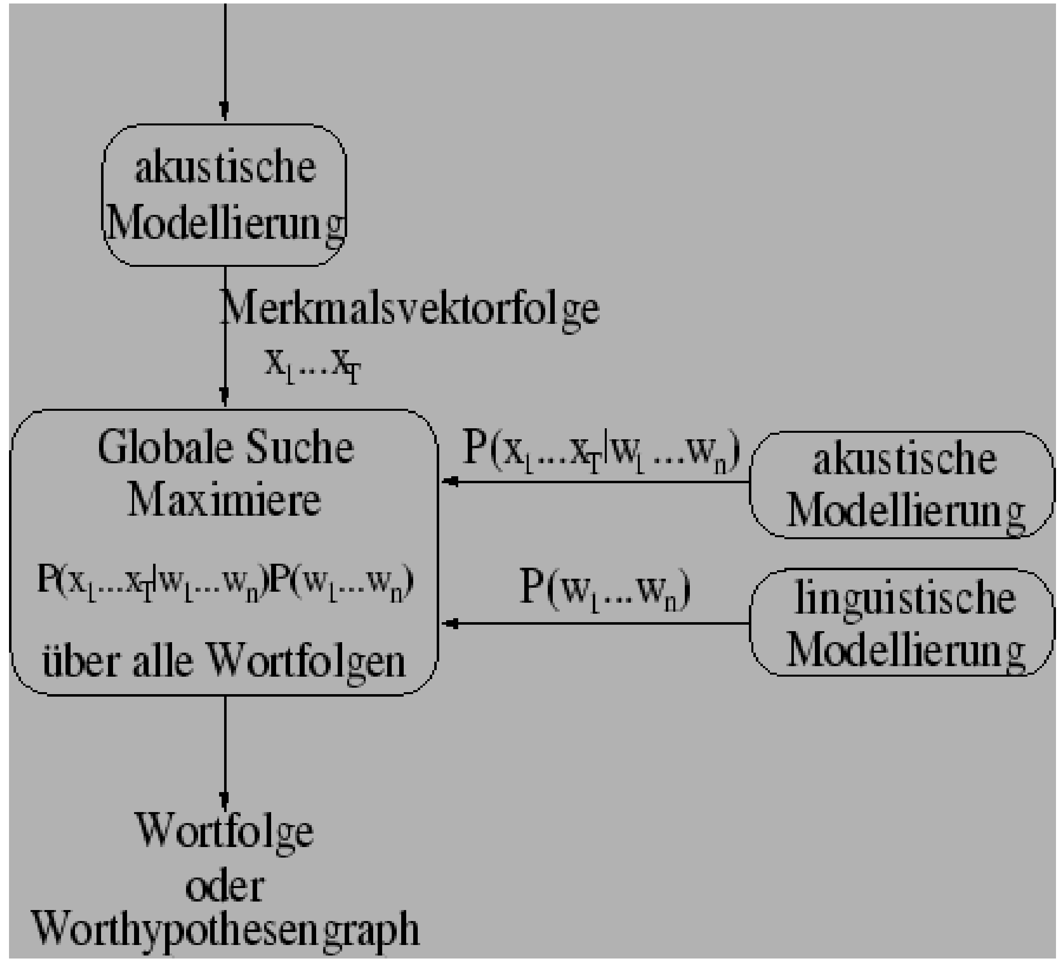
# Einführung in die Computerlinguistik

## Spracherkennung und Hidden Markov Modelle

Dozentin: Wiebke Petersen

WS 2004/2005

# Spracherkennung



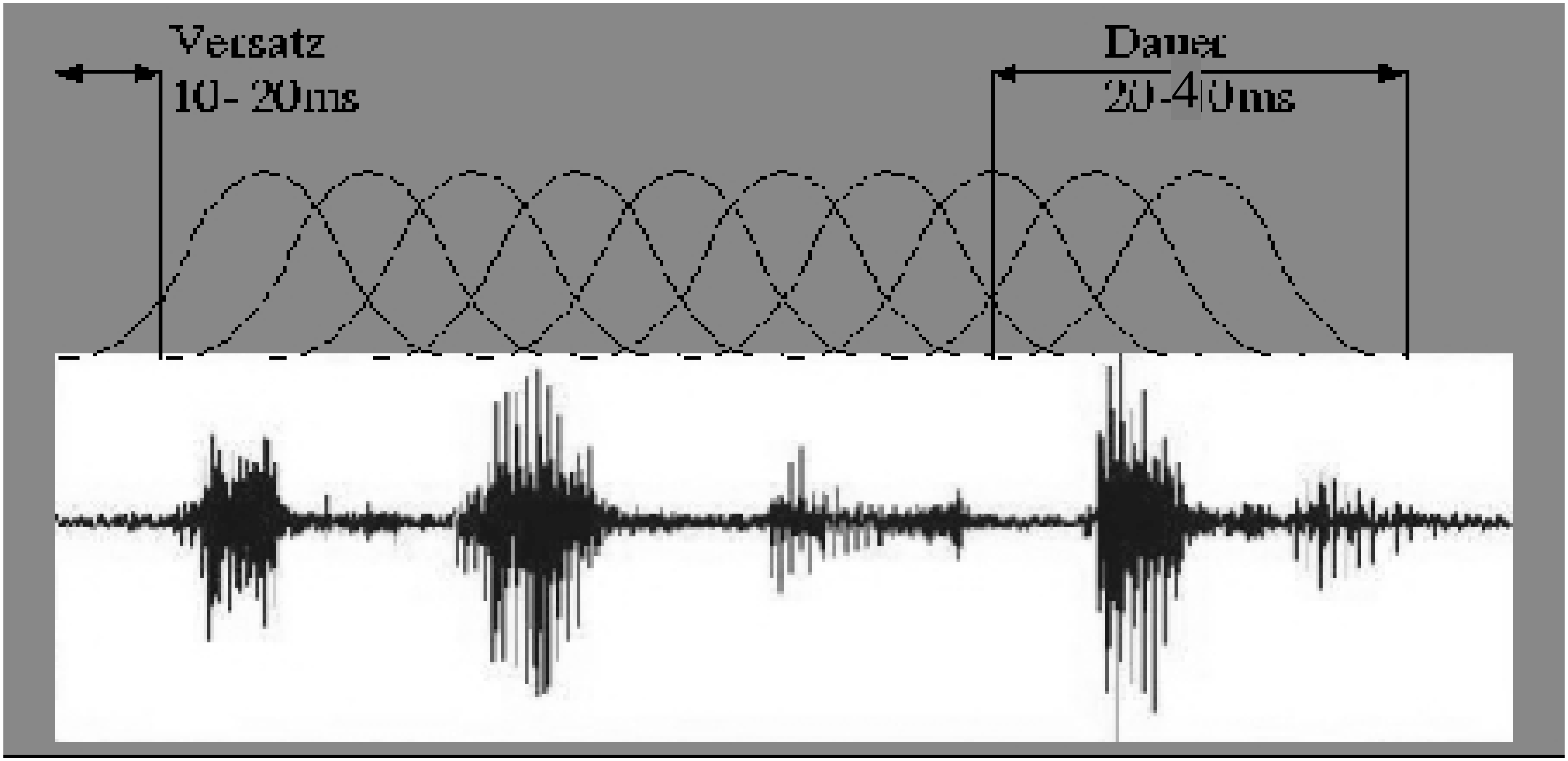

- die klassenbedingte Verteilung $P(x_1^T|w_1^N)$

  Dies ist die Wahrscheinlichkeit, die Vektoren $x_1^T$ zu beobachten, gegeben die