** were reliable. (easy)
    - After the audience applauded **the actor** left the stage. (hard)
- Puzzle:
  - ▣ Both sentences are similar in that the ambiguous NP
    - wants to be attached as the *Direct Object* of the preceding verb
    - must be re-analysed as the *Subject* of the following clause.
  - ▣ *What makes them different?*

Psycholinguistik

Lars Konieczny

## Weak garden-path

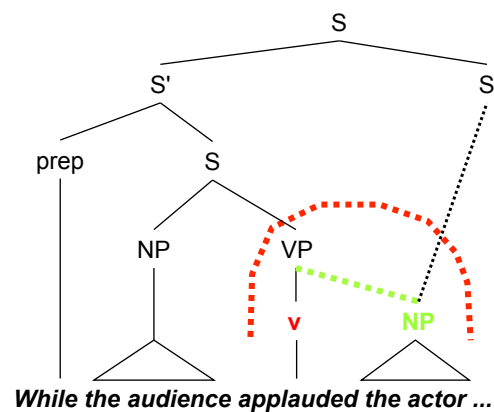
□ Pritchett (1988): Thematic domains



Psycholinguistik

## Severe garden-path

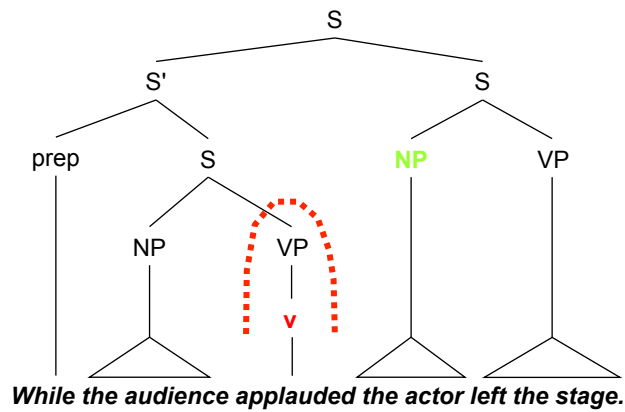
□ Pritchett (1988): Thematic domains



Psycholinguistik

## Severe garden-path

- Pritchett (1988): Thematic domains



Psycholinguistik

## Problems

- Pritchett's model centers around *thematic assigners*
  - ▣ Implies *head-driven parsing*
  - ▣ Therefore fails to account for findings on *incrementality*

Psycholinguistik

Lars Konieczny

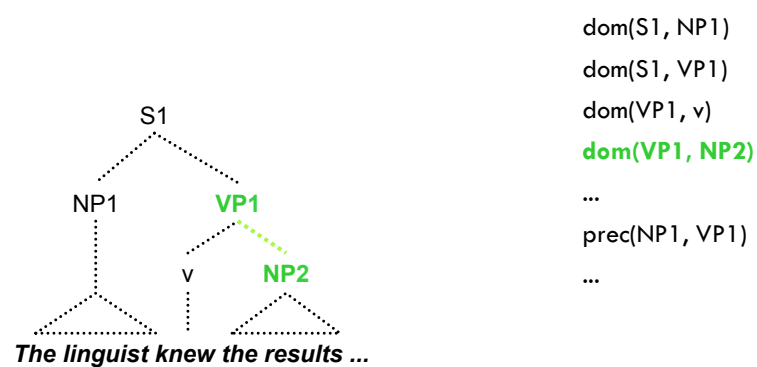


# Underspecification

- References
  - Marcus, Hindle and Fleck (1983): D-theory
  - Weinberg (1994): Minimal Commitment
  - Gorrell (1995), Sturt & Crocker (1998): Structural Determinism
- Idea: Underspecified tree descriptions (instead of actual trees), using predicates
  - Dominance (vs immediate dominance)
    - $\text{dom}(x, y)$  means: Node  $y$  is somewhere below  $x$
  - Precedence
    - $\text{prec}(x, y)$  means:  
Node  $x$  (and its daughters) precedes  $y$  (and its daughters)
  - Important: A node cannot be both *dominated* **and** *preceded* by another node

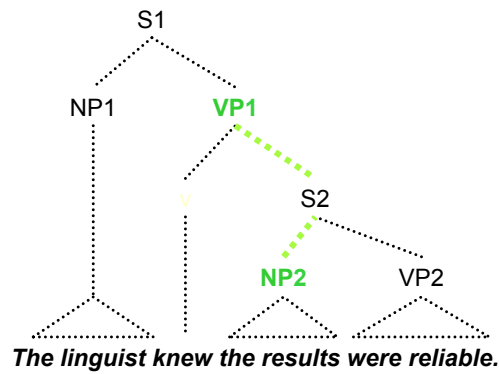
Psycholinguistik

# Structural Determinism (SD)



Psycholinguistik

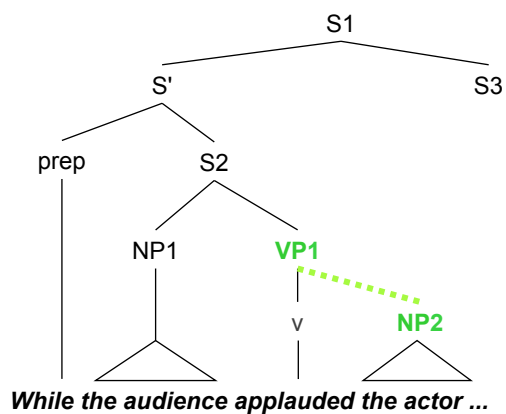
## SD: Easy "reanalysis"



dom(S1, NP1)  
 dom(S1, VP1)  
 dom(VP1, v)  
**dom(VP1, NP2)**  
 dom(VP1, S2)  
 dom(S2, NP2)  
 dom(S2, VP2)  
 ...  
 prec(NP1, VP1)  
 ...

Psycholinguistik

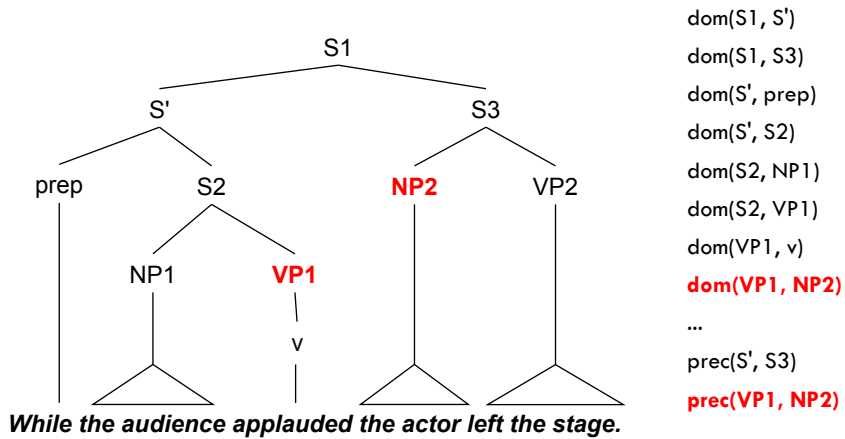
## SD: Severe garden-path



dom(S1, S')  
 dom(S1, S3)  
 dom(S', prep)  
 dom(S', S2)  
 dom(S2, NP1)  
 dom(S2, VP1)  
 dom(VP1, v)  
**dom(VP1, NP2)**  
 ...  
 prec(S', S3)  
 ...

Psycholinguistik

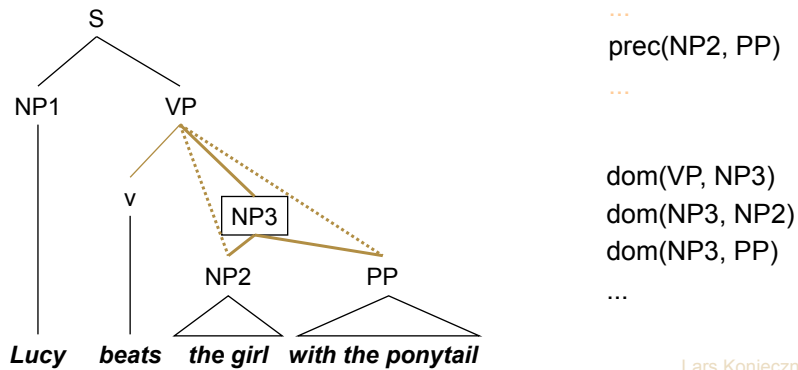
## SD: Severe garden-path



Psycholinguistik

## SD: PP-attachment

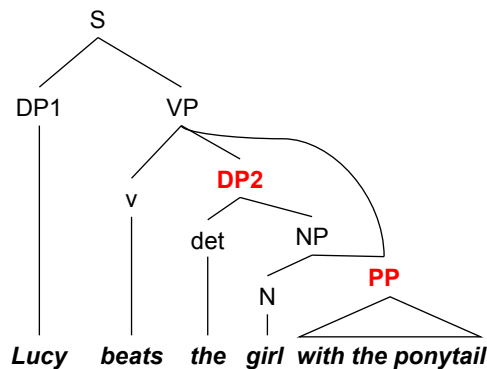
- *Lucy beats the girl with the ponytail*
  - ▣ Mild GP
  - ▣ No problem, when a simple grammar is used



