

## Sprachorientierte KI: Syntax und Parsing

- Syntax als Untersuchungsgegenstand
- Wortartendisambiguierung
- Phrasenstrukturgrammatiken
- Parsing mit Phrasenstrukturgrammatiken
- Restrikingierte Phrasenstrukturgrammatiken
- Unifikationsgrammatiken
- Constraint-basierte Grammatiken
- Robustes Parsing



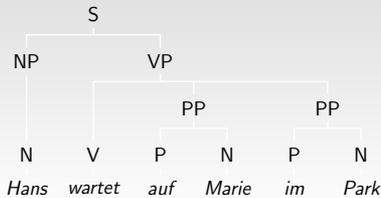
## Wortartendisambiguierung

- Lexikalische Kategorien
- Regelbasierte Tagger
- Stochastische Tagger
- Transformationsbasierte Tagger
- Anwendungen



## Lexikalische Kategorien

- Ziel: syntaktische Regeln sollen generalisierbare Zusammenhänge beschreiben
- Zusammenfassen der terminalen Symbole zu Klassen mit äquivalentem syntaktischen Verhalten.  
~ Wortarten



## Lexikalische Kategorien

- Kriterien für Wortkategorien (RADFORD 1988)
- Phonologische Evidenz: Erklärung systematischer Aussprachevarianten

*We need to **increase** productivity.*  
*We need an **increase** in productivity.*  
*Why do you **torment** me?*  
*Why do you leave me in **torment**?*  
*We might **transfer** him to another club.*  
*He's asked for a **transfer**.*



## Lexikalische Kategorien

- Kriterien für Wortkategorien (2)
  - Semantische Evidenz: Erklärung struktureller Ambiguitäten
- Mistrust wounds.*  
*..., wo die wilden tiere jagen.*  
*Er hat liebe genossen.*

Semantische Eigenschaften sind irrelevant:

Verben	Aktionen	laufen, tragen, lachen, ...
Nomen	Objekte	Tisch, Pferd, Hans, ...
Adjektive	Zustände	krank, glücklich, ...
...	...	...



## Lexikalische Kategorien

- Kriterien für Wortkategorien (3)
- Morphologische Evidenz
- verschiedene Flexionsmuster für Verben, Substantive, Adjektive
  - aber: irreguläre Flexion: starke Verben, *ist*
- unterschiedliche Wortbildungsmuster
  - Steigerungsformen für Adjektive
  - Deverbalisierung: -ung
  - Denominalisierung: -ier, -eln (?)
  - Deadjektivierung: -heit, -keit
  - keine Ableitungen für Präpositionen und Hilfsverben



## Lexikalische Kategorien

- Kriterien für Wortkategorien (4)
- Syntaktische Evidenz: Distributionsklassen
- Nomen
  - Linguistics can be a pain in the neck.*
  - John can be a pain in the neck.*
  - Girls can be a pain in the neck.*
  - Television can be a pain in the neck.*
  - \*Went can be a pain in the neck.*
  - \*For can be a pain in the neck.*
  - \*Older can be a pain in the neck.*
  - \*Conscientiously can be a pain in the neck.*
  - \*The can be a pain in the neck.*



## Lexikalische Kategorien

- Kriterien für Wortkategorien (4)
- Distributionsklassen (2)
- Im Deutschen: Abstraktion von flexivischen Anpassungen erforderlich
- analoge Analysen für Verben, Modalverben, Adjektive, Präpositionen, ...



## Lexikalische Kategorien

- Inventar lexikalischer Kategorien:

N	Nomen	Haus, Hund, Lehrer, ...
V	Verb	suchen, fragen, werden, sein, ...
P	Präposition	auf, unter, zwischen, nach, ...
A	Adjektiv	schön, gut, rot, ...
ADV	Adverb	abends, anders, vielleicht, ganz, ...
M	Modalverben	wollen, dürfen, sollen, ...
D	Determiner	der, diese, ihr, alle, genug, ...



