

7. Übungsblatt - Abgabe: 16.12.2014

Aufgabe 7.1 - Merkmalsstrukturen - Unifikation

Betrachten Sie die folgenden Merkmalsstrukturen:

- (a)
$$\left[\begin{array}{ll} \text{VORNAME} & \text{max} \\ \text{NACHNAME} & \text{müller} \\ \text{VATER} & \left[\begin{array}{ll} \text{VORNAME} & \text{heinz} \\ \text{NACHNAME} & \text{müller} \end{array} \right] \\ \text{ALTER} & 15 \end{array} \right]$$
- (b)
$$\left[\begin{array}{ll} \text{VORNAME} & \text{max} \\ \text{MUTTER} & \left[\begin{array}{ll} \text{VORNAME} & \text{anna} \\ \text{NACHNAME} & \text{schmidt} \end{array} \right] \\ \text{BRUDER} & \left[\begin{array}{ll} \text{VORNAME} & \text{karl} \\ \text{ALTER} & 5 \end{array} \right] \end{array} \right]$$
- (c)
$$\left[\begin{array}{ll} \text{VORNAME} & \text{max} \\ \text{VATER} & \left[\text{ALTER } 43 \right] \\ \text{BRUDER} & \left[\begin{array}{ll} \text{VORNAME} & \text{hans} \\ \text{NACHNAME} & \text{schmidt} \\ \text{ALTER} & 13 \end{array} \right] \end{array} \right]$$
- (d)
$$\left[\begin{array}{ll} \text{NACHNAME} & \text{müller} \\ \text{VATER} & \left[\text{ALTER } 43 \right] \\ \text{BRUDER} & \left[\begin{array}{ll} \text{VORNAME} & \text{max} \\ \text{ALTER} & 15 \end{array} \right] \end{array} \right]$$
- (e)
$$\left[\begin{array}{ll} \text{NACHNAME} & \text{schmidt} \\ \text{VATER} & \left[\text{ALTER } 43 \right] \\ \text{MUTTER} & \left[\text{NACHNAME } \text{schmidt} \right] \\ \text{ALTER} & 15 \end{array} \right]$$

- (a) Betrachten sie alle möglichen Paarungen der Merkmalsstrukturen. Welche Paare sind konsistent (d.h. unifizierbar), welche Paare nicht?
Hinweis: Notieren Sie Ihre Ergebnisse in einer Tabelle.
- (b) Wählen Sie zwei unifizierbare Paare und führen Sie die beiden Unifikationen durch. Stellen Sie die Ergebnisse wieder als Merkmalsstruktur dar.

Aufgabe 7.2 - Grammatik mit Merkmalen I

- (a) Geben Sie die Merkmalsstrukturen der Lexikoneinträge an, wie sie für eine Analyse des folgenden Satzes nötig sind (analog zur Beispielgrammatik 4).

Die Linguisten schreiben eine Grammatik

Geben Sie -abweichend von den Beispielen in Grammatik 4- lexikalische Einträge, die unterspezifiziert sind, in allgemeiner Form an: Bei dem Artikel „eine“ können wir z.B. nur mit Bestimmtheit sagen, dass das Genus feminin und der Numerus Singular ist, können aber den Kasus nicht eindeutig angeben (es könnte sich um Akkusativ oder Nominativ handeln.) Wir würden hier also folgenden Eintrag angeben:

DET → eine
⟨DET num⟩ = sg
⟨DET gen⟩ = fem

- (b) Kann es mit dieser Art von Lexikoneinträgen Probleme auf dem Gebiet der Abdeckung oder der Korrektheit geben? Erläutern Sie an einem Beispiel.
- (c) Analysieren Sie obigen Satz mit Ihren Merkmalsstrukturen und der Beispielgrammatik 4. Geben Sie den Ableitungsbaum an, schreiben Sie dann die durch die Constraints erhaltenen Merkmalsstrukturen an die Knoten des Baumes.
- (d) Geben Sie Merkmalsstrukturen für die neu auftretenden Wörter für den folgenden (ungrammatischen) Satz an (ebenfalls analog zu Beispielgrammatik 4 und mit allgemeinen Merkmalen wie oben beschrieben):