Wortfolge $w_1^N$. Dies wird als akustisches Modell bezeichnet.

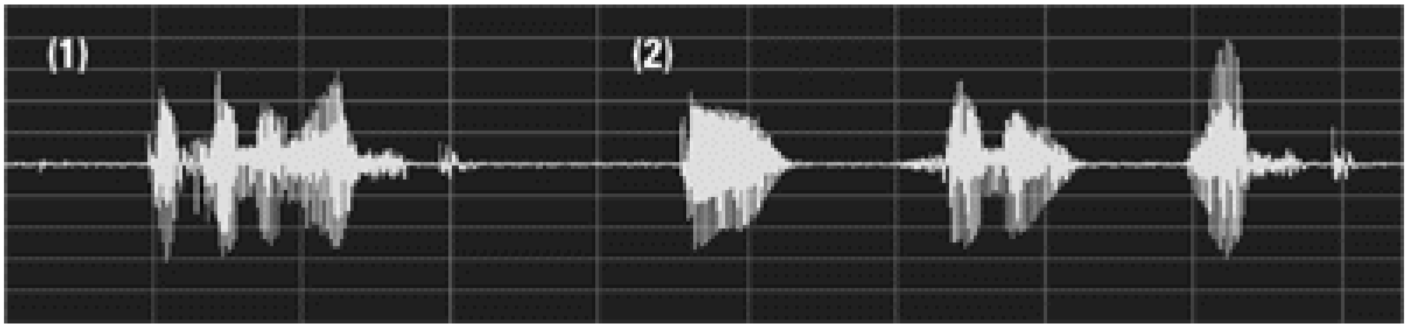
- die a-priori-Wahrscheinlichkeit einer Wortfolge $P(w_1^N)$.

  Dies ist die Wahrscheinlichkeit, mit der die Wortfolge $w_1^N$ auftreten kann. Sie ist unabhängig von den aktuell gemachten Beobachtungen. Dies wird linguistisches Modell bzw. Sprachmodell (LM) genannt.

