

8. Übungsblatt - Abgabe: 10.01.2012

Dieses Blatt müssen die meisten von Ihnen nicht abgeben, aber Sie können durch Abgabe des Blattes Zusatzpunkte für die Klausurzulassung erhalten. Diejenigen, für die dieses Blatt obligatorisch ist, werden von uns per Email benachrichtigt.

Aufgabe 8.1 - Semantische Relationen

Die folgenden Daten stammen aus einem Assoziationsexperiment, das in Saarbrücken durchgeführt wurde: Versuchspersonen wurde ein Wort präsentiert (Triggerwort), und sie sollten semantisch ähnliche Wörter dazu aufzählen. Im folgenden sind drei Triggerwörter, jeweils mit 5 häufig genannten assoziierten Wörtern, angegeben:

- **Fahrrad:** Fahrzeug, Sport, Rad, Pedal, Moped
 - **Erdbeere:** Kuchen, Frucht, Sahne, Garten, süß
 - **Buch:** Seite, Roman, Autor, Bibliothek, lesen
- (a) Beurteilen Sie für alle Paare von Triggerwort und Assoziation, ob die zwischen ihnen bestehende semantische Relation durch eine der WordNet-Relationen (Hyponymie/Hyperonymie, Meronymie/Holonymie, Antonymie) plausibel beschrieben wird.
- (b) Geben Sie in den Fällen, in denen dies Ihrer Ansicht nach der Fall ist, die Relation an (z.B.: „Fahrrad – Fahrzeug: Hyponymie“) und versuchen Sie die Beziehung in der WordNet-Datenbank aufzufinden. Problem ist, dass es kein deutsches WordNet mit einer frei zugänglichen Web-Schnittstelle gibt. Nehmen Sie deswegen den Umweg über das Englische: Passende englische Übersetzung wählen (ggf. mithilfe von Wörterbuch oder LEO) und im englischen WordNet (<http://wordnetweb.princeton.edu/perl/webwn>) recherchieren. Geben Sie die Resultate der Recherche an.
- (c) Versuchen Sie in den Fällen, in denen Ihrer Ansicht nach keine WordNet-Relation vorliegt, die semantische Beziehung mit eigenen Worten zu charakterisieren.

Aufgabe 8.2 Der Katakansisator

Katakana ist eine japanische Silbenschrift, in der im Japanischen z.B ausländische Namen und Fremdwörter geschrieben werden. In der folgenden Aufgabe sollen Sie einen Transducer (vgl. Aufgabe 3.3) schreiben, der aus einem deutschen Wort in Lautschrift dessen Katakana-Darstellung generiert.

Dabei gelten folgende Regeln (für unsere Aufgabe, im realen Leben gibt es einige Besonderheiten mehr):

- Wörter bestehen aus Silben. Eine Silbe im Japanischen besteht entweder aus einem einzelnen Vokal oder aus einem Konsonanten, gefolgt von einem Vokal oder aus einem einzelnen n.
- Mögliche Konsonanten sind k, g, s, z (stimmloses [s] und stimmhaftes [z] s), t, d, m, n, h, b, p, y (ähnlich einem deutschen j), r, w, sh (ähnlich dem deutschen sch [ʃ]), ch (ähnlich einem tsch, können wir hier vernachlässigen). mögliche Vokale sind a, i, u, e, o. Vokale können lang oder kurz sein. Standardmäßig sind sie kurz, lange Vokale (in der Lautschrift im Deutschen durch Doppelpunkt gekennzeichnet) werden im Japanischen durch einen Längungsstrich angezeigt.
- Treffen mehrere Konsonanten in einem deutschen Wort aufeinander, so wird aus jedem Konsonanten, der von einem anderen Konsonanten gefolgt wird, im Japanischen eine eigene Silbe, indem an diesen Konsonanten ein u angefügt wird, bzw im Fall von t und d ein o. Nur ein n darf alleine stehen und muss nicht unbedingt von einem Vokal gefolgt werden.
- Es gibt nur 5 verschiedenen Vokale: a, i, u, e, o. Die verschiedenen e-Varianten im Deutschen werden alle auf das japanische e abgebildet, entsprechend die anderen Vokale. ö und ä wird als e gehandhabt, ein ü wird als i übersetzt.
- f-Laute werden zu einem h. Der deutsche ng-Laut [ŋ] wird als ng wiedergegeben.
- Die verschiedenen r- und l-Laute im Deutschen ([r] [l] [ʀ] [ʁ]) werden alle durch das japanisch r repräsentiert.
- Der deutsche ich-Laut [ç] wird zu einem h, bzw am Silbenende zu einem hi, der ach-Laut [x] zu einem h am Silbenende zu ha.
- Ein w-Laut v wird zu einem b.

Die Katakana-Umschrift orientiert sich dabei an der Aussprache und nicht an der Schreibweise eines deutschen Wortes. Sie können daher davon ausgehen, dass Sie als Input bereits eine Lautschriftrepräsentation des deutschen Begriffes vorliegen haben. Die Graphik zeigt eine Lautschrifttabelle, mit den deutschen Lauten die sie berücksichtigen sollten.

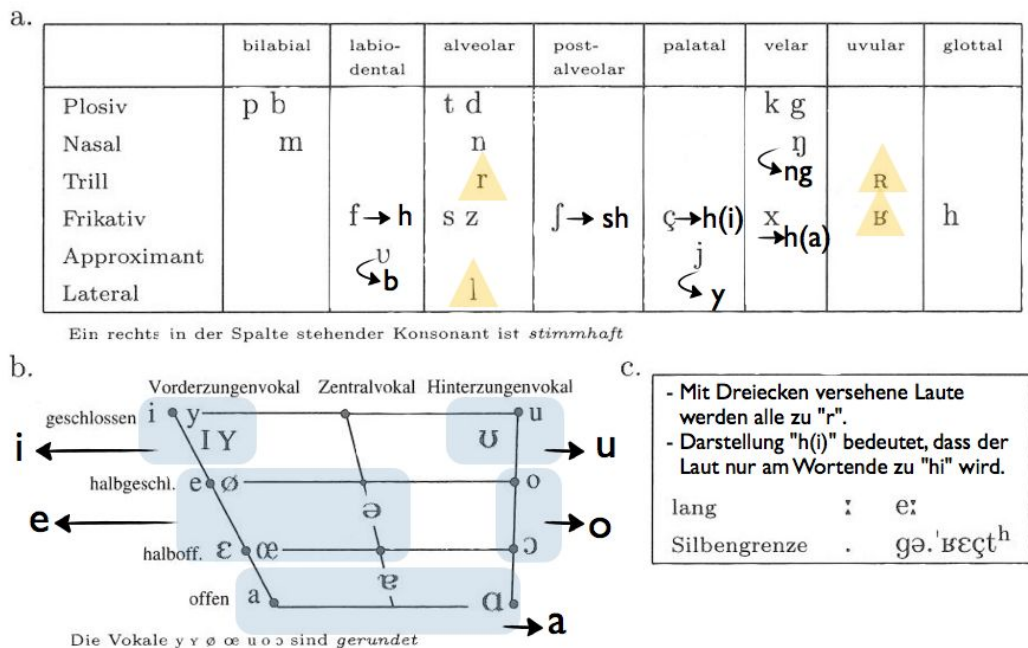


Tabelle 3.2: Das Internationale Phonetische Alphabet (für Deutsch)

Die folgende Tabelle zeigt beispielhaft einige deutsche Wörter, und ihre Katakana-Repräsentation.

dt. Wort	Lautschrift	Katakana
Saarbrücken	zaː.bʁʏ.kən.	za-buriken
Linguistik	lɪŋ.gu.ɪs.tɪk.	ringuguisutiku
Universität	u.ni.vɛʁ.zi.tɛ:t.	uniberuzite-to
Sauerkraut	za.vaː.kraʊt.	zaba-kurauto
Baumkuchen	baʊm.ku.xən.	baumukuhēn
Lumpenproletariat	lʊm.pən.pro.le.ta.ʀɪ.a:t	rumupenpuroretaria-to
Bastian Schweinsteiger	bas.ti.an.ʃvam.ʃtaɪ.gaː.	basutian shuwainshutaiga-

Sie sollen also einen Transducer schreiben, der die 2. Spalte in die 3. Spalte übersetzt und dabei zumindest die Beispielwörter abdeckt, er sollte also Lautschriftzeichen einlesen (ein . markiert dabei eine Silbengrenze im Deutschen) und lateinische Schriftzeichen ausgeben. Er sollte sich nur am Ende einer deutschen Silbe in einem akzeptierenden Zustand befinden.

Hinweis: Lassen Sie sich nicht vom Umfang der Aufgabenstellung abschrecken, die Aufgabe kann mit einem Automaten mit einer überschaubaren Zustandsmenge gelöst werden.

Aufgabe 8.3 Adjektivattribute und Präpositionalphrasen mit Merkmalsconstraints

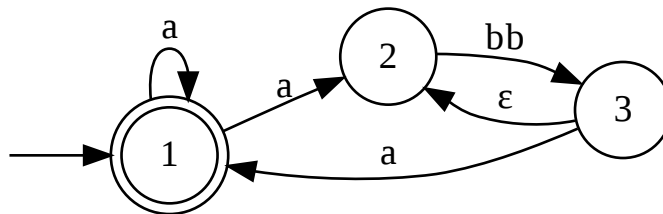
Modifizieren Sie die Grammatik 7 vom Foliensatz 7 so, dass auch Adjektive und Präpositionalphrasen als Modifier einer NP erlaubt sind. Gehen Sie folgendermaßen vor:

- Beschreiben Sie die Phänomene linguistisch: Geben Sie an, wo es jeweils Agreement gibt, bzw welche Anforderungen es außerdem an die neuen Komponenten gibt, jeweils mit Beispiel zur Verdeutlichung. Sie können annehmen, dass Präpositionen immer eindeutig einen bestimmten Kasus nach sich ziehen.
- Geben Sie passende Regeln an (diese sollten bereits aus der Vorlesung bekannt sein) und annotieren Sie diese mit Merkmalsconstraints. Nehmen Sie an, dass eine Präposition für eine NP im passenden Kasus subkategorisiert (analog zu Verben und deren Argumenten).
- Geben Sie geeignete Lexikoneinträge an und leiten Sie damit den Satz „Der fleißige Student liest das große, dicke Buch in der Bibliothek.“ ab.

Hinweis: Sie dürfen annehmen, dass die erforderlichen Lexikoneinträge voll ausspezifiziert sind.

Aufgabe 8.4 Determinisierung von Automaten

Gegeben sei der folgende NEA über dem Alphabet $\Sigma = \{a, b\}$.



- Beschreiben Sie die von diesem NEA akzeptierte Sprache informell.
- Machen Sie den NEA deterministisch: Gehen Sie dabei nach dem in der Vorlesung besprochenen Verfahren vor: Lösen Sie Mehrsymbolkanten auf, beseitigen Sie Epsilon-kanten und konstruieren Sie die Übergangstabelle,

Abgabe in Gruppen von bis zu drei Studierenden bis **10.01.2012** 10 Uhr entweder als Email im pdf-Format an i2cl@coli.uni-sb.de oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.