



Automatisches Lernen von Regeln zur quellseitigen Umordnung

EIN ANSATZ VON DMITRIY GENZEL

Duwaraka Murugadas

Fortgeschrittene Methoden der statistischen maschinellen Übersetzung

(Miriam Käshammer)

06.01.2015

Agenda

2

Das Problem

Der Lösungsansatz

Der Algorithmus

Evaluation

Agenda

3

Das Problem

Der Lösungsansatz

Der Algorithmus

Evaluation

Das Problem I

4

Er sagte mir, dass er gehen wollte.

He told me that he wanted to leave.

Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Er sagte mir, dass er gehen wollte.

Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Er sagte mir, dass er gehen wollte.



Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Er sagte mir, dass er gehen wollte. ✓

Er sagte mir, dass er wollte gehen.

Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Er sagte mir, dass er gehen wollte. ✓

Er sagte mir, dass er wollte gehen. ✓

Das Problem II

5

Er sagte mir, dass er gehen wollte.



He told me that he wanted to leave.

Er sagte mir, dass er gehen wollte. ✓

Er sagte mir, dass er wollte gehen. ✓

...

Das Problem II

5

Er sagte mir, dass er gehen wollte.

Er sagte mir, dass er gehen wollte.

Welche Reihenfolge ist die richtige?

He told me that he wanted to leave. ...

AGENDA

6

Das Problem

Der Lösungsansatz

Der Algorithmus

Evaluation

Pre-ordering

7

- + Unabhängig vom eigentlichen Übersetzungssystem
- + Verkürzt die Dekodierungszeit
- Umordnung kann im Nachhinein nicht mehr vorgenommen werden

Ziel: automatisches Lernen von Pre-Ordering Regeln für möglichst viele Sprachpaare

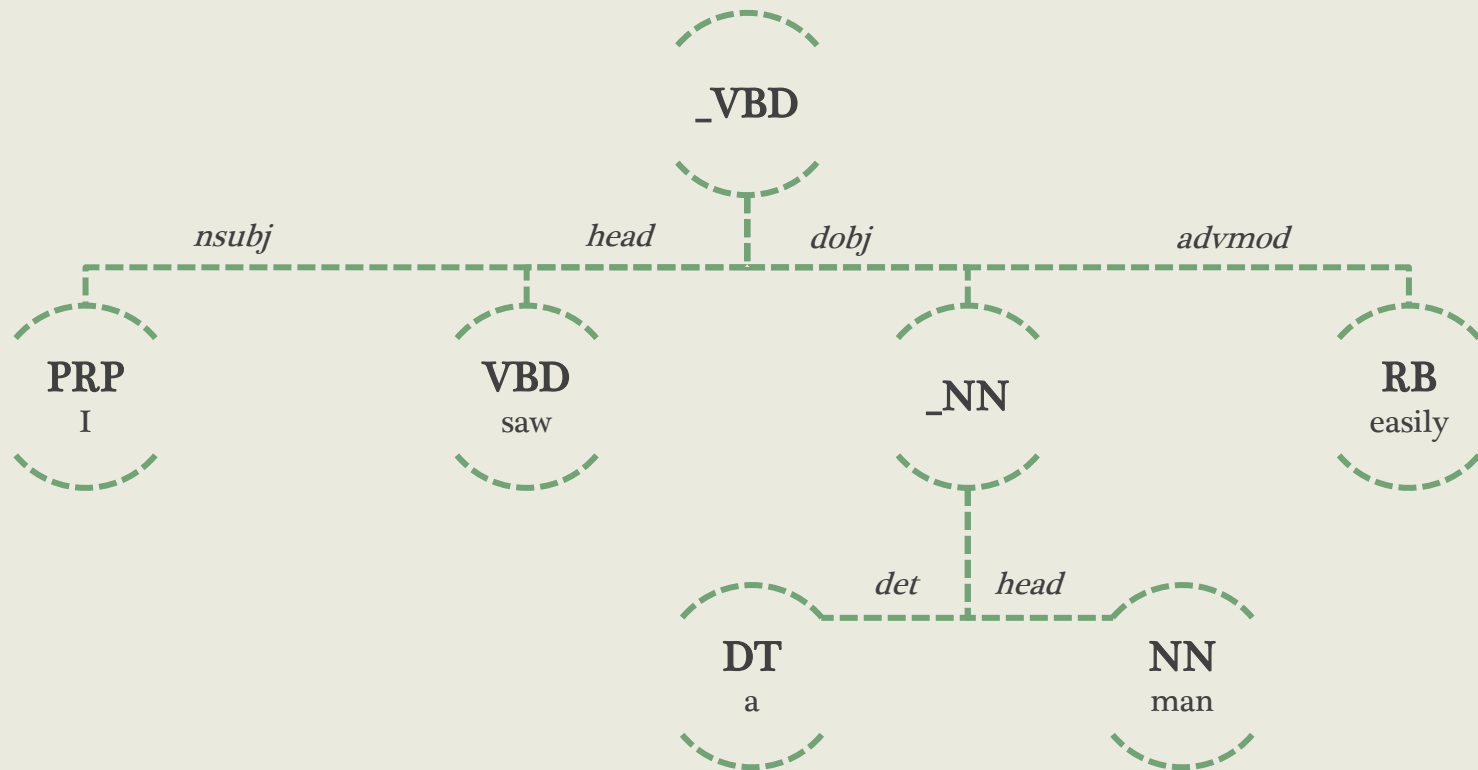
Das Prinzip

8

- Umordnungen der Quellseite der Trainingsdaten
- Darstellung als Parsebaum

Parsebaum

9



Der Regelraum

10

- Jede Regel besteht aus zwei Teilen:
 - Conditioning Context
 - Aktion
- Für jeden Knoten wird bestimmt, ob ein Conditioning Context ein Match ist, und wenn ja, wird die entsprechende Aktion durchgeführt
- Jede Regel kann nur einmal auf einen Knoten angewendet werden