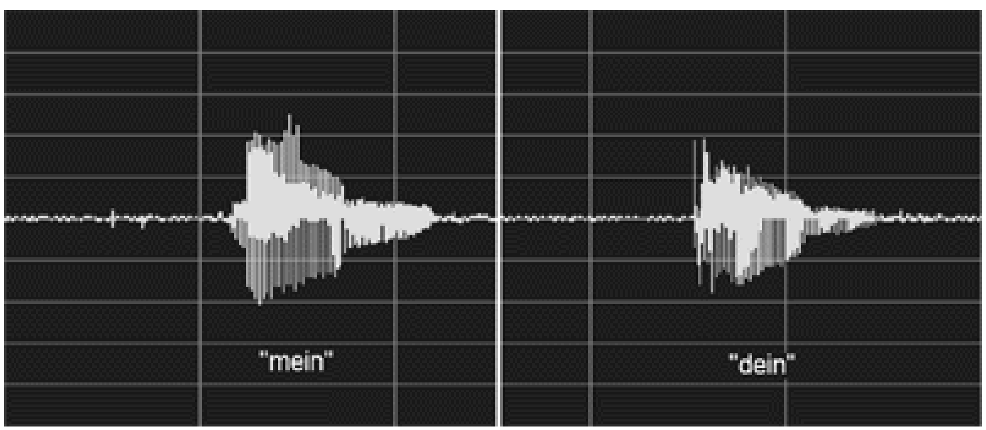
# Merkmalsextraktion



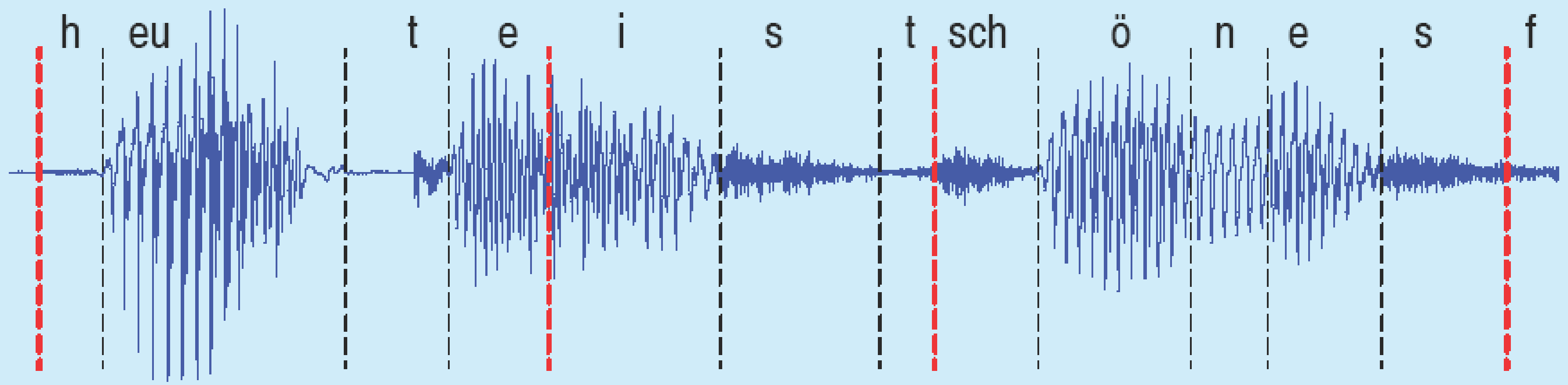
# Frequenzdiagramme (Beispiele)



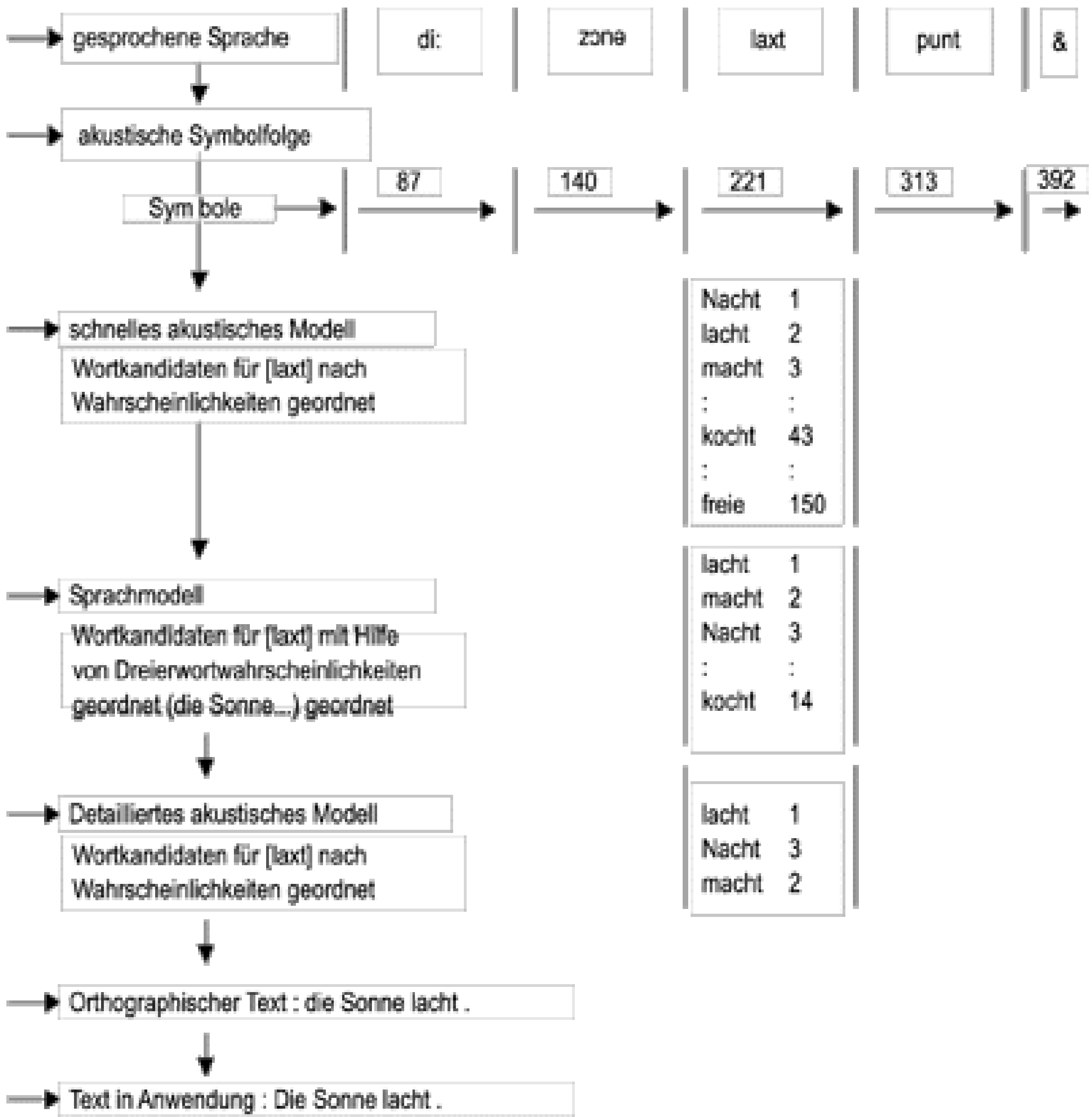

Die Abbildung zeigt den Unterschied der Frequenzdiagramme für den kontinuierlich (1) und den diskret (2) gesprochenen Satz "Die Sonne lacht".



# Frequenzdiagramme (Segmentierungsproblem)

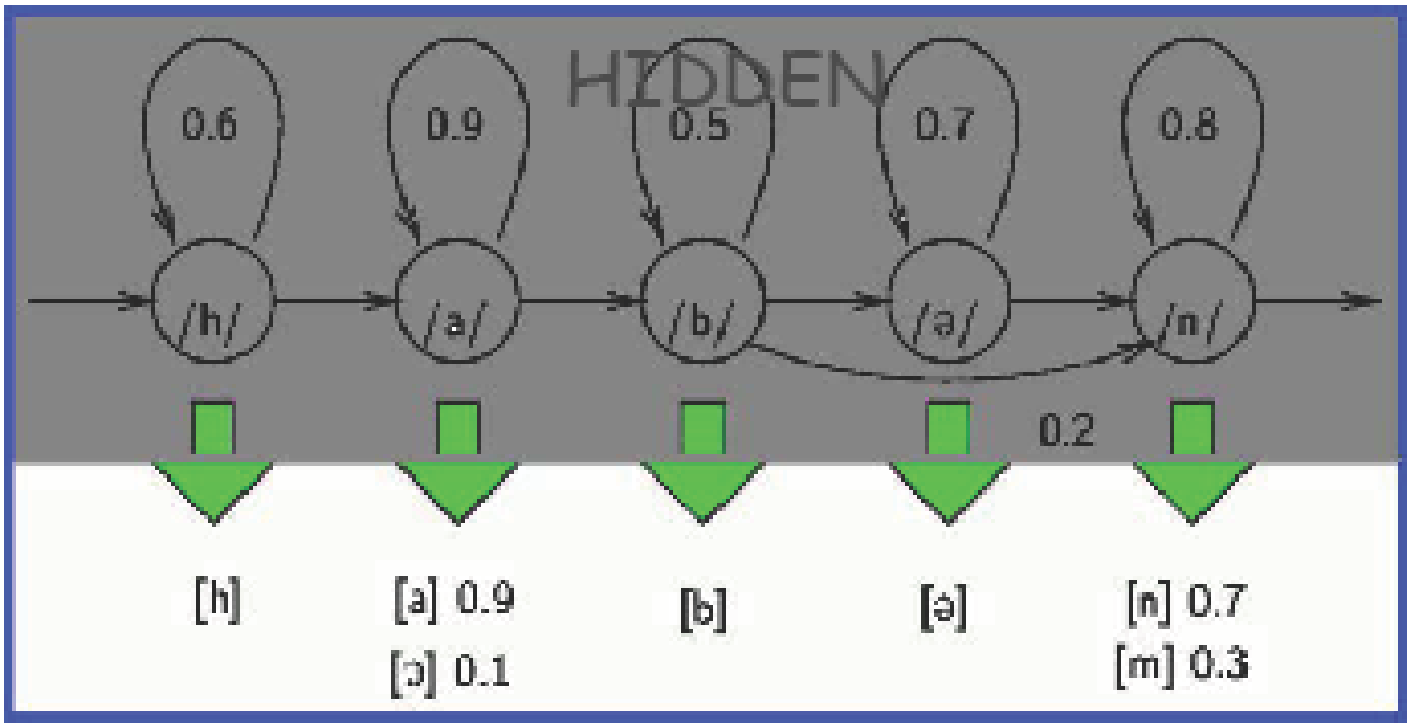



*Beispiel:* Segmentierung in Phone bzw. Wörter


gesprochene Sprache
di:
zone
laxt
punt
&
akustische Symbolfolge
Sym bole
87
140
221
313
392
schnelles akustisches Modell
Wortkandidaten für [laxt] nach Wahrscheinlichkeiten geordnet
Nacht 1
lacht 2
macht 3
kocht 43
freie 150
Sprachmodell
Wortkandidaten für [laxt] mit Hilfe von Dreierwortwahrscheinlichkeiten geordnet (die Sonne...) geordnet
lacht 1
macht 2
Nacht 3
kocht 14
Detailliertes akustisches Modell
Wortkandidaten für [laxt] nach Wahrscheinlichkeiten geordnet
lacht 1
Nacht 3
macht 2
Orthographischer Text : die Sonne lacht .
Text in Anwendung : Die Sonne lacht .

# Hidden Markov Modell HMM für 'haben'

Der Beobachter sieht beim HMM von außen nur die Folge der ausgegebenen Symbole. Die Folge der Zustände bleibt ihm verborgen.



# Markov-Kette

Definition 3

Eine Markov-Kette ist ein spezieller stochastischer Prozess,
bei dem zu jedem Zeitpunkt
die Wahrscheinlichkeiten aller zukünftigen Zustände
nur vom momentanen Zustand abhängen
(= Markov-Eigenschaft)

