

POS Tagging

Katja Markert, mit einigen Folien von Yannick Versley

Institut für Computerlinguistik
Uni Heidelberg
markert@cl.uni-heidelberg.de

December 4, 2019

- 1 Bisher: Wortbasierte Verfahren
- 2 Bisher: Generatives Modell zur Textklassifikation
- 3 Jetzt: Übergang zu Grammatik
- 4 Jetzt: Start mit Wortartenklassifikation
- 5 Auch mit generativen Verfahren

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging
- 3 Regelbasiertes Tagging
- 4 Die beiden wichtigsten Informationsarten im Tagging
- 5 HMMs: Hidden Markov Modelle für Tagging

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging
- 3 Regelbasiertes Tagging
- 4 Die beiden wichtigsten Informationsarten im Tagging
- 5 HMMs: Hidden Markov Modelle für Tagging

Terminologie

Wortarten, syntaktische Kategorie, Parts of Speech, POS

POS als Generalisierung über Wörter

- Dionysos Thrax von Alexandria (100 v.Chr): 8 Wortarten
- Wortarten in CL: 50 - 200

8 klassische Wortarten: Nomen, Verben, Adjektive, Adverbien, Präpositionen, Determinanten, Konjunktionen, Pronomen

Welche Eigenschaften definieren ein Verb?
Und ein Adjektiv?

Semantische Eigenschaften: Referenz auf Dinge → Nomen?

Distributionelle Eigenschaften: Kookkurrenz mit Artikeln,
Possessiva, Adjektiven → Nomen?

Morphologische Eigenschaften: Können Plural formen → Nomen?

Wortform

- 1 *Er/PER hat von meinem Tellerchen gegessen*
- 2 *Dieser/PDS hat von meinem Tellerchen gegessen*
- 3 *Seiner/POSS Mutter glaube ich kein Wort*

Substitution und Distribution:

- 1 *Der/PDS da hat mich gebissen.* (dieser da)
- 2 *Der/ART Ball ist rund.* (ein)
- 3 *Peter, der/PRELS gestern Geburtstag hatte.* (, welcher)

Substitution:

- 1 *Der Mann will gefallen/VVINP* (reisen)
- 2 *Der Mann ist gefallen/VVPP*. (gereist)
- 3 *Diese Kekse gefallen/VVFIN uns*. (Dieser Keks gefällt uns)
- 4 *The man has walked/VBN*. (fallen)
- 5 *The man walked/VBD*. (fell)

Wortart	Beispiele (DE)	Wortart	Beispiele (DE)
Noun Nomen	Holz, Peter	Det Artikel	der, die, einem
Verb Verben	gegeben, will	Num Zahlen	23, fünf
Adj Adjektive	grün, anschiemig	Pron Pronomen	er, meiner
Adv Adverbien	so, manchmal	Prt Partikel	(bringt mich) um (das ist) ja (prima)
Adp Prä-/Adpositionen	vor, um..herum	X Abkürzungen etc.	XY12, ami, Lach- (und Sachgeschichten)
Conj Konjunktionen	und, aber, weil	. Zeichensetzung	. , ! ? () :

Petrov/Das/McDonald LREC 2012: A Universal Part-of-Speech Tagset

Wortart	Beispiele (EN)	Wortart	Beispiele (EN)
Noun Nomen	castle, Susan	Det Artikel	the, a, an
Verb Verben	gave, can	Num Zahlen	23, five
Adj Adjektive	green, slow	Pron Pronomen	he, her, who, whose
Adv Adverbien	slowly	Prt Partikel	(look sth) up
Adp Prä-/Adpositionen	in, after, before	X Abkürzungen etc.	XY132, mulher, uh
Conj Konjunktionen	and, &, but	. Zeichensetzung	. , ! ? () :

Petrov/Das/McDonald LREC 2012: A Universal Part-of-Speech Tagset

Geschlossen/closed

- Klassen mit wenigen Elementen, die selten neue Elemente bekommen
- Beispiel: Präpositionen

Offen/open

- Klassen mit vielen Elementen, oft neue Elemente
- Beispiel: Nomen, Verben, Adjektive, Adverbien

Tagset

- Liste aller möglichen Wortarten in einer Sprache
- Enkodiert oft morphologische Kategorie, wie Person, Genus, Tempus etc.

