

# Einführung in die Computerlinguistik – Morphologie (morphembasiert mit Transduktoren)

Dozentin: Wiebke Petersen

8.12.2009

# Morphologische Grundbegriffe

**Wort / Lexem:** abstrakte Einheit, die verschiedenen Formen zugrunde liegt.

**Wortform:** verschiedene einem Lexem zugrundeliegenden Formen

**Paradigma:** Menge von Wortformen eines Lexems.

**Synkretismus:** Zusammenfallen verschiedener Wortformen

# Beispiel

## Lexem: GEHEN

gehe gehen

gehst geht

geht gehen

...

ging gingen

gingst gingt

ging gingen

...

# Beispiel

## Lexem: GEHEN

gehe gehen

gehst geht

geht gehen

...

ging gingen

gingst gingt

ging gingen

...

# Wortbildung

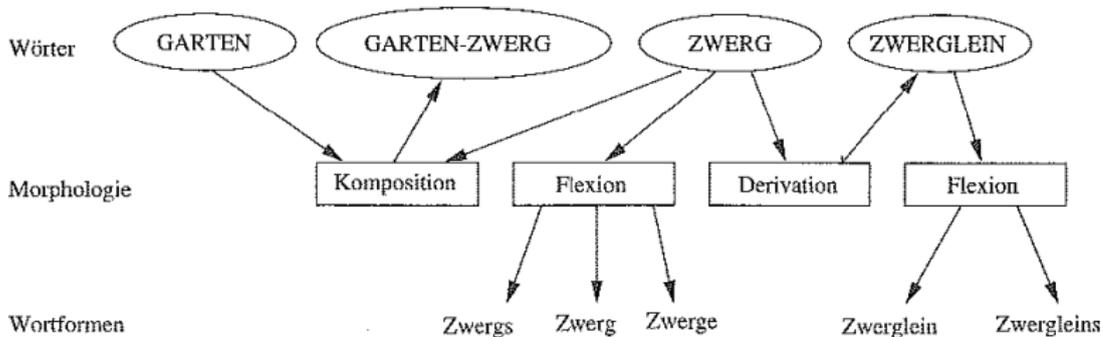


Abbildung 3.4: Flexion und Wortbildung

**Komposition:** Wort + Wort  $\mapsto$  Wort

**Derivation:** Wort + (gebundenes Morphem)  $\mapsto$  Wort

**Flexion:** Wort + Flexionsmorphem  $\mapsto$  Wortform

(Abbildungen aus Carstensen et. al. 2004)

# Morpheme

**Morphem:** Kleinste bedeutungstragende Einheit (genauer, Menge von bedeutungsgleichen Einheiten in komplementärer Verteilung)

**freies Morphem:** Hund, Kind, auf, und, ...

**gebundenes Morphem:** -s, ver-, -lein, ...

**Wurzel:** (Morphem – Ausgangspunkt für Flexion und Derivation)  
lauf, Kind, Hund, schwarz, ...

**Stamm:** (Morphemcluster ohne Flexionsaffixe) vergleich, verkauf, enthält, ...

**Allomorph:** Bsp. Pluralmorphem:  $\{-e, -er, -s, \dots\}$  (-e, -er und -s sind allomorph)

# Regularitäten, Subregularitäten und Ausnahmen

## Beispiel: Pluralbildung im Englischen

### regulär

- door – doors
- dog – dogs
- cat – cats

### subregulär

- fox – foxes
- ibis – ibises
- trush – trushes

### Ausnahmen

- child – children
- goose – geese
- mouse – mice

# Sprachen mit reicher Morphologie

Türkisch:

uygarlaştıramadıklarımızdanmışsınızcasına

*uygar* +*laş* +*tır* +*ama* +*dık* +*lar* +*ımız* +*dan* +*mış* +*sınız* +*casına*

civilized +BEC +CAUS +NABL +PART +PL +P1PL +ABL +PAST +2PL +AsIf

“(behaving) as if you are among those whom we could not civilize”

- +BEC “become”
- +CAUS the causative verb marker (‘cause to X’)
- +NABL “not able”
- +PART past participle form
- +P1PL 1st person pl possessive agreement
- +2PL 2nd person pl
- +ABL ablative (from/among) case marker
- +AsIf derivationally forms an adverb from a finite verb

# Sprachen mit reicher Morphologie

Walisisch:

**LLANFAIRPWLLGWYNGYLLGOGERYCHWYRNDROBWLLLLANTYSILIOGOGOGOCH**

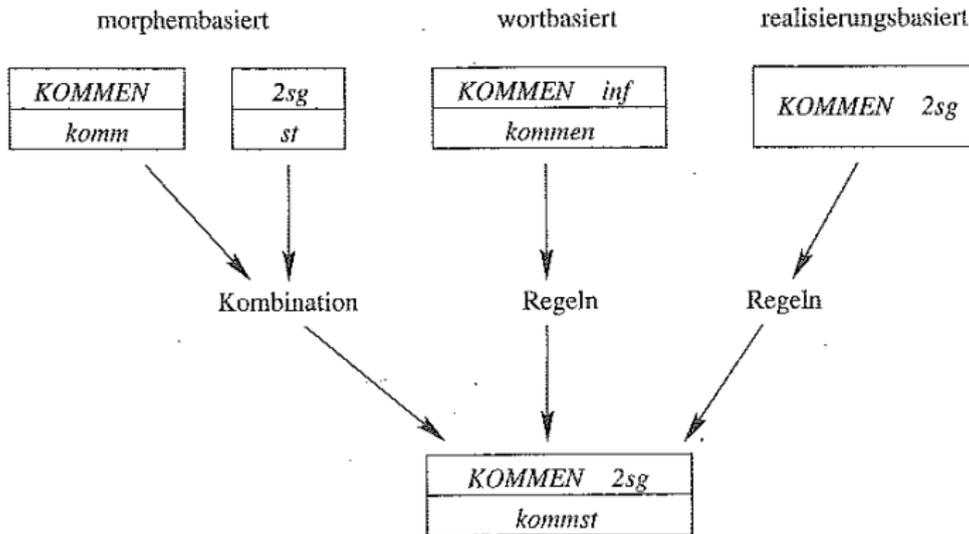
**ST MARYS CHURCH IN THE HOLLOW OF THE WHITE HAZEL NEAR TO THE RAPID WHIRLPOOL OF LLANTYSILIO OF THE RED CAVE**

St Mary's church in the hollow of the white hazel near to the rapid  
whirlpool and the church of St Tysilio of the red cave

# Bedeutung der Morphologie in der CL

- Systematische Erfassung der Beziehungen zwischen Wörtern und Wortformen
- Vermeidung von Vollformlexika (Rechtschreibkorrektur, ...)
- Vereinfachung der sprachlichen Verarbeitung (z.B. durch Lemmatisierung)

# Grundmodelle der generativen Morphologie



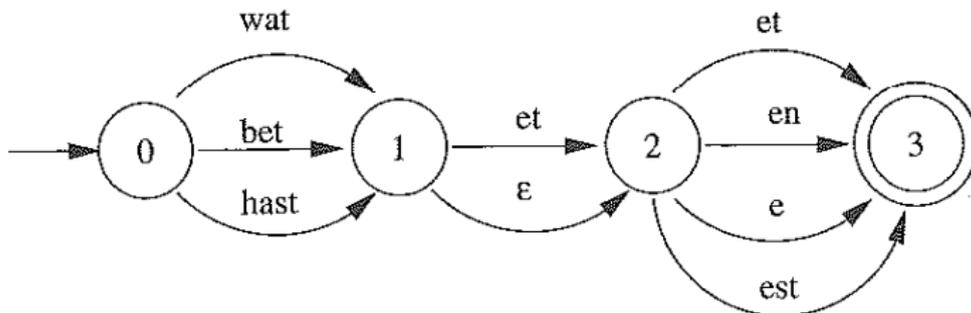
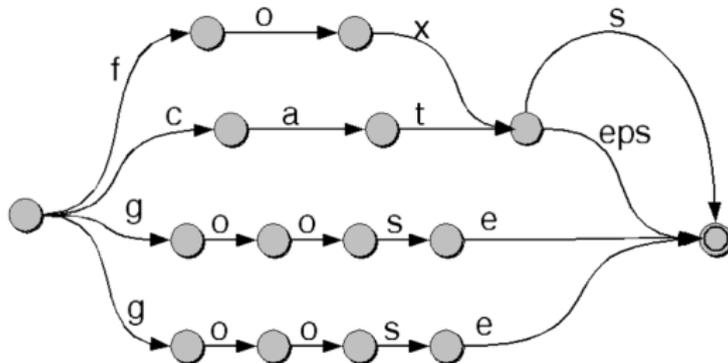
**morphembasiert:** Morpheme kombiniert zu Wortformen

