

Einführung in die Computerlinguistik

Tutorium

Seminar für Computerlinguistik

Dozentin: Prof. Anette Frank

20.01.2009

- 1 Merkmalsstrukturen
- 2 Baumbanken und PCFGs
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik
- 4 Prädikatenlogik
 - Syntax
 - Semantik
- 5 Diskurssemantik

- 1 Merkmalsstrukturen
- 2 Baumbanken und PCFGs
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik
- 4 Prädikatenlogik
 - Syntax
 - Semantik
- 5 Diskurssemantik

Aufgabe: Unifikation von MS

Sind die folgenden Attribut-Wert-Paar miteinander unifizierbar (jeweils paarweise)?

1

$$\left[\begin{array}{l} \text{SUBJ} \\ \left[\begin{array}{ll} \text{GEN} & \text{mask} \\ \text{NUM} & \text{sg} \end{array} \right] \end{array} \right]$$

2

$$\left[\begin{array}{l} \text{SUBJ} \\ \text{OBJ} \end{array} \left[\begin{array}{ll} \text{KAS} & \text{nom} \\ \text{KAS} & \text{akk} \end{array} \right] \right]$$

3

$$[\text{AGR} \left[\text{SUBJ} \left[\text{Gen} \text{ fem} \right] \right]]$$

4

$$\left[\begin{array}{l} \text{AGR} \\ \text{OBJ} \end{array} \left[\begin{array}{l} \text{SUBJ} \\ \left[\begin{array}{ll} \text{KAS} & \text{nom} \\ & \text{hund} \end{array} \right] \end{array} \right] \right]$$

Lösung: Unifikation

- S_1 und S_2

$$\left[\begin{array}{l} \text{SUBJ} \\ \text{OBJ} \end{array} \left[\begin{array}{l} \text{GEN} \text{ mask} \\ \text{NUM} \text{ sg} \\ \text{KAS} \text{ nom} \\ \text{KAS} \text{ akk} \end{array} \right] \right]$$

- S_1 und S_3

$$\left[\begin{array}{l} \text{SUBJ} \\ \text{AGR} \end{array} \left[\begin{array}{l} \text{GEN} \text{ mask} \\ \text{NUM} \text{ sg} \\ \text{SUBJ} \left[\text{Gen} \text{ fem} \right] \end{array} \right] \right]$$

- S_1 und S_4

$$\left[\begin{array}{l} \text{SUBJ} \\ \text{AGR} \\ \text{OBJ} \end{array} \left[\begin{array}{l} \text{GEN} \text{ mask} \\ \text{NUM} \text{ sg} \\ \text{SUBJ} \left[\text{KAS} \text{ nom} \right] \\ \text{hund} \end{array} \right] \right]$$

...Lösung: Unifikation

- S_2 und S_3

$$\left[\begin{array}{l} \text{SUBJ} \quad [KAS \quad nom] \\ \text{OBJ} \quad [KAS \quad akk] \\ \text{AGR} \quad [SUBJ \quad [Gen \quad fem]] \end{array} \right]$$

- S_2 und $S_4 = \perp$

- S_3 und S_4

$$\left[\begin{array}{l} \text{AGR} \quad [SUBJ \quad [Gen \quad fem] \\ \text{OBJ} \quad [KAS \quad nom]] \\ \text{AGR} \quad [SUBJ \quad [Gen \quad fem] \\ \text{OBJ} \quad hund] \end{array} \right]$$

- 1 Merkmalsstrukturen
- 2 **Baumbanken und PCFGs**
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik
- 4 Prädikatenlogik
 - Syntax
 - Semantik
- 5 Diskurssemantik

Aufgabe: Baumbanken und PCFGs

- 1 Erstelle zu folgenden Sätzen eine Baumbank:
 - 1 Die kleinen Narren und Närrinnen lachen.
 - 2 Die kleine Frau und Ulla kaufen ein kleines Auto.
 - 3 Die alten Männer und Frauen machen eine Kaffeefahrt.
- 2 Erstelle auf Basis der Baumbank eine probabilistische kontextfreie Grammatik (PCFG).
(In der Lösung sind nur die Regeln und die Wahrscheinlichkeiten angegeben.)