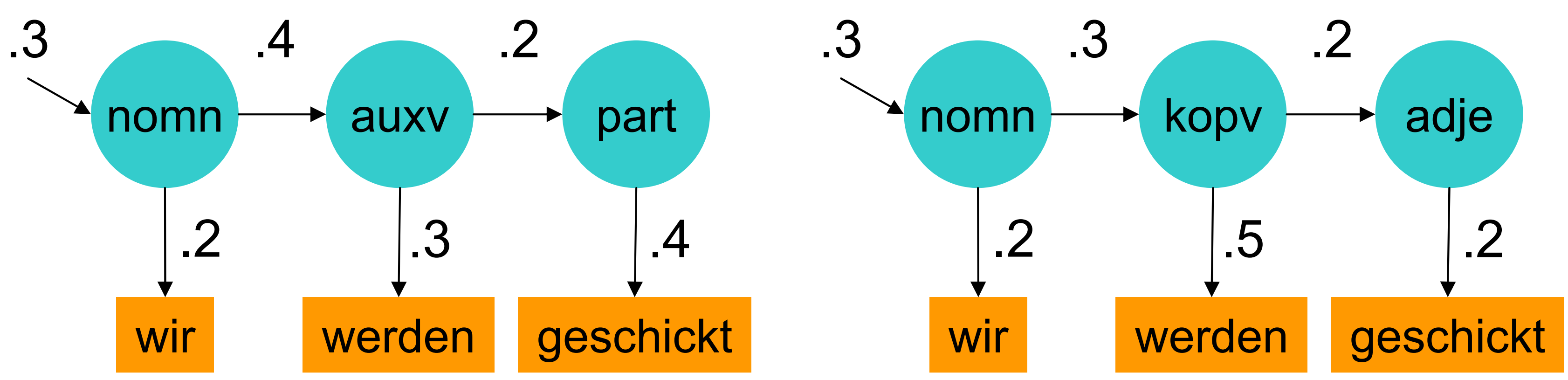
d.h. es gilt:

$$P(X_{t+1} = x_{t+1} \mid X_1 = x_1, X_2 = x_2, \ldots X_t = x_t) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

Brants, 1999: 30

# Hidden Markov-Modell: Beispiel

- in einem Text lassen sich nur die Ausgaben (= produzierte Wörter) beobachten (visible)
- die Sequenz von Zuständen (= Wortarten), die die Wörter ausgeben, (Satzmuster) lässt sich nicht beobachten (hidden)
- mehrere Sequenzen können dieselbe Ausgabe erzeugen:




.3 x .2 x .4 x .3 x .2 x .4 =0.000576

.3 x .2 x .3 x .5 x .2 x .2 =0.000360

# Ein Hidden Markov-Modell

| | Übergangsmatrix | | | | | Emissionsmatrix | | | | Startwahrscheinlichkeit |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_t$ | $X_{t+1}$ | | | | | $o_t$ | | | | $\pi$ |
| | Adje | AuxV | KopV | Nomn | Part | geschickt | werden | wir | ... | |
| Adje | .2 | .1 | .1 | .4 | .2 | .2 | 0 | 0 | .8 | .3 |
| AuxV | .2 | .3 | .1 | .2 | .2 | 0 | .3 | 0 | .7 | .2 |
| KopV | .2 | .2 | .1 | .4 | .1 | 0 | .5 | 0 | .5 | .1 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 | 0 | .2 | .8 | .3 |
| Part | .3 | .1 | .2 | .1 | .3 | .4 | 0 | 0 | .6 | .1 |

# Hidden Markov-Modell: Gewinnung der Daten – Übersicht

- Annotation eines Corpus
- Auszählung der Sequenzen
- Umrechnung der Häufigkeiten in prozentuale Anteile

# Hidden Markov-Modell: Gewinnung der Daten (1)

- Annotation eines Corpus
- Auszählung der Sequenzen
- Umrechnung der Häufigkeiten in prozentuale Anteile

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Wir* | *werden* | *geschickt* | *.* | *Wir* | *werden* | *geschickt* | *.* | *...* | *...* |
| nomn | auxv | part | $\Omega$ | nomn | kopv | adje | $\Omega$ | … | … |

# Hidden Markov-Modell: Gewinnung der Daten (2)

- Annotation eines Corpus
- Auszählung der Sequenzen
- Umrechnung der Häufigkeiten in prozentuale Anteile

| | Adje | AuxV | KopV | Nomn | Part | Ω | geschickt | werden | wir | . |
|---|---|---|---|---|---|---|---|---|---|---|
| Adje | - | - | - | - | - | 1 | 1 | - | - | - |
| AuxV | - | - | - | - | 1 | - | - | 1 | - | - |
| KopV | 1 | - | - | - | - | - | 1 | - | - | - |
| Nomn | - | 1 | 1 | - | - | - | - | - | 2 | - |
| Part | - | - | - | - | - | 1 | - | - | - | - |
| Ω | - | - | | 1 | - | - | - | - | - | 2 |

# Hidden Markov-Modell: Gewinnung der Daten (3)

- Annotation eines Corpus
- Auszählung der Sequenzen
- Umrechnung der Häufigkeiten in prozentuale Anteile

| | Adje | AuxV | KopV | Nomn | Part | Ω | geschickt | werden | wir | . |
|---|---|---|---|---|---|---|---|---|---|---|
| Adje | - | - | - | - | - | 1.0 | 1.0 | - | - | - |
| AuxV | - | - | - | - | 1.0 | - | - | 1.0 | - | - |
| KopV | 1.0 | - | - | - | - | - | 1.0 | - | - | - |
| Nomn | - | 0.5 | 0.5 | - | - | - | - | - | 1.0 | - |
| Part | - | - | - | - | - | 1.0 | - | - | - | - |
| Ω | - | - | | 1.0 | - | - | - | - | - | 1.0 |

# Drei grundlegende Aufgaben, die mit HMMs bearbeitet werden

1. Dekodierung: Wahrscheinlichkeit einer Beobachtung finden
   - brute force
   - Forward-Algorithmus / Backward-Algorithmus
2. Beste Pfad-Sequenz finden
   - brute force
   - Viterbi-Algorithmus
3. Training: Aufbau des besten Modells aus Trainingsdaten





