

Merkmalsstrukturen: Äquivalenz, Subsumtion und Unifikation

Lea Frermann

Universität des Saarlandes

May 16, 2011

Probleme mit Phrasenstrukturgrammatiken

① Übergenerierung

- ▶ Eine Phrasenstrukturgrammatik die alle gültigen Sätze eine Sprache beschreibt, wird immer übergenerieren (→ Ausnahmen, Selektive Präferenzen)

② Untergenerierung

- ▶ Fernabhängigkeiten, Kongruenz sind nur schwer mit PSGs zu beschreiben
zB.: This is the book_t which I think you said John recommended_t.

Probleme mit Phrasenstrukturgrammatiken

① Übergenerierung

- ▶ Eine Phrasenstrukturgrammatik die alle gültigen Sätze eine Sprache beschreibt, wird immer übergenerieren (\rightarrow Ausnahmen, Selektive Präferenzen)

② Untergenerierung

- ▶ Fernabhängigkeiten, Kongruenz sind nur schwer mit PSGs zu beschreiben
zB.: This is the book_t which I think you said John recommended_t.

Lösung: Annotation von Grammatiken mit Merkmalstrukturen

Merkmalsstrukturen

- ...sind Mengen von Merkmalen mit zugeordneten Werten

Merkmalsstrukturen

- ...sind Mengen von Merkmalen mit zugeordneten Werten
- Merkmale sind implizit konjunktiv verknüpft

Merkmalsstrukturen

- ...sind Mengen von Merkmalen mit zugeordneten Werten
- Merkmale sind implizit konjunktiv verknüpft
- **Unterspezifikation:** Über ein Merkmal, das nicht in der Struktur auftritt wird keine Aussage getroffen

Merkmalsstrukturen

- ...sind Mengen von Merkmalen mit zugeordneten Werten
- Merkmale sind implizit konjunktiv verknüpft
- **Unterspezifikation:** Über ein Merkmal, das nicht in der Struktur auftritt wird keine Aussage getroffen
- Werte können atomar oder komplex sein (\rightarrow der Wert des Merkmals ist selber eine Merkmalsstruktur)

Merkmalsstrukturen

- ...sind Mengen von Merkmalen mit zugeordneten Werten
- Merkmale sind implizit konjunktiv verknüpft
- **Unterspezifikation:** Über ein Merkmal, das nicht in der Struktur auftritt wird keine Aussage getroffen
- Werte können atomar oder komplex sein (\rightarrow der Wert des Merkmals ist selber eine Merkmalsstruktur)
 - ▶ Dies erlaubt zusammengehörige Merkmale zusammenzufassen, die Merkmalsstruktur zu strukturieren (z.B. können "Kasus", "Numerus" und "Genus" als komplexer Wert des Merkmals "Kongruenz" zusammengefasst werden)

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)
 - ▶ zB. müssen Subjekt und Verb eines Satzes in Numerus übereinstimmen

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)
 - ▶ zB. müssen Subjekt und Verb eines Satzes in Numerus übereinstimmen
- Koreferenzen werden in der Regel durch gleiche Indexe gekennzeichnet

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)
 - ▶ zB. müssen Subjekt und Verb eines Satzes in Numerus übereinstimmen
- Koreferenzen werden in der Regel durch gleiche Indexe gekennzeichnet
- Werte können allerdings auch zufällig den gleichen Wert aufweisen (**Typen-Identität**)

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)
 - ▶ zB. müssen Subjekt und Verb eines Satzes in Numerus übereinstimmen
- Koreferenzen werden in der Regel durch gleiche Indexe gekennzeichnet
- Werte können allerdings auch zufällig den gleichen Wert aufweisen (**Typen-Identität**)
 - ▶ zB. können Obj1 und Obj2 des Verbs “geben” den gleichen Numerus haben

Koreferenz/Pfadäquivalenz vs. zufällige Wert-Übereinstimmung

- Koreferenzen erzwingen die Identität bestimmter Merkmalswerte (**Token-Identität**)
 - ▶ zB. müssen Subjekt und Verb eines Satzes in Numerus übereinstimmen
- Koreferenzen werden in der Regel durch gleiche Indexe gekennzeichnet
- Werte können allerdings auch zufällig den gleichen Wert aufweisen (**Typen-Identität**)
 - ▶ zB. können Obj1 und Obj2 des Verbs “geben” den gleichen Numerus haben
- In diesem Fall werden beide Werte unabhängig voneinander angegeben und nicht weiter gekennzeichnet

Äquivalenz

- Merkmalsstrukturen sind äquivalent, g.d.w. sie die gleichen Informationen enthalten

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)
- Allgemeinere Merkmalsstrukturen subsumieren speziellere Merkmalsstrukturen

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)
- Allgemeinere Merkmalsstrukturen subsumieren speziellere Merkmalsstrukturen
- **Def.:** $t1 \sqsubseteq t2$, g.d.w. die Menge der von $t2$ beschriebenen Elemente eine Teilmenge der von $t1$ beschriebenen Elemente ist.

