

Sprachverarbeitung: Übung 16

Syntaktische Analyse deutscher Sätze

Das Thema dieser Übung ist die syntaktische Analyse deutscher Sätze. Die Aufgabe besteht darin, die folgende Grammatik für einfache deutsche Sätze hinsichtlich Vokabular und Satzkonstruktionen zu erweitern.

$$\begin{aligned} S &\rightarrow NG\ VG \\ NG &\rightarrow Art\ N \\ VG &\rightarrow V \\ Art &\rightarrow \text{der} \mid \text{die} \\ N &\rightarrow \text{baum} \mid \text{wiese} \\ V &\rightarrow \text{blueht} \end{aligned}$$

Diese Grammatik ist in `Uebung16/Gegebenes/beispiel.dat` zu finden. Es ist gut zu sehen, dass diese Grammatik nur sehr wenige Sätze akzeptiert. Sie können dies auch mit dem Parser testen. Starten Sie ihn mit dem Matlab-Skript `start_parser_gui.m`.

Aufgabe 1: Erweiterung des Lexikons und der Grammatik

Erweitern Sie die gegebene kontextfreie Grammatik (d.h. das Lexikon und die Grammatikregeln) so, dass auch die folgenden Sätze akzeptiert werden:

- (a) der grosse baum blueht
- (b) das kind sieht den baum
- (c) die baeume bluehen auf der grossen wiese
- (d) die kleinen kinder sehen die grossen baeume
- (e) der kleine baum des kindes blueht
- (f) die kinder sehen auf der wiese den baum
- (g) das kleine kind sieht den baum auf der wiese
- (h) die baeume bluehen auf der grossen wiese der kinder

Gehen Sie dabei wie folgt vor: Kopieren Sie die Beispielgrammatik in eine neue Datei, beispielsweise `ueb16_1.dat`, tragen Sie die fehlenden Wörter ins Lexikon ein und erweitern Sie dann die Regeln der Grammatik.

Fügen Sie eine zusätzliche NG-Regel mit einem Adjektiv **Adj** ein, damit Satz (a) akzeptiert wird.

Damit Satz (b) akzeptiert wird, müssen Sie eine weitere VG-Regel einsetzen, die festlegt, dass eine VG nebst einem Verb auch ein Objekt enthalten kann. Ein Objekt ist wiederum eine Nominalgruppe.

Satz (c) enthält eine Präpositionalgruppe PNG, also eine Nominalgruppe mit vorangehender Präposition **Praep**. Es werden somit zwei neue Regeln benötigt, eine PNG- und eine weitere VG-Regel.

Gehen Sie schrittweise so weiter und testen Sie nach jedem Erweiterungsschritt die Grammatik mit ein paar positiven und negativen Beispielen. Achten Sie insbesondere darauf, dass die Testbeispiele, die in einem früheren Schritt akzeptiert worden sind, nach einem Erweiterungsschritt immer noch akzeptiert werden.

Mit zunehmender Komplexität der Grammatik wächst dabei die Wahrscheinlichkeit, dass der Parser mehrere Lösungen findet. Schauen Sie diese an und überlegen Sie sich, ob diese Mehrdeutigkeiten durch Ändern der Grammatik zu vermeiden sind.

Wenn die Sätze (a) bis (h) richtig analysiert werden, also mindestens ein Syntaxbaum pro Satz so aussieht, wie Sie es erwarten, dann gehen Sie zur nächsten Aufgabe.

Aufgabe 2: Spezifikation von Übereinstimmungen

Wahrscheinlich haben Sie bereits festgestellt, dass die Grammatik von Aufgabe 1 auch ungrammatikalische Sätze wie die folgenden akzeptiert:

- (i) das grossen baum blueht
- (j) das kind sieht der baum
- (k) die baeume bluehen auf die grossen wiese
- (l) die kleinen kinder sieht die grossen baeume
- (m) der kleine baum das kind blueht
- (n) die kinder sehen auf dem wiese den baum
- (o) das kleine kind sieht der baeume auf der wiese
- (p) die baeume bluehen auf der grossen wiesen der kinder

Erweitern Sie die Grammatik und das Lexikon mit Attributen so, dass die Sätze (a) bis (h) akzeptiert werden, die Sätze (i) bis (p) jedoch nicht. Um weniger schreiben zu müssen, nutzen Sie dabei vorteilhafterweise die Möglichkeiten des DCG-Formalismus aus. So können Sie beispielsweise die drei Lexikoneinträge

```
Art(nom,pl,m)  "die"
Art(nom,pl,f)  "die"
Art(nom,pl,n)  "die"
```

zu einem einzigen Eintrag zusammenfassen, d.h. für die konstanten Genus-Attribute **m**, **f** und **n** die Variable **?gen** einsetzen, was den folgenden Eintrag ergibt:

```
Art(nom,pl,?gen) "die"
```

Auch bei Adjektiven können Einträge teilweise zusammengefasst werden. Beispielsweise ist es möglich, sämtliche Pluralformen von "klein" (insgesamt 12)

```

Adj(nom,pl,m)  "kleinen"
Adj(nom,pl,f)  "kleinen"
Adj(nom,pl,n)  "kleinen"
Adj(gen,pl,m)  "kleinen"
Adj(gen,pl,f)  "kleinen"
Adj(gen,pl,n)  "kleinen"
Adj(dat,pl,m)  "kleinen"
⋮

```

zum folgenden Lexikoneintrag zusammenzufassen:

```

Adj(?kas,pl,?gen)  "kleinen"

```

Gehen Sie bei der Erweiterung wiederum schrittweise vor, d.h. erweitern Sie eine Regel mit Attributen und testen Sie diese sofort mit positiven und negativen Beispielen. Sie fangen vorteilhafterweise mit den einfachen Konstituenten an, also z.B. mit der NG. Entsprechende Satzteile können Sie mit dem Parser einfach testen, indem Sie im Eingabefeld **Constituent** die Konstituente ändern, die der Parser als Wurzel des Ableitungsbaumes benutzen soll.

Achten Sie bei der Präpositionalgruppe PNG darauf, dass der Kasus der NG von der Präposition bestimmt (bzw. eingeschränkt wird). In einer PNG mit der Präposition "auf" kann beispielsweise nur eine NG im Dativ oder im Akkusativ stehen.

Aufgabe 3: Erweiterung für Fragesätze

Überlegen Sie, wie die Grammatik von Aufgabe 2 zu erweitern ist, dass auch die folgenden Fragesätze akzeptiert werden:

- (q) blueht der grosse baum
- (r) sieht das kind den baum
- (s) wo bluehen die baeume
- (t) was sehen die kleinen kinder

Die Sätze (q) und (r) sind Fragen mit invertierter Stellung von Subjekt und Prädikat, und die Sätze (s) und (t) werden zusätzlich mit einem W-Fragepartikel **WFP** eingeleitet.

Erweitern Sie sodann Lexikon und Grammatik von Aufgabe 2.