Lars Konieczny

## SD: PP-attachment

- BUT:  
With a more realistic grammar, SD predicts a severe garden-path!



...  
dom(VP, DP2)  
dom(VP, PP)

...  
**prec(DP2, PP)**  
...

dom(DP2, NP)  
dom(NP, PP)  
**dom(DP2, PP)**  
...

Psycholinguistik

Lars Konieczny

## A limited parallel model

(Gibson, 1991)

- Multiple structures are built in parallel and *weighed*
  - ▣ processing load units, PLUs
- When the cost difference (in PLUs) between two constructions exceeds a constant P the worse alternative is discarded.
- Properties of constructions are derived from the *Theta criterion* (Chomsky, 1981):
  - ▣ Each argument bears one and only one theta-role and each theta role is assigned to one and only one argument.

Psycholinguistik

Lars Konieczny

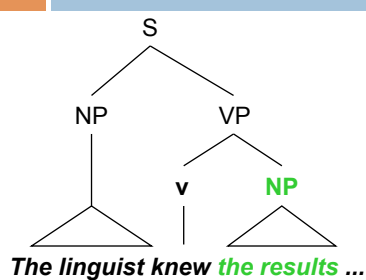
## Processing Load Units (PLUs)

- Property of thematic reception
  - ▣ “Associate a load of X(tr) PLUs to each constituent that is in a position to receive a theta role in some coexisting structure, but whose theta assigner is not unambiguously identifiable in the structure in question”.
- Property of Lexical Requirement
  - ▣ “Associate a load of X(lr) PLUs to each lexical requirement that is obligatory in some coexisting structure, but ... [there is] ... no thematic element in the structure in question.”

Psycholinguistik

Lars Konieczny

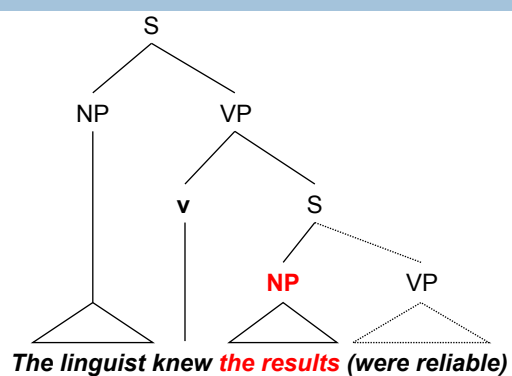
## Weak garden-path



0 PLUs:

Every lexical requirement is satisfied (0 PLU)  
 Every NP gets a theta-role (0 PLU)

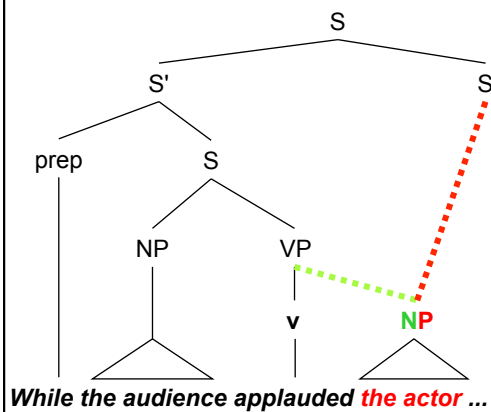
Psycholinguistik



1 PLU:

Every lexical requirement is satisfied (0 PLU)  
 But the NP doesn't receive a theta-role yet! (1 PLU)

## Severe garden-path



### 0 PLUs for **VP-attachment**:

Every lexical requirement is satisfied (0 PLU)

Every NP gets a theta-role (0 PLU)

### 2 PLU for **S-attachment**:

The lexical requirement of *applauded* is not satisfied (1 PLU)

The NP doesn't receive a theta-role yet! (1 PLU)