**wortbasiert:** Wortformen werden regelhaft aus Wortformen abgeleitet

**realisierungsbasiert:** Regeln legen fest, wie Wortform aus Bedeutung und Funktion realisiert werden

# Morphologie mit endlichen Automaten

## Erkennung von Wortformen



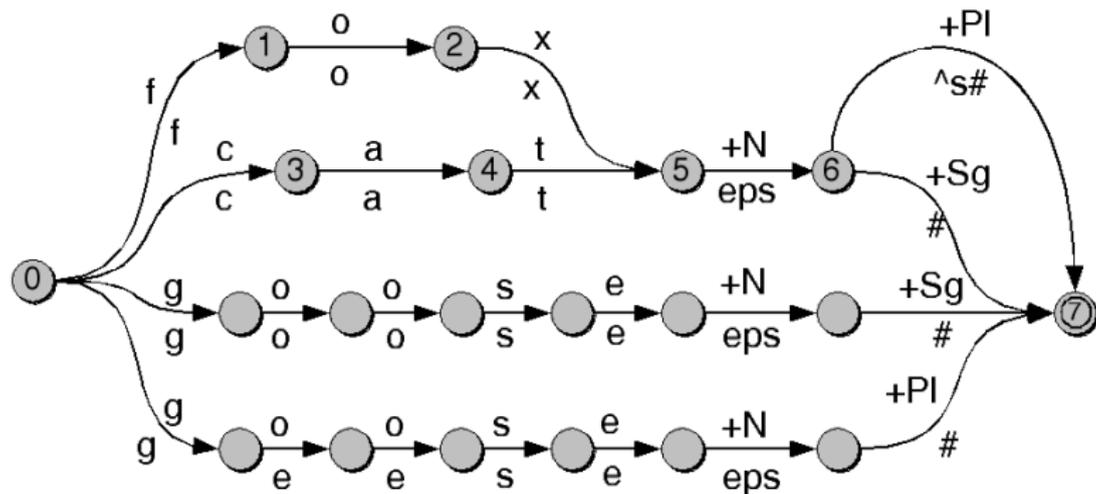
# Parsing / Generierung vs. Erkennung

- Mit endlichen Automaten können wir nur Wörter erkennen
- Wir können nicht Parsen
  - Parsen: Aufbau von Struktur
  - Gewöhnlich haben wir einen String einer Sprache und suchen seine Struktur (Parsing)
  - Oder wir haben eine Struktur und möchten die Oberflächenform generieren (Produktion/Generierung)
- Beispiel
  - von “cats” zu “cat +N +PL” (Parsing)
  - von “cat +N +PL” zu “cats” (Generierung)

# Lösung: Endliche Transduktoren

- Füge ein extra Band / Tape hinzu
- Füge extra Symbole zu den Übergängen hinzu
- Auf dem einen Band wird “cats” gelesen, auf das andere wird “cat +N +PL” geschrieben (oder umgekehrt)

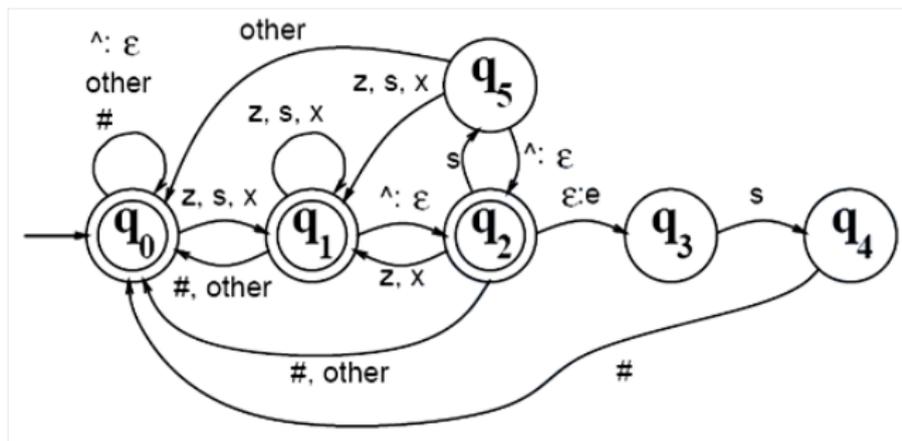
# Transduktor: Englischer Plural 1



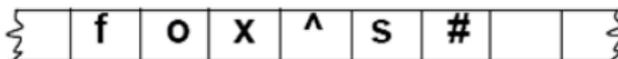
Lexical { f o x +N +PL }

Intermediate { f o x ^ s # }

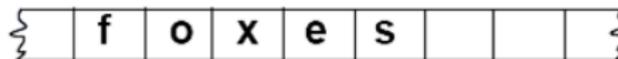
# Transduktor: Englischer Plural 2 (e-Epenthese)



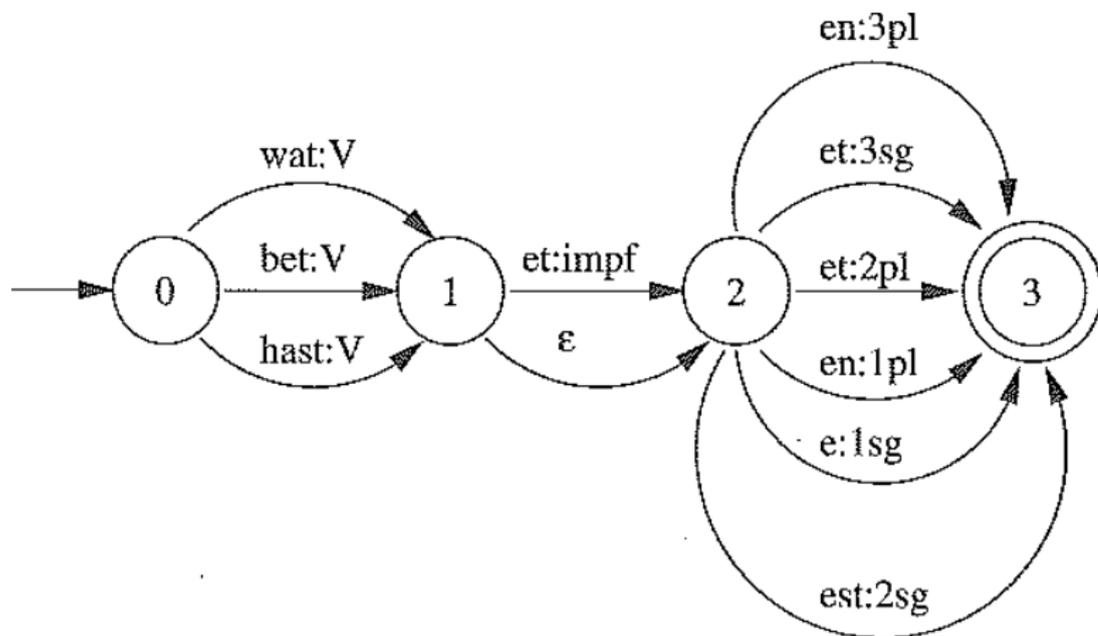
Intermediate



Surface



# Transduktor: Deutsche Verbformen



# Übungsaufgabe

- 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 für die reguläre Flexion der Verben im Englischen.

# Hausaufgaben:

(Abgabe 15.12. um 14 Uhr ; BN: Aufgabe 1 und 2)

- 1 Erklären Sie anhand der Wörter “fox” und “cat”, wie die Transduktoren von Folie 14 und 15 arbeiten.
- 2 Geben Sie für den Transduktor von Folie 15 die Ableitungen für die Ketten “bus<sup>^</sup>s#” und “hat<sup>^</sup>s#” an.
- 3 Arbeiten Sie **eine** der Aufgaben der vorangegangenen Folie aus.