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)
- Allgemeinere Merkmalsstrukturen subsumieren speziellere Merkmalsstrukturen
- **Def.:** $t1 \sqsubseteq t2$, g.d.w. die Menge der von $t2$ beschriebenen Elemente eine Teilmenge der von $t1$ beschriebenen Elemente ist.
- also: $t1 \sqsubseteq t2$ gdw. $||t1|| \supseteq ||t2||$

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)
- Allgemeinere Merkmalsstrukturen subsumieren speziellere Merkmalsstrukturen
- **Def.:** $t1 \sqsubseteq t2$, g.d.w. die Menge der von $t2$ beschriebenen Elemente eine Teilmenge der von $t1$ beschriebenen Elemente ist.
- also: $t1 \sqsubseteq t2$ gdw. $||t1|| \supseteq ||t2||$
- Das Topelement (T) enthält keine Informationen (subsumiert alle Merkmalsstrukturen)

Subsumtion

- Subsumtion (\sqsubseteq) definiert eine partielle Ordnung über Merkmalsstrukturen, nach Grad der Allgemeinheit (warum partiell?)
- Allgemeinere Merkmalsstrukturen subsumieren speziellere Merkmalsstrukturen
- **Def.:** $t1 \sqsubseteq t2$, g.d.w. die Menge der von $t2$ beschriebenen Elemente eine Teilmenge der von $t1$ beschriebenen Elemente ist.
- also: $t1 \sqsubseteq t2$ gdw. $||t1|| \supseteq ||t2||$
- Das Topoelement (T) enthält keine Informationen (subsumiert alle Merkmalsstrukturen)
- Das Bottomelement (\perp) ist eine inkonsistente Merkmalstruktur (subsumiert keine Merkmalstruktur)

Unifikation

- Unifikation (\sqcup) bildet eine neue Merkmalsstruktur durch konjunktive Verknüpfung zweier Merkmalsstrukturen, wenn die Merkmalsstrukturen kompatibel sind

Unifikation

- Unifikation (\sqcup) bildet eine neue Merkmalsstruktur durch konjunktive Verknüpfung zweier Merkmalsstrukturen, wenn die Merkmalsstrukturen kompatibel sind
- **Def.:** Die Unifikation von t_1 und t_2 ist die allgemeinste Struktur, die sowohl von t_1 als auch von t_2 subsumiert wird.

Unifikation

- Unifikation (\sqcup) bildet eine neue Merkmalsstruktur durch konjunktive Verknüpfung zweier Merkmalsstrukturen, wenn die Merkmalsstrukturen kompatibel sind
- **Def.:** Die Unifikation von t_1 und t_2 ist die allgemeinste Struktur, die sowohl von t_1 als auch von t_2 subsumiert wird.
- also: $\|t_1 \sqcup t_2\| = \|t_1\| \cap \|t_2\|$

Unifikation

- Unifikation (\sqcup) bildet eine neue Merkmalsstruktur durch konjunktive Verknüpfung zweier Merkmalsstrukturen, wenn die Merkmalsstrukturen kompatibel sind
- **Def.:** Die Unifikation von t_1 und t_2 ist die allgemeinste Struktur, die sowohl von t_1 als auch von t_2 subsumiert wird.
- also: $\|t_1 \sqcup t_2\| = \|t_1\| \cap \|t_2\|$
- Das Ergebnis ist die Schnittmenge der Elemente der eingehenden Merkmalsstrukturen

Unifikation

- Unifikation (\sqcup) bildet eine neue Merkmalsstruktur durch konjunktive Verknüpfung zweier Merkmalsstrukturen, wenn die Merkmalsstrukturen kompatibel sind
- **Def.:** Die Unifikation von t_1 und t_2 ist die allgemeinste Struktur, die sowohl von t_1 als auch von t_2 subsumiert wird.
- also: $\|t_1 \sqcup t_2\| = \|t_1\| \cap \|t_2\|$
- Das Ergebnis ist die Schnittmenge der Elemente der eingehenden Merkmalsstrukturen
- \perp , wenn die Merkmalsstrukturen inkompatibel sind

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.
- **Def.:** Die Generalisierung von t_1 und t_2 ist die spezielleste Struktur, die sowohl t_1 als auch t_2 subsumiert.

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.
- **Def.:** Die Generalisierung von t_1 und t_2 ist die speziellste Struktur, die sowohl t_1 als auch t_2 subsumiert.
- also: $\|t_1 \sqcap t_2\| = \|t_1\| \sqcup \|t_2\|$

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.
- **Def.:** Die Generalisierung von t_1 und t_2 ist die spezielleste Struktur, die sowohl t_1 als auch t_2 subsumiert.
- also: $\|t_1 \sqcap t_2\| = \|t_1\| \sqcup \|t_2\|$
- In der resultierenden Menge sind also sowohl t_1 als auch t_2 enthalten

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.
- **Def.:** Die Generalisierung von t_1 und t_2 ist die spezielleste Struktur, die sowohl t_1 als auch t_2 subsumiert.
- also: $\|t_1 \sqcap t_2\| = \|t_1\| \cup \|t_2\|$
- In der resultierenden Menge sind also sowohl t_1 als auch t_2 enthalten
- Das Resultat entspricht jedoch oft nicht unserer Intuition

Generalisierung

- Generalisierung (\sqcap / \sqcup) bestimmt jene Informationen, die zwei Merkmalsstrukturen gemeinsam ist.
- **Def.:** Die Generalisierung von t_1 und t_2 ist die spezielleste Struktur, die sowohl t_1 als auch t_2 subsumiert.
- also: $\|t_1 \sqcap t_2\| = \|t_1\| \cup \|t_2\|$
- In der resultierenden Menge sind also sowohl t_1 als auch t_2 enthalten
- Das Resultat entspricht jedoch oft nicht unserer Intuition
 - ▶ \rightarrow Generalisierung mit Disjunktion