Manning/Schütze, 2000: 325

# A1: Wahrscheinlichkeit einer Beobachtung finden

gegeben <u>eine Sequenz von Beobachtungen</u> $O = (o_1, \ldots, o_T)$

O=(*wir,werden,geschickt*)

<u>ein Modell</u> $\mu = (A, B, \Pi)$

| | Adje | AuxV | KopV | Nomn | Part | g‘schickt | werden | wir | .. | $\pi$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | .2 | 0 | 0 | .8 | .3 |
| AuxV | .2 | .3 | .1 | .2 | .2 | 0 | .3 | 0 | .7 | .2 |
| KopV | .2 | .2 | .1 | .4 | .1 | 0 | .5 | 0 | .5 | .1 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 | 0 | .2 | .8 | .3 |
| Part | .3 | .1 | .2 | .1 | .3 | .4 | 0 | 0 | .6 | .1 |

gesucht <u>die Wahrscheinlichkeit</u> $P(O \mid \mu)$

$$P(wir, werden, geschickt \mid \mu)$$

# A1: Wahrscheinlichkeit einer Beobachtung finden

<u>Lösungsweg 1: brute force</u>

Für alle möglichen Zustandsfolgen

- Berechnung der Wahrscheinlichkeit der Beobachtungen
- Summierung der Wahrscheinlichkeiten

$$P(O \mid \mu) = \sum_{X} P(O \mid X, \mu) P(X \mid \mu)$$

$$= \sum_{X_1 \ldots X_{t+1}} \pi_{X_1} \prod_{t=1}^{T-1} \underset{\substack{\uparrow \\ \text{state transition}}}{a_{X_t X_{t+1}}} \; \underset{\substack{\uparrow \\ \text{symbol emission}}}{b_{X_t X_{t+1} o_t}}$$





Manning/Schütze, 2000: 326

# A1: Wahrscheinlichkeit einer Beobachtung finden

## Lösungsweg 1: brute force: Beispiel

$$P(O|\mu) = \sum_{X_1 \ldots X_{t+1}} \pi_{X_1} \prod_{t=1}^{T-1} a_{X_t X_{t+1}} b_{X_t X_{t+1} o_t}$$

Beispiel für
state emission HMM

| | | |
|---|---|---|
| P(*wir,werden,geschickt* \| Adje Adje Adje, μ) | | =0.0 |
| + P(*wir,werden,geschickt* \| Adje Adje AuxV, μ) | | |
| + … | | |
| + P(*wir,werden,geschickt* \| Nomn AuxV Part, μ) | .3 x .2 x .4 x .3 x .2 x .4 | =0.000576 |
| + … | | |
| + P(*wir,werden,geschickt* \| Nomn KopV Adje, μ) | .3 x .2 x .3 x .5 x .2 x .2 | =0.000360 |
| + … | | |
| + P(*wir,werden,geschickt* \| Part Part Part, μ) | | =0.0 |
| = … | | =0.000936 |

# HMM [Aufgabe]: Beste Pfadsequenz finden

gegeben <u>eine Sequenz von Beobachtungen</u> $O = (o_1, \ldots, o_T)$

O=(*wir,werden,geschickt*)

<u>ein Modell</u> $\mu = (A, B, \Pi)$

| | Adje | AuxV | KopV | Nomn | Part | g‘schickt | werden | wir | .. | $\pi$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | .2 | 0 | 0 | .8 | .3 |
| AuxV | .2 | .3 | .1 | .2 | .2 | 0 | .3 | 0 | .7 | .2 |
| KopV | .2 | .2 | .1 | .4 | .1 | 0 | .5 | 0 | .5 | .1 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 | 0 | .2 | .8 | .3 |
| Part | .3 | .1 | .2 | .1 | .3 | .4 | 0 | 0 | .6 | .1 |

gesucht <u>die wahrscheinlichste Pfadsequenz</u>

$P(wir, werden, geschickt \mid \mu)$ $\arg_x \max P(O \mid \mu)$

# HMM [Aufgabe]: Beste Pfadsequenz finden

Modellvariante:
mit Startsymbol
statt mit Tabelle der Startwahrscheinlichkeiten)

| | Adje | AuxV | KopV | Nomn | Part | Ω |
|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | 0 |
| AuxV | .2 | .3 | .1 | .2 | .2 | 0 |
| KopV | .2 | .2 | .1 | .4 | .1 | 0 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 |
| Part | .3 | .1 | .2 | .1 | .3 | 0 |
| Ω | .3 | .2 | .1 | .3 | .1 | 0 |

| g‘schickt | werden | wir | ... | . |
|---|---|---|---|---|
| .2 | 0 | 0 | .8 | 0 |
| 0 | .3 | 0 | .7 | 0 |
| 0 | .5 | 0 | .5 | 0 |
| 0 | 0 | .2 | .8 | 0 |
| .4 | 0 | 0 | .6 | 0 |
| 0 | 0 | 0 | 0 | 1 |

# Ineffiziente Lösung

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | $\Omega \rightarrow$ | *wir* \| Adje $\rightarrow$ | *werden* \| Adje | $\rightarrow$ | *geschickt* \| Adje |
| 2 | | | | | *geschickt* \| AuxV |
| 3 | | | | | *geschickt* \| KopV |
| 4 | | | | | *geschickt* \| Nom |
| 5 | | | | | *geschickt* \| Part |
| 6 | | | *werden* \| AuxV | $\rightarrow$ | *geschickt* \| Adje |
| 7 | | | | | *geschickt* \| AuxV |
| .. | | | | | … |
| .. | | … | … | | … |
| 85 | | *wir* \| Nomn $\rightarrow$ | *werden* \| AuxV | $\rightarrow$ | *geschickt* \| Part |
| 86 | | *wir* \| Nomn $\rightarrow$ | *werden* \| KopV | $\rightarrow$ | *geschickt* \| Adje |
| .. | | | … | | |
| 125 | | *wir* \| Part $\rightarrow$ | *werden* \| Part | $\rightarrow$ | *geschickt* \| Part |

Für 3 Beobachtungen und 5 Kategorien $5^3$ Schritte

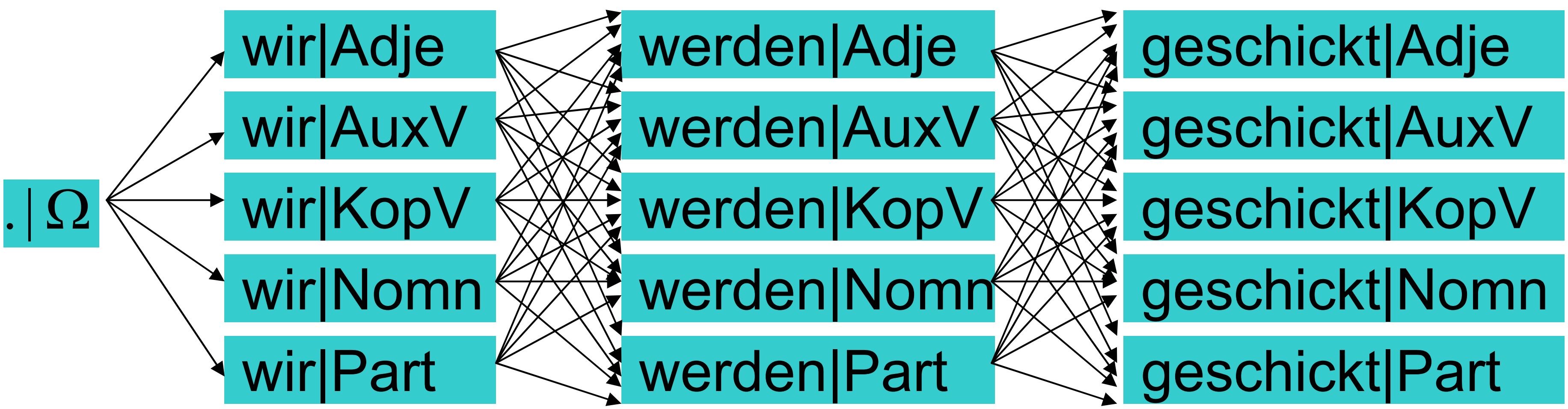
# Beispiel

ineffiziente Suche der besten Lösung

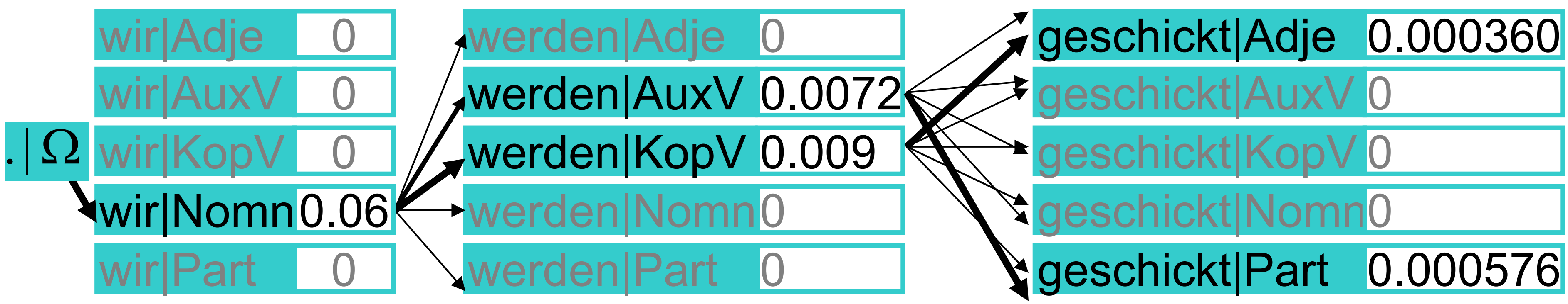
P(Adje Adje Adje | *wir,werden,geschickt*, μ) =0.0
P(Adje Adje AuxV | *wir,werden,geschickt*, μ)
…
P(Nomn AuxV Part | *wir*,*werden*,*geschickt*, μ) .3 x .2 x .4 x .3 x .2 x .4 =0.000576
…
P(Nomn KopV Adje | *wir*,*werden*,*geschickt*, μ) .3 x .2 x .3 x .5 x .2 x .2 =0.000360
…
P(Part Part Part | *wir,werden,geschickt*, μ) =0.0

# Viterbi-Lösung

- Kompakte Darstellung der Pfade als Gitter (Trellis)
- Wiederverwendung partieller Ergebnisse statt Neuberechnung
- Speichert für jeden Zeitpunkt t
  - die Wahrscheinlichkeit des wahrscheinlichsten Pfades, der zu einem Knoten führt
  - den Vorgängerknoten auf diesem Pfad

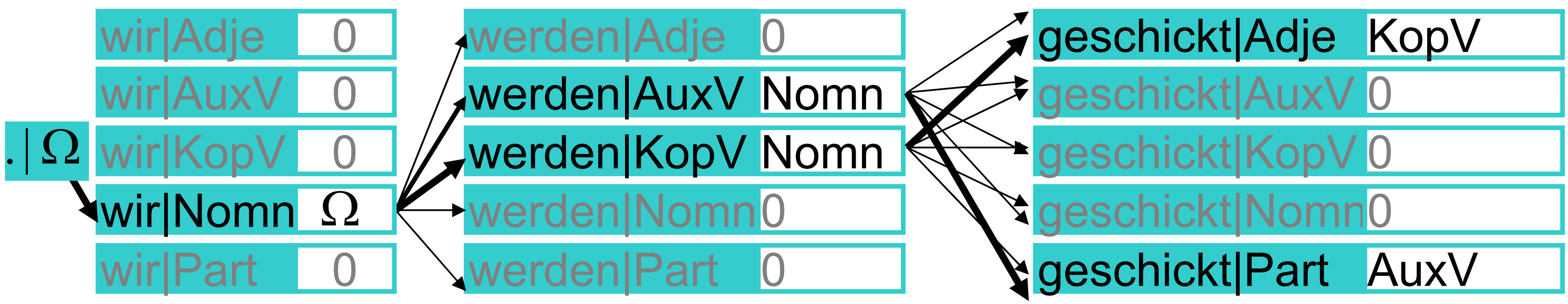


# Daten 1: Wahrscheinlichkeit des wahrscheinlichsten Pfades




| | |
|---|---|
| wir\|Nomn | Knoten |
| >0 | Wahrscheinlichkeit des wahrscheinlichsten Pfades |
| 0 | unwahrscheinlicher Pfad |

# Daten 2: Vorgängerknoten auf wahrscheinlichstem Pfad




| | |
|---|---|
| wir\|Nomn | Knoten |
| Ω | Vorgänger-Knoten auf wahrscheinlichstem Pfad |
| 0 | Vorgänger-Knoten auf wahrscheinlichstem Pfad bei $P(X_i) = 0$ (= unwahrscheinlicher Pfad) |

# Daten 1 und 2: Übersicht

# Tracing (1): Initialisierung

Setzen der Startknoten für Verfolgung

- Wahrscheinlichkeit der wahrscheinlichsten Pfade
- Vorgängerknoten auf den wahrscheinlichsten Pfaden

$.|\Omega$

$.|\Omega$

# Tracing (2): 1. Iteration

| | | |
|---|---|---|
| | wir\|Adje | 0 |
| | wir\|AuxV | 0 |
| .\|Ω | wir\|KopV | 0 |
| | wir\|Nomn | 0.06 |
| | wir\|Part | 0 |

Funktion $\delta$

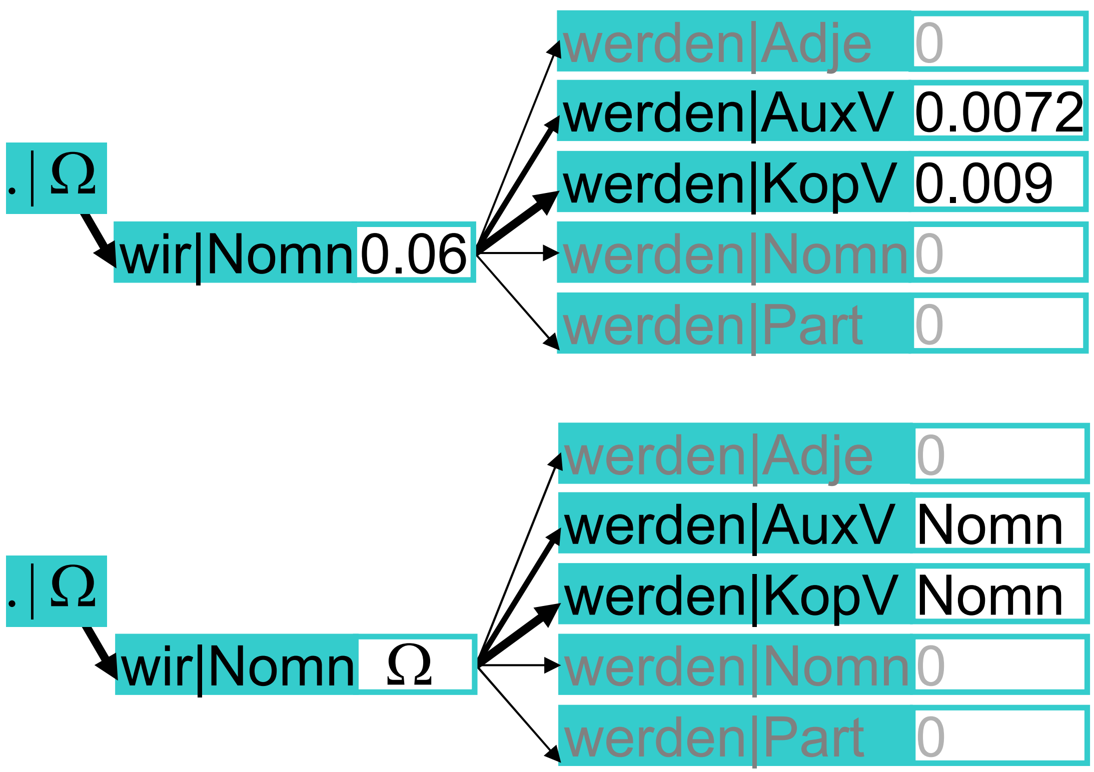
Berechnung der Wahrscheinlichkeit des wahrscheinlichsten Pfades

| | | |
|---|---|---|
| | wir\|Adje | 0 |
| | wir\|AuxV | 0 |
| .\|Ω | wir\|KopV | 0 |
| | wir\|Nomn | Ω |
| | wir\|Part | 0 |

Funktion $\psi$

Ermittlung des Vorgängerknotens auf dem wahrscheinlichsten Pfad

# Tracing (3): 2. Iteration



# Tracing (4): 3. Iteration

# Tracing (5): Terminierung und Pfadausgabe