Conditioning Context I

11

Konjunktion
von
Bedingungen

Bedingung =
(*Feature*,
Value)-Paar

Conditioning Context I

11

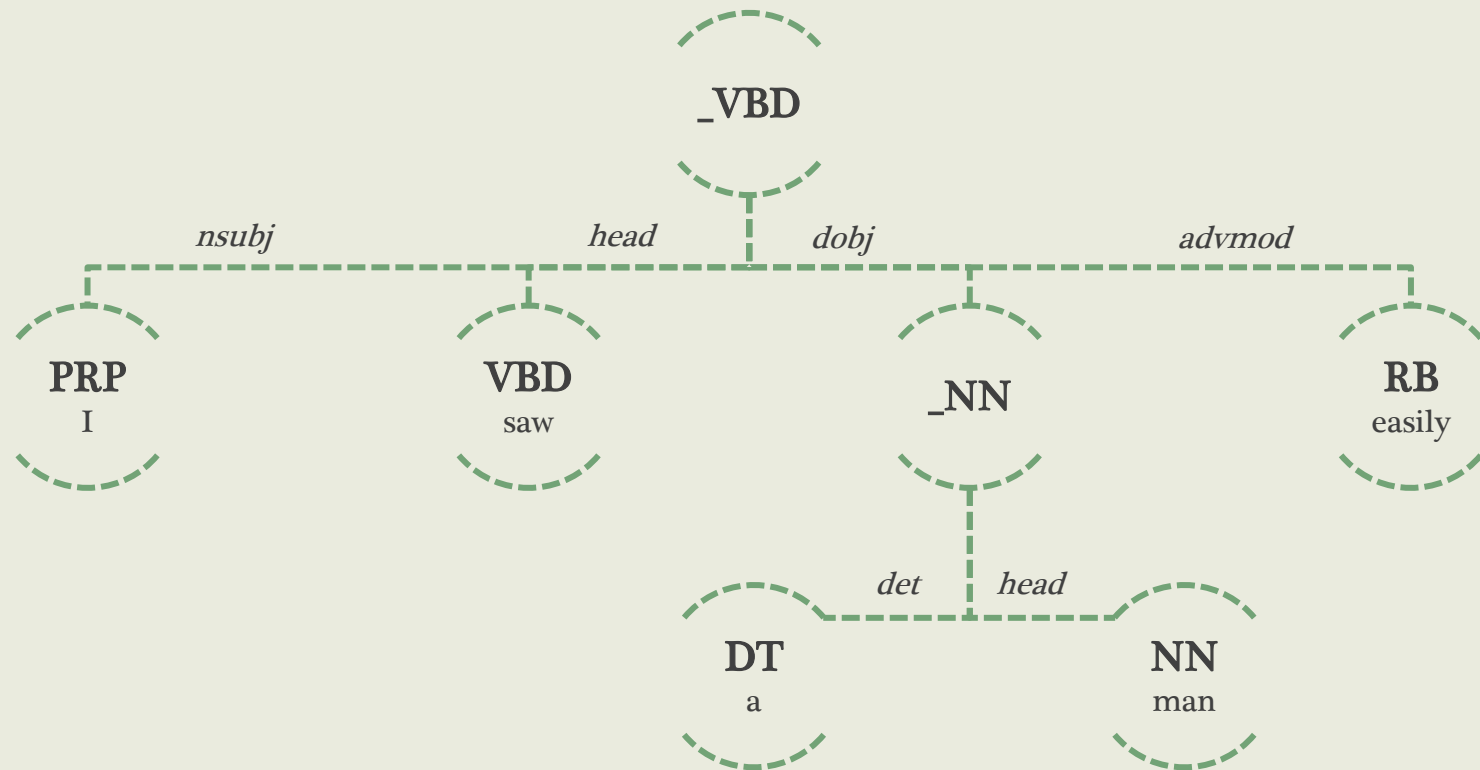
Konjunktion
von
Bedingungen

Bedingung =
(*Feature*,
Value)-Paar

Beschreibung
eines Knotens

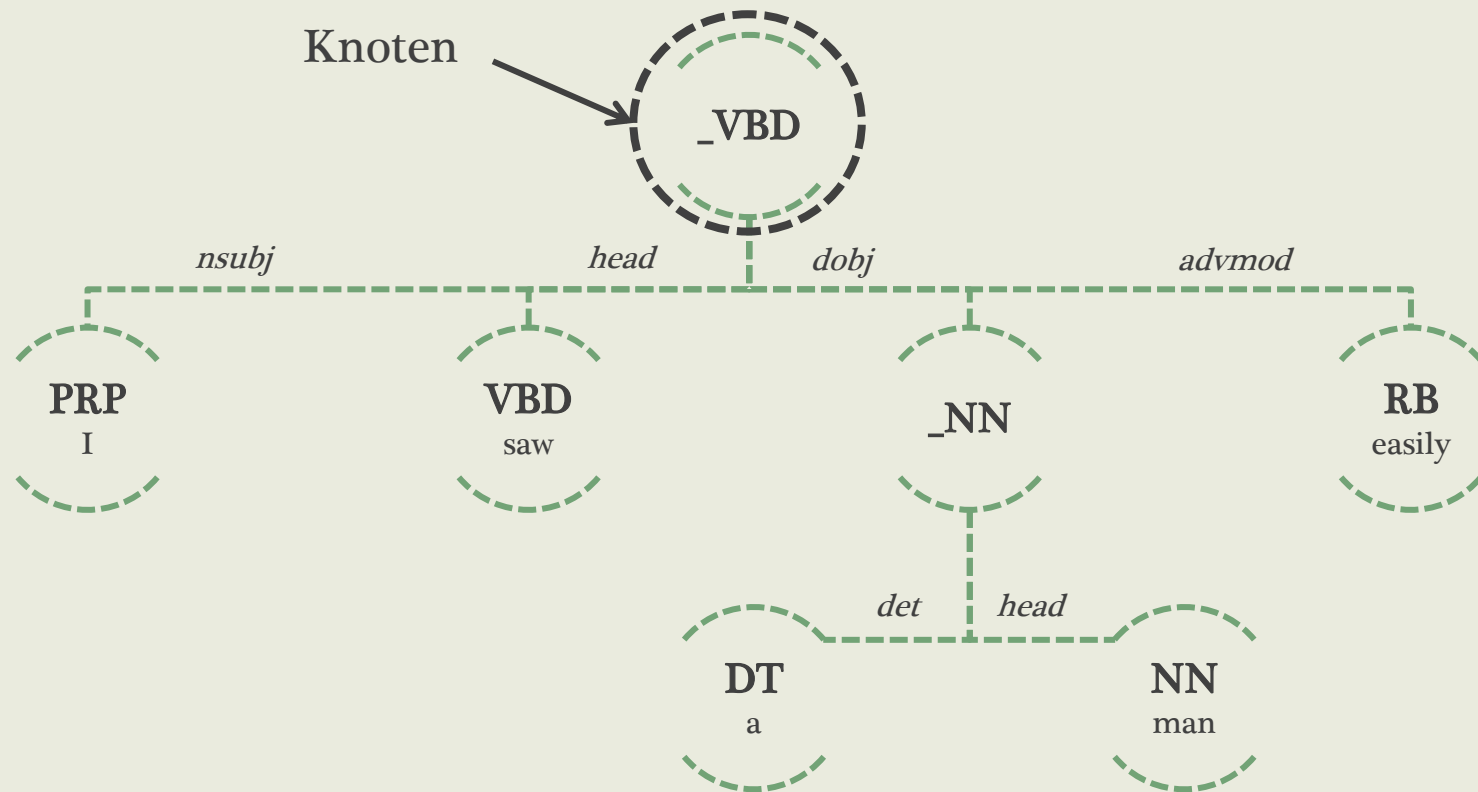