**1 PLU < P < 2 PLUs.**

Psycholinguistik

## Probleme

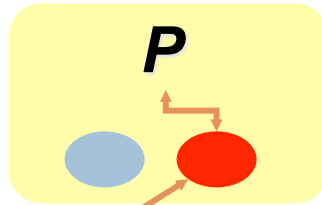
- Empirisch zu findende Verarbeitungsschwierigkeiten bei unbewussten Holzwegeffekten sind im Rahmen des parallelen Modells nur schwer zu erklären.
- Die Menge der frei schätzbaren Parameter des Modells (P, die Zuordnung von PLUs zu den verschiedenen Prinzipien etc.) reduziert den theoretischen Wert.
- Die Unterscheidung zwischen bewussten and unbewussten Holzwegeffekten erscheint oft willkürlich (und theoriegeleitet).

Psycholinguistik

Lars Konieczny

## Referential Success (Crain & Steedman, 1985)

- A psychologist was counselling **a married couple**. One of them was nice to him, but the other one was fighting with him.



The psychologist told the woman that he was having trouble with

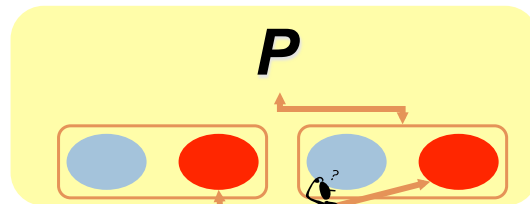
- a. her husband.
- b. to leave her husband.



Lars Konieczny

## Referential Success (Crain & Steedman, 1985)

- A psychologist was counselling **two married couples**. One of them was nice to him, but the other one was fighting with him.



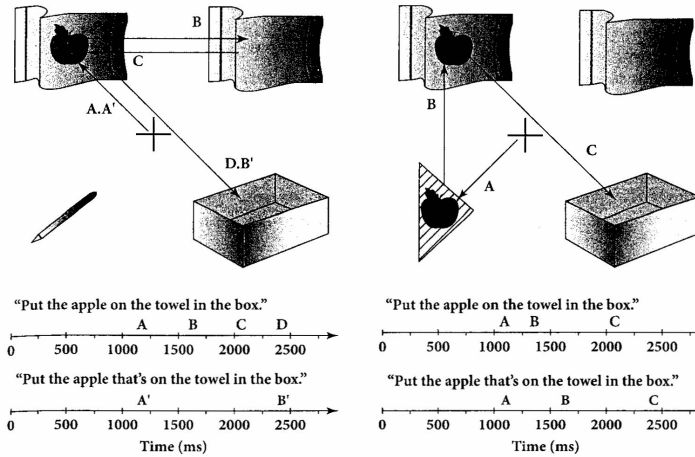
The psychologist told the woman that he was having trouble with

- a. her husband.
- b. to leave her husband.



Lars Konieczny

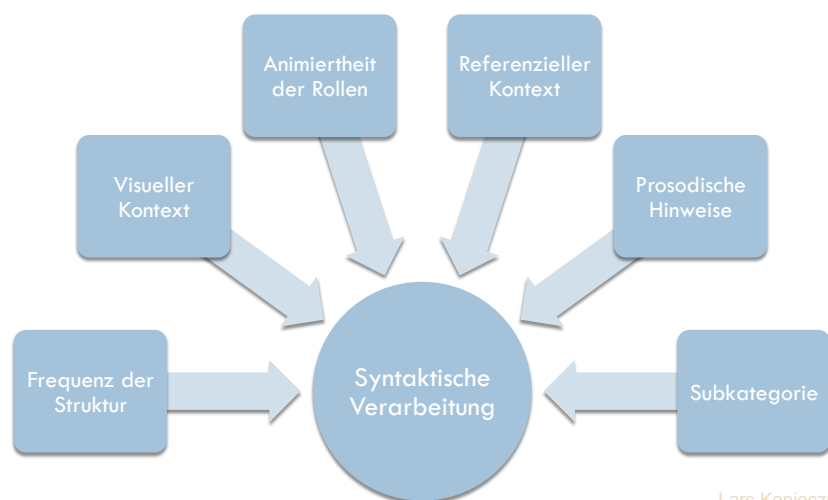
# Visual World Paradigm (Tanenhaus et al., 1995)



Sprache und Kognition: Satzverarbeitung

Lars Konieczny

# Ein interaktiver Ansatz



Lars Konieczny