

Schlussprüfung Morphologie und Lexikographie FS 08

Aufgabenstellung: Simon Clematide

Prüfung vom 3. Juni 2008
Institut für Computerlinguistik
Universität Zürich

Vorname _____ Matrikelnummer _____

Nachname _____

Für Studierende der folgenden Studiengänge:

- ☐ BA - Studiengang Computerlinguistik (Phil. Fakultät)
- ☐ BA - Studiengang Computerlinguistik und Sprachtechnologie (Phil. Fakultät)
- ☐ BA-Studierende (Wirtschaftswissenschaftliche Fakultät)
- ☐ Studierende des Nebenfachs Informatik mit Studienbeginn ab WS 04/05
- ☐ Multidisziplinär (ETH)
- ☐ Andere:

Nur für Lizentiatsstudierende der Computerlinguistik als ein Fach aus der Phil. Fakultät:

Strasse: _____ Hauptfach: _____

PLZ/Ort: _____ E-Mail: _____

Erreichte Punktzahl: _____

Teilnoten: _____

Endnote: _____ Bestanden: ☐ Ja ☐ Nein

Auf jedes Blatt mit Lösungen den Nachnamen schreiben!

Viel Erfolg!

Wichtige Hinweise

Punkte-Maximum: 90 (pro Minute 1 Punkt)

Hinweis: Bitte schreiben Sie in einem überlegten und knappen, aber verbalen Stil (keine Stichwortsammlungen). Bei inhaltlichen Auswahlendungen, wo einfach mal alles spontan hingeschrieben wird und Falsches wie Korrektes munter vermischt sind, behalte ich mir Abzüge vor. Das Zeitbudget ist so berechnet, dass man häufig überlegen und schreiben kann.

1 Morphologische Analyse (10 Punkte)

Das Morphologie-Analyse-System GERTWOL gibt für die Wortform “erfolgreichen” unter vielen anderen die folgenden Analysen zurück:

"er|folg|reich" A POS SG AKK MASK STARK

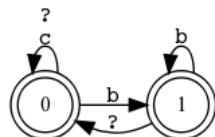
"erfolg|reich" A POS SG GEN MASK STARK

Stellen Sie Hypothesen auf, aufgrund welcher Morphe und Wortbildungsregeln diese Analysen entstanden sind. Sind Bildungsregeln wie für die 1. Analyse sinnvoll? Argumentieren Sie dafür (mit Beispielen) oder dagegen.

Lösungen/Diskussion Die 1. Analyse entsteht aus dem komplexen Verb “erfolgen” (nur Stamm) und dem Adjektiv “reich” (mit schwacher Kompositionsgrenze). Wortbildungen der Form V+A gibt es nicht so viele. Die denkbare Variante mit “er” als Pronomen müsste allerdings schon eine starke Kompositionsgrenze aufweisen. Die 2. Analyse ist eine Wortbildung vom Typ N+A (hier in Gertwol mit schwacher Kompositionsgrenze), welche häufig auftritt.

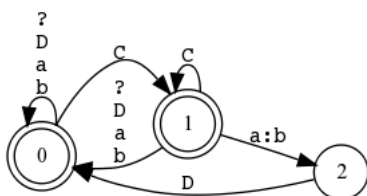
2 Reguläre Ausdrücke und Transduktoren in XFST (10 Punkte)

- a) Zeichnen Sie das Zustandsdiagramm eines minimalen Automaten, der die Sprache $\sim \$[b\ c]$ akzeptiert (4 Punkte).



Lösungen/Diskussion Es muss ein Automat gezeichnet werden, der alle Zeichenketten akzeptiert, welche kein “bc” enthalten. Für “bc” wird deshalb erst eine Sackgasse konstruiert, wobei man den Sackgassenzustand dann auch gleich wieder entfernen kann, weil daraus nie wieder etwas akzeptiert werden kann.

- b) Schreiben Sie einen regulären Ausdruck, der einen Automaten ergibt, wie in der folgenden Abbildung (6 Punkte):



Lösungen/Diskussion

$[a (->) b \mid c _ D]$

Die Formulierung als Restriktion $a:b \Rightarrow c _ D$ wird von XFST nicht unterstützt, da nur Sprachen, aber keine Relationen restringiert werden können. In anderen Formalismen wie SFST könnte dies so gemacht werden, solange die “feasible pairs” gegeben sind.