Beispiele:

- Penn tag set (45 tags) \implies Penn Treebank
- Brown tag set (87 tags) \implies Brown Korpus
- CLAWS c5 tag set (62 tags) \implies BNC
- CLAWS7 tag set (132 tags)
- SSTS: Stuttgart Tübingen Tagset für das Deutsche. 55 Tags. \implies Tiger Korpus

Das komplette Penn Treebank Tagset

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	"to"	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential 'there'	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VBN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WPS	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolmas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	"	Left quote	<i>' or "</i>
POS	Possessive ending	<i>'s</i>	"	Right quote	<i>' or "</i>
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>[, (, {, <</i>
PRP\$	Possessive pronoun	<i>your, one's</i>)	Right parenthesis	<i>],), }, ></i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>! ?</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>: ; ... --</i>
RP	Particle	<i>up, off</i>			

Figure 5.6 Penn Treebank part-of-speech tags (including punctuation).

In dieser Vorlesung benutzt

Category	Example	Claws c7	Brown	Penn
Adjective	happy	JJ	JJ	JJ
Noun	aircraft	NN	NN	NN
Noun singular	woman, book	NN1	NN	NN
Noun plural	women, books	NN2	NNS	NNS
Preposition	for	IF, II, IO, IW	IN	IN, TO
To (before inf)	to	TO	TO	TO

Finde Taggingfehler in folgenden Sätzen (Penn Treebank Set)

- 1 I/PRP need/VBP a/DT flight/NN from/IN Atlanta/NN
- 2 Does/VBZ this/DT flight/NN serve/VB dinner/NNS

- Adverbien (RB), Partikel (RP), Präpositionen (IN)
 - 1 *Mrs./NNP Shaefer/NNP never/RB got/VBD **around/RP** to/TO joining/VBG*
 - 2 *All/DT we/PRP gotta/VBN do/VB is/VBZ go/VB **around/IN** the/DT corner/NN*
 - 3 *Cateau/NNP Petrus/NNP costs/VBZ **around/RB** 250/CD*
- Partizipien – Adjektive
 - 1 *They were **married/VBN** by the Justice of the Peace yesterday.*
 - 2 *At the time, she was already **married/JJ***

Übung (von Naoaki Okazaki): Annotiere die markierten Wörter

- *the **near** side of the moon.*
- *They had approached quite **near**.*
- *We were **near** the station.*
- *I dont trust him **at all**.*
- ***All right!***
- *He wore a **striking** hat.*
- *The **striking** teachers protested.*
- *The **reading** for this class is difficult.*
- *He was **invited** by some friends of hers.*
- *He was very **surprised** by her remarks.*

- Annotationsschema für Penn Treebank: *Beatrice Santorini. 1990. Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision), Technical Report, University of Pennsylvania.*
- Annotationsschemas beinhalten Definitionen, Tests, Entscheidungsbäume, Ausnahmen
- Wird meist umfangreich und durchläuft mehrere Versionen
- Mehrere Menschen annotieren gleichen Text und deren Annotationen werden verglichen, um Reproduzierbarkeit zu messen.

- Wichtigste Tags
- Tags werden durch Morphologie und Distribution (und manchmal Semantik) definiert
- Tagsets variieren nach Sprache, Aufgabe und Körnigkeit
- Tagging auch für Menschen teilweise schwer entscheidbar
- Aufgabenblatt 7

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging**
- 3 Regelbasiertes Tagging
- 4 Die beiden wichtigsten Informationsarten im Tagging
- 5 HMMs: Hidden Markov Modelle für Tagging

Input : Wortsequenz zumeist schon tokenisiert e.g., *Trump fights press.*

Output : Wortsequenz mit POS tags e.g. *Trump/NNP fights/VBZ press/NN*

Alternativen?

POS Tagging

Weise jedem Wort (und Punctuation) einer Wortsequenz genau eine Wortklasse aus einem gegebenem Tagset zu. Nach Tokenisierung.