## Lexikalische Kategorien

- Distributionsanalyse lässt Gestaltungsspielraum
    - Engl.: Partikel und Konjunktionen als Präpositionen
    - Engl.: Adjektive und Adverbien als positionelle Varianten der gleichen Kategorie
      - Adjektive modifizieren Nomen
      - Adverbien modifizieren Adjektive, Adverbien, Präpositionen und Verben
- There is a real crisis.*  
*He is really nice.*  
*He walks really slowly.*  
*He is really down.*  
*He must really squirm.*



## Lexikalische Kategorien

- Hauptkategorien (major categories): N, V, A, P
  - Merkmalsrepräsentation für die Hauptkategorien:  $\pm V, \pm N$

	[V +]	[V -]
[N +]	Adjektiv	Nomen
[N -]	Verb	Präposition



## Lexikalische Kategorien

- Ziel: Formulierung kategorienübergreifender Generalisierungen
  - Engl.: nur [N -]-Wortformen erlauben Nominalgruppenkomplemente
    - John loves [Mary] (V + NP)*
    - John bought a present for [Mary] (P + NP)*
    - \* *John's admiration [Mary] (N + NP)*
    - \* *John is fond [Mary] (A + NP)*
  - Ital.: [N +] flektiert nach dem Genus, [N -] nicht
    - bravo ragazzo (guter Junge)*
    - brava ragazza (gutes Mädchen)*
    - bravi ragazzi (gute Jungen)*
    - brave ragazze (gute Mädchen)*



## Lexikalische Kategorien

- feinere Unterteilung der Verben

	[AUX +]	
[AUX -]	[M +]	[M -]
schlafen	wollen	haben
gehen	können	sein
sagen	dürfen	
...	...	



## Lexikalische Kategorien

- offene Wortklassen: produktiv, Neubildungen möglich
  - Nomen, Verben, Adjektive, Adverbien
- geschlossene Wortklassen: relativ fester Bestand, Funktionswörter
  - Präpositionen, Artikel, Pronomen, Konjunktionen, Hilfsverben, Partikel, Zahlwörter



## Lexikalische Kategorien

- Tagsets
  - Kategorieninventare zur Annotation von Korpora
  - teilweise auch morphosyntaktische Subkategorisierung
  - "technische" Tags
    - Fremdwörter, Symbole, Interpunktion, ...



## Lexikalische Kategorien

■ Tagsets		
Penn-Treebank	Marcus et al. (1993)	45
British National Corpus (C5)	Garside et al. (1997)	61
British National Corpus (C7)	Leech et al. (1994)	146
Tiger (STTS)	Schiller, Teufel (1995)	54
Prague Treebank	Hajic (1998)	3000/1000



## Lexikalische Kategorien

### ■ Penn-Treebank (Marcus, Santorini, Marcinkiewicz 1993)

CC	Coordinating conjunction	<i>and, but, or, ...</i>
CD	Cardinal Number	<i>one, two, three, ...</i>
DT	Determiner	<i>a, the</i>
EX	Existential <i>there</i>	<i>there</i>
FW	Foreign Word	<i>a priori</i>
IN	Preposition or subord. conjunction	<i>of, in, by, ...</i>
JJ	Adjective	<i>big, green, ...</i>
JJR	Adjective, comparative	<i>bigger, worse</i>
JJS	Adjective, superlative	<i>lowest, best</i>
LS	List Item Marker	<i>1, 2, One, ...</i>
MD	Modal	<i>can, could, might, ...</i>
NN	Noun, singular or mass	<i>bed, money, ...</i>
NNP	Proper Noun, singular	<i>Mary, Seattle, GM, ...</i>
NNPS	Proper Noun, plural	<i>Koreas, Germanies, ...</i>
NNS	Noun, plural	<i>monsters, children, ...</i>

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 17

## Lexikalische Kategorien

### ■ Penn-Treebank (2)

PDT	Predeterminer	<i>all, both, ... (of the)</i>
POS	Possessive Ending	<i>'s</i>
PRP	Personal Pronoun	<i>I, me, you, he, ...</i>
PRP\$	Possessive Pronoun	<i>my, your, mine, ...</i>
RB	Adverb	<i>quite, very, quickly, ...</i>
RBR	Adverb, comparative	<i>faster, ...</i>
RBS	Adverb, superlative	<i>fastest, ...</i>
RP	Particle	<i>up, off, ...</i>
SYM	Symbol	<i>+, %, &amp; ...</i>
TO	to	<i>to</i>
UH	Interjection	<i>uh, well, yes, my, ...</i>
VB	Verb, base form	<i>write, ...</i>
VBD	Verb, past tense	<i>wrote, ...</i>
VBG	Verb, gerund	<i>writing</i>
VBN	Verb, past participle	<i>written, ...</i>

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 18

## Lexikalische Kategorien

### ■ Penn-Treebank (3)

VBP	Verb, non-3rd singular present	<i>write, ...</i>
VBZ	Verb, 3rd person singular present	<i>writes, ...</i>
WDT	Wh-determiner	e.g. <i>which, that</i>
WP	Wh-pronoun	e.g. <i>what, whom, ...</i>
WP\$	Possessive wh-pronoun	<i>whose, ...</i>
WRB	Wh-adverb	e.g. <i>how, where, why</i>
\$	Dollar sign	<i>\$</i>
#	Pound sign	<i>#</i>
"	left quote	<i>"</i>
'	right quote	<i>'</i>
(	left parentheses	<i>(</i>
)	right parentheses	<i>)</i>
,	comma	<i>,</i>
.	sentence final punct.	<i>., !, ?</i>
:	mid-sentence punct.	<i>;, ,, —, ...</i>

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 19

## Lexikalische Kategorien

### ■ Beispiele

Book/NN/VB that/DT/WDT flight/NN ./.

Book/VB that/DT flight/NN ./.

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 20

## Lexikalische Kategorien

### ■ das Stuttgart-Tübinger Tagset (STTS) (Schiller und Teufel 1995)

ADJA	attributives Adjektiv	das <i>große</i> Haus
ADJD	adverbiales oder prädikatives Adjektiv	er fährt/ist <i>schnell</i>
ADV	Adverb	<i>schon, bald, doch</i>
APPR	Präposition; Zirkumposition links	<i>in</i> der Stadt, <i>ohne</i> mich
APPRART	Präposition mit Artikel	<i>im</i> Haus, <i>zur</i> Sache
APPO	Postposition	ihm <i>zufolge</i> , der Sache <i>wegen</i>
APZR	Zirkumposition rechts	von jetzt <i>an</i>
ART	bestimmter oder unbestimmter Artikel	<i>der, die, das, ein, eine, ...</i>
CARD	Kardinalzahl	<i>zwei</i> Männer, im Jahre <i>1994</i>
FM	Fremdsprachliches Material	Es wird mit "A big fish" übersetzt
ITJ	Interjektion	mhm, ach, tja
ORD	Ordinalzahl	[der] <i>neunte</i> [August]

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 21

## Lexikalische Kategorien

### ■ das Stuttgart-Tübinger Tagset (STTS)(2)

KOUI	unterordn. Konjunktion mit "zu" + Infinitiv	<i>um/anstatt zu</i> leben
KOUS	unterordnende Konjunktion mit Satz	<i>weil, dass, damit, wenn, ob</i>
KON	nebenordnende Konjunktion	<i>und, oder, aber</i>
KOKOM	Vergleichskonjunktion	<i>als, wie</i>
NN	normales Nomen	<i>Tisch, Herr, das Reisen</i>
NE	Eigennamen	<i>Hans, Hamburg, HSV</i>
PDS	substituierendes Demonstrativpronomen	<i>dieser, jener</i>
PDAT	attribuierendes Demonstrativpronomen	<i>jener</i> Mensch
PIS	substituierendes Indefinitpronomen	<i>keiner, viele, man, niemand</i>
PIAT	attrib. Indefinitpron. ohne Determiner	<i>kein/irgendein</i> Mensch,
PIDAT	attrib. Indefinitpron. mit Determiner	ein <i>wenig</i> Bier, <i>beide</i> Brüder
PPER	irreflexives Personalpronomen	<i>ich, er, ihm, mich, dir</i>
PPOSS	substituierendes Possessivpronomen	<i>meins, deiner</i>
PPOSAT	attribuierendes Possessivpronomen	<i>mein</i> Buch, <i>deine</i> Mutter

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 22

## Lexikalische Kategorien

### ■ das Stuttgart-Tübinger Tagset (STTS)(3)

PRELS	substituierendes Relativpronomen	der Hund, <i>der</i>
PRELAT	attribuierendes Relativpronomen	der Mann, <i>dessen</i> Hund
PRF	reflexives Personalpronomen	<i>sich, einander, dich, mir</i>
PWS	substituierendes Interrogativpronomen	<i>wer, was</i>
PWAT	attribuierendes Interrogativpronomen	<i>welche</i> Farbe, <i>wessen</i> Hut
PWAV	adverbiales Interrogativoder Relativpronomen	<i>warum, wo, wann, worüber</i>
PAV	Pronominaladverb	<i>dafür, deswegen, trotzdem</i>
PTKZU	"zu" vor Infinitiv	zu gehen
PTKNEG	Negationspartikel	<i>nicht</i>
PTKVZ	abgetrennter Verbzusatz	er kommt <i>an</i> , er fährt <i>rad</i>
PTKANT	Antwortpartikel	<i>ja, nein, danke, bitte</i>
PTKA	Partikel bei Adjektiv oder Adverb	<i>am</i> schönsten, <i>zu</i> schnell
SGML	SGML Markup	<turnid=n022k_TS2004>
SPELL	Buchstabierfolge	S-C-H-W-E-I-K-L

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 23

## Lexikalische Kategorien

### ■ das Stuttgart-Tübinger Tagset (STTS)(4)

TRUNC	Kompositions-Erstglied	An- und Abreise
VVFIN	finites Verb, voll	du <i>gehst</i> , wir <i>kommen</i> [an]
VVIMP	Imperativ, voll	<i>komm</i> !
VVINF	Infinitiv, voll	<i>gehen, ankommen</i>
VVIZU	Infinitiv mit "zu", voll	<i>anzukommen, loszulassen</i>
VVPP	Partizip Perfekt, voll	<i>gegangen, angekommen</i>
VAFIN	finites Verb, aux	du <i>bist</i> , wir <i>werden</i>
VAIMP	Imperativ, aux	<i>sei</i> ruhig !
VAINF	Infinitiv, aux	<i>werden, sein</i>
VAPP	Partizip Perfekt, aux	<i>gewesen</i>
VMFIN	finites Verb, modal	<i>dürfen</i>
VMINF	Infinitiv, modal	<i>wollen</i>
VMPP	Partizip Perfekt, modal	<i>gekonnt</i> , er hat gehen <i>können</i>
VMXY	Nichtwort, Sonderzeichen enthaltend	<i>3:7, H2O, D2XW3</i>

Wolfgang Menzel: Sprachorientierte KI: Syntax und Parsing – p. 24

## Lexikalische Kategorien

### das Stuttgart-Tübinger Tagset (STTS)(5)

\$	Komma	,
\$	Satzbeendende Interpunktion	. ? ! ; :
\$()	sonstige Satzzeichen; satzintern	- [ ] ( )

### Beispiele (Tiger-Korpus)

Werden/VAFIN sie/PPER diesmal/ADV lachen/VVINFIN //\$.  
kreischen/VVINFIN ?\$.

Mehr/PIAT Zeit/NN wenden/VVFIN die/ART  
US-Bürger/NN nur/ADV für/APPR Arbeiten/NN und/KON  
Schlafen/NN auf/PTKVZ ./\$.



## Regelbasierte Tagger

### ENGTWOL, Universität Helsinki (Voutilainen 1995)

#### zweistufiger Ansatz

- Zuweisung von Wortarthypothesen
- Selektion von Wortarthypothesen

#### reichhaltiges Lexikon mit morphosyntaktischen Merkmalen

#### zwei Stufen

- Morphologische Analyse
- Morphologische Disambiguierung



## Regelbasierte Tagger

### Morphologischer Analysator

- Zweiebenen-Morphologie
- reichhaltige morphosyntaktische Information verfügbar

```
("<round>"  
 ("round" <SVO><SV> V SUBJUNCTIVE VFIN (@+FMAINV))  
 ("round" <SVO><SV> V IMP VFIN (@+FMAINV))  
 ("round" <SVO><SV> V INF)  
 ("round" <SVO><SV> V PRES -SG3 VFIN (@+FMAINV))  
 ("round" PREP)  
 ("round" N NOM SG)  
 ("round" A ABS)  
 ("round" ADV ADVL (@ADVL))
```



## Regelbasierte Tagger

### 35-45% der Token sind mehrdeutig: 1.7-2.2 Analysealternativen pro Wortform

### Hypothesenselektion mit Constraints (1100)

- lineare Abfolge von morphologischen Merkmalen

### Beispiel

- Eingabesatz: *a reaction to the ringing of a bell*

### Lexikoneintrag:

```
("<to>"  
 ("to" PREP)  
 ("to" INFMARK> (@INFMARK>))
```



## Regelbasierte Tagger

### Beispiel

- Constraint

```
("<to>" =0 (INFMARK>) (NOT 1 INF)  
 (NOT 1 ADV)  
 (NOT 1 QUOTE)  
 (NOT 1 EITHER)  
 (NOT 1 SENT-LIM))
```

Streiche die Infinitivlesart, wenn unmittelbar rechts von to  
kein Infinitiv, Adverb, Zitat, *either*, *neither*, *both* oder  
Satzende vorkommt.



## Regelbasierte Tagger

### Qualitätsmessung

- Messung auf einem annotierten Testset ("gold standard")

$$\text{Recall} = \frac{\text{ermittelte korrekte Kategorien}}{\text{gesuchte korrekte Kategorien}}$$

- Recall < 100%: fehlerhafte Klassifikationen

$$\text{Precision} = \frac{\text{ermittelte korrekte Kategorien}}{\text{Gesamtzahl der ermittelten Kategorien}}$$

- Recall < Precision: unvollständige Kategorisierung
- Recall = Precision: vollständig disambiguiertes Output  
→ accuracy, Genauigkeit
- Recall > Precision: unvollständige Disambiguierung



## Regelbasierte Tagger

### ENGTWOL:

- Testset: 2167 Wortformtoken
- Recall: 99.77 %
- Precision: 95.94 %

→ unvollständige Disambiguierung



## Regelbasierte Tagger

### Wie gut sind die Ergebnisse?

#### 1. Obere Schranke: Wie gut ist die Annotation?

- 96-97% Übereinstimmung zwischen Annotatoren  
(MARCUS ET AL. 1993)
- fast 100% Übereinstimmung bei gegenseitiger  
Abstimmung (VOUTILAINEN 1995)
- für deutsche Texte 98.6% (Brants 2000)

#### 2. Untere Schranke: Wie gut ist die Klassifikation?

- Baseline:  
z.B. häufigster Tag (Unigram-Wahrscheinlichkeit)
- Beispiel:  $P(NN|race) = 0.98$   $P(VB|race) = 0.02$
- 90-91% Precision/Recall (CHARNIAK ET AL. 1993)



## Stochastische Tagger

- manuelle Erstellung von Regelsystemen
  - aufwendig
  - fehleranfällig
- Ausweg: trainierbare Verarbeitungskomponenten
  - freie Parameter eines Modells werden aufgrund von Beobachtungsdaten eingestellt
  - Maschinelles Lernen



## Stochastische Tagger

- zwei grundlegende Klassen von Lernverfahren
  - überwachtes Lernen (Lernen mit Lehrer)
    - Ausgangspunkt: Beispielsammlung
    - Paare aus Eingabedaten und gewünschten Verarbeitungsergebnissen
    - annotierte Korpora
  - unüberwachtes Lernen (Lernen ohne Lehrer)
    - Lernprozess extrahiert eigenständig regelhafte Strukturen aus den Daten
    - Beispiel: Clustern von Daten aufgrund inhärenter Regelmäßigkeiten



## Stochastische Tagger

- stochastische Lernverfahren:
  - Schätzen von Wahrscheinlichkeiten in einem stochastischen Modell
- stochastisches Modell versucht die Entstehung der Beobachtungsdaten durch einen stochastischen Prozess zu beschreiben
  - Modellvorstellung: gestörter Kanal



## Stochastische Tagger

- Kanalmodell
  - Zuordnung von Tags zu Wortformen ist nicht deterministisch
  - "Störung" im Kanal ist vom Kontext abhängig
  - Modell mit Gedächtnis: Markov-Modell
  - Gedächtnis wird durch Zustände modelliert
  - Modellparameter beschreiben die Wahrscheinlichkeit eines Zustandsübergangs
    - Transitionswahrscheinlichkeiten:  $P(s_i | s_1 \dots s_{i-1})$



## Stochastische Tagger

- Markov-Modell einer Münze
  - Diagram showing states W and Z with transition probabilities of 0.5.
- Markov-Modell einer gezinkten Münze
  - Diagram showing states W and Z with transition probabilities of 0.7 and 0.3.



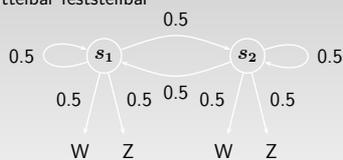
## Stochastische Tagger

- Hidden-Markov-Modelle
  - Beobachtungen sind nicht mehr strikt an Zustände gekoppelt
  - Zustandsfolge beeinflusst die Beobachtungsfolge nur stochastisch
  - Emissionswahrscheinlichkeiten:  $P(o_i | s_1 \dots s_{i-1})$



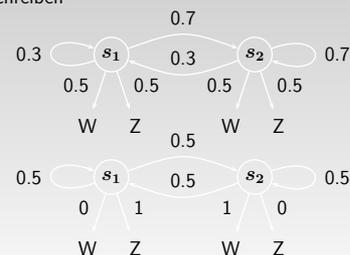
## Stochastische Tagger

- Hidden-Markov-Modelle
  - für einen externen Beobachter ist die Zustandsfolge nicht unmittelbar feststellbar



## Stochastische Tagger

- verschiedene HMMs können die gleichen Beobachtungsdaten beschreiben



- große Flexibilität beim Schätzen der Parameter



## Stochastische Tagger

- Vorgehen:
  - Entwurf: Wahl einer geeigneten Modellstruktur: Zustände und Wahrscheinlichkeiten
  - Training: Schätzen der Wahrscheinlichkeiten auf den Beobachtungsdaten (Lernstichprobe)
  - Klassifikation: Ermittlung der wahrscheinlichsten Zustandsfolge
    - durch welche Zustandsfolge hat das Modell die gegebene Beobachtung vermutlich erzeugt?
  - Evaluation: Messen der Verarbeitungsqualität auf einer separaten Teststichprobe



## Stochastische Tagger

- Modellstruktur eines HMM-Taggers
  - Beobachtung: Wortformen  $w_i$
  - Zustände: Tags  $t_i$
  - Transitionswahrscheinlichkeiten:  $P(t_i | t_1 \dots t_{i-1})$
  - Emissionswahrscheinlichkeiten:  $P(w_i | t_1 \dots t_{i-1})$



## Stochastische Tagger

- Klassifikation: Ermitteln der wahrscheinlichsten Tagsequenz

$$t_j[1, n] = \arg \max_{t[1, n]} P(t[1, n] | w[1, n])$$

- Bayessche Regel

$$t_j[1, n] = \arg \max_{t[1, n]} \frac{P(t[1, n]) \cdot P(w[1, n] | t[1, n])}{p(w[1, n])}$$

- Wahrscheinlichkeit der Wortformenfolge ist für gegebene Beobachtung konstant und beeinflusst daher die Auswahl der Tagfolge nicht

$$t_j[1, n] = \arg \max_{t[1, n]} P(t[1, n]) \cdot P(w[1, n] | t[1, n])$$



## Stochastische Tagger

- Kettenregel für Wahrscheinlichkeiten

$$P(t[1, n]) \cdot P(w[1, n] | t[1, n])$$

$$= \prod_{i=1}^n P(t_i | w_1 t_1 \dots w_{i-1} t_{i-1}) \cdot P(w_i | w_1 t_1 \dots w_{i-1} t_{i-1} t_i)$$

$$t_j[1, n] = \arg \max_{t[1, n]}$$

$$\prod_{i=1}^n P(t_i | w_1 t_1 \dots w_{i-1} t_{i-1})$$

$$\cdot P(w_i | w_1 t_1 \dots w_{i-1} t_{i-1} t_i)$$



## Stochastische Tagger

- 1. vereinfachende Annahme: Wortform ist nur vom aktuellen Tag abhängig

$$t_j[1, n] = \arg \max_{t[1, n]}$$

$$\prod_{i=1}^n P(t_i | w_1 t_1 \dots w_{i-1} t_{i-1}) \cdot P(w_i | t_i)$$

- 2. vereinfachende Annahme: aktuelles Tag ist nur von seinen Vorgängern (nicht den Wortformen!) abhängig

$$t_j[1, n] = \arg \max_{t[1, n]} \prod_{i=1}^n P(t_i | t_1 \dots t_{i-1}) \cdot P(w_i | t_i)$$



## Stochastische Tagger

- 3. vereinfachende Annahme: aktuelles Tag ist nur von den zwei unmittelbaren Vorgängern abhängig

- begrenztes Gedächtnis (Markov-Annahme):  
Trigramm-Modell

$$t_j[1, n] = \arg \max_{t[1, n]} \prod_{i=1}^n P(t_i | t_{i-1} t_{i-2}) \cdot P(w_i | t_i)$$

→ Markov-Prozess 2. Ordnung



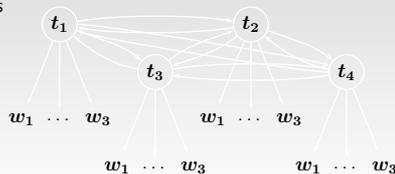
## Stochastische Tagger

- weitere Vereinfachung zum Bigram-Modell
  - stochastische Abhängigkeit ist auf den unmittelbaren Vorgänger beschränkt

$$t_j[1, n] = \arg \max_{t[1, n]} \prod_{i=1}^n P(t_i | t_{i-1}) \cdot P(w_i | t_i)$$

→ Markov-Prozess

1. Ordnung



## Stochastische Tagger

- iterative Berechnungsvorschrift ist nicht praktikabel

$$|T[1, n]| = |T|^n$$

$T[1, n]$  Menge aller Tagsequenzen,  $T$  Tagset,  $n$  Satzlänge

- rekursive Reformulierung: VITERBI, BELLMANN-FORD  
Prinzip der "dynamischen Programmierung"

$$\alpha_n = \max_{t[1, n]} \prod_{i=1}^n P(t_i | t_{i-1}) \cdot P(w_i | t_i)$$

$$\alpha_n = \max_{t_{n-1}} P(t_n | t_{n-1}) \cdot P(w_n | t_n) \cdot \alpha_{n-1}$$

Scores können als Konfidenzabschätzung ausgegeben werden



## Stochastische Tagger

- Schätzen der Wahrscheinlichkeiten:

- Transitionswahrscheinlichkeiten

$$P(t_i | t_{i-2}t_{i-1}) = \frac{c(t_{i-2}t_{i-1}t_i)}{c(t_{i-2}t_{i-1})}$$

- Emissionswahrscheinlichkeiten

$$P(w_i | t_i) = \frac{c(w_i, t_i)}{c(t_i)}$$



## Stochastische Tagger

- ungesehene Übergangswahrscheinlichkeiten

- Backoff: Rückgriff auf Bigram- bzw. Unigram-Wahrscheinlichkeiten

$$P(t_i | t_{i-2}t_{i-1}) = \begin{cases} P(t_i | t_{i-2}t_{i-1}) & \text{falls } c(t_{i-2}t_{i-1}t_i) > 0 \\ P(t_i | t_{i-1}) & \text{falls } c(t_{i-2}t_{i-1}t_i) = 0 \\ & \text{und } c(t_{i-1}t_i) > 0 \\ P(t_i) & \text{sonst} \end{cases}$$



## Stochastische Tagger

- ungesehene Übergangswahrscheinlichkeiten

- Interpolation: Mischen der Trigram mit den Bigram- und Unigram-Wahrscheinlichkeiten

$$P(t_i | t_{i-2}t_{i-1}) = \lambda_1 P(t_i | t_{i-2}t_{i-1}) + \lambda_2 P(t_i | t_{i-1}) + \lambda_3 P(t_i)$$

- $\lambda_1$ ,  $\lambda_2$  und  $\lambda_3$  sind kontextabhängige Parameter

- Randbedingung:  $\lambda_1 + \lambda_2 + \lambda_3 = 1$

- werden auf einer separaten Datenmenge trainiert



## Stochastische Tagger

- ungesehene Wortformen

- Schätzen der Tagwahrscheinlichkeiten aus Suffixen (ggf. auch Präfixen)

- ungesehene Wortartzuordnung

- Glättung (smoothing)

- Umverteilen von Wahrscheinlichkeitsmasse von den gesehenen auf die ungesehenen Ereignisse (discounting)

- z.B. WITTEN-BELL discounting (WITTEN-BELL 1991)

- Wahrscheinlichkeitsmasse der einmal gesehenen Ereignisse wird auf alle ungesehenen Ereignisse verteilt



## Stochastische Tagger

- Beispiel: TnT (BRANTS 2000)

Korpus	Anteil unbek. Wortf.	Genauigkeit		Gesamt
		bekannte Wortformen	unbekannte Wortformen	
PennTB (engl.)	2.9%	97.0%	85.5%	96.7%
Negra (dt.)	11.9%	97.7%	89%	96.7%
Heise (dt.)*				92.3%

\*) Trainingskorpus  $\neq$  Testkorpus

- Maximum entropy tagger (RATNAPARKHI 1996): 96.6%



## Transformations-basierte Tagger

- Idee: schrittweise Korrektur von fehlerhaften Zwischenresultaten (BRILL 1995)

- kontextsensitive Regeln, z.B.

- Change NN to VB when the previous tag is TO

- Regeln werden aus einem Korpus gelernt

1. Initialisierung: Wähle die Tagsequenz mit der höchsten Unigram-Wahrscheinlichkeit
2. Vergleiche das Ergebnis mit der Korpusannotation
3. Generiere eine Regel, die die meisten Fehler beseitigt
4. erneutes Tagging und weiter mit 2.

Abbruch, falls Verbesserung nur noch unbedeutend



## Transformations-basierte Tagger

- Regelgenerierung wird durch Templates gesteuert

- Change tag  $a$  to tag  $b$  when ...

- ... the preceding/following word is tagged  $z$ .

- ... the word two before/after is tagged  $z$ .

- ... one of the two preceding/following words is tagged  $z$ .

- ... one of the three preceding/following words is tagged  $z$ .

- ... the preceding word is tagged  $z$  and the following word is tagged  $w$ .

- ... the preceding/following word is tagged  $z$  and the word two before/after is tagged  $w$ .



## Transformations-basierte Tagger

- Ergebnis des Trainings: geordnete Liste von Transformationsregeln

von	nach	Bedingung	Beispiel
NN	VB	vorheriges Tag ist TO	to/TO race/NN $\rightarrow$ VB
VBP	VB	eines der 3 vorh. Tags ist MD	might/MD vanish/VBP $\rightarrow$ VB
NN	VB	eines der 2 vorh. Tags ist MD	might/MD not reply/NN $\rightarrow$ VB
VB	NN	eines der 2 vorh. Tags ist DT	
VBD	VBN	einer der 3 vorh. Tags ist VBZ	



## Transformations-basierte Tagger

- 97.0% Genauigkeit, wenn nur die ersten 200 Regeln verwendet werden
- 96.8% Genauigkeit mit den ersten 100 Regeln
- Qualität eines stochastischen Parsers auf den gleichen Daten (96.7%) wird mit 82 Regeln erreicht
- extrem lange Trainingszeiten  
≈ 10<sup>6</sup>-fache eines HMM-Taggers



## Anwendungen

- Wortbetonung in der Sprachsynthese  
'content/NN    con'tent/JJ  
'object/NN    ob'ject/VB  
'discount/NN    dis'count/VB
- Ermittlung des Wortstamms (z.B. Textrecherche)
- Klassenbasierte Sprachmodelle für die Spracherkennung
- "flache" Analyse, z.B. zur Informationsextraktion
- Vorstufe zum Parsing, insbesondere für stochastische Parser