3 Pluralbildung im Englischen (15 Punkte)

Schreiben Sie ein regelbasiertes xfst-Programm, das zumindest Pluralformen von Wörtern wie (“spies”, “toys”, “shelves”, “wives”, “potatoes”, “foxes”, “cups”) und Singularformen mit den korrekten morphologischen Analysen verknüpft.

```

apply down spies
spy+PL
apply down spy
spy+SG
apply down spys

define NLex {spy}|{toy}|{shelf}|{wife}|{potato}|{cup};

define C [b|c|d|f|g|h|j|k|l|m|n|p|q|r|s|t|v|w|x|y|z];

define NInfl [NLex "+SG"]| [NLex 0:s "+PL"];

define R1 [ y   -> {ie} || C _ s "+PL"];
define R2 [ {f} -> {ve} ||   _ s "+PL"];
define R3 [ [..] -> e   || o _ s "+PL"];

define Cleanup [["+SG"]|["+PL"] -> 0];

read regex NInfl .o. R1 .o. R2 .o. R3 .o. Cleanup;

print words

```

4 Affigierung in Bontoc (10 Punkte)

In der philippinischen Sprache Bontoc ergibt eine Infigierung mit dem Morph “um” ein inchoatives Verb.

Adjektiv	Bedeutung	Verb	Bedeutung
antjoak	gross	um antjoak	grösser werden
kawisat	gut	kum awisat	besser werden
pusiak	arm	pum usiak	ärmer werden

Falls das Adjektiv mit einem Vokal beginnt, wird das Affix “um” vorangestellt, sonst direkt nach dem 1. Konsonanten eingefügt. Implementieren Sie diese Operation in xfst, so dass die Adjektive in die entsprechenden Verben deriviert werden inklusive dem Mehrzeichensymbol +VINCH für inchoatives Verb. Folgende Relation soll erscheinen:

```

apply down kawisat
kumawisat+VINCH

```

Lösung/Diskussion Ein einfacher, aber nicht minimaler Ansatz:

```

define ADJ {antjoak}|{kawisat}|{pusiak};

define V [a|e|i|o|u];

define C [b|c|d|f|g|h|j|k|l|m|n|p|q|r|s|t|v|w|x|y|z];

define R1 [ [..] -> {um} || .#. C _ ];
define R2 [ [..] -> {um} || .#. _ V ];

read regex [ADJ 0:"+VINCH"] .o. R1 .o. R2;

print words

```

5 Typologie (10 Punkte)

a) Fügen Sie den Sprachtypus in die 3. Spalte ein. (4 Punkte)

Nr.	Beispiel (mit engl. Glosse)	Übersetzung	Sprachtypus
1	laudabo praise/3/Sg/Future/Ind	Ich werde loben.	flektierend
2	cöpluklerimizdekiledenmiydi garbage Aff Pl 1/Pl Loc Rel Pl Abl Int Aux Past	War es von denen, was in unserem Müll war?	agglutinierend
3	gou bu ai chi quingcai dog not like eat vegetable	z.B. Hunde mögen kein Gemüse essen	isolierend
4	qaya:liyu:lu:ni kayaks make excellent he Past	Er war hervorragend im Kayak Herstellen.	polysynthetisch

b) Welches sind die unterschiedlichen morphologischen Herausforderungen von Sprachen des 1. bzw. 2. Typs? (6 Punkte)

Lösung/Diskussion Flektierend: Formensynkretismus, d.h. hohe Mehrdeutigkeit bei der Analyse; Allomorphe Stämme und Affixe, d.h. bestimmte Affixe verknüpfen sich nur mit bestimmten Stämmen (z.B. starke Verben, Nomen mit Pluralstämmen)

Agglutinierend: Vokalharmonie; Lange Ketten von Affixen, welche z.T. tausende von möglichen Formen für ein Wort ergibt.

6 Computermorphologie (15 Punkte)

Zählen Sie 5 verschiedene Anwendungen der Computermorphologie auf. Geben Sie ein typisches Beispiel für die Art und benötigte Detailliertheit dieser Anwendung.

Lösung/Diskussion Siehe Skript 3.3.

7 Terminologische Klärungen (10 Punkte)

7.1 Worin unterscheiden sich Stamm und Wurzel? (5 Punkte)

Geben Sie Beispiele!

Lösung/Diskussion Siehe Skript 2.4

7.2 Worin unterscheiden sich Derivation und Konversion? (5 Punkte)

Geben Sie Beispiele!

Lösung/Diskussion Siehe Skript 2.4

8 Computergestützte Lexikographie (10 Punkte)

Welche Unterstützung für die lexikographische oder terminologische Arbeit kann die Computerlinguistik und Sprachtechnologie bieten?

Lösung Einige Stichwörter: Erstellung, Verwaltung und Annotation (Tagging, Lemmatisierung, (partielle) syntaktische Analyse) von Korpora; Abfragesprachen; Konkordanzen (KWIC); Häufigkeits- und Verwendungsanalysen; Kollokationsanalysen; Automatisches Erschliessen von lexikalischen Eigenschaften; Automatische Erkennung von Fachbegriffen und Eigennamen; Automatische Erkennung von Übersetzungskandidaten; ...