

Hausaufgabe 3

— Probabilistische Methoden, Generierung, Anaphern —

Wer einen Anglistikschein machen möchte, sollte bitte jeweils immer die englischen Sätze analysieren—bitte “English” auf die Hausarbeit schreiben (N.B.: wer möchte, kann alle Aufgaben auf English beantworten.) Wer einen Germanistikschein machen möchte, sollte bitte jeweils immer die deutschen Sätze analysieren (bitte “Deutsch” auf die Hausarbeit schreiben). Ansonsten kann man frei wählen.

1 N-Grams

1.1 Deutsch

Nehmen wir an, daß uns der folgende annotierte Text zur Verfügung steht (der einfache Text ist in (1)) um einen eigenen Tagger zu trainieren. Dieser Text wurde mit Helmut Schmid's *Tree-Tagger* annotiert (und noch manuell aufbereitet).

Es/PPER/es gab/VVFIN/geben einmal/ADV/einmal eine/ART/ein
kleine/ADJA/klein Hexe/NN/Hexe, die/PRELS/d konnte/VMFIN/können
sehr/ADV/sehr schnell/ADV/schnell fliegen/VVFIN/fliegen und/KON/und
vor/APPR/vor allem/PIS/alle war/VAFIN/sein sie/PPER/sie wegen/APPR/wegen
ihren/PPOSAT/ihr schnellen/ADJA/schnell Zauberkünsten/NN/Zauberkunst
sehr/ADV/sehr bekannt/ADJD/bekannt. Sie/PPER/sie besaß/VVFIN/besitzen
einen/ART/ein Raben/NN/Rabe, der/PRELS/d auch/ADV/auch schnell/ADV/schnell
fliegen/VVINF/fliegen konnte/VMFIN/können und/KON/und außer-
dem/ADV/außerdem ziemlich/ADV/ziemlich schnell/ADV/schnell und/KON/und
frech/ADJD/frech antwortete/VVFIN/antworten. Von/APPR/von dem/ART/d
schnellen/ADJA/schnell Besen/NN/Besen, auf/APPR/auf dem/PRELS/d
die/ART/d beiden/PIS/beide oft/ADV/oft flogen/VVFIN/fliegen ganz/ADV/ganz

zu/PTKZU/zu schweigen/VVINF/schweigen. Kurzum/ADV/kurzum
 , die/ART/d Hexe/NN/Hexe war/VAFIN/sein sehr/ADV/sehr schnell/ADJD/schnell
 , aber/KON/aber dafür/KOUS/dafür auch/ADV/auch ungenau/ADJD/ungenau.
 Wenn/KOUS/wenn sie/PPER/sie mal/ADV/mal schnell/ADV/schnell
 Tee/NN/Tee kochen/VVINF/kochen wollte/VMFIN/wollen , kam/VVFIN/kommen
 oft/ADV/oft stattdessen/ADJA/;unknown; Kaffee/NN/Kaffee dabei/PAV/dabei
 raus/PTKVZ/raus. Wenn/KOUS/wenn sie/PPER/sie schnell/ADV/schnell
 ein/ART/ein Brot/NN/Brot backen/VVINF/backen wollte/VMFIN/wollen,
 gab/VVFIN/geben es/PPER/es Apfelkuchen/NN/Apfelkuchen. Wenn/KOUS/wenn
 sie/PPER/sie schnell/ADV/schnell den/ART/d Raben/NN/Rabe füttern/VVINF/füttern
 wollte/VMFIN/wollen, bekam/VVFIN/bekommen er/PPER/er manch-
 mal/ADV/manchmal nur/ADV/nur Wasser/NN/Wasser.

- (1) Es gab einmal eine kleine Hexe, die konnte sehr schnell fliegen und vor allem war sie wegen ihren schnellen Zauberkünsten sehr bekannt. Sie besaß einen Raben, der auch schnell fliegen konnte und ausserdem ziemlich schnell und frech antwortete. Von dem schnellen Besen, auf dem die beiden oft flogen ganz zu schweigen. Kurzum, die Hexe war sehr schnell, aber dafür auch ungenau. Wenn sie mal schnell Tee kochen wollte, kam oft stattdessen Kaffee dabei raus. Wenn sie schnell Brot backen wollte, gab es Apfelkuchen. Wenn sie schnell den Raben füttern wollte, bekam er manchmal nur Wasser.

Nehmen wir jetzt weiter an, daß wir mit dem folgenden Satz konfrontiert sind und daß unser neuer POS-tagger entscheiden soll, ob *schnell* in diesem Satz ein Adjektiv oder ein Adverb ist.

- (2) Die Hexe war sehr schnell.

Fragen:

1. Wenn wir mit einem ganz einfachen Unigram-Tagger anfangen, was wird unser POS-Tagger als Analyse vorschlagen und warum?
2. Können bessere Resultate erzielt werden, wenn man Bigramme oder sogar Trigramme mit einbezieht (d.h., Informationen über die Nachbarworte)?

1.2 English

Assume that you have the following annotated text to train a POS-tagger with (the simple text is given in (3)). This text was annotated with Helmut Schmid's *Tree Tagger* and has been reworked manually.

Canseco/NP/Canseco was/VBD/be famous/JJ/famous for/IN/for his/PP/his
kick/NN/kick ./SENT/. He/PP/he could/MD/could kick/VB/kick from/IN/from
one/CD/one end/NN/end of/IN/of the/DT/the field/NN/field to/TO/to
the/DT/the other/JJ/other without/IN/without any/DT/any effort/NN/effort
./SENT/. His/PP/his kick/NN/kick tended/VBD/tend to/TO/to con-
found/VB/confound goal-keepers/NNS/goal-keeper ./,/, as/IN/as the/DT/the
ball/NN/ball would/MD/would curve/NN/curve away/RB/away at/IN/at
the/DT/the last/JJ/last second/NN/second ./SENT/. When/WRB/when
he/PP/he would/MD/would kick/VB/kick a/DT/a penalty/NN/penalty
./,/, the/DT/the goal/NN/goal keeper/NN/keeper never/RB/never
had/VBD/have any/DT/any hope/NN/hope ./SENT/. A/DT/a kick/NN/kick
such/JJ/such as/IN/as Canseco/NP/Canseco 's/POS/'s is/VBZ/be
rare/JJ/rare and/CC/and so/RB/so it/PP/it is/VBZ/be no/DT/no
wonder/NN/wonder that/IN/that his/PP/his kick/NN/kick alone/RB/alone
has/VBZ/have earned/VBN/earn him/PP/him millions/NNS/million
./SENT/. But/CC/but in/IN/in recent/JJ/recent times/NNS/time ./,/,
his/PP/his kick/NN/kick has/VBZ/have begun/VBN/begin to/TO/to
be/VB/be a/DT/a liability/NN/liability and/CC/and it/PP/it is/VBZ/be
not/RB/not clear/JJ/clear whether/IN/whether his/PP/his kick/NN/kick
will/MD/will carry/VB/carry him/PP/him through/RP/through ./SENT/.

- (3) Canseco was famous for his kick. He could kick from one end of the field to the other without any effort. His kick tended to confound goal-keepers, as the ball would curve away at the last second. When he would kick a penalty, the goal keeper never had any hope. A kick such as Canseco's is rare and so it is no wonder that his kick alone has earned him millions. But in recent times, his kick has begun to be a liability and it is not clear whether his kick will carry him through.

Now assume that our new POS-tagger is confronted with the new sentence in (4) and that it needs to decide whether *kick* is a noun or a verb.

- (4) Canseco could kick into the top left corner of a goal.

Questions:

1. If we have built a simple Unigram tagger to start with, what would our POS-Tagger guess and why?
2. Could better results be achieved with a bigram or trigram version of our tagger (that is, if one had information about the neighbors of *kick*)?

2 Deep Parsing

1. Ist Sprache kontext-frei (context-free)?
2. Wofür wird Unifikation beim Parsen eingesetzt? (Nicht mehr als ein Absatz).
3. Warum sollte man überhaupt *deep parsing* Strategien noch anstreben, wenn *shallow parsing* (z.B., POS-tagger, Chunk Parser) doch schon relativ gut funktioniert? (Am Besten geben Sie ein Beispiel und erläutern es kurz.)

3 Generierung

Wie unterscheiden sich Systeme wie SHRDLU, FOG oder Eliza von Systemen, die in Zusammenhang mit grammatischen Modellen, wie z.B. der LFG, natürlichsprachliche Sätze generieren? Was ist der wesentliche Unterschied zwischen den zwei verschiedenen Generierungsmethoden? (Nicht mehr als 2 Absätze.)

4 Anapherauflösung

4.1 Deutsch

Diese Aufgabe legt den folgenden Text zugrunde.

- (5) Der Adler arbeitete wieder an dem Pfeil. Endlich gelang es ihm, die Spitze herauszuziehen. Der Pfeil fiel ins feuchte Gras. Harka sprang hin und holte ihn. Als er wieder bei seinem Vater saß, saß auch der Adler auf seinem Stein, ruhig wie ein Bild. Das Fleisch hielt er zwischen den Fängen.
[Liselotte Welskopf-Henrich, *Der Weg in die Verbannung*]

Der Text in (5) enthält mehrere Pronomen.

1. Identifizieren Sie die Pronomen.
2. Stellen Sie einen Algorithmus auf, der ermittelt, auf welchen der Diskurspartizipanten sich die jeweiligen Pronomen beziehen. Der Algorithmus kann auf Centering Theory, oder einem anderem etabliertem Verfahren basieren, es kann aber auch frei ein völlig neuer Algorithmus erfunden werden.

4.2 English

In this exercise, we are working with the text in (6).

- (6) Brat could see Simon's bulk against the sky. From the shape of it, he was half-kneeling on the edge, by the rope. Brat could touch him by putting out a hand. Fool that he had been to underrate Simon. Simon had taken no chances. He hadn't even taken the chance of following him. [Josephine Tey, *Brat Farrar*]

1. Identify all the pronouns in the text.
2. Apply an algorithm which identifies the antecedents of each of the pronouns (i.e., figure out which entity the pronoun refers to). You could use Centering Theory or another well-established algorithm. You could also make up your own algorithm.