

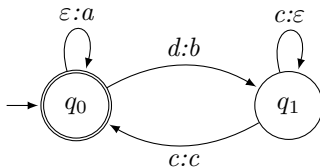
Einführung in die Computerlinguistik

Hausaufgabe 3 (FST), Abgabe 11.05.2014

Laura Kallmeyer

SoSe 2015, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 Betrachten Sie folgenden FST:



- Auf welche Strings bildet der FST die folgenden Eingaben ab? (Es können mehrere Strings sein, eventuell auch unendlich viele, die Abbildung ist nicht immer eindeutig. Man kann den entsprechenden regulären Ausdruck angeben.)
 (a) ε (b) dc (c) $dccc$ (d) $dedcc$
- Welche Strings akzeptiert dieser FST und wie transformiert er sie? (Eine Beschreibung der Transformation in Worten genügt.)

Lösung:

- Angabe der entsprechenden regulären Ausdrücke, da es immer unendlich viele Bilder gibt:
 (a) a^* (b) a^*bca^* (c) a^*bca^* (d) $a^*bca^*bca^*$
- Akzeptiert wird die von $(dc^+)^*$ denotierte reguläre Sprache. Ausgabe ist jeweils ein Wort, in dem jedes d in der Eingabe durch ein b ersetzt wird, von jeder c -Folge wird nur das letzte c in die Ausgabe übernommen, die anderen werden gelöscht, und es werden am Wortanfang und Ende und vor jedem b im Ausgabewort beliebig viele (eventuell auch 0) a s eingefügt.

Aufgabe 2

Erstellen Sie Finite State Transducer, die folgende Formen von "venir" (frz. "kommen") erkennen und deren Zwischenebene bzw. deren Analyse ausgeben:

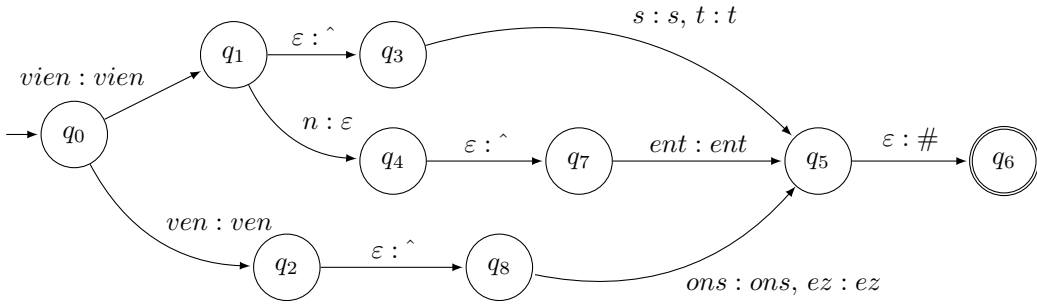
Wortform	Zwischenebene	Analyse	Wortform	Zwischenebene	Analyse
viens	$vien \hat{s}\#$	venir V 1.Sg	venons	$ven \hat{ons}\#$	venir V 1.Pl
viens	$vien \hat{s}\#$	venir V 2.Sg	venez	$ven \hat{ez}\#$	venir V 2.Pl
vient	$vien \hat{t}\#$	venir V 3.Sg	viennent	$vien \hat{ent}\#$	venir V 3.Pl

Zunächst sollen in einem ersten FST die Symbole $\hat{}$ und $\#$ für Morphem- und Wortgrenze eingefügt werden, ein zweiter FST macht dann die morph. Analyse. (Beide FSTs können für Generierung natürlich auch in der anderen Richtung betrieben werden.)

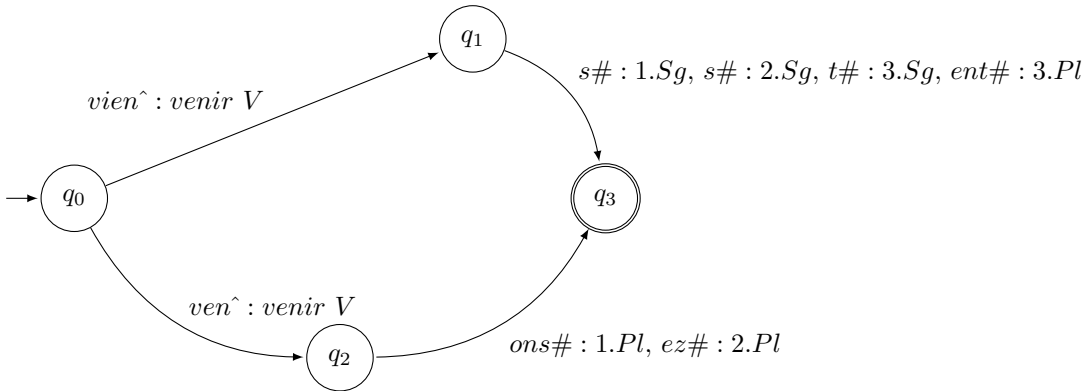
Vorsicht: In der 3.Person Plural gibt es in der Wortform ein zweites n zwischen Lexem und Flexionsmorphem, das in der Zwischenebene nicht mehr da ist.

Lösung:

FST, der Morphemgrenzen markiert:



FST für die morph. Analyse:



Aufgabe 3 Geben Sie einen FST an, der die von $(ab|c)^*$ denotierte Sprache erzeugt, und der 1. Jede Folge ab umdreht, also für jede solche Folge ba ausgibt, und der 2. bei jeder (maximal langen) Folge von cs in der Eingabe das erste c durch ein d ersetzt und die folgenden cs jeweils nichtdeterministisch entweder löscht oder durch ein a ersetzt.

Bsp.: $\langle \text{cccabc}, \text{dabad} \rangle, \langle \text{abccababcabcc}, \text{badbabadbada} \rangle, \langle \text{abcccababcabcc}, \text{badababadbada} \rangle, \dots$

Lösung:

