

Aufgabenblatt 1

Syntax natürlicher Sprachen

Universität München, CIS, WS 2016/17

Hans Leiß

Abgabetermin: Mi, 28.10.2017

Aufgabe 1.1 Überlegen Sie sich für die Wortarten Nomen, Artikel und (Voll-)Verb, welche Formdimensionen die Wörter dieser Art haben, und welche unveränderlichen Merkmale man unterscheiden sollte.

Sie können sich eine Grammatik des Deutschen nehmen (z.B. Helbig/Buscha oder eine andere der auf den Folien genannten) und nachsehen, welche Wortarten sie für das Deutsche angibt, jeweils mit den Formdimensionen und -werten. Vergleichen Sie, welche der Wortarten in der Klassifizierung von Dionysos Thrax vorkamen und welche nicht.

Wenn Deutsch nicht Ihre Muttersprache ist, können Sie die Aufgabe auch für die Wortarten Ihrer Muttersprache machen.

Aufgabe 1.2 Auf der Kursseite finden Sie unter `beispielsaetze.txt` eine alphabetisch sortierte Liste von 8350 deutschen Sätzen und unter `beispielsaetze.woerter.txt` die sortierte Liste der darin vorkommenden Wörter (im Sinne von Zeichenfolgen ohne Leerzeichen, Zeilenumbruch, Interpunktionszeichen). Angenommen, wir wollen die verschiedenen Wortformen eines Wortes finden und zu einem Paradigma des Wortes zusammenstellen.

- (a) Wählen Sie je 5 Ihrer Meinung nach häufige Nomen, Adjektive und Vollverben und prüfen Sie an der sortierten Wortliste, wieviele der möglichen Formen des Worts im Text vorkamen.
- (b) Wie könnte man rein formal (durch Operationen mit Zeichenreihen) versuchen, Flexionsaffixe (Präfixe und Suffixe) zu finden und vom möglichen Stamm zu unterscheiden?
- (c) Wie viele der Flexionsendungen von Nomen, Adjektiven und Vollverben haben Sie für denselben „Stamm“ für Ihre Beispielwörter finden können? (Angenommen, Sie erkennen die Stämme, auch mit Stammvarianten.)
- (d) Wie könnte man rein formal versuchen, Stammvarianten zu finden?
- (e) Können Sie Kontexte angeben, mit denen man an Hand der Satzliste ermitteln könnte, welche der Vollformen eines Nomens welcher abstrakten Form zuzuordnen sind? Für welche Ihrer Beispielwörter können Sie damit die Formen für unterschiedliche Kasus unterscheiden, d.h. in welchen Kontexten findet man den Nominativ Singular, in welchen den Genitiv Singular, usw.? Welche Kontexte liefern bestimmte Verbformen?

Als Kontext von w können Sie $(*u, v*)$ erlauben, d.h. Sie geben nur die direkten Links- und Rechts-„Nachbarn“ von w wieder, nicht alles bis zum Satzanfang oder -ende. Sie können auch mit dem Unix-Befehl `grep` versuchen, Wörter in einem bestimmten Kontext zu finden, z.B.

```
grep -e "des [A-Z][a-z]*s " beispieleaetze.txt
```

um gewisse (Adjektive mit) Nomen im Genitiv zu finden (falls Sie `grep` schon ein bißchen benutzt haben). Bei einem so kleinen Text dürfte es nur wenige Wörter geben, die im gleichen 1-Wort-Kontext (d.h. ein vorgegebenes Wort vor und ein vorgegebenes Wort nach dem Vorkommen) auftreten. (Finden Sie ein Beispiel?)

Aufgabe 1.3 Aufgabe zu SWI-Prolog:

- (a) Prüfe (mit `which pl` oder `which prolog`), ob ein Prolog auf Ihrem Rechner installiert ist. (Am Anfang reicht jedes Prolog, aber vermutlich laufen manche Programme, die ich zur Verfügung stelle, nur unter SWI-Prolog.)

Falls kein Prolog installiert ist, holen Sie sich SWI-Prolog von www.swi-prolog.org und installieren es. Wenn es Probleme gibt, schreiben Sie sich auf, was Sie getan haben und was nicht funktioniert, vielleicht können wir es damit schon klären.

- (b) Erweitere das Programm `Beispiele/einbruch.pl` um Fakten und Regeln, die besagen:

Maria ist unverdächtig.

Wer einen Entlasteten beschuldigt, ist verdächtig.

Wenn zwei Personen einander belasten, sind beide verdächtig.

- (c) Ändere das Programm `Beispiele/einbruch.pl` so, daß es zu den Prädikaten `verdaechtig/1` und `entlastet/1` jeweils nur eine Regel gibt, die dasselbe wie das bisherige Programm ausdrückt. (Die Regel muß im Wenn-Teil eine Alternative der bisherigen Wenn-Teile enthalten.)
- (d) Lade von der Kursseite das `graphviz.hl.pl` in Prolog und wenn `dot` und `gv` installiert sind, teste mit `why/1` oder `why2/1`, ob die Anzeige des Suchbaums für eine Anfrage funktioniert, z.B. `?- why2(verdaechtig(emil)).`