

Stochastische Verfahren für Semantik

Beispiel: R. Bonnema, R. Bod, R. Scha: A DOP Model for Semantic Interpretation, ACL '97

<http://earth.let.uva.nl/bonnema/acl-DOP-sem/acl-DOP-sem.html>

Idee: *Data-oriented Processing* auf Semantik erweitern

1. Was ist DOP?

Korpus mit syntaktisch annotierten, d.h. gelabelten Bäumen, als Operation darauf Komposition \circ . Neue Eingabe wird analysiert, indem vorhandene Teilbäume komponiert werden.

Model: Einschätzung der W'keit, dass ein Teilbaum t an einer bestimmten Stelle eingesetzt wird: Anzahl der Vorkommen von t im Korpus / Anzahl aller Teilbäume mit dem gleichen Wurzellabel:

$$P(t) = |t| / \sum_{t': R(t')=R(t)} |t'|$$

W'keit einer Ableitung:

$$P(t_1 \circ \dots \circ t_n) = \prod_i P(t_i)$$

W'keit eines Baumes, t_{id} der i -te Teilbaum in der Ableitung d von Baum T :

$$P(T) = \sum_d \prod_i P(t_{id})$$

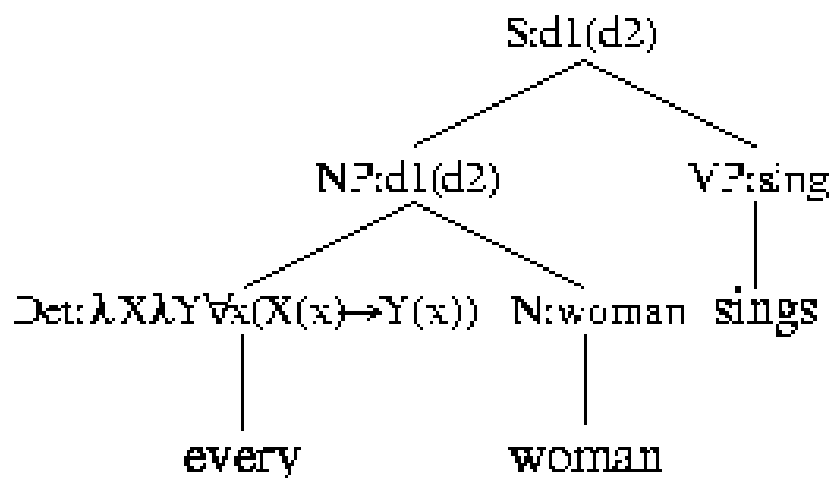
2. Erweiterung um semantische Annotationen (weiterer Semantik-Formalismus: Extensionale Typentheorie (Prädikatenlogik höherer Ordnung, Typen, λ -Abstraktion)

Jeder syntaktisch annotierte Knoten wird jetzt ebenfalls semantisch annotiert:

(pre-)lexikalische mit typen-logischer Formel, die die Bedeutung repräsentiert.

nicht-lexikalische mit Formel-Schema, mit dem man die Formeln der Tochterknoten zusammensetzen kann.

Problem: kompositionale Semantik angenommen



Modifikation der Komposition bzw. Extraktion:

Wenn ein Teilbaum extrahiert wird, wird das entstehende Blatt im Oberbaum mit einer Unifikationsvariablen gelabelt. Bei Komposition wird sie mit der Formel am eingefügten Wurzelknoten unifiziert.

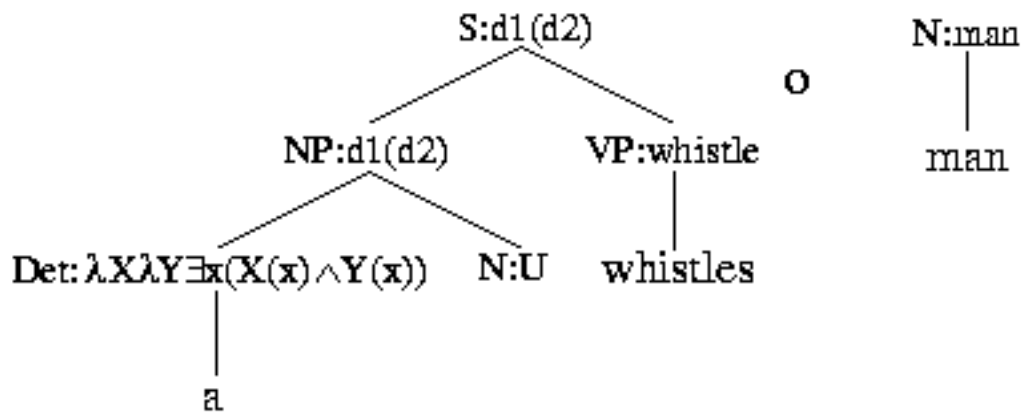
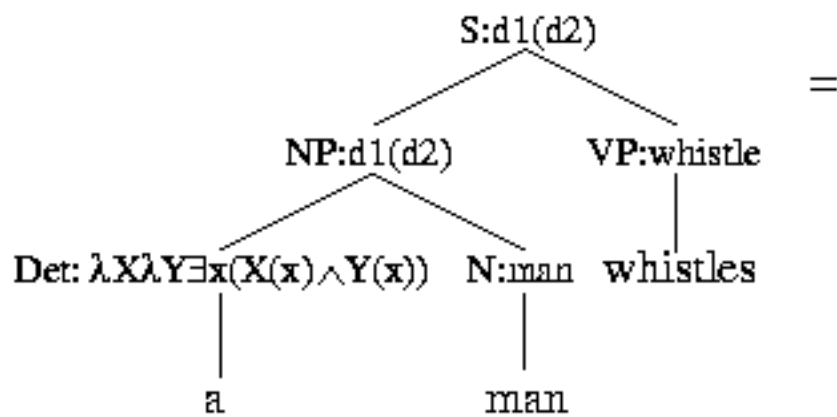


Figure: Decomposing a tree into subtrees with unification variables.

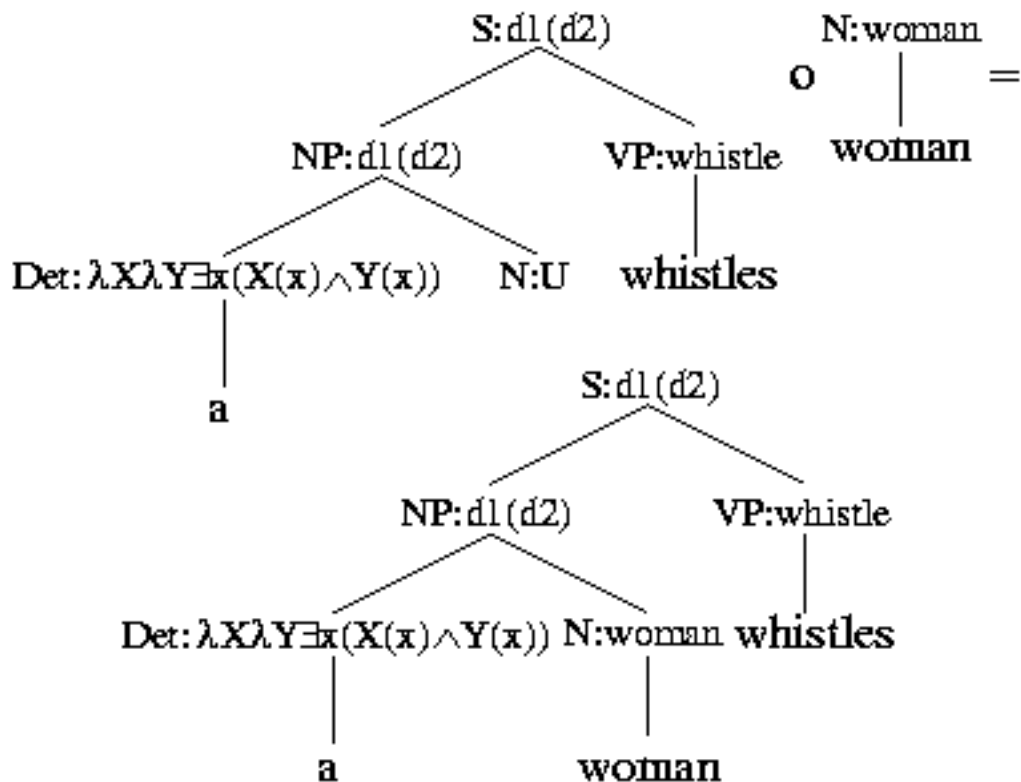


Figure: Generating an analysis for "A woman whistles".

gleich dem syntaktischen:

$$P(t) = |t| / \sum_{t': R(t')=R(t)} |t'|$$

W'keit einer Ableitung:

$$P(t_1 \circ \dots \circ t_n) = \prod_i P(t_i)$$

$$P(T) = \sum_d \prod_i P(t_{id})$$

Interpretation eines Strings: Semantische Annotation der Wurzel des Parsebaums.

W'keit einer Interpretation, p diejenigen Parses, die eine äquivalente Interpretation liefern:

$$P(I) = \sum_p \sum_d \prod_i P(t_{idp})$$

Systemevaluierungen

Bsp:Text Retrieval: entspricht ein gefundenes Dokument der Anfrage?

Eventualitätstabelle zur Klassifikation von Beobachtungen ($a + b + c + d = n$, a echte positive, A positive Entscheidung; n echte negative, N negative Entscheidung):

	'Ja' ist richtig	'Nein' ist richtig
gefunden	$a, P(A a), \text{Treffer}$	$b, P(A n), \text{false pos.}$
nicht gef.	$c, P(N a), \text{false neg.}$	$d, P(N n), \text{true neg.}$

Weitere Masse

$$\text{Precision} = \frac{a}{a + b}$$

$$\text{Recall} = \frac{a}{a + c}$$

$$\text{Fallout} = \frac{b}{b + d}$$