Die Studenten schreiben dem Hausarbeit

- (e) Welche Constraints werden bei der Analyse dieses Satzes verletzt? An welcher Stelle im Ableitungs-/Unifikationsprozess treten die Constraintverletzungen auf?
- (f) Analysieren Sie den Satz aus Aufgabe (a) noch einmal mit Beispielgrammatik 7 (mit Subkategorisierung und Kopfmerkmalen) und entsprechenden lexikalischen Einträgen, d.h. geben Sie den entstehenden Baum an und annotieren Sie die Knoten mit den entsprechenden AVMS.

Aufgabe 7.3 - Grammatik mit Merkmalen II: Relativsätze

Betrachten Sie Relativsätze wie z.B. *Der Student sieht das Kind, das schläft.*

Zur Erzeugung von (einer bestimmten Sorte von) Relativsätzen können die beiden folgenden Regeln genutzt werden:

NP → Det N SRel

SRel → RPro VP

- (a) Wovon hängen die grammatischen Merkmale (Kasus, Numerus, Genus) eines Relativpronomens ab? Erklären Sie anhand des Beispielsatzes für jedes einzelne dieser drei Merkmale. (Hinweis: Es ist ein Zusammenspiel aus Kongruenz und Subkategorisierung.)
- (b) Ergänzen Sie die einfachen Merkmalsconstraints (vgl Grammatik 4) zu den beiden oben angegebenen Regeln.
- (c) Die beiden obigen Regeln decken nicht alle Arten von Relativsätzen ab. Geben Sie 2 Beispielsätze an, die nicht von den obigen Regeln abgedeckt werden und begründen Sie. (Es geht dabei nicht um fehlende Lexikoneinträge und auch nicht darum, dass wir mit der Grammatik beispielsweise keine NPs mit Adjektiv ableiten können und diese also auch nicht im Relativsatz auftauchen können, sondern um die generelle Struktur von Relativsätzen.)
- (d) **Zusatzaufgabe:** Leiten Sie (mit Hilfe der neuen Regeln aus b) mit Merkmalsstrukturen, entsprechenden lexikalischen Einträgen, sowie den Regeln aus der Beispielgrammatik 4 den Beispielsatz ab.

Aufgabe 7.4 - Tiger mit Merkmalen

Auch das Tigerkorpus ist mit Merkmalen annotiert. Eine Version, bei der die grammatischen Merkmale, tabsepariert in der 7. Spalte, angegeben sind, finden Sie auf der Vorlesungsseite unter `tiger_merkmale.txt`. Zusätzlich ist die grammatische Funktion des Wortes in der 11. Spalte angegeben. Eine Liste der vorkommenden Funktionen finden Sie unter <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/kanten.html>.

Nehmen Sie an, sie müssten für eine Anwendung wissen, mit welchen grammatischen Merkmalen Nomen als Subjekte (SB) und Akkusativobjekte (AO) in Tiger realisiert sind. Ihre Aufgabe ist es also, herauszufinden, welche Merkmalsmuster im Tigerkorpus für Nomen in Subject- bzw Akkusativobjekt-Position wie oft vorkommen.

Benutzen Sie in Ihrer Pipeline den Befehl `awk`, um die relevanten Spalten auszufiltern: der Befehl `awk -F ';' '{print $2 "\t" $3}' datei.txt`, druckt beispielweise die 2. und 3. Spalte, getrennt durch einen Tab, in einem Dokument, das mit Semicolons in Spalten getrennt ist. Außerdem können Sie wieder die alten Bekannten `sort`, `uniq` und `grep` benutzen.

Geben Sie Ihre(n) Befehl(e) an, und die Ergebnisse, die Sie erhalten haben. Sie können Merkmalsmuster mit weniger als 20 Vorkommen ignorieren.

Dr. Vera Demberg

Andrea Horbach

Einführung in die Computerlinguistik

<http://www.coli.uni-saarland.de/courses/I2CL-14/>

WS 2014/2015

Abgabe in Gruppen von bis zu drei Studierenden am **16.12.2014** vor der Vorlesung.