Zwischenebene zwischen vollem Parsing und Wortebene:

- Sprachsynthese und Spracherkennung: *content*
- Eigennamenerkennung und Chunking in Informationsextraktion:
The Uranian government declared war on Beloria.
- Bedeutungsdisambiguierung
- Parsingvorläufer
- Textklassifikation

Textklassifikation (siehe letzte Woche)

Beobachtung:	Dokument
Kategorien:	Textthemen, Sentiment, ...
Restriktionen der Klassen:	Thementags
Methoden:	überwachtes ML, regelbasiert ...
Merkmale/Wissensquellen:	Worthäufigkeiten; ...

POS Tagging

Beobachtung:	Satz
Kategorien:	alle möglichen Tagsequenzen
Restriktionen der Klassen:	Tagsets
Methoden:	überwachtes ML, regelbasiert ...
Merkmale/Wissensquellen:	???

Welche Informationen können beim Tagging helfen?

- Ordne einer Sequenz von Beobachtungen Klassen zu (eine Klasse pro Element der Sequenz)
- Klassenzuordnung eines Elements kann von der Klassenzuordnung eines anderen Elements abhängig sein
- Globale Optimierung

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging
- 3 Regelbasiertes Tagging**
- 4 Die beiden wichtigsten Informationsarten im Tagging
- 5 HMMs: Hidden Markov Modelle für Tagging

Voutilainen, A. (1999). Handcrafted rules. In: *Syntactic wordclass tagging*

Zum einen: Lexikon mit Liste aller möglichen PoS für ein Wort, 65.000 Einträge (Wortstämme mit Merkmalen)

Pavlov	PAVLOV N NOM SG PROPER
had	HAVE V PAST VFIN SVO HAVE PCP2 SVO
shown	SHOW PCP2 SVOO SVO SV
that	ADV PRON DEM SG DET CENTRAL DEM SG CS
salivation	N NOM SG

Voutilainen, A. (1999). Handcrafted rules. In: *Syntactic wordclass tagging*

Zum zweiten: 3,744 manuelle entwickelte Constraints.

ADVERBIAL-THAT RULE

Given input: “that”

if

(+1 A/ADV/QUANT); / * if next word is adj, adverb, or quantifier */

(+2 SENT-LIM); / * and following which is a sentence boundary, */

(NOT -1 SVOC/A); / * and the previous word is not a verb like */

/ * ‘consider’ which allows adjs as object complements */

then eliminate non-ADV tags

else eliminate ADV tag

Voutilainen, A. (1999). Handcrafted rules. In: *Syntactic wordclass tagging*

Dann: Eliminierung inkorrektter Tags durch Constraints

Pavlov	PAVLOV N NOM SG PROPER
had	HAVE V PAST VFIN SVO
	HAVE PCP2 SVO
shown	SHOW PCP2 SVOO SVO SV
that	ADV
	PRON DEM SG
	DET CENTRAL DEM SG
	CS
salivation	N NOM SG

- Erstellung komplexer Regelsysteme
- Aufwendige Erstellung und Pflege
- Sprachspezifische Lexika und Tagging-Regeln

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging
- 3 Regelbasiertes Tagging
- 4 Die beiden wichtigsten Informationsarten im Tagging**
- 5 HMMs: Hidden Markov Modelle für Tagging

- N mögliche Tags, Größe Tagset.
- n Länge der Wortsequenz w_1, \dots, w_n
- w_i ein bestimmtes Wort in der Sequenz
- t tag
- \hat{t} von Algorithmus gewählter tag

A Priori Frequenz von Tags für ein Wort

Idee: Für ein gegebenes Wort sind oft manche Tags viel wahrscheinlicher als andere

- Beispiel: *flour/NN* vs. *flour/VB*
- Einfaches überwachtes Verfahren:

$$\hat{t}_i = \underset{t_j}{\operatorname{argmax}} P(t_j | w_i) \text{ für alle } i \in \{1, \dots, n\}$$

- Dabei wird $p(t|w)$ auf Trainingsdaten geschätzt. Wie?
- Probleme?

Syntagmatische Info

Bestimmte Tagfolgen sind wahrscheinlicher als andere

- Beispiel: *DT JJ NN* vs. *DT JJ VB*
- Wir brauchen ein Verfahren, das Word-Tag Häufigkeiten und syntagmatische Information integriert.