Conditioning Context II

12



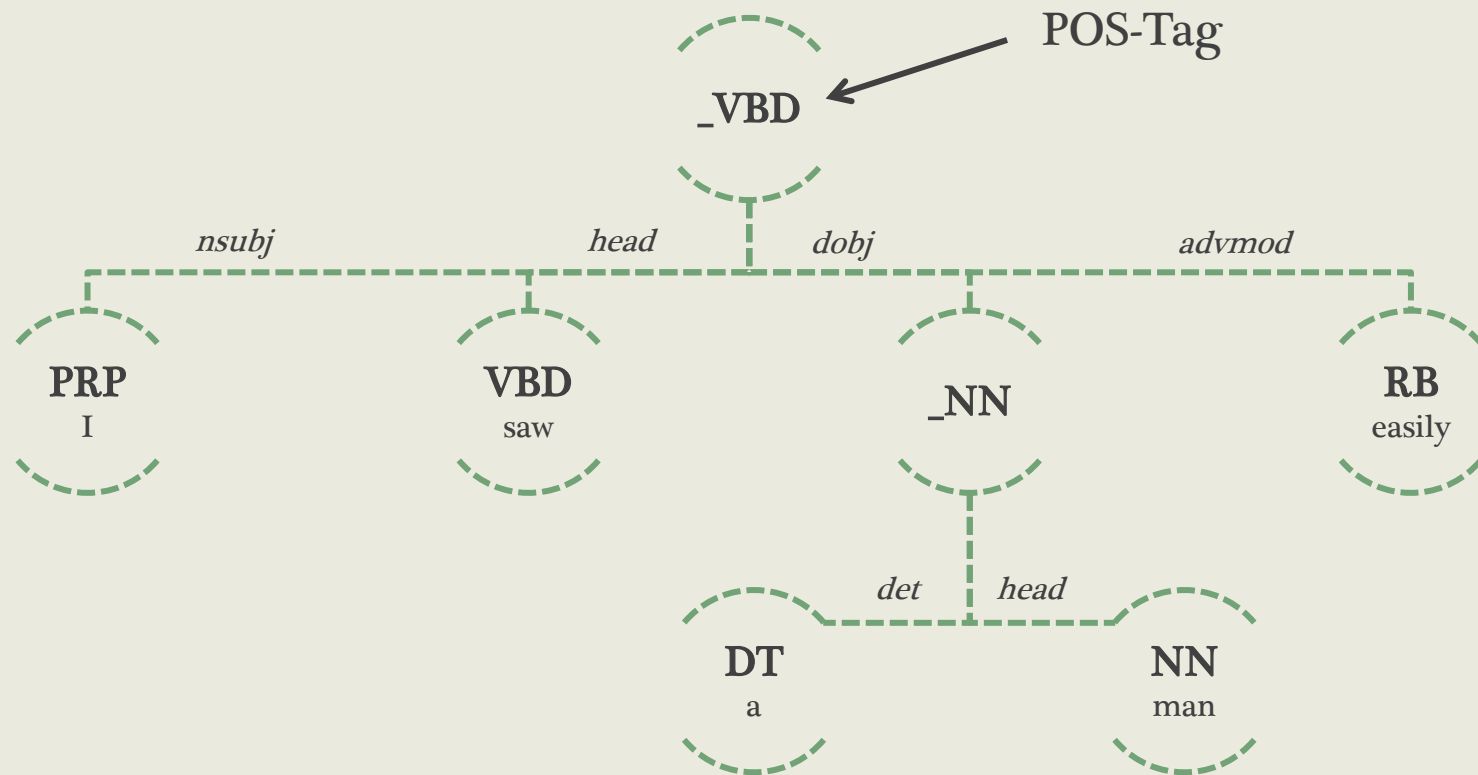
Conditioning Context II

12



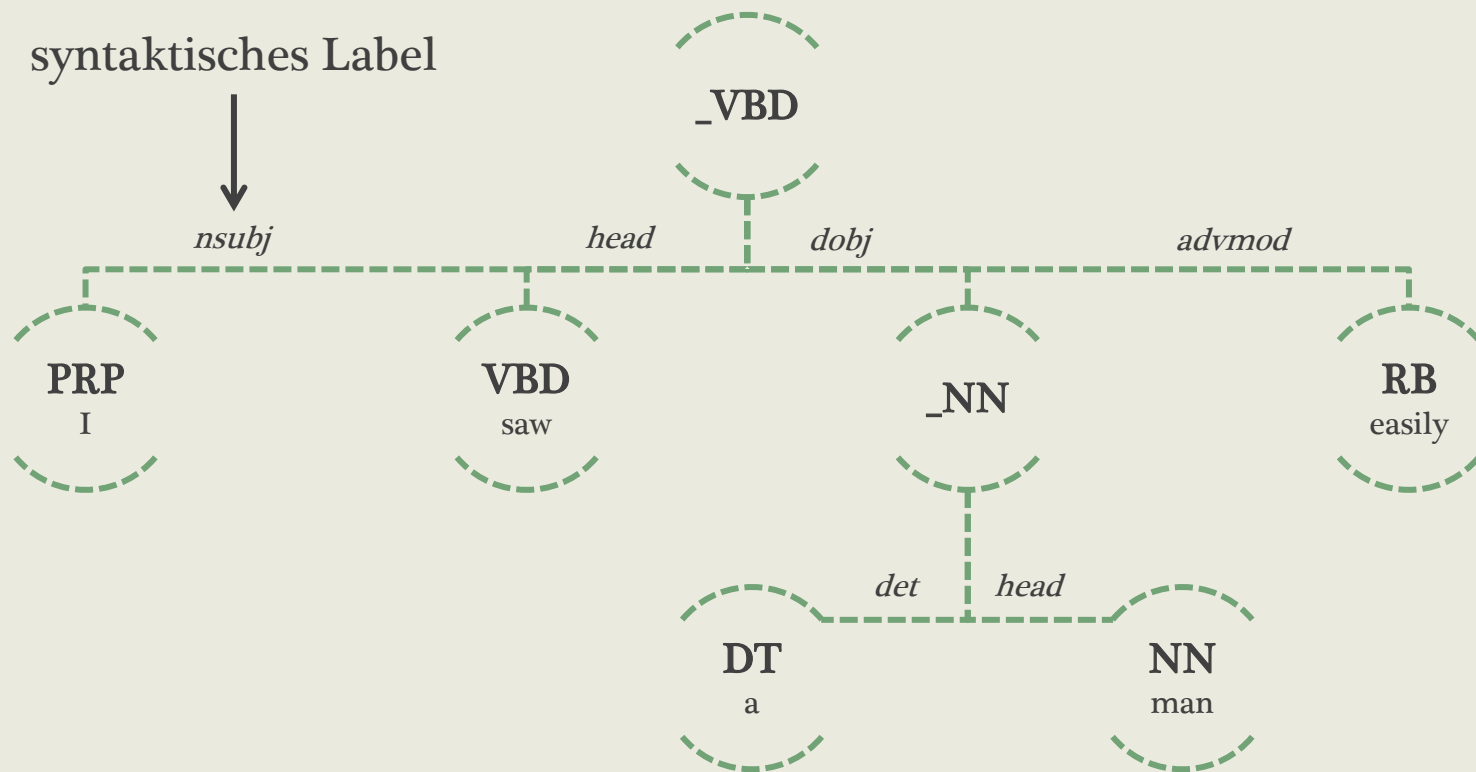
Conditioning Context II

12



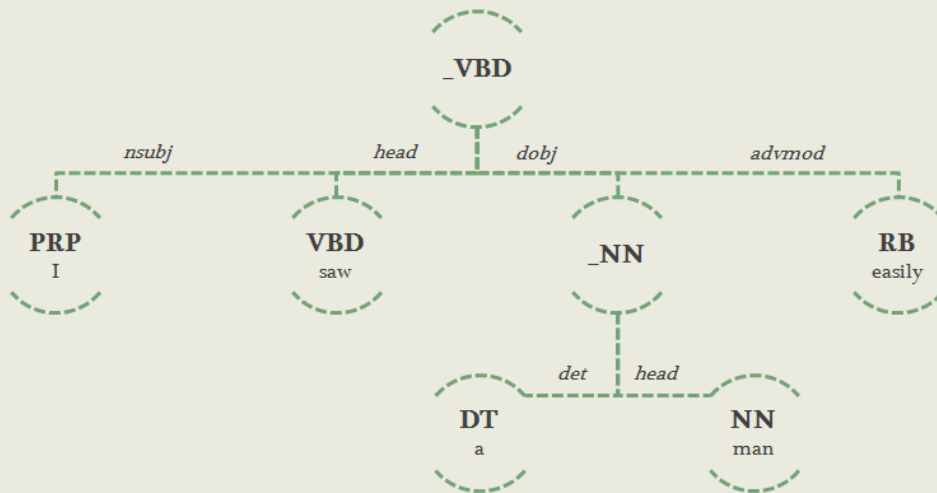
Conditioning Context II

12



Conditioning Context III

