

Von Sätzen und ihren Bedeutungen – und wie Computer sie berechnen

Studium Generale
Vortragsreihe “Fenster zur Forschung”

Laura Kallmeyer
Universität Tübingen

Tübingen, 11. Juli 2007

Inhalt

1. Thema und Zielsetzung des Projekts
2. Syntax und Semantik
 - (a) Was ist Bedeutung?
 - (b) Wie berechnet man Bedeutung?
3. Automatische Syntaxanalyse
 - (a) Baumgrammatiken
 - (b) Deutsche Wortstellung
 - (c) Multikomponenten
4. Automatische Berechnung der Semantik
5. Zusammenfassung
6. Demonstration am Englischen

Thema und Zielsetzung des Projekts (1)

Thema: Eine lexikalisierte Baumgrammatik für ein Fragment des Deutschen unter Einbeziehung von Syntax und Semantik

Fachgebiet: Computerlinguistik

Emmy-Noether Nachwuchsgruppe, gefördert von der DFG.

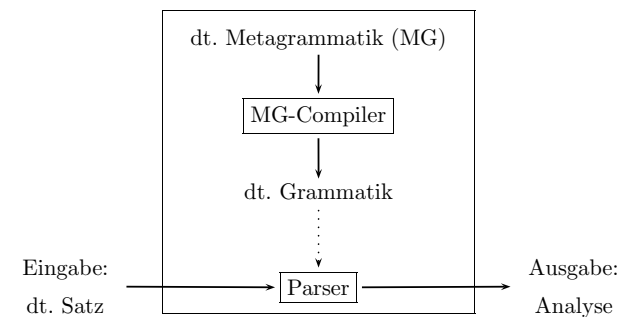
Laufzeit: Juli 2005 – September 2009.

Mitarbeiter: PD Dr. Laura Kallmeyer, Timm Lichte, Wolfgang Maier, Yannick Parmentier

Ziel des Projekts: Automatische Analyse der Syntax und Semantik deutscher Sätze

Thema und Zielsetzung des Projekts (2)

Komponenten:



Syntax und Semantik: Was ist Bedeutung? (1)

(1) Peter isst einen Apfel

Frage: Was bedeutet dieser Satz?

Wahrheitsbedingensemantik: Um diesen Satz zu verstehen, muss man verstehen, was gegeben sein muss, damit dieser Satz wahr wird.

Die Bedeutung eines Satzes beschreibt man mithilfe eines logischen Ausdrucks.

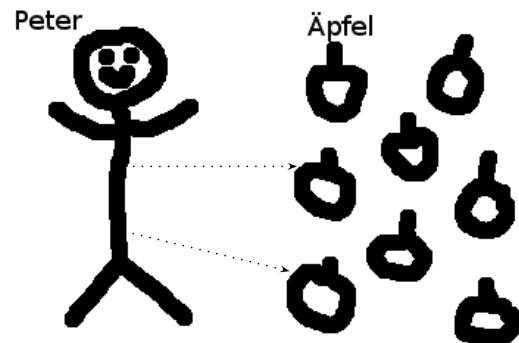
Syntax und Semantik: Was ist Bedeutung? (3)

Bedeutung von (1) *Peter isst einen Apfel*:

Es gibt ein eindeutig bekanntes Individuum namens Peter und es gibt ein Element x , so dass das unäre Prädikat *Apfel* auf x zutrifft und so dass die binäre Relation *essen* zwischen Peter und x besteht.

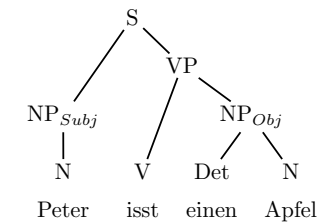
Entsprechender logischer Ausdruck:

$$\exists x[\text{Apfel}'(x) \wedge \text{essen}'(\text{Peter}', x)]$$

Syntax und Semantik: Was ist Bedeutung? (2)**Syntax und Semantik: Berechnung von Bedeutung (1)**

Frage: Wie kann man die Bedeutung berechnen?

1. Schritt: Analyse der Struktur



Syntaktische Analysen werden durch Bäume repräsentiert.