- 1 Wortarten
- 2 Wortartenambiguität und Definition von Tagging
- 3 Regelbasiertes Tagging
- 4 Die beiden wichtigsten Informationsarten im Tagging
- 5 HMMs: Hidden Markov Modelle für Tagging**

- Output (Wörter) ist sichtbar, aber die internen Zustände (tags) sind “versteckt” (hidden)
- Die internen Zustände haben die Wörter generiert, nach bestimmten Wahrscheinlichkeiten.
- Für manche Outputs gibt es mehrere mögliche Sequenzen von internen Zuständen. Wir suchen die insgesamt wahrscheinlichste Sequenz \implies Globale Optimierung
- FSAs mit Wahrscheinlichkeiten . . .
- Wir schauen uns das für POS Tagging an und formalisieren und generalisieren danach!

Finde die wahrscheinlichste Tag-Sequenz $\hat{t}_{1,n}$ für eine Wortsequenz w_1, \dots, w_n

HMMs sind generativ, also Bayes'sches Gesetz (wieder einmal)

$$\begin{aligned}\hat{t}_{1,n} &= \operatorname{argmax}_{t_{1,n}} p(t_{1,n} | w_{1,n}) \\ &= \operatorname{argmax}_{t_{1,n}} \frac{p(w_{1,n} | t_{1,n}) p(t_{1,n})}{p(w_{1,n})} \\ &= \operatorname{argmax}_{t_{1,n}} p(w_{1,n} | t_{1,n}) p(t_{1,n})\end{aligned}$$

- 1 Markovannahme für Tags!

$$p(t_i | t_{1,i-1}) = p(t_i | t_{i-1})$$

- 2 Worte sind unabhängig voneinander (gegeben die Tags):

$$p(w_{1,n} | t_{1,n}) = \prod_{i=1}^n p(w_i | t_{1,n})$$

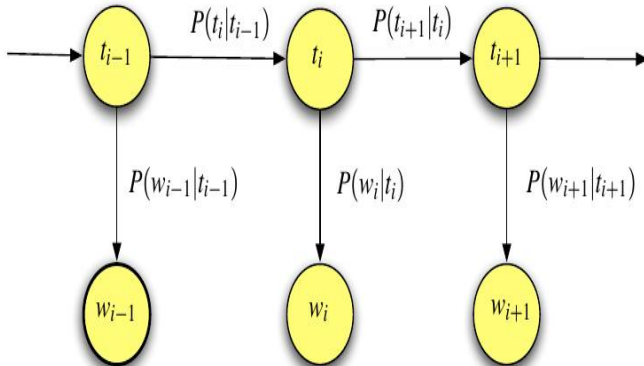
- 3 Wort hängt nur von seinem eigenem Tag ab, nicht von anderen Tags

$$p(w_i | t_{1,n}) = p(w_i | t_i)$$

- 4 Zeitinvarianz

$$\hat{t}_{1,n} := \operatorname{argmax}_{t_{1,n}} p(w_{1,n}|t_{1,n})p(t_{1,n})$$

$$\begin{aligned} p(w_{1,n}|t_{1,n})p(t_{1,n}) &\stackrel{(2)}{=} \prod_{i=1}^n p(w_i|t_{1,n})p(t_{1,n}) \\ &= \prod_{i=1}^n p(w_i|t_{1,n}) \\ &\quad \times p(t_n|t_{1,n-1}) \times p(t_{n-1}|t_{1,n-2}) \times \dots \times p(t_2|t_1) \\ &\stackrel{(1,3)}{=} \prod_{i=1}^n p(w_i|t_i) \\ &\quad \times p(t_n|t_{n-1}) \times p(t_{n-1}|t_{n-2}) \times \dots \times p(t_2|t_1) \\ &= \prod_{i=1}^n [p(w_i|t_i) \times p(t_i|t_{i-1})] \end{aligned}$$



- Schätzung/Training: wie schätze ich die Wahrscheinlichkeiten $p(w_i|t_i)$ sowie $p(t_i|t_{i-1})$?
- Wie berechne ich das Maximum effizient? bzw. Dekodierung: Wie weise ich eine Tagsequenz einer gesehene Wortsequenz zu?
- Was mache ich mit unbekanntem Wörtern (*XXXtentacion*)?

Übergangswahrscheinlichkeiten: $p(t_i|t_{i-1})$

Aus (erfundenem) Korpus. Erster Tag in Zeile. VBE spezieller Tag für manche Verbformen von *to be*.