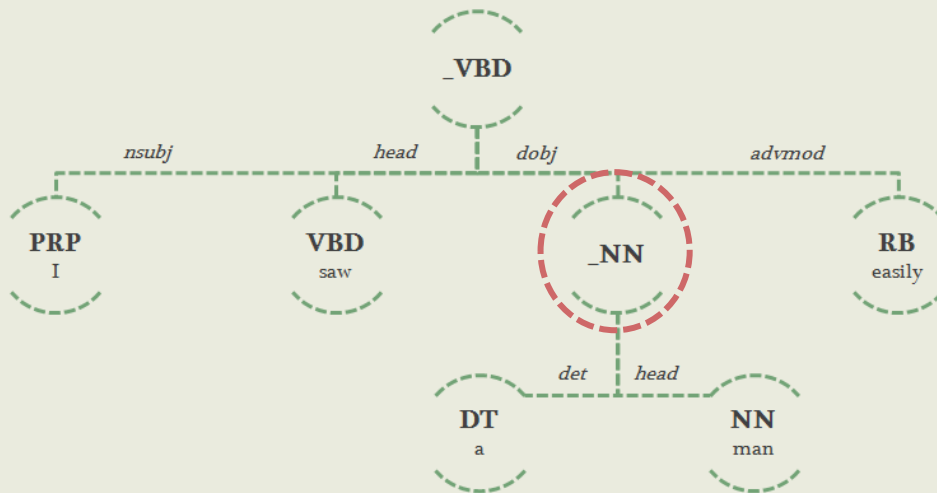
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

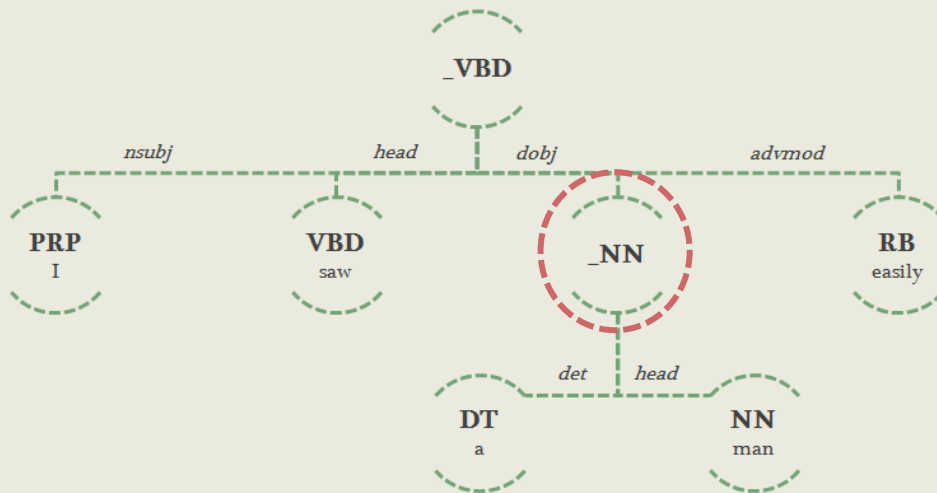
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

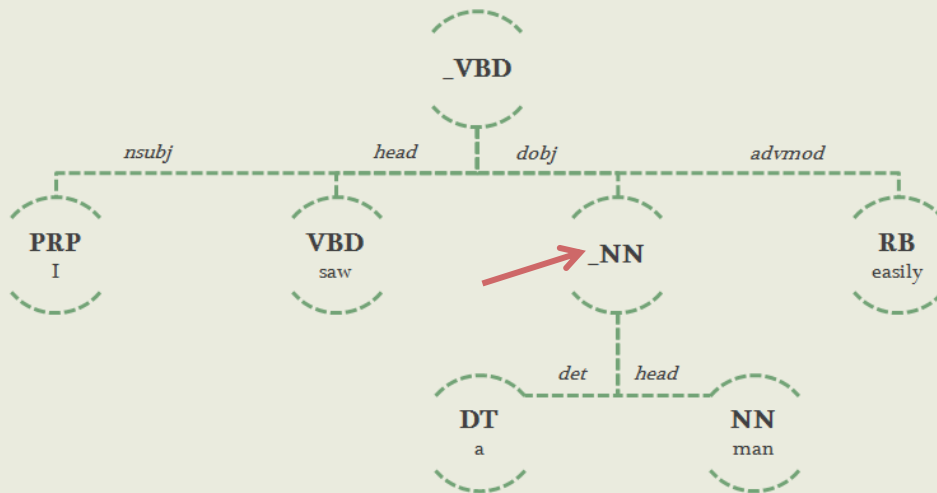
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

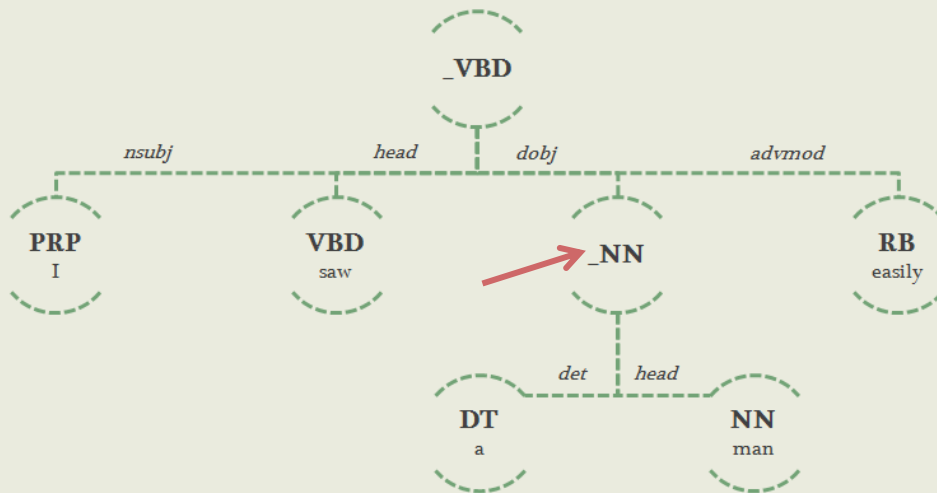
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13



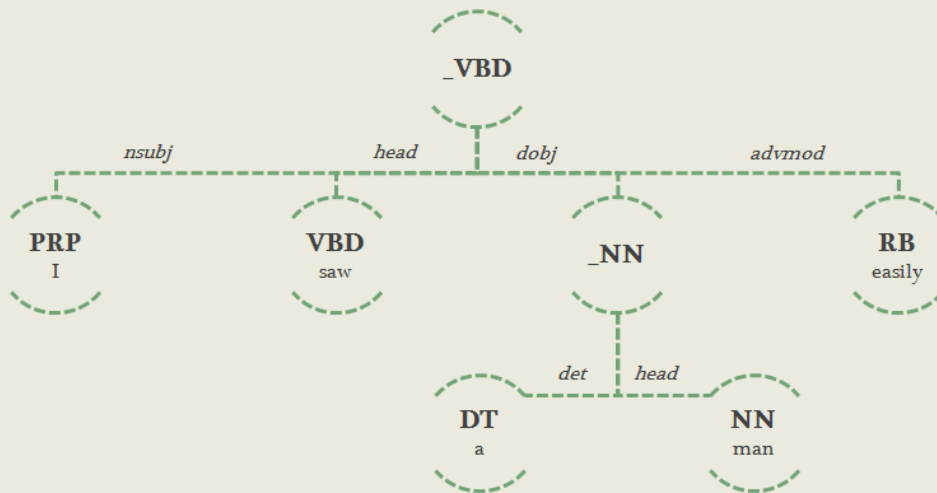
Bedingung:

$nT = NN$

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

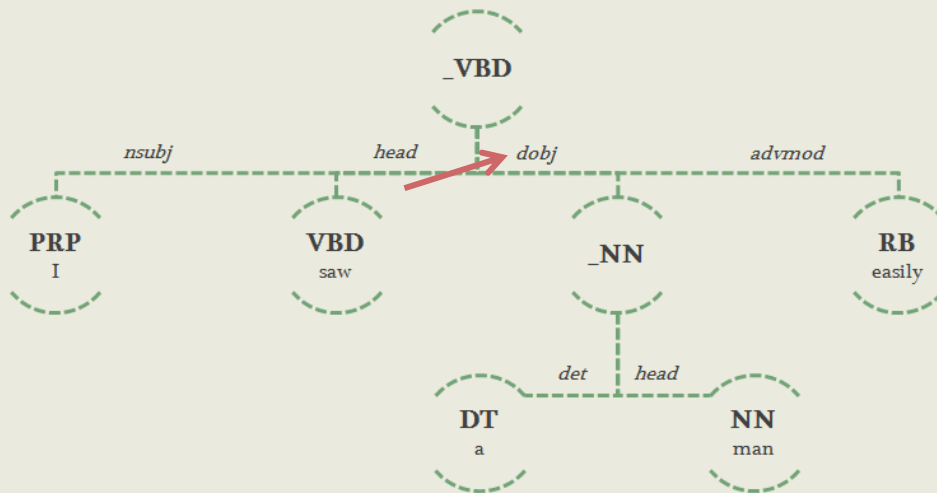
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

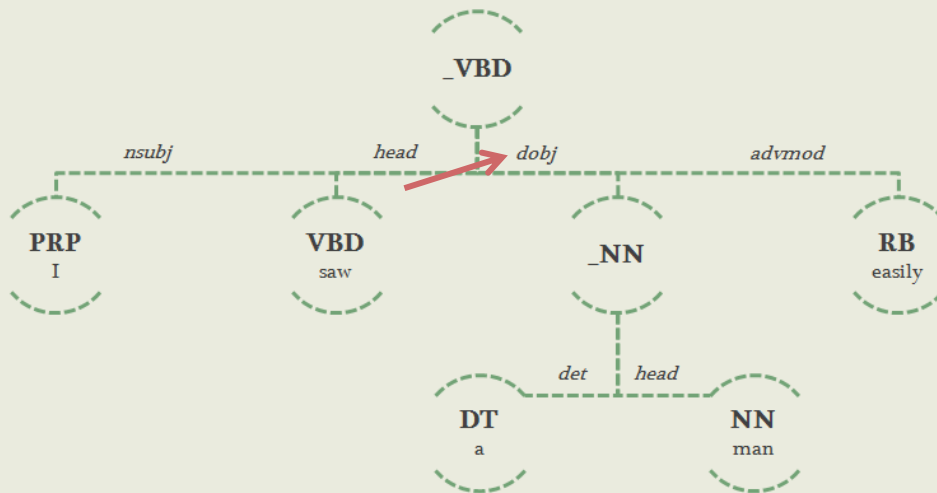
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13



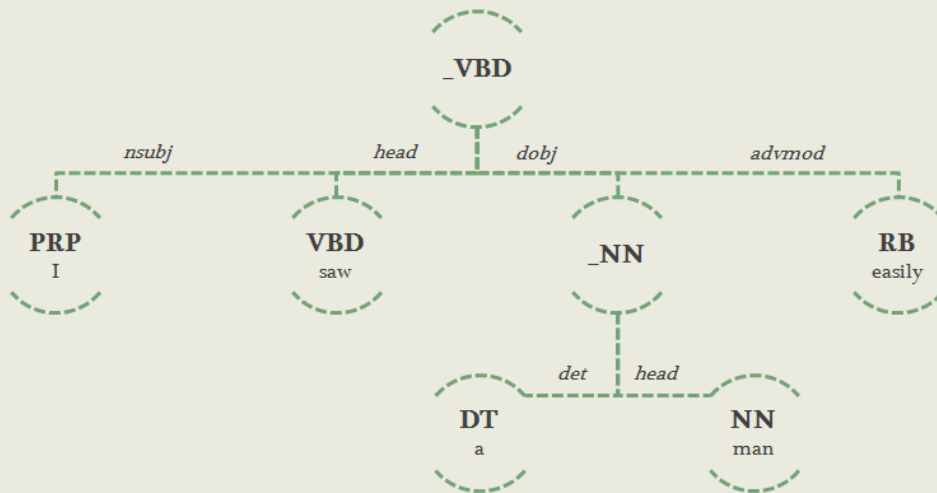
Bedingung:

nL = dobj

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

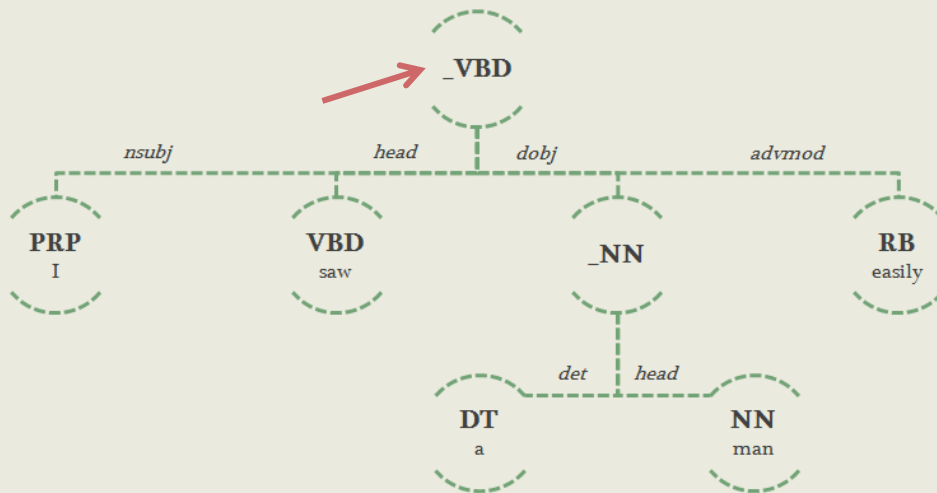
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

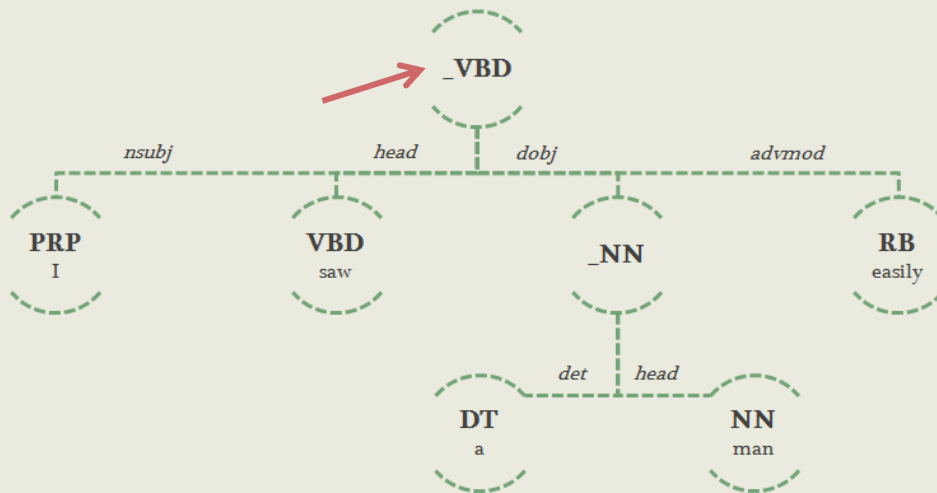
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13



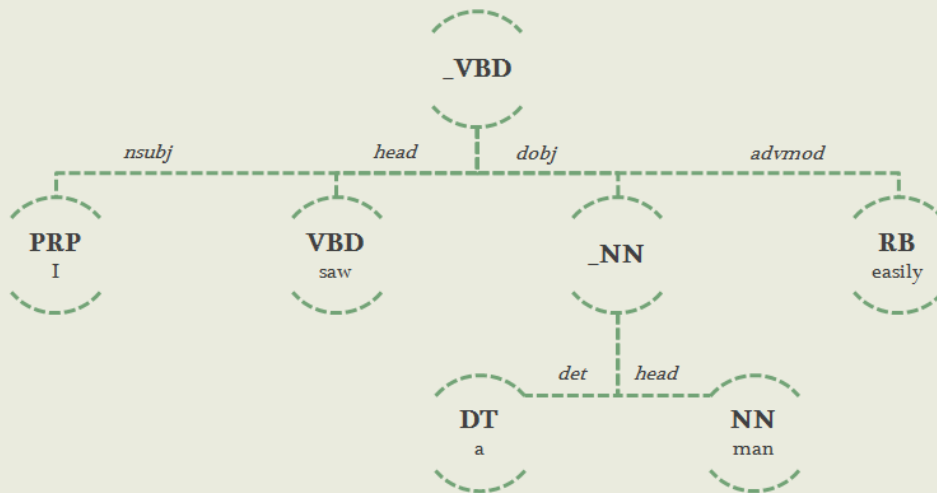
Bedingung:

pT = _VBD

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13

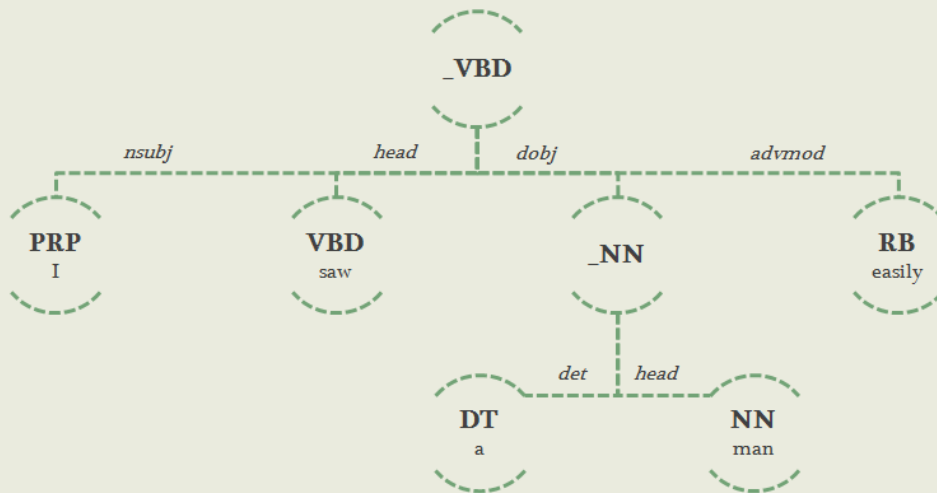


Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

13



Bedingung:

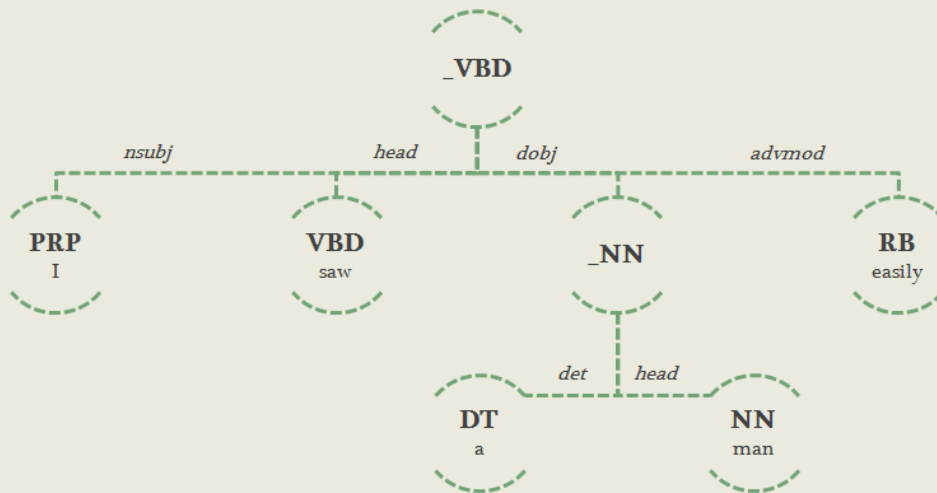
pL = none

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

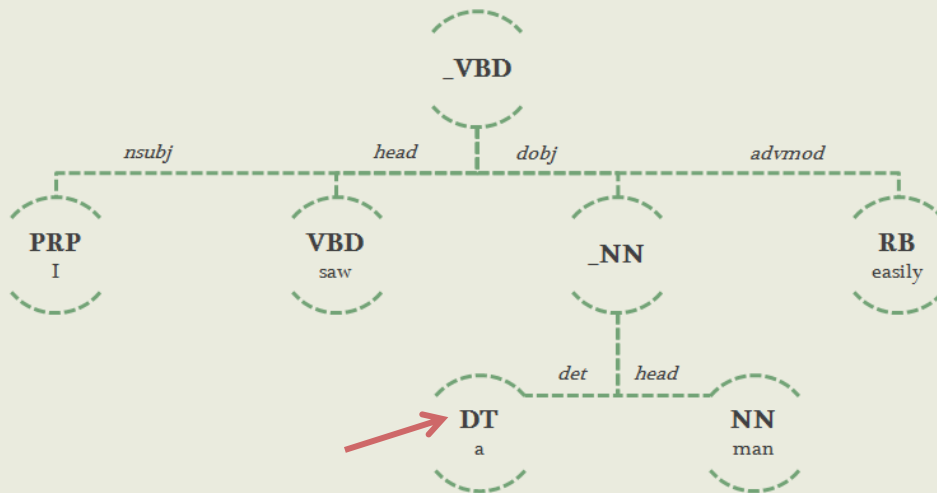
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

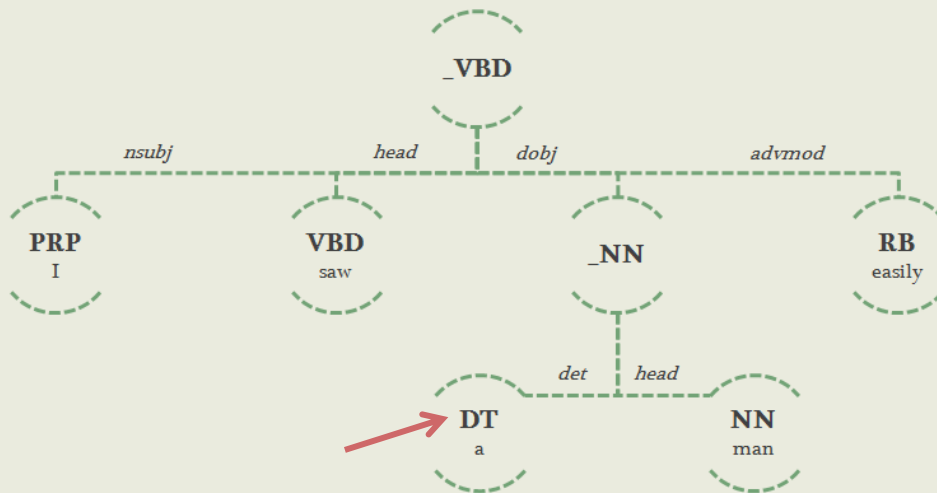
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13



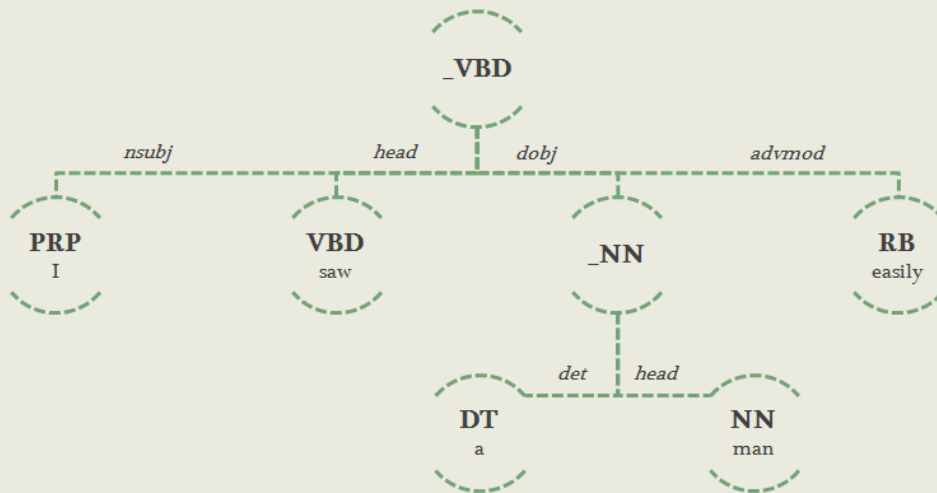
Bedingung:

1T = DT

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13

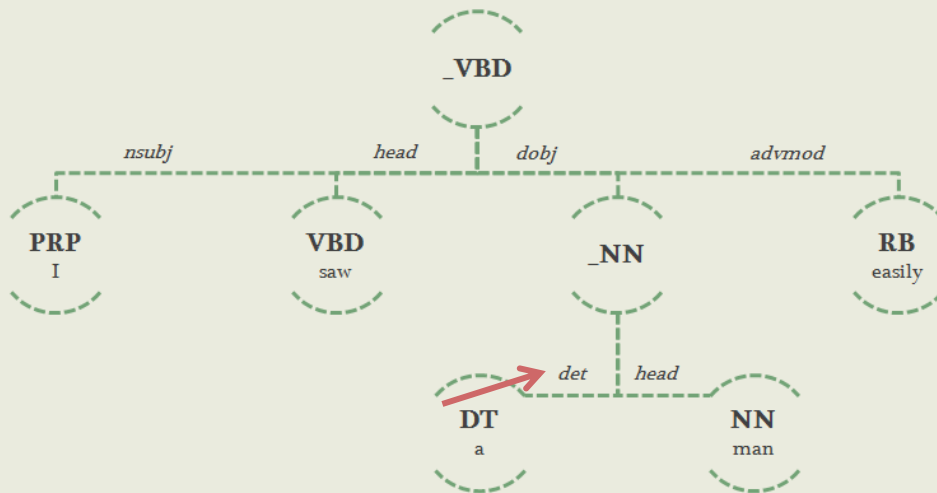


Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

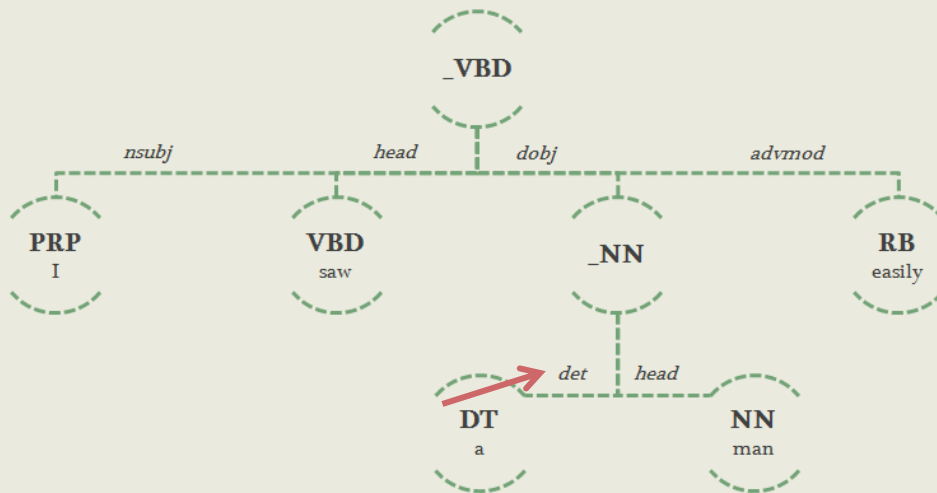
13



Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13



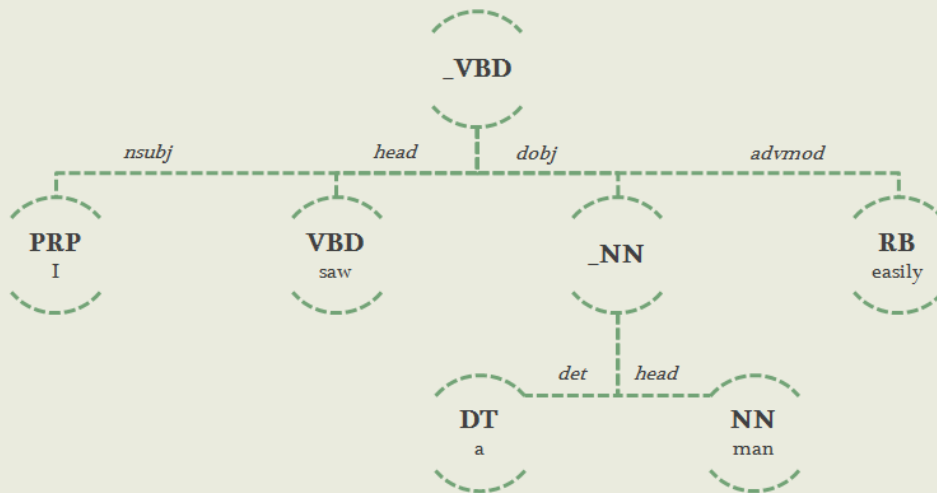
Bedingung:

1L = det

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13

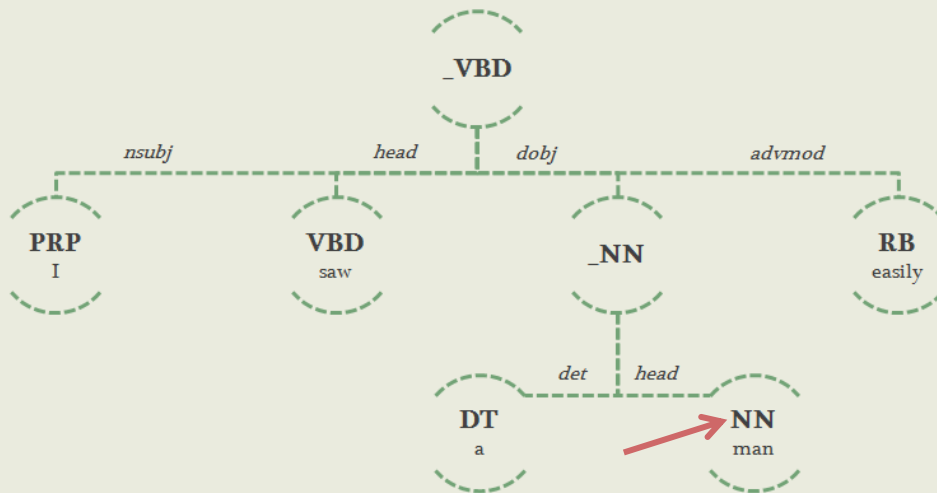


Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

13