Lösung: Baumbank

- Die kleinen Narren und Närrinnen lachen.

```
[S [NP [NP [D die]
           [AP [Adj kleinen]]
           [NP [N Narren]] ]
     [Conj und]
     [NP [N Närrinnen]] ]
  [VP [V lachen ] ] ]
```

- Die alten Männer und Frauen machen eine K-fahrt.

```
[S [NP [D die ]
     [AP [Adj alten]]
     [NP [NP [N Maenner]]
         [Conj und]
         [NP [N Frauen]] ] ] ]
  [VP [V machen]
     [NP [D eine]
         [N Kaffeefahrt]]] ]
```

- Die kleine Frau und Ulla kaufen ein kleines Auto.

```
[S [NP [NP [D die]
           [AP [Adj kleine]]
           [NP [N Frau]] ]
     [Conj und]
     [NP [N Ulla]] ]
  [VP [V kaufen]
     [NP [D ein]
         [AP [Adj kleines]]
         [NP [N Auto]] ]]] ]
```

Lösung: PCFG - $P(\alpha \rightarrow \beta) = (*)$

α	$C(\alpha)$	β	$C(\beta)$	(*)	α	$C(\alpha)$	β	$C(\beta)$	(*)
S	3	NP VP	3	1	N	8	Närrinnen	1	1/8
NP	15	NP Conj NP	3	1/5	N	8	Quatsch	0	0
NP	15	D AP NP	4	4/15	N	8	Frau	1	1/8
NP	15	D N	1	1/15	N	8	Männer	1	1/8
NP	15	N	7	7/15	N	8	Ulla	1	1/8
VP	3	V NP	2	2/3	N	8	Auto	1	1/8
VP	3	V	1	1/3	N	8	Frauen	1	1/8
AP	4	Adj	4	1	N	8	Kaffeefahrt	1	1/8
D	5	ein	1	1/5	N	8	Narren	1	1/8
D	5	eine	1	1/5	Adj	4	kleine	1	1/4
D	5	die	3	3/5	Adj	4	alten	1	1/4
V	3	machen	1	.33	Adj	4	kleines	1	1/4
V	3	kaufen	1	.33	Adj	4	kleinen	1	1/4
V	3	lachen	1	.33	Conj	3	und	3	1

Aufgabe: Evaluierung

- 1 Berechne die Wahrscheinlichkeiten der Lesarten für den Satz *die alten Männer und Ulla kaufen ein Auto*
- 2 Wähle eine Lesart als Goldstandard und evaluiere damit eine andere Analyse mit den Maßen labelled precision und recall.

Lösung: Evaluierung

0 die 1 alten 2 Männer 3 und 4 Ulla 5 kaufen 6 ein 7 Auto 8

Lesart 1 (reference parse):

```
(S (NP (NP (D die)
           (AP (JJ alten))
           (NP (N Männer))))
   (CC und)
   (NP (N Ulla))
   (VP (V kaufen)
        (NP (D ein)
             (N Auto))))))
```

- Konstituenten(rp):
S(0:8), NP(0:5), AP(1:2),
NP(2:5), NP(2:3), NP(4:5),
VP(5:8), NP(6:8)
- #((korr.) Konstituenten(rp)) = 8

Lesart 2 (hypothesis parse):

```
(S (NP (D die)
           (AP (JJ alten))
           (NP (NP (N Männer))
               (CC und)
               (NP (N Ulla))))
   (VP (V kaufen)
        (NP (D ein)
             (N Auto))))))
```

- Konstituenten(hp):
S(0:8), NP(0:5), NP(0:3),
AP(1:2), NP(2:3), NP(4:5),
VP(5:8), NP(6:8)
- #((korr.) Konstituenten(hp)) = 7
#((Konstituenten(hp))) = 8

...Lösung: Evaluierung

- $\#((\text{korr.}) \text{Konstituenten}(rp)) = 8$
 - $\#(\text{korr. Konstituenten}(hp)) = 7$
 - $\#(\text{Konstituenten}(hp)) = 8$
- ⇒ labelled precision = $\frac{(\text{korr. Konst.}(hp))}{(\text{Konst.}(hp))} = \frac{7}{8}$
- ⇒ labelled recall = $\frac{(\text{korr. Konst.}(hp))}{(\text{Konst.}(rp))} = \frac{7}{8}$

- 1 Merkmalsstrukturen
- 2 Baumbanken und PCFGs
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik**
- 4 Prädikatenlogik
 - Syntax
 - Semantik
- 5 Diskurssemantik

Beweisverfahren: Reductio ad Absurdum

Zeige die Allgemeingültigkeit der folgenden Formeln:

- 1 $(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$ (Kettenschluss)
- 2 $(p \rightarrow q) \rightarrow \neg(p \wedge \neg q)$
- 3 $\neg p \rightarrow (p \rightarrow q)$

Lösung: Reductio ad Absurdum

$$\textcircled{1} (A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$$

$((A \rightarrow B) \wedge (B \rightarrow C))$	\rightarrow	$(A \rightarrow C)$	
	0		(1)
1		0	(2)
1	1	1	0
1		0	(4)
1	0		(5)

1. Annahme, dass Formel falsch ist.
 2. Das ist nur der Fall, wenn $(A \rightarrow B) \wedge (B \rightarrow C)$ wahr und $(A \rightarrow C)$ falsch.
 3. Damit müssen $(A \rightarrow B)$ und $(B \rightarrow C)$ wahr sein und A wahr und C falsch sein.
 4. Also ist auch das erste A wahr bzw. das erste C in Formel falsch.
 5. Damit $(A \rightarrow B)$ nach (3) und (4) also wahr sein kann, muss B wahr sein. Bzw. damit $(B \rightarrow C)$ nach (3) und (4) wahr sein kann, muss B falsch sein. Widerspruch!
- Also ist Formel allgemeingültig (Tautologie).

Beweisverfahren: Tableaux-Methode

Zeige die Allgemeingültigkeit der folgenden Formeln:

- 1 $A \wedge B \rightarrow A$
- 2 $(p \rightarrow q) \rightarrow \neg(p \wedge \neg q)$
- 3 $\neg p \rightarrow (p \rightarrow q)$
- 4 $(p \leftrightarrow q) \rightarrow (p \wedge q) \vee (\neg p \wedge \neg q)$
- 5 $p \wedge \neg q \rightarrow \neg(p \rightarrow q)$
- 6 $(p \wedge \neg q) \rightarrow \neg(p \rightarrow q)$

Lösung: Tableaux-Methode

$$6 \quad (p \wedge \neg q) \rightarrow \neg(p \rightarrow q)$$

$$(1) \quad \neg((p \wedge \neg q) \rightarrow \neg(p \rightarrow q))$$

$$(2) \quad (p \wedge \neg q) \quad F_{\rightarrow}(1)$$

$$(3) \quad \neg\neg(p \rightarrow q) \quad F_{\rightarrow}(1)$$

$$(4) \quad p \quad W_{\wedge}(2)$$

$$(5) \quad \neg q \quad W_{\wedge}(2)$$

$$(6) \quad (p \rightarrow q) \quad N(3)$$

$$(7a) \quad \swarrow \quad \neg p \quad (7b) \quad \searrow \quad q \quad W_{\rightarrow}(6)$$

$$(8a) \quad p \quad (8b) \quad \neg q \quad \text{Wiederholung von (4)}$$

⇒ In allen Zweigen ist ein Widerspruch (Formel und ihre Negation sind beide in einem Zweig), d.h. alle Zweige sind geschlossen.

⇒ Tableaux ist geschlossen.

⇒ Formel ist allgemeingültig.

Wahrheitstafeln

Entscheide mittels Wahrheitstafeln, ob folgende Formeln allgemeingültig, kontingent oder kontradiktorisch sind:

- 1 $A \rightarrow (B \rightarrow C) \leftrightarrow A \wedge B \rightarrow C$
- 2 $(p \rightarrow q) \rightarrow \neg(p \wedge \neg q)$
- 3 $\neg p \rightarrow (p \rightarrow q)$
- 4 $(q \rightarrow r) \wedge (p \rightarrow \neg q) \leftrightarrow (q \rightarrow r \wedge \neg p)$
- 5 $p \vee q \leftrightarrow \neg p \wedge \neg q$

- 1 Merkmalsstrukturen
- 2 Baumbanken und PCFGs
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik
- 4 Prädikatenlogik**
 - Syntax
 - Semantik
- 5 Diskurssemantik

Formeln und Terme

Sind die folgenden Ausdrücke prädikatenlogisch wohlgeformte

a) Terme, b) Formeln oder c) nichts von beidem?

a, b, c sind Konstanten, x, y, z Variablen, t, s, r Terme, ϕ, ψ, δ Formeln, P, Q, R, T Relationszeichen (Prädikate) (Stelligkeit ergibt sich aus dem Kontext)

- 1 x bzw. a
- 2 $\forall y(P(x) \vee Q(z))$
- 3 $P(a, b, c) \rightarrow R(a)T(x)$
- 4 $\neg P(a = c, x)$
- 5 $R(t, s, r)$, R 2-stelliges Relationszeichen
- 6 $\exists x$
- 7 $(\phi \vee \psi) \rightarrow \exists x\delta$
- 8 $((\delta \leftrightarrow (\phi \wedge \psi)) \vee \forall x\forall y(R(x, y) \rightarrow \delta))$

Lösung: Formeln und Terme

Sind die folgenden Ausdrücke **prädikatenlogisch** wohlgeformte

a) Terme, b) Formeln oder c) nichts von beidem?

- | | | |
|---|---|------------|
| 1 | x bzw. a | a) bzw. a) |
| 2 | $\forall y(P(x) \vee Q(z))$ | b) |
| 3 | $P(a, b, c) \rightarrow R(a)T(x)$ | c) |
| 4 | $\neg P(a = c, x)$ | c) |
| 5 | $R(t, s, r)$, R 2-stelliges Relationszeichen | c) |
| 6 | $\exists x$ | c)!! |
| 7 | $(\phi \vee \psi) \rightarrow \exists x\delta$ | c) |
| 8 | $((\delta \leftrightarrow (\phi \wedge \psi)) \vee \forall x\forall y(R(x, y) \rightarrow \delta))$ | b) |

Aufgabe: Modelle

Modell = $\langle D, F \rangle$

Sei M ein Modell, dessen Individuenbereich aus

$D = \{go, mc, mi, do, da, tr\}$ besteht. Die Individuen werden durch die Individuenkonstanten *Goofy*, *Mickey*, *Minnie*, *Donald*, *Dagobert*, *Trick* bezeichnet.

- Definiere die Interpretationsfunktion F für die Individuenkonstanten, die Prädikate $P(hund)$, $N(maus)$, $E(ente)$ und die Relationen $H(ist_haustier_von)$, $M(mag)$, $V(ist_verwandt_mit)$

(Nach dem Modell, das von Mickey Maus und Ducktales bekannt ist.)

- Welche der folgenden Aussagen (Formeln) sind wahr im Modell M ?
 - $\forall x(E(x) \rightarrow \neg P(x))$
 - $\forall x \neg \exists y(N(x) \wedge N(y) \rightarrow M(x, y))$
 - $\exists x \exists y(P(x) \wedge N(y) \wedge H(x, y))$
 - $V(Minnie, Trick) \vee \neg E(Goofy)$
 - $\exists x \exists y(P(x) \wedge E(y) \wedge V(x, y))$

- 1 Merkmalsstrukturen
- 2 Baumbanken und PCFGs
 - Baumbanken
 - PCFGs
 - Evaluierung
- 3 Aussagenlogik
- 4 Prädikatenlogik
 - Syntax
 - Semantik
- 5 Diskurssemantik

Wichtige Begriffe

Erkläre die Bedeutung folgender Begriffe:

- Diskurs
- Kohärenz
- Koreferenz
- Koreferenzketten
- Anaphorische Ausdrücke, Antezedenten
- Kohärenzrelation (Diskursrelation): Result, Explanation, Parallel, Elaboration, Occasion