	NNP	NN	NNS	VBZ	VBE	JJ	< \s >	Sum
< s >	405 180	222 742	135 307	4753	9097	10 300	0	6 053 093
NNP	814 194	261 241	62 256	52 989	9508	150	143 003	3 851 890
NN	129 364	1 159 757	153 168	180 813	39 366	340	2 843 000	14 423 629
NNS	18 473	62 613	27 334	14 343	157 286	210	1 521 300	5 216 164
VBZ	15 246	28 310	14 615	374	46	87 993	21 000	709 004
VBE	2 611	5 030	14 335	21	253	15 400	3 457	638 576
JJ	23 001	3 580 455	2 500 342	130	20	3 550	15 340	6 385 186

Damit z.B. $p(NN|JJ) = \frac{3580455}{6385186} = 0.56$

	NNP	NN	NNS	VBZ	VBE	JJ	< \s >
< s >	0.067	0.037	0.022	0.0008	0.0015	0.0017	0
NNP	0.211	0.068	0.016	0.014	0.0025	0.00004	0.037
NN	0.009	0.080	0.011	0.012	0.0027	0.00002	0.197
NNS	0.0035	0.012	0.005	0.0027	0.03	0.00004	0.292
VBZ	0.0215	0.039	0.021	0.0005	0.00006	0.124	0.029
VBE	0.004	0.0079	0.022	0.00003	0.0004	0.024	0.0054
JJ	0.0036	0.56	0.39	0.00002	0.000003	0.0005	0.0024

Zu lesen als $p(\text{Spaltentag} | \text{Zeilentag})$

- Wir haben angenommen **Zeitinvarianz**
- **Zeitinvarianz**: $p(t_i|t_{i-1}) = p(t_2|t_1)$
- Übergangswahrscheinlichkeiten werden in der Realität meist mit Interpolation gesmoothed

Emissionswahrscheinlichkeiten: $p(w_i|t_i)$

	Trump	deals	are	legal	Sum
NP	8	0	0	0	3 851 890
NN	11	0	0	0	14 423 629
NNS	0	790	0	0	5 216 164
VBZ	0	7450	0	0	709 004
VBE	0	0	454 316	0	638 576
JJ	0	0	0	555	6 385 186

$$p(\text{Trump}|\text{NNP}) = \frac{8}{3851890} = 0.000002$$

	Trump	deals	are	legal
NNP	0.000002	0	0	0
NN	$7.6 \cdot 10^{-7}$	0	0	0
NNS	0	0.00015	0	0
VBZ	0	0.01	0	0
VBE	0	0	0.71	0
JJ	0	0	0	0.00008

Zu lesen als $p(w_i | t_i)$

Beispiel: *Trump deals are legal* hat 4 mögliche Tagsequenzen (siehe Emissionstabelle).

- Option 1: $p(\text{NNP NNS VBE JJ}|\text{Trump deals are legal}) \propto$
 $p(\text{Trump}|\text{NNP})p(\text{NNP}|\langle s \rangle) \cdot p(\text{deals}|\text{NNS})p(\text{NNS}|\text{NNP}) \cdot$
 $p(\text{are}|\text{VBE})p(\text{VBE}|\text{NNS}) \cdot p(\text{legal}|\text{JJ})p(\text{JJ}|\text{VBE}) \cdot$
 $p(\langle /s \rangle|\text{JJ}) = (0.000002 \cdot 0.067) \cdot (0.00015 \cdot 0.016) \cdot (0.71 \cdot$
 $0.03) \cdot (0.00008 \cdot 0.024) \cdot (0.0024) = 3.15 \cdot 10^{-23}$
- Option 2: $p(\text{NN NNS VBE JJ}|\text{Trump deals are legal}) \propto$
 $p(\text{Trump}|\text{NN})p(\text{NN}|\langle s \rangle) \cdot p(\text{deals}|\text{NNS})p(\text{NNS}|\text{NN}) \cdot$
 $p(\text{are}|\text{VBE})p(\text{VBE}|\text{NNS}) \cdot p(\text{legal}|\text{JJ})p(\text{JJ}|\text{VBE}) \cdot$
 $p(\langle /s \rangle|\text{JJ}) = (7.6 \cdot 10^{-7} \cdot 0.037) \cdot (0.00015 \cdot 0.011) \cdot (0.71 \cdot$
 $0.03) \cdot (0.00008 \cdot 0.024) \cdot (0.0024) = .4.967 \cdot 10^{-24}$