Syntax und Semantik: Berechnung von Bedeutung (2)

2. Schritt: Bedeutungsbeitrag der einzelnen Wörter

- Peter: $Peter'$
- isst: $essen'(\dots, \dots)$
- Apfel: $Apfel'(\dots)$
- einen: $\exists x[\dots \wedge \dots]$

Automatische Syntaxanalyse: Baumgrammatiken (1)

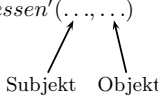
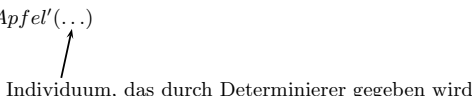
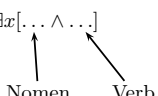
Zentrale Frage: Wie muss die Grammatik aussehen, damit man für sie ein Analyseprogramm (einen Parser) schreiben kann?

Mit anderen Worten: Welchen Formalismus wählt man für die Grammatik?

Ein Grammatikformalismus ist eine mathematisch definierte Spezifikationsprache, die es erlaubt, Grammatiken aufzuschreiben, die wiederum Sprachen beschreiben.

Syntax und Semantik: Berechnung von Bedeutung (3)

3. Schritt: Berechnung der Bedeutung aufgrund der Struktur

- $essen'(\dots, \dots)$

- $Apfel'(\dots)$

- $\exists x[\dots \wedge \dots]$


Ergibt $\exists x[Apfel'(x) \wedge essen'(Peter', x)]$

Automatische Syntaxanalyse: Baumgrammatiken (2)

Ziel: Der Formalismus sollte

- ausdrucksstark genug sein, um eine Beschreibung aller natürlichsprachlichen Phänomene zu erlauben;
- eingeschränkt genug sein, um zu garantieren, dass der Parser in einer akzeptablen Zeit ein Ergebnis liefert.

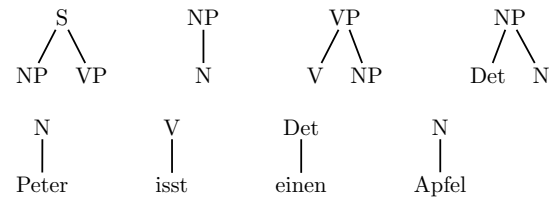
Automatische Syntaxanalyse: Baumgrammatiken (3)

Bsp.: Eine **kontextfreie Grammatik** ist eine Menge von Bäumen der Höhe 1. Die Menge der von ihr erzeugten syntaktischen Strukturen enthält alle Bäume,

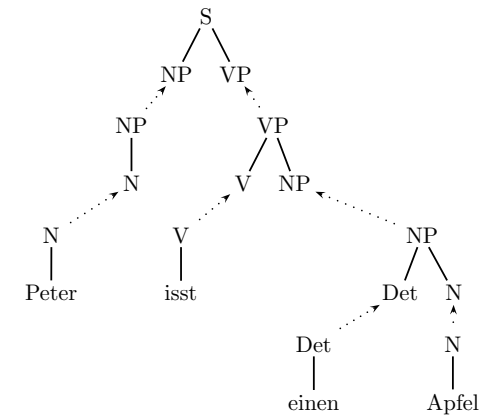
- die ausgehend von einem Baum mit Wurzelsymbol S durch sukzessive Substitution (Ersetzen von Blättern durch neue passende Bäume aus der Grammatik) gewonnen werden können, und
- die nur terminale Blätter enthalten.

Automatische Syntaxanalyse: Baumgrammatiken (4)

Eine kontextfreie Grammatik:



(Üblichere Schreibweise: $S \rightarrow NP VP$, $NP \rightarrow N$, $VP \rightarrow V NP$, ...)

Automatische Syntaxanalyse: Baumgrammatiken (5)**Automatische Syntaxanalyse: Baumgrammatiken (6)**

Kontextfreie Grammatiken reichen nicht aus, um natürliche Sprachen zu beschreiben.

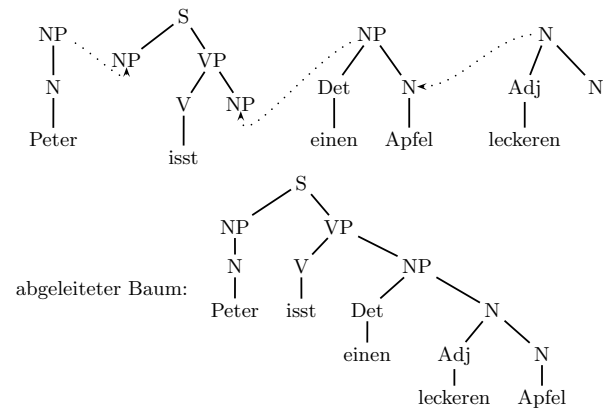
Erweiterung:

- größere Bäume in der Grammatik;
- neben Substitution (Ersetzen von Blättern durch neue Bäume) auch Adjunktion (Ersetzen von internen Knoten durch neue Bäume).

Baumadjunktionsgrammatiken

Automatische Syntaxanalyse: Baumgrammatiken (7)

(2) Peter isst einen leckeren Apfel

**Automatische Syntaxanalyse: Deutsche Wortstellung (2)**

Es gibt sogar Verben, die erlauben, die eigenen Argumente und die eines eingebetteten Infinitivs beliebig zu permutieren.

- (8) Tim hat seiner Mutter das Fahrrad zu reparieren versprochen
- (9) Tim hat das Fahrrad seiner Mutter zu reparieren versprochen
- (10) Das Fahrrad hat Tim seiner Mutter zu reparieren versprochen
- (11) Das Fahrrad hat seiner Mutter Tim zu reparieren versprochen
- (12) Seiner Mutter hat Tim das Fahrrad zu reparieren versprochen
- (13) Seiner Mutter hat das Fahrrad Tim zu reparieren versprochen

Automatische Syntaxanalyse: Deutsche Wortstellung (1)

Baumadjunktionsgrammatiken sind ausdrucksstärker als kontextfreie Grammatiken, aber **nicht ausdrucksstark genug für natürliche Sprache**.

Problem: Wortstellungsvariation im Deutschen.

- (3) Peter isst einen Apfel
- (4) Einen Apfel isst Peter
- (5) Der Junge gibt dem Mädchen einen Apfel
- (6) Der Junge gibt einen Apfel dem Mädchen
- (7) Einen Apfel gibt der Junge dem Mädchen

...

⇒ Einzige Einschränkung: genau ein Element steht vor dem Verb.

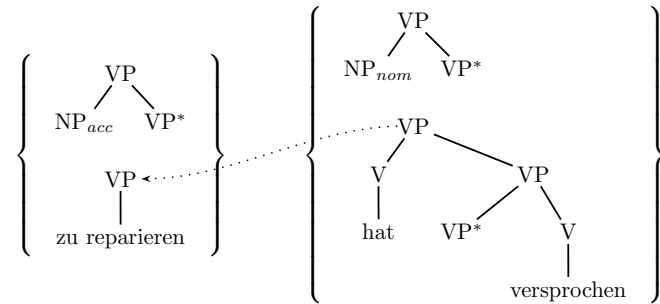
Automatische Syntaxanalyse: Multikomponenten (1)

Nimmt man an, dass der Lexikoneintrag (Baum) eines Verbs Blätter für dessen Argumente enthält, so können diese Strukturen mit Baumadjunktionsgrammatiken nicht beschrieben werden.

Erweiterung: Aufbrechen der Bäume in mehrere Teilbäume (**Multikomponenten**), die im Laufe der Ableitung durch Adjunktion zusammengefügt werden müssen.

Automatische Syntaxanalyse: Multikomponenten (2)

- (14) a. Peter hat das Fahrrad zu reparieren versprochen
- b. Das Fahrrad hat Peter zu reparieren versprochen

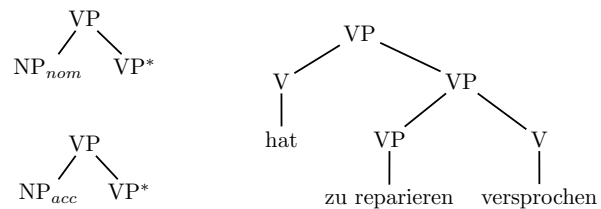


Automatische Berechnung der Semantik (1)

Idee:

- Jedem Baum in der Grammatik ist ein logischer Ausdruck zugeordnet.
- Dieser Ausdruck kann Variablen für noch zu füllende Argumente enthalten.
- Diese Variablen können bestimmten Knoten im Baum zugeordnet sein und werden beim Verknüpfen der Bäume mit entsprechenden Werten identifiziert.

Automatische Syntaxanalyse: Multikomponenten (3)

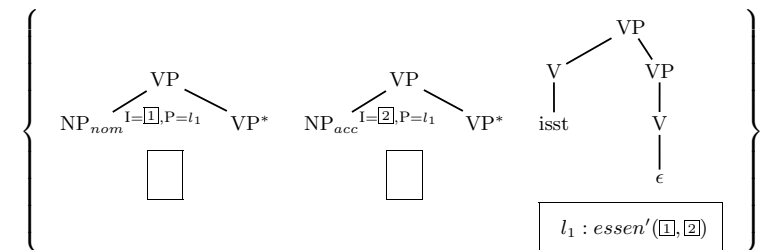


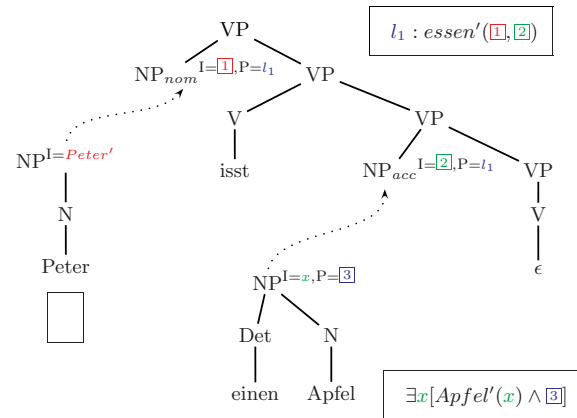
Die beiden Argumente können an jeden der VP-Knoten im verbalen Baum adjungiert werden.

Automatische Berechnung der Semantik (2)

- (15) Peter isst einen Apfel

Lexikoneintrag für *isst*:



Automatische Berechnung der Semantik (3)

Ergibt $\exists x[Apfel'(x) \wedge essen'(Peter', x)]$.

Zusammenfassung (2)

Noch nicht abgeschlossene Teile:

- Parser (Analyseprogramm) für die Syntax; für die Konstruktion der Bedeutung gibt es schon ein Tool (LORIA, Nancy)
- Implementierung der Metagrammatik; bisher stehen einige Klassen für Verben in der Grammatik, allerdings ohne Semantik

Zusammenfassung (1)

Projektziel: Syntax- und Semantikanalyse deutscher Sätze

Bisher erreichte Ergebnisse:

- Grammatikformalismus (Syntax) und Analysen der wichtigsten Wortstellungsvariationen
- Architektur zur Berechnung der Bedeutung (Zusammenarbeit mit University of Pennsylvania)
- Metaformat für diesen Formalismus und Compiler, der eine Metagrammatik in die entsprechende Grammatik umwandelt (Zusammenarbeit mit LORIA, Nancy)

Zusammenfassung (3)

Das Projekt ist ein Grundlagenforschungsprojekt.

Unmittelbarer Erkenntnisgewinn:

- Wissen über die Komplexität deutscher Syntax und damit über die mindestnotwendige Komplexität von Grammatikformalismen, die deutsche Syntax beschreiben.
- Neue Techniken zur Berechnung von Bedeutung auf der syntaktischen Struktur, die eine computationale Verarbeitung erlauben.
- Konkrete Algorithmen, die die Syntax- und Semantikanalyse durchführen.

Zusammenfassung (4)

Mögliche Anwendungen der im Projekt gewonnenen Erkenntnisse:

- Maschinelle Übersetzung.
- Dialogsysteme.
- Automatische Extraktion von Information aus geschriebenen Texten.

Demonstration am Englischen

Kleine Spielzeuggrammatik für das Englische
(Baumadjunktionsgrammatik).

- Von der Metagrammatik zur Grammatik: XMG.
- Parsing (Syntax): llp2.
- Parsing und Konstruktion der Semantik: SemConst.