

Einführung in die Computerlinguistik

Morphologische Verarbeitung

Teil 2: Morphologische Verarbeitung

1. Motivationen und Ziele
2. Morphologische Grundbegriffe
3. Komponenten und mögliche Probleme der morphologischen Verarbeitung
4. Morphologische Modelle

2.1 Motivation (Computerlinguistik)

- Wie können die systematischen Beziehungen zwischen Wörtern und Wortformen erfasst werden? Wie könnten Regeln zur Wortbildung aussehen?
- Geschickte Speicherung des Lexikons (kein Vollformenlexikon)
⇒ Formalisierung morphologischer Phänomene
- Vereinfachung der weiteren Verarbeitung (z.B. durch Lemmatisierung)

2.1 Motivation (Psycholinguistik)

- Wie ist unser Lexikon organisiert? Vollformen- oder minimales Lexikon?
 - Weder das eine noch das andere (belegt durch Priming-Experimente)
 - Versprecher:
 - Da müssen noch Einkünfte ausgeholt werden.
 - Ich weiß ja nicht, was da im Koch topft.
 - easy enoughly
- ⇒ Mentales Lexikon scheint eine Repräsentation der morphologischen Struktur von Wörtern zu beinhalten

2.1 Ziele

- Morphologisches Parsing von Wortformen (Analyse)
- Erzeugen entsprechender Wortformen anhand der morphologischen Bestandteile (Generierung)

Wortform	Morphemanalyse
<i>Hunde</i>	<i>Hund + N(mask) + PL</i>
<i>Mütter</i>	<i>Mutter + N(fem) + PL</i>
<i>gehst</i>	<i>geh + V + 2.Sg.Präs.</i>
<i>ansprichst</i>	<i>an + sprech + V + 2.Sg.Präs.</i>

2.2 Morphologische Grundbegriffe

- **Wort / Lexem:** abstrakte Einheit, die verschiedenen Formen zugrunde liegt und dem Eintrag im Lexikon entspricht
- **Wortformen:** verschiedene Formen, denen ein Lexem zugrundeliegt
- **Paradigma:** (Teil-)Menge der Wortformen eines Lexems

	sg.	pl.
1	<i>gehe</i>	<i>gehen</i>
2	<i>gehst</i>	<i>geht</i>
3	<i>geht</i>	<i>gehen</i>

2.2 Morphologische Grundbegriffe

- **Flexion:** Zwerg-s; Zwerg-e
- **Wortbildung:**
 - **Komposition:** Garten-Zwerg
 - **Derivation:** zwerg-enhaft; Zwerg-lein
- **Morphem:** Kleinste bedeutungstragende Einheit
 - **Freie Morpheme:** können ohne den Kontext anderer Morpheme geäußert werden [Zwerg]
 - **Gebundene Morpheme:** können nicht ohne den Kontext anderer Morpheme geäußert werden [Zwerg-e]

2.2 Morphologische Grundbegriffe

- **Morphem:**

- **Grundmorphem / Wurzel:** Ausgangspunkt für Derivation und Flexion [**Zwerg**]
- **Affixe:** sind periphere gebundene Morpheme:
 - * **Präfixe:** Affixe, die **vor** der Wurzel stehen [**Ver-**kauf]
 - * **Suffixe:** Affixe, die **nach** der Wurzel stehen [kauf-**en**]
 - * **Circumfixe:** Affixe, die sowohl **vor** als auch **nach** der Wurzel stehen [**ge-**bet-**et**]

2.2 Morphologische Grundbegriffe

- **Stamm:** Morphemkombination ohne Flexionssuffix [**kauf**; **ver-kauf**]
- **Allomorphe:** unterschiedliche Realisierung desselben Morphems (**Morphe**) in verschiedenen Kontexten [z.B. Pluralmorphem: Kind-**er**, Text-**e**]
- **Nicht-Konkatenativität:** Ausdruck einer Kategorie wie *Plural* oder *Imperfekt* nicht durch ein Morphem, sondern durch andere Mittel wie Umlautungsprozess [Mutter → **Mütter**] oder Ablautung [stoßen → **stieß**]

2.3 Komponenten

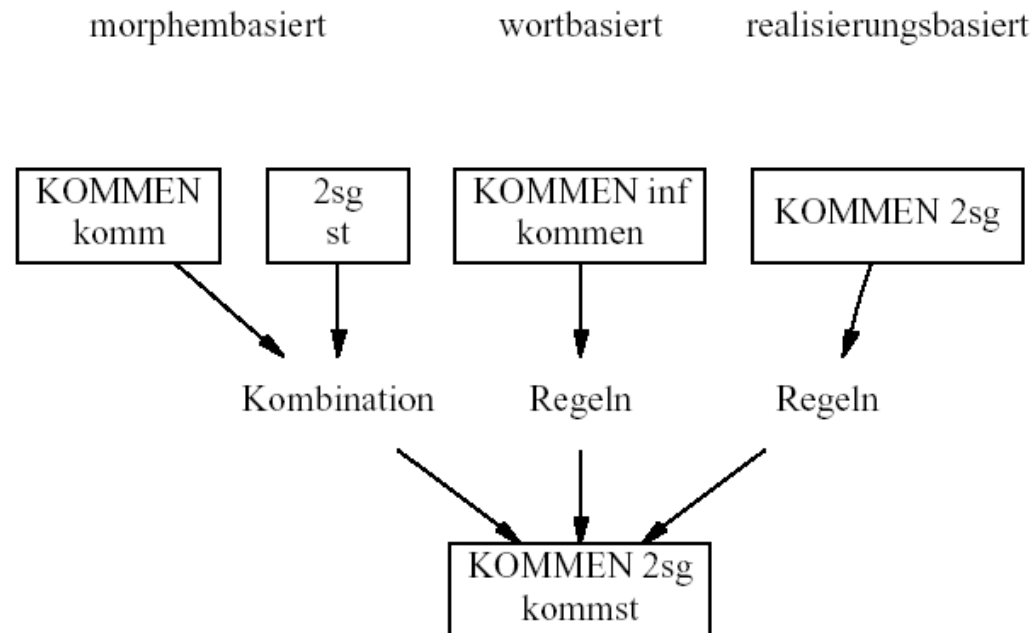
1. **Lexikon** mit Liste von Stämmen und Affixen; Information über deren Beschaffenheit [z.B. Verb, Nomen etc.]
2. **Morphotaktik**: Modell der Morphemverknüpfungsmöglichkeiten
3. **phonologische Regeln**: z.B. Auslautverhärtung, Ablaut, Umlaut
4. **orthographische Regeln**: z.B. *city* → *cities*

2.3 Mögliche Probleme

- **Neutralisierung:** ein Kontrast wird bisweilen neutralisiert, d.h. Wortformen spiegeln nur einen Teil ihrer Funktion wieder [Haken = sg./pl.] (Synkretismus)
⇒ Wie kann man die „gemeinte“ Funktion ermitteln?
- **Nichtkonkatenativität:** Wie lassen sich diese Phänomene erfassen?
- **Wann laufen morphologische Prozesse regelgeleitet ab? Wann gibt es Ausnahmen?**
- **Wie lassen sich Allomorphien erfassen?**

2.4 Morphologische Modelle

- Morphembasierte Ansätze
- Wortbasierte Ansätze
- Realisierungsbasierte Ansätze



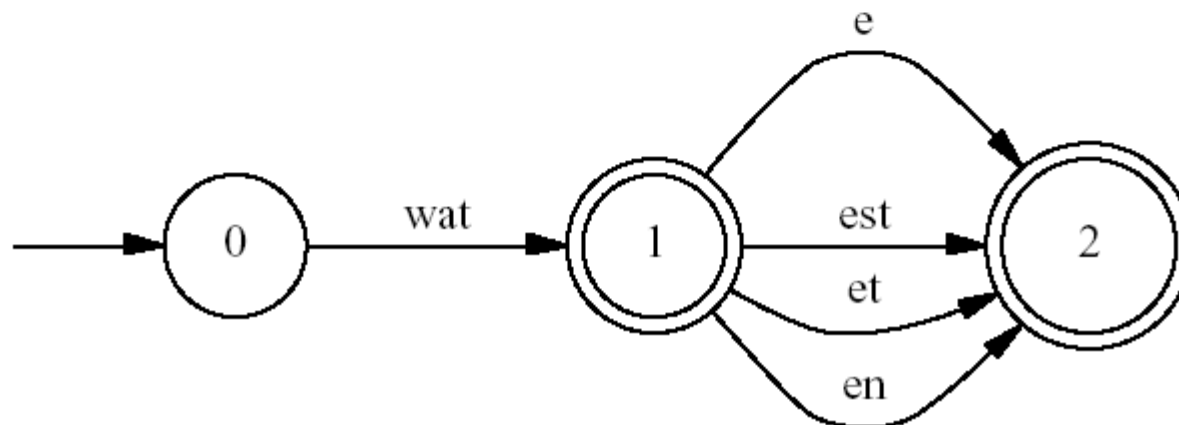
2.4 Wortbasierte/realisierungsbasierte Ansätze

- **Wortbasierte Ansätze:** Bilden neuer Wörter aus anderen Wörtern durch Anwendung von Regeln

- **Realisierungsbasierte Ansätze:**
 - Ausgegangen wird von der vorgegebenen Bedeutung oder Funktion einer Wortform
 - Regeln legen die Realisierung der Wortform fest
 - Nicht-konkatenative Phänomene sind unproblematisch
 - Regelwerk durch Anwendung von Defaultregeln unkompliziert

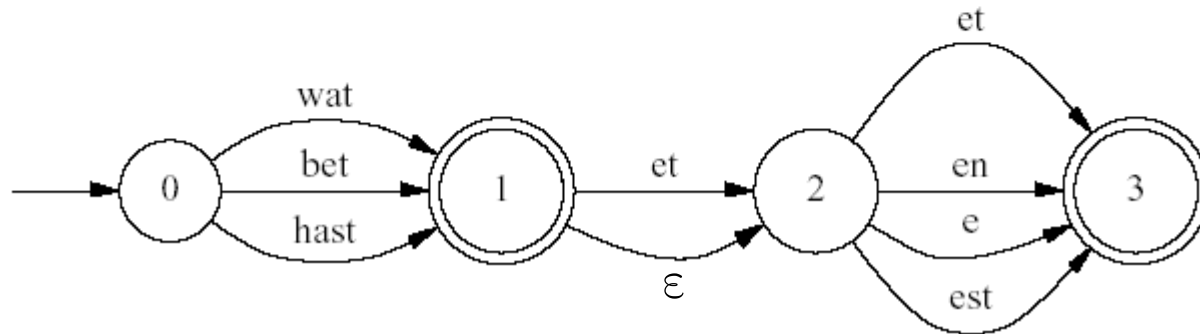
2.4 Morphembasierte Ansätze: FSA

- Formalismus zur Modellierung morphologischer Prozesse: **Endliche Automaten (FSA)**
- FSA kodiert Mengen von (wohlgeformten) Ketten
- Beispiel: Rechtschreibprüfung



2.4 Morphembasierte Ansätze: FSA

Einfacher FSA zur Erkennung der Präsens- und Präteritumformen von *waten*, *beten* und *hasten*:



2.4 Two-Level-Morphology

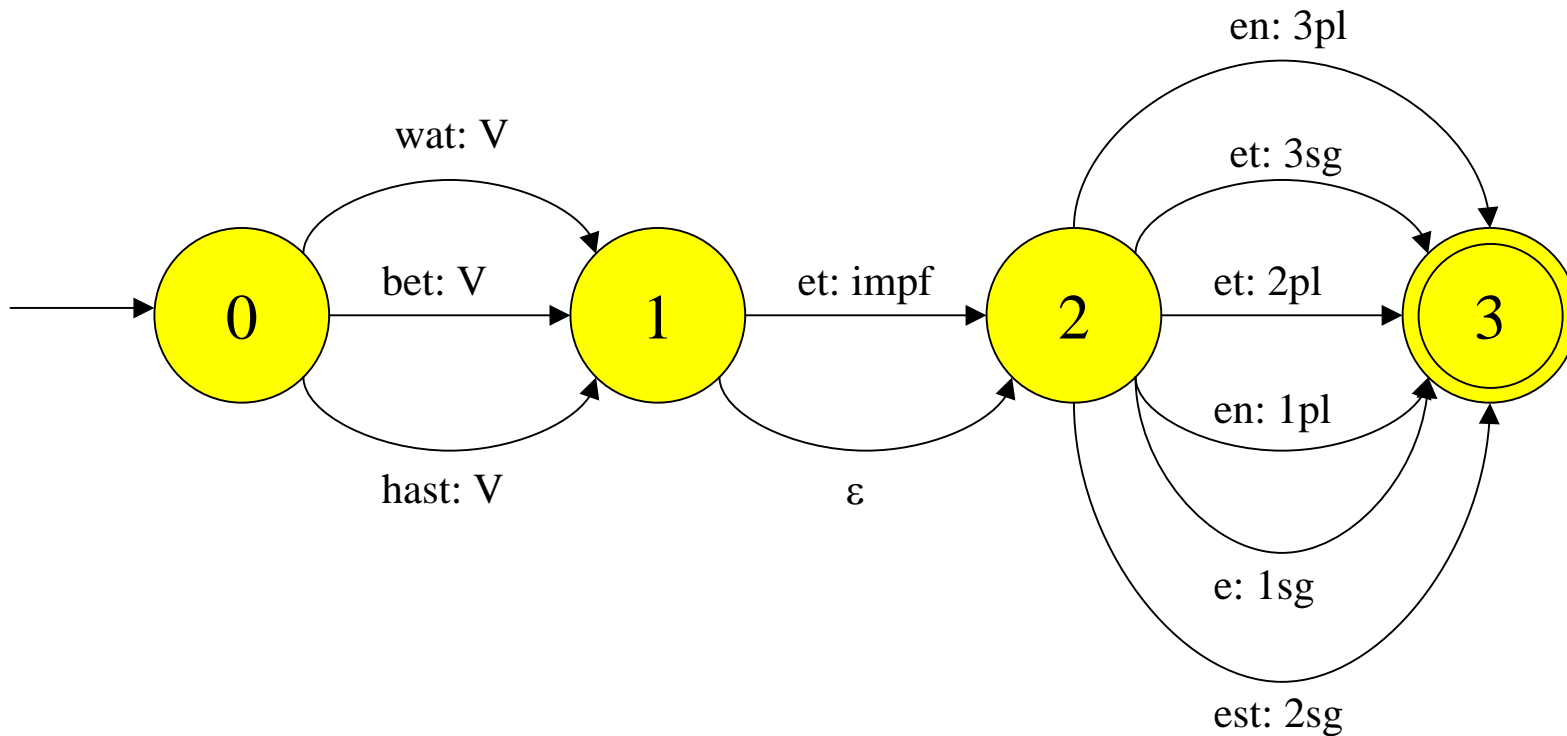
- Wortformen sollen nicht nur auf Wohlgeformtheit überprüft, sondern auch analysiert werden [*beteten* = bet- Imp 1./3.p1]
- Mittels **Two-Level-Morphology** (Kaskenniemi 83) kann eine Beziehung zwischen der **lexikalischen Ebene** (Aneinanderreihung einzelner Morpheme) und der **Oberflächenebene** (entsprechende Wortform) hergestellt werden

Lexikalische Ebene	c	a	t	+N	+PL
Oberflächenebene	c	a	t	s	

2.4 Morphembasierte Ansätze: FST

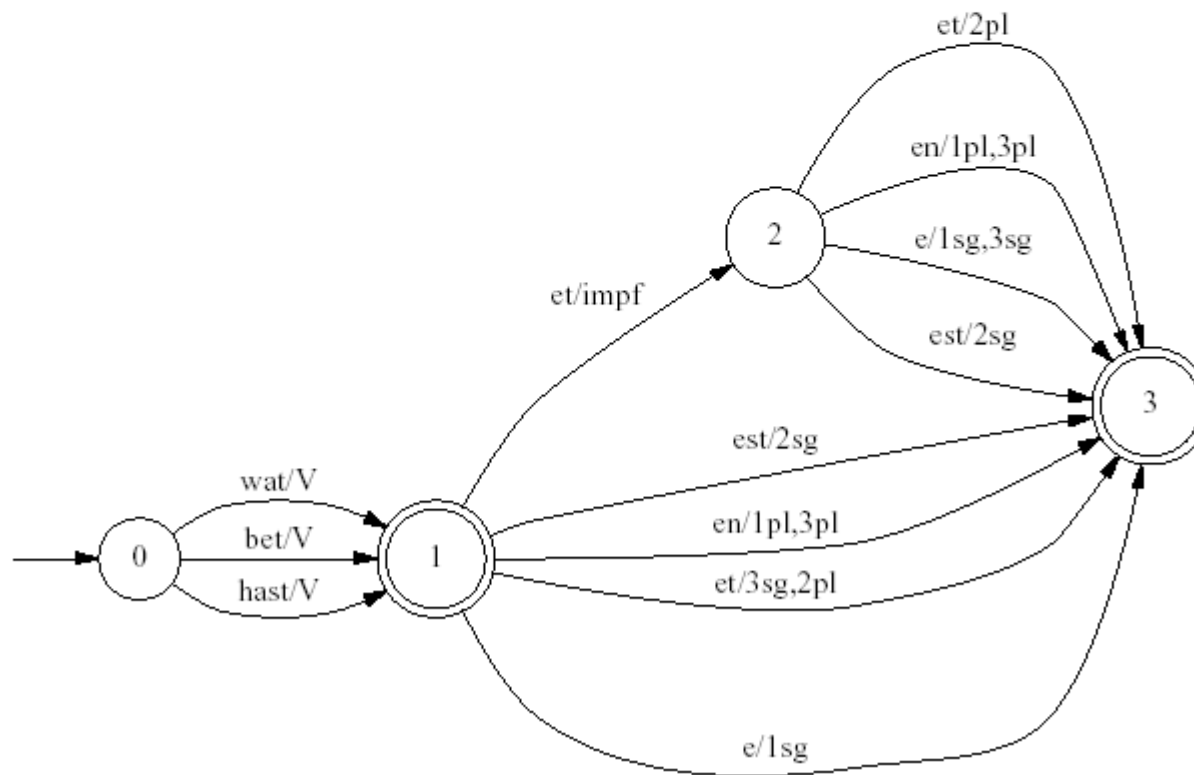
- Automat, der diese Verbindung herstellen kann, ist ein **Finite State Transducer (FST)**
- FST kodiert Abbildungen von Ketten auf andere Ketten
- FST kann sowohl aus der Analyserichtung [*beteten* → *bet- Imp 1./3.p1*] wie auch aus der Generierungsrichtung angewandt werden [*bet- Imp 1./3.p1* → *beteten*]

Verbformen in einem FST (fehlerhaft)



2.4 Morphembasierte Ansätze: FST

Verbformen von *waten*, *beten* und *hasten* in einem FST

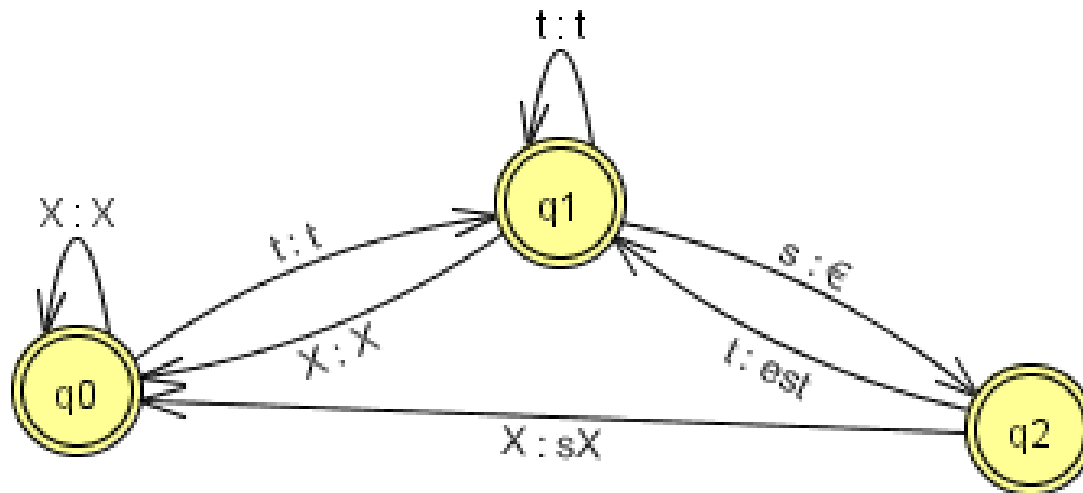


Codierung einer morphophonologischen Regel

hol-st wat-est
schau-st hast-est

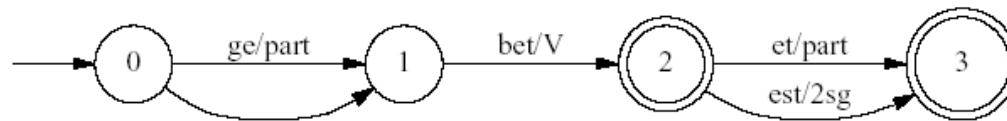
(phonologisch konditionierte Allomorphie)

$\varepsilon \rightarrow e : t_st$

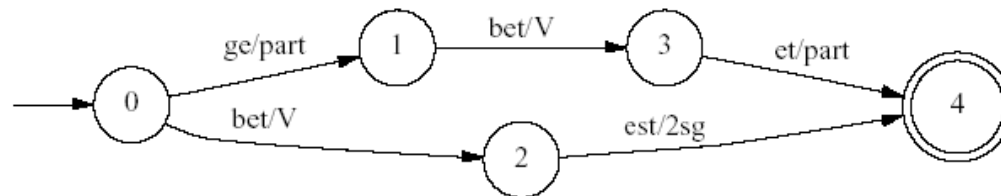


nichtlokale Abhängigkeiten (Zirkumfixe)

Verbflexion mit Partizipialformen (FST ohne Beschränkungen):

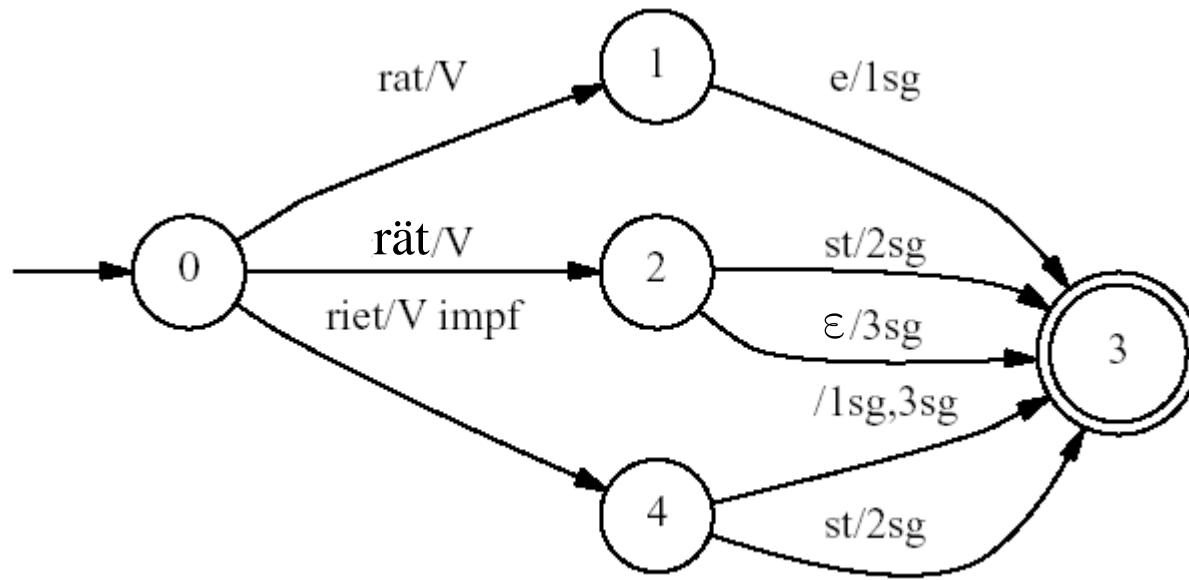


Verbflexion mit Partizipialformen (korrekter FST):



2.4 Morphembasierte Ansätze: FST

Präsens- und Imperfektformen (Sg) von *raten* im
FST_{Ablaut}



2.4 Morphembasierte Ansätze: FST

Problem des FST_{Ablaut}: Zugehörigkeit der Stämme zu einem zugrundeliegenden Morphem wird nicht erkannt

Idee: Einfügen einer Zwischenebene

Lexikalische Ebene	rAt	3.Sg
Zwischenebene	rAt	@
Oberflächenebene	rät	

Zwischen jeder Ebene agiert ein FST

Übungsaufgaben zu FST

I) 1) Erstellen Sie einen Finite State Transducer, der die Steigerungsformen folgender Adjektive erkennt:

schlau - schlauer - schlausten

klein - kleiner - kleinsten

fein - feiner - feinsten

Bei Eingabe von *schlauer* soll der FST beispielsweise Adj + Komp (für *Komparativ*) ausgeben, bei Eingabe von *feinsten* etwa Adj + Sup (für *Superlativ*).

2) Erstellen Sie einen Finite State Transducer, der möglichst viele flektierte Formen des Verbs *bauen* erkennt.

3) Erstellen Sie einen Finite State Transducer, der die deklinierten Formen folgender Nomen erkennt: Kind, Hund, Frau, ...

II) 1) Erstellen Sie einen Finite State Transducer, der die e-Epenthese in *nettesten* und *sanftesten* vollzieht.

2) Erstellen Sie einen Finite State Transducer, der die Auslautverhärtung im Deutschen am Wortende vollzieht.

Übungsaufgaben zu EA

Machen Sie sich mit dem Programm JFLAP vertraut und konstruieren Sie folgende endliche Automaten:

Unser Alphabet bestehe nur aus 'a' und 'b'

- 1) Erstellen sie einen endlichen Automaten, der die Sprache der Wörter erkennt, in denen eine ungerade Anzahl von a's vorkommt.
- 2) Erstellen sie einen endlichen Automaten, der die Sprache der Wörter erkennt, in denen höchstens 5 a's vorkommen.
- 3) Erstellen sie einen endlichen Automaten, der die Sprache der Wörter gerader Länge erkennt.
- 4) Erstellen sie einen endlichen Automaten, der die Sprache aller Wörter außer 'abb' erkennt.