- Option 3: $p(\text{NNP VBZ VBE JJ}|\text{Trump deals are legal}) \propto$
 $p(\text{Trump}|\text{NNP})p(\text{NNP}|\langle s \rangle) \cdot p(\text{deals}|\text{VBZ})p(\text{VBZ}|\text{NNP}) \cdot$
 $p(\text{are}|\text{VBE})p(\text{VBE}|\text{VBZ}) \cdot p(\text{legal}|\text{JJ})p(\text{JJ}|\text{VBE}) \cdot p(\langle /s \rangle|\text{JJ}) =$
 $(0.000002 \cdot 0.067) \cdot (0.01 \cdot 0.014) \cdot (0.71 \cdot 0.00006) \cdot (0.00008 \cdot$
 $0.024) \cdot (0.0024) = 3.68 \cdot 10^{-24}$
- Option 4: $p(\text{NN VBZ VBE JJ}|\text{Trump deals are legal}) \propto$
 $p(\text{Trump}|\text{NN})p(\text{NN}|\langle s \rangle) \cdot p(\text{deals}|\text{VBZ})p(\text{VBZ}|\text{NN}) \cdot$
 $p(\text{are}|\text{VBE})p(\text{VBE}|\text{VBZ}) \cdot p(\text{legal}|\text{JJ})p(\text{JJ}|\text{VBE}) \cdot p(\langle /s \rangle|\text{JJ}) =$
 $(7.6 \cdot 10^{-7} \cdot 0.037) \cdot (0.01 \cdot 0.012) \cdot (0.71 \cdot 0.00006) \cdot (0.00008 \cdot$
 $0.024) \cdot (0.0024) = 6.62 \cdot 10^{-25}$

Option 1 wird gewählt! NB: Wir haben global optimiert!

Naiv

- 1 Schreibe alle möglichen Tagsequenzen einer Wortsequenz auf
- 2 Berechne Wahrscheinlichkeit jeder Tagsequenz einzeln
- 3 Nimm maximale Wahrscheinlichkeit und die korrespondierende Tagsequenz

Problem: Zu viele Möglichkeiten $O(N^n)$, wobei N die Anzahl der Tags und n die Länge der Eingabe

Beispiel: *Trump deals are legal and the bank confirmed that* bräuchte dann (je nach Tagset) $2 \cdot 2 \cdot 2 \cdot 3$ Berechnungen

Nicht die Lösung: Greedy

Man könnte einfach nach jedem Unigram die bis dahin gesehene höchste Wahrscheinlichkeit behalten, und immer nur mit einem Pfad weiterrechnen.

	Trump	deals	are	legal
Richtig	NNP	NNS	VBE	JJ
Greedy T1	NNP ($0.000002 \cdot 0.067$)	?	?	?
Greedy T1	NN ($7.6 \cdot 10^{-7} \cdot 0.037$)	?	?	?
Wählt	NNP	?	?	?
Greedy T2	NNP	NNS ($0.00015 \cdot 0.016$)	?	?
Greedy T2	NNP	VBZ ($0.01 \cdot 0.014$)	?	?
Wählt	NNP	VBZ	?	?
...

Wir wählten den falschen Tag (im Vergleich zur gesamten Lösung).
Warum: Entscheidung ein Wort zu früh. Damit wurde die niedrige Übergangswahrscheinlichkeit $p(VBE|VBZ)$ nicht angeschaut. Keine globale Optimierung!

- POS Tagging als Sequenzklassifikation
- Zwei wichtige Informationsarten: tag-per-word probabilities und pos sequence probabilities
- überwachte Baseline: Unigramtagger. Weise jedem Wort (unabhängig vom Kontext) sein häufigstes Tag aus dem Trainingskorpus zu
- HMM Tagger: generatives Modell, das tag-per-word probs und pos sequence probabilities kombiniert (unter vielen Unabhängigkeitsannahmen)
- Problem: Effiziente Dekodierung (= Zuweisung einer Tagsequenz) unter Beibehaltung der globalen Optimierung