Einführung in die Computerlinguistik **WS 2003/2004**
M. Pinkal / S. Pado

Übungsaufgaben 3

(1) Das Adjektiv-Endungs-Diagramm auf der Vorlesungsfolie akzeptiert das Wort *st*. Die Superlativendung *kleinst* kommt aber für sich allein gar nicht vor. Korrigieren Sie das Diagramm mit möglichst wenig Aufwand so, daß *st* für sich genommen nicht mehr akzeptiert wird (alle übrigen Endungen dagegen wie bisher). **(1P)**

(2) Entwerfen Sie ein möglichst einfaches Diagramm, das die einfachen flektierten Formen des Verbs *reden* spezifiziert (Präsens und Präteritum, ohne Partizipialendungen!). **(1P)**

(3) (a) In der Vorlesung haben wir ein Diagramm betrachtet, das einige zulässige Wortartketten für Nominalausdrücke spezifiziert. Das Diagramm spezifiziert „volle“ Nominalausdrücke mit Artikel (Determinator), Gattungssubstantiv (Nomen), und einem potentiellen pränominalen adjektivischen Attribut. Erweitern bzw. modifizieren Sie das Diagramm so, daß es auch Nominalausdrücken mit post-nominalen Präpositionalausdrücken umfaßt. Also z.B.:

ART NN (*das Auto)*
ART ADJA ADJA NN (*das neue schnelle Auto*)
ART ADJA NN APPR ART NN (*das grüne Auto auf dem Parkplatz*)
ART NN APPR ART NN APPR ART ADJA NN (*das Auto auf dem Parkplatz an der neuen Mensa*)

Diese Beispiele verwenden das deutsche Standard-Tagset, des *Stuttgart-Tübingen-Tagset (STTS)* zur Beschreibung der Wortarten. Die genauen Bedeutungen der einzelnen Tags können Sie unter http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html nachlesen. **(2P)**

(b) Versuchen Sie, Nominalausdrücke mit Personalpronomen (PPER) und Eigennamen (NE) im Diagramm mit zu berücksichtigen. Dabei sollen Ausdrücke wie **Er unter dem Tisch* oder *Der Fabian* nicht spezifiziert werden. **(1P)**

(4) (a) Entwerfen Sie ein Zustandsdiagramm, das alle durch 2 teilbaren natürlichen Zahlen (in Dezimalschreibweise) beschreibt. Um es einfacher zu machen, können Sie Nullen als Präfix zulassen. **(2P)**

(b) Entwerfen Sie ein entsprechendes Diagramm für durch 4 teilbare Zahlen (Hinweis: Die letzten beiden Ziffern müssen eine durch vier teilbare Zahl ergeben). **(1P)**

(c) Bonusaufgabe (nur für Interessierte): Entwerfen Sie ein entsprechendes Diagramm für durch 3 teilbare Zahlen (Hinweis: Teilbarkeit durch 3 ist aus der Quersumme ablesbar!). **(2P)**

(5) Lesen Sie das Kapitel 5.7 (Sprachsynthese) im Carstensen et al.. **(2P)**

(a) Text-To-Speech-Synthese : Warum sind Komposita schwierig? Warum Zahlausdrücke?

(b) Welche Aussprachevarianten sind für folgende Ausdrücke denkbar? Staubecken, Gestern, 13., 1984

(6) **Bonusaufgabe:** Testen Sie die deutsche Text-To-Speech-Sprachsynthese von Rhetorical Systems (http://www.rhetoricalsystems.com/cgi-bin/demo.cgi), Logox (http://www.logox.de/cgi-bin/speechform.cgi), und AT&T (http://www.research.att.com/projects/tts/demo.html) in folgenden Kategorien:

- Aussprache von Zahlen und Quantitäten (13.000, 75%, ¾, ...)
- Aussprache von Abkürzungen (IBM, UNO)
- Aussprache von Eigennamen (Celle, Cochem, Bush, Lafontaine, Rhône, ...)
- Aussprache von verschiedenen Satzmodi (Fragesatz, Aussagesatz, Befehl)
- Aussprache von komplexen Sätzen
- Aussprache von verschiedenen Arten Diskurs (Listen, Fließtext, Theaterstücke...)

Bewerten Sie für jede Kategorie die segmentale Qualität („Aussprache“) und suprasegmentale Qualität („Prosodie“). Bestimmen Sie am Ende Ihr Lieblingssystem (mit Begründung). **(3P)**