

Einführung in die Computerlinguistik

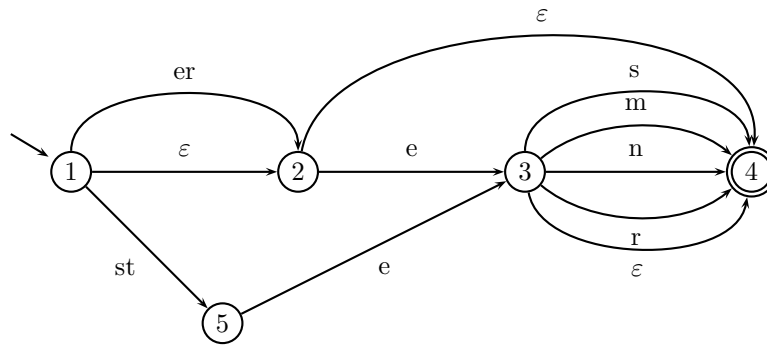
WS 06/07

Übungsblatt 4

Musterlösung

(Mit herzlichem Dank an Jan Hendrik Dithmar für den TeX-Code seiner
Lösungen und den Support!)

1. (a) Das Adjektiv-Ednungs-Diagramm auf der Vorlesungsfolie akzeptiert das Wort *st*. Die superlativendung „kleinst“ kommt aber für sich alleine gar nicht vor. Korrigieren Sie das Diagramm mit möglichst wenig Aufwand so, dass *st* für sich genommen nicht mehr akzeptiert wird (alle anderen Endungen aber wie bisher).



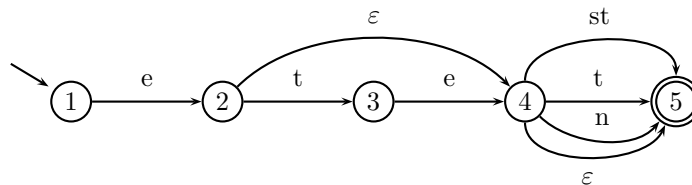
Die einfachste Möglichkeit besteht im Einführen eines neuen Zustandes (5), zu dem eine 'st'-Kante geht. (Von 2 geht eine ϵ -Kante weg zum Endzustand - die wird für die Endung -er und die 'leere' Endung gebraucht.) Es wird sozusagen eine „Umleitung“ von (1) nach (3) gebaut, die -ste erkennt (wie vorher), aber verhindert, dass man nach -st über eine ϵ -Kante in den Endzustand kommt.

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

Die Endungen zum Stamm 'red' :

-e	-ete
-est	-etest
-et	-etet
-en	-eten

Die Endungen haben alle gemeinsam, dass sie mit e Beginnen. Am Ende steht entweder ε (also nichts), st, t oder n, für die Präteritum-Form gibt es das Infix 'te'. Da kein deterministischer Automat verlangt war, kann der DEA unten genau nach diesen Vorgaben konstruiert werden.



2. (a) *In der Vorlesung haben wir ein Diagramm betrachtet, das einige zulässige Wortartketten für Nominalausdrücke spezifiziert. Das Diagramm spezifiziert Nominalausdrücke mit Artikel (Determinator), Gattungssubstantiv (Nomen) und einem potentiellen pränominalen adjektivischen Attribut. Erweitern bzw. modifizieren sie das Diagramm so, dass es auch Nominalausdrücke mit post-nominalen Präpositionalausdrücken umfasst. Also z.B.:*

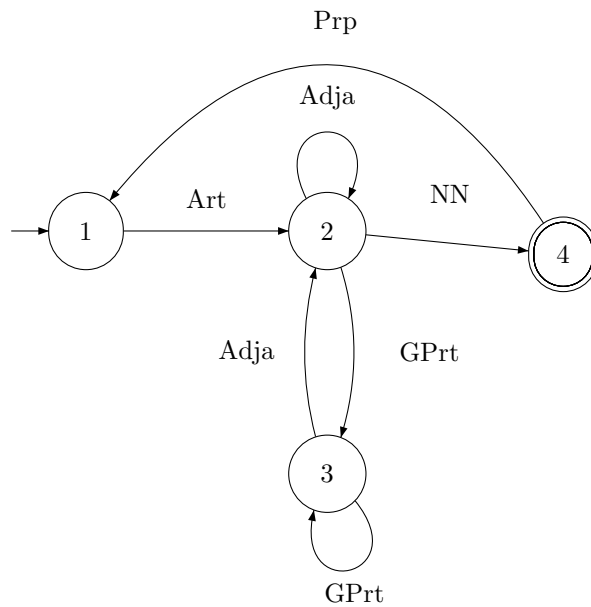
Det N (das Auto)

Det A A N (das neue schnelle Auto)

Det A N Prp Det N (das grüne Auto auf dem Parkplatz)

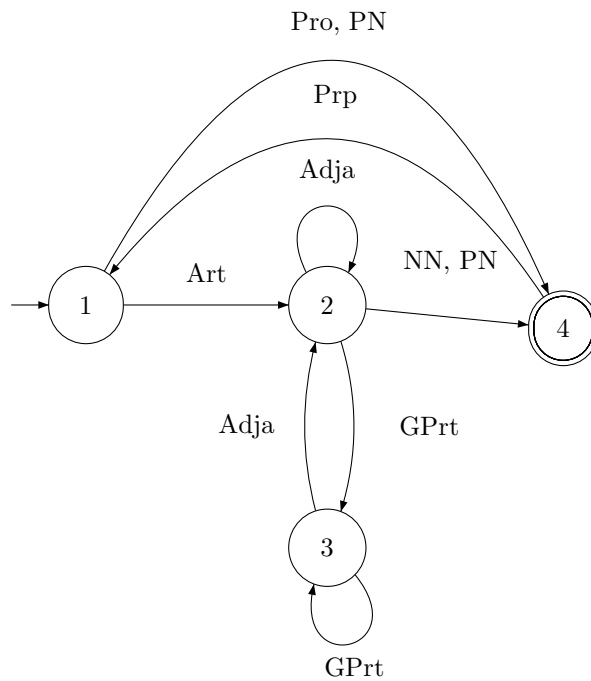
Det N Prp Det N Prp A N (das Auto auf dem Parkplatz beim neuen Institutsgebäude)

Präpositionalausdrücke sind Ausdrücke, bei denen eine Präposition von einem Nominalausdruck gefolgt wird. Daher wird vom Endzustand eine Kante mit einer Präposition zum Anfangszustand gezogen:



- (b) Versuchen sie, Nominalausdrücke mit Personalpronomen (*Pro*) und Eigennamen (*PN* für „proper noun“) im Diagramm mit zu berücksichtigen.

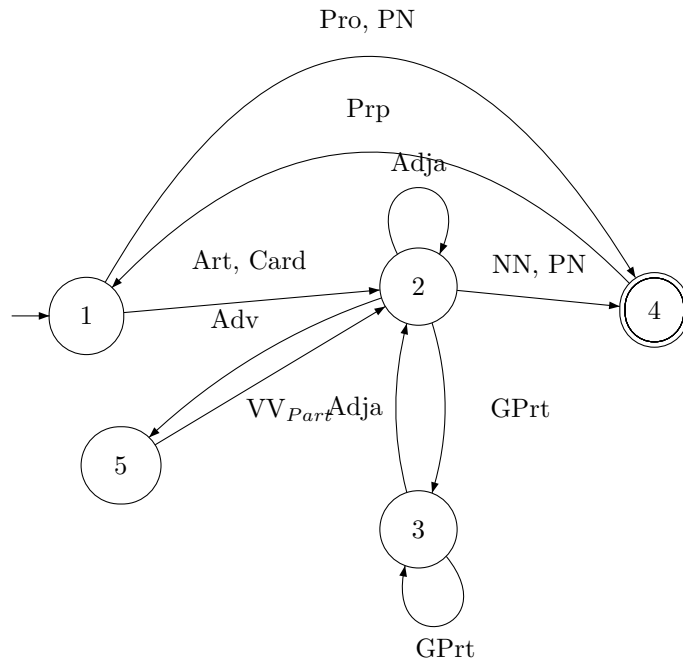
Die Idee ist hier, dass Pronomen alleine als Nominalausdruck stehen können (statt *das Auto fährt es fährt*), Eigennamen genauso (*Peter fährt*). Es gibt aber auch Nominalausdrücke - auch z.B. solche mit Adjektiven - wo Eigennamen nach einem Artikel stehen können (*Die kleine Maria möchte bitte vom Kinderparadies abgeholt werden.*)



- (c) *Geben Sie zwei unterschiedliche Beispiele für Nominalkausdrücke, die durch das Zustandsdiagramm nicht abgedeckt sind. Erweitern Sie das Diagramm so, dass es auch diese Fälle akzeptiert.*

Ein hier gegebenes Beispiel sind Sätze mit Numeralia (*zwei* kleine Autos), wobei die Artikel durch Numeralia ersetzt werden können (*Card*, für cardinal number).

Weiterhin wurden Sätze mit Nominalphrasen mit Partizipien und Adverbien nicht akzeptiert (der *laut pfeifende* Vogel). Dies wird durch die Schleife über den neuen Zustand (5) gewährleistet (Adv = Adverb, VV_{Part} = (Voll-)Verb im Partizip Präsens)



3. (a) (Lese-Task)
- (b) *Eine ältere Version der MS Word-Rechtschreibprüfung hat mir vor Jahren als Alternativen für ?semantisch? die Wörter ?seemännisch? und ?romantisch? angegeben. Bestimmen Sie die im Text genannte Levenshtein-Distanz zwischen den drei Wörtern (also auch zwischen ?seemännisch? und ?romantisch?).*

Berechnung der Levensthein-Distanzen:

semantisch	s → r	semantisch	+e	seemännisch	n → t
remantisch	e → o	seemantisch	a → ä	seemäntisch	ä → a
romantisch		seemäntisch	t → n	seemantisch	s → r
		seemännisch		reemantisch	e → o
				roemantisch	-e
				romantisch	
2		3		5	

Die Reihenfolge der Schritte und der Begriffe spielt natürlich keine Rolle.

- (c) *Sagen Sie in kurzen Worten, wozu man bei der Rechtschreibprüfung kontextabhängige Korrektur benötigt.*

Viele Rechtschreibfehler führen zu existierenden, aber im Kontext falschen Worten, wie z.B. fehlende Kongruenz (*das kleines Haus*), Missachtung von Verb-Valenz (*der Junge schläft den Tisch*) und andere mehr.