Lösungsvorschlag für das Übungsblatt 7.

Aufgabe 1.

a)

Die Wortformen, die zwei oder mehr Tags haben, führen zu Ambiguitäten, die mithilfe von POS-Tagging aufgelöst werden können. Dazu gibt es zwei Herangehensweisen: Statistische (Markov-Modelle) und regelbasierte Kontextvergleiche, die auch der Brill-Tagger verwendet. Es funktioniert folgendermaßen: Der Brill-Tagger fängt mit einem initialen Tag für jedes Wort an und wendet daraufhin Transformationsregeln an, um eine verbesserte Tagfolge zu erhalten.

Die Regeln für unbekannte Wörter sind dergestalt, dass sie auf Basis von bestimmten Eigenschaften des Wortes oder nebenstehenden Wörtern den Tag des Wortes verändern können. Die Transformationsregeln ändern den Tag eines Wortes, wenn im Kontext des Wortes bestimmte Tags oder Worte auftreten.

Das Training funktioniert folgendermaßen: Aus Regeltemplates werden Regeln erstellt und angewendet. Die Regel, die das Ergebnis am meisten verbessert wird an das Ende der Liste der gelernten Regeln angefügt. Dies wird solange wiederholt, bis es keine oder nur noch eine minimale Verbesserung gibt. Es werden zuerst die Regeln für die unbekanntesten Wörter und dann die Transformationsregeln trainiert.

Die Regeln für unbekannte Wörter versuchen, (a) mit Affixen den Tag zu erraten (z.B. "blablahous" wird als ADJ getaggt, da es der übliche Tag für Wörter mit der Endung "ous" ist) (b) durch Hinzunahme oder Wegnahme von Affixen zu einem dem Tagger bereits bekannten Wort zu gelangen (c) durch nebenstehende Wörter den Tag zu erraten (d) dies durch ein Zeichen im Wort zu tun.

Die Transformationsregeln ändern einen Tag von X nach Y, wenn (a) vor oder nach dem jetzigen Wort ein bestimmtes Wort steht oder (b) für ein Wort vor und/oder nach dem jetzigen Wort ein bestimmter Tag vergeben wurde.

Funktionsweise:

Bei der Initialisierung erhält jedes bekannte Wort den Tag, den es im Trainingskorpus am häufigsten hatte. Nach dieser Initialisierung werden zuerst alle Regeln für unbekannte Wörter in der gelernten Reihenfolge nacheinander angewandt. Daraufhin werden alle gelernten Transformationsregeln nacheinander angewandt.

b) *aus Ihren Beiträgen:*

his/pp\$ thesis/nn is/bez so/ql restricting/jj as/cs to/to deny/vb ...

"restricting" würde eher als Verbverlaufsform (vbg) statt als Adjektiv (jj) getaggt.

Templates:

1. The preceding/following word is tagged c.

--> Instanz: alter(vbg, jj, prevtag(ql)) < vermutlich hilfreich

--> Instanz: alter(vbg, jj, nexttag(cs))

5. The preceding word is tagged c and the following word is tagged d.

--> Instanz: alter(vbg, jj, prevtag(ql), nexttag(cs))

6. The preceding/following word is tagged c and the word two before/after is tagged d.

--> Instanz: alter(vbg, jj, prevtag(ql) & prev2tag(bez)) < vlt. noch präziser als die Instanz vom Template 1, könnte aber auch

Verlaufsformen falsch identifizieren

--> Instanz: alter(vbg, jj, prevtag(ql) & next2tag(to))

--> Instanz: alter(vbg, jj, nexttag(cs) & prev2tag(bez)) < könnte auch erfolgreich sein.

--> Instanz: alter(vbg, jj, nexttag(cs) & next2tag(to))

which/wdt are/ber genuine/jj protests/nns of/in our/pp\$ own/jj

"protests" würde vermutlich als Verbform 3.Pers.Sing. (vbz) statt als nominaler Plural getaggt.

Templates:

1. The preceding/following word is tagged c.

--> Instanz: alter(vbz, nns, prevtag(jj)) < ziemlich sicher, da Verbformen Adverbien statt Adjektive als Attribute bekommen

--> Instanz: alter(vbz, nns, nexttag(in))

2. The word two before/after is tagged c.

--> Instanz: alter(vbz, nns, prev2tag(ber)) < wohl auch hilfreiche Regel, da eine Verbkonstruktion unter solchen Umständen Nonsens wäre

--> Instanz: alter(vbz, nns, next2tag(pp\$))

3. One of the two preceding/following words is tagged c.

--> Instanz: alter(vbz, nns, prev[2]tag(ber)) < allgemeinere Version der Instanz von Template 2

--> Instanz: alter(vbz, nns, prev[2]tag(jj))

--> Instanz: alter(vbz, nns, next[2]tag(pp\$))

--> Instanz: alter(vbz, nns, next[2]tag(in))

Aufgabe 2.

a)

Chunk Parsing ist die Vorgehensweise, nicht ganze Sätze direkt zu parsen, sondern zunächst den Text in kleinere sinntragende Einheiten aus mehreren Wörtern (Chunks) zu zerlegen. Jeder Chunk gruppiert sich dabei um genau einen "Major Head", den Kern des Zusammenhangs: "Major Heads" nach Abney sind alle inhaltstragenden Wörter mit Ausnahme derer, die zwischen einem Funktionswort und dem inhaltstragenden Wort stehen, auf das das Funktionswort verweist; oder aber Pronomen, auf die sich eine Präposition bezieht.

Mit Chunk Parsing werden auf jeden Fall Zusammenhänge eher erfasst als mit anderen Methoden des flachen regelbasierten Natural Language Processing. Von daher wäre es denkbar, dass es speziell zum "Verständnis" komplexerer Äußerungen beitragen könnte, also bei der Kontextinterpretation in Dialogsystemen oder Deutungen der Äußerungsintention, wie sie etwa bei Watson nötig war.

Das in der Aufgabestellung angesprochene Problem des Chunk Parsing wiederum besteht darin, dass die "Hierarchie" der Chunks innerhalb eines Satzes unklar sein und sich je nach Lesart verschieben kann, wenn der Kontext bzw. die Intention des betreffenden Satzes nicht bekannt ist oder (vom Parser) nicht verstanden werden kann. Dies wird besonders deutlich, wenn man die auf Seite 15 dargestellte Ambiguität ins Deutsche übersetzt: "Er traf die Frau, die er in Italien heiratete" vs. "Er traf die Frau, die er heiratete, in Italien" sind zwei deutlich unterscheidbare Lesarten ein- und desselben englischen Satzes. Eine solche Situation rührt daher, dass die Chunks als praktisch eigenständige Sinneinheiten nicht zwingend in ihrer kontextuellen Bedeutung determiniert sind, was den Grund dafür darstellt, dass die Verb-Objekt-Beziehung von so zentraler Bedeutung ist, denn ihr werden ja die Modifikationen der einzelnen Satzglieder (bzw. weitere Chunks) untergeordnet.

Motivationen für Chunk Parsing:

- Intonation : sprachliche Einheiten mit jeweils einer stärkeren Betonung
- Effizienz : Teilstrukturen lassen sich effizienter berechnen.
- Minimaler Aufwand : Für gewisse Anwendungen braucht es keine vollständige syntaktische Analyse.
- Robustheit : Für syntaktisch fehlerhaften Input braucht es tolerante, oberflächlichere Analysemethoden.

b)

Der Regelapparat gibt vor, dass der Parser ein Wort aus dem Eingabestring in einen Stapel (stack) fügt (->SH für shift) bis es merkt, dass eine Sequenz von Wörtern gefunden wurde, die nach den Regeln der Grammatik vollständig ist. Dann wird das Wort zurückgeführt und

genau analysiert (->RE für reduce) zu welcher Wortklasse es gehört. Ist der stack vollständig, so wird „S“ für Accept ausgegeben.

Chunk-Parser basieren dabei auf getaggetem Text.

Input des Chunk-Parsers	Output des Chunk-Parsers
[KON und]	[KON und]
[PROAV außerdem]	[PROAV außerdem]
[VVFIN sorgt]	[VVFEX [VVFIN sorgt]]
[PPOSAT mein]	[NX [PPOSAT mein]
[NN Agent]	[NN Agent]]
[\$, ,]	[\$, ,]
[PRELS der]	[NXPRO [PRELS der]]
[PPOSAT meine]	[NX [PPOSAT meine]
[NN Eigenheiten]	[NN Eigenheiten]]
[VVFIN kennt]	[VVFEX [VVFIN kennt]]
[\$, ,]	[\$, ,]
[APPR für]	[PX [APPR für]
[ARTIND eine]	[NX [ARTIND eine]
[ADJA gewisse]	[AJXatt [ADJA gewisse]]
[NN Reibungslosigkeit]	[NN Reibungslosigkeit]]]
[\$. .]	[\$. .]

Während Chunks, die linear aufeinander folgen, durch die wiederholte Anwendung des Chunk-Parsers erkannt werden, werden eingebettete Chunks vom System anhand kaskadierter Parser-Module analysiert, wobei der Output jeweils als Input eines weiteren Chunk-Parsers dient:

- Ebene 0 [APPR für] [ARTIND eine] [ADJA gewisse]
- Ebene 1 [APPR für] [ARTIND eine] [AJXatt [ADJA gewisse]]
- Ebene 2 [APPR für] [NX [ARTIND eine] [AJXatt [ADJA gewisse]]]
- Ebene 3 [PX [APPR für] [NX [ARTIND eine] [AJXatt [ADJA gewisse]]]

Zuerst wird der adjektivische Chunk analysiert, da er nicht komplex ist. Dies gehorcht dem in Abney beschriebenen Prinzip des easy-first parsings: diejenigen Strukturen, die mit dem geringsten Aufwand erkannt werden können, werden zuerst ausgezeichnet. Danach werden komplexere Strukturen auf höheren Ebenen identifiziert, was zu weiteren islands of certainty führt. Die Anzahl der zu behandelnden Einheiten reduziert sich auf jeder Ebene, wodurch auch der Grad an Ambiguität im Text abnimmt.

Noch ein Beispiel:

Satz:

mit dem ICE geht das heute ziemlich schnell wenn man keinen Unfall hat
 with the ICE goes that today pretty fast if one no accident has
 appr art nn vvfin pds adv adv adjd kous pis piat nn vvfin
 'taking the ICE is pretty fast if one does not have an accident'

After 1st FSA (finite state automata) – adverbial phrases

Mit dem ICE geht das heute ziemlich schnell wenn man keinen Unfall hat
 appr art nn vvfin pds [AV AV] kous pis piat nn vvfin

after 2nd FSA – noun phrases:

mit dem ICE geht das heute ziemlich schnell wenn man keinen Unfall hat
 appr [NP NP] vvfin [NP] [AV AV] kous [NP] [NP NP] vvfin

after 3rd FSA – prepositional phrases:

mit dem ICE geht das heute ziemlich schnell wenn man keinen Unfall hat
 [PP PP] vvfin [NP] [AV AV] kous [NP] [NP NP] vvfin

after 4th FSA – clauses:

mit dem ICE geht das heute ziemlich schnell wenn man keinen Unfall hat

[SX] [SX] [SX]

c)

If/cs he/pps is/bez a/at traditionalist/nn ./, he/pps is/bez an/at eclectic/jj traditionalist/nn ./.

Nominalphrasen:

If/cs [NP he/pps] is/bez [NP a/at traditionalist/nn] ./, [NP he/pps] is/bez [NP an/at eclectic/jj traditionalist/nn] ./.

Verbalphrasen:

If/cs [NP he/pps] [VP is/bez] [NP a/at traditionalist/nn] ./, [NP he/pps] [VP is/bez] [NP an/at eclectic/jj traditionalist/nn] ./.

Sätze:

[S If/cs he/pps is/bez a/at traditionalist/nn] ./, [S he/pps is/bez an/at eclectic/jj traditionalist/nn] ./.

Aufgabe 3.

Die Diskussion auf Linguistlist nachlesen – auch der Wikipedia Artikel gibt guten Überblick zu dem Streit.

Im Vortrag spricht Chomskys Paper über „Poverty of the stimulus“. Er behandelt damit die Vernachlässigung der Betrachtung der natürlichen Sprachentwicklung im Bereich der kognitiven bzw. linguistischen Wissenschaften. Anhand verschiedener Beispiele soll „bewiesen“ werden, dass die Fähigkeit des Spracherwerbs mit einem internen und komplexern Mechanismus zu tun hat der anderen Arten fehlt, die ansonsten sehr wohl die Fähigkeit zu kognitiver Entwicklung haben. Insbesondere sei auch die Menge an Daten und Ihre Struktur – nur positive Daten nicht geeignet um das Erlernen der Sprache zu erklären. Insoweit müsse eine interne Struktur mit so etwas wie Grammatikregeln im Menschen gegeben sein. Vgl auch den Artikel von Pullum et al. auf der Webseite.