

# Einführung in die Computerlinguistik

## Hausaufgabe zu FST und Morphologie, Abgabe 07.05.2018

Laura Kallmeyer

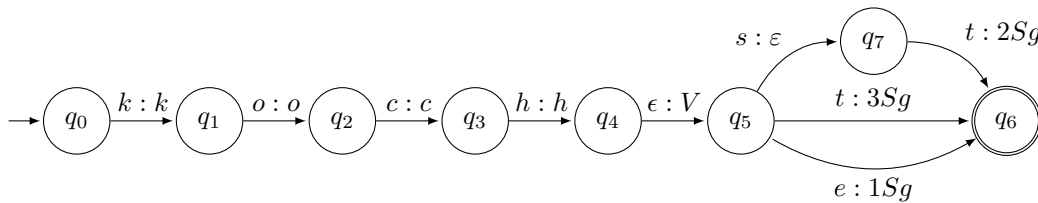
Sommer 2018, Heinrich-Heine-Universität Düsseldorf

**Aufgabe 1** Gegeben ist das folgende Tupel für einen finite state transducer:

- $Q = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7\}$
- $\Sigma = \{k, o, c, h, e, s, t\}$
- $\Delta = \{k, o, c, h, V, 3Sg, 2Sg, 1Sg\}$
- 1.  $\delta(q_0, k, k) = \{q_1\}$  2.  $\delta(q_1, o, o) = \{q_2\}$  3.  $\delta(q_2, c, c) = \{q_3\}$  4.  $\delta(q_3, h, h) = \{q_4\}$   
 5.  $\delta(q_4, \epsilon, V) = \{q_5\}$  6.  $\delta(q_5, e, 1Sg) = \{q_6\}$  7.  $\delta(q_5, t, 3Sg) = \{q_6\}$  8.  $\delta(q_5, s, \epsilon) = \{q_7\}$   
 9.  $\delta(q_7, t, 2Sg) = \{q_6\}$
- Startzustand ist  $q_0$ .
- $F = \{q_6\}$

1. Geben Sie die graphische Darstellung dieses FST an.
2. Geben Sie für zwei der drei Ausdrücke, die der Automat generieren kann, die lexikalische Ebene und die Oberflächenebene an.

Lösung:



1.

Lexikalische Ebene: 

...	k	o	c	h	V	1Sg	...
-----	---	---	---	---	---	-----	-----

2.

Oberfläche: 

...	k	o	c	h	e	...
-----	---	---	---	---	---	-----

Lexikalische Ebene: 

...	k	o	c	h	V	2Sg	...
-----	---	---	---	---	---	-----	-----

Oberfläche: 

...	k	o	c	h	s	t	...
-----	---	---	---	---	---	---	-----

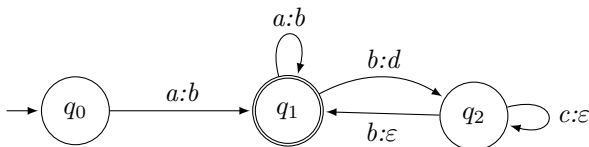
Lexikalische Ebene: 

...	k	o	c	h	V	3Sg	...
-----	---	---	---	---	---	-----	-----

Oberfläche: 

...	k	o	c	h	t	...
-----	---	---	---	---	---	-----

**Aufgabe 2** Betrachten Sie folgenden FST:



1. Auf welche Strings bildet der FST die folgenden Eingaben ab:

(a) aa

(b) abbaaa

(c) aabcbcb

(d) aabbabba

2. Welche Strings akzeptiert dieser FST und wie transformiert er sie? (Eine Beschreibung der Transformation in Worten genügt.)

Lösung:

- (a)  $bb$                                   (b)  $bdbbb$                                   (c)  $bbdd$                                   (d)  $bbdbdb$
- Akzeptiert wird die von  $a(bc^*b)^*$  denotierte Sprache. Als Ausgabe ergibt sich in derselben Reihenfolge für jedes  $a$  aus der Eingabe ein  $b$  und für jeden Teilstring der Form  $bc^*b$  ein  $d$ .

### Aufgabe 3

Erstellen Sie einen Finite State Transducer, der alle flektierten finiten Formen des Verbs "sprechen" im Präsens Indikativ erkennt und deren morphologische Analyse ausgibt. Sie dürfen mehrere aufeinander folgende Buchstaben zusammenfassen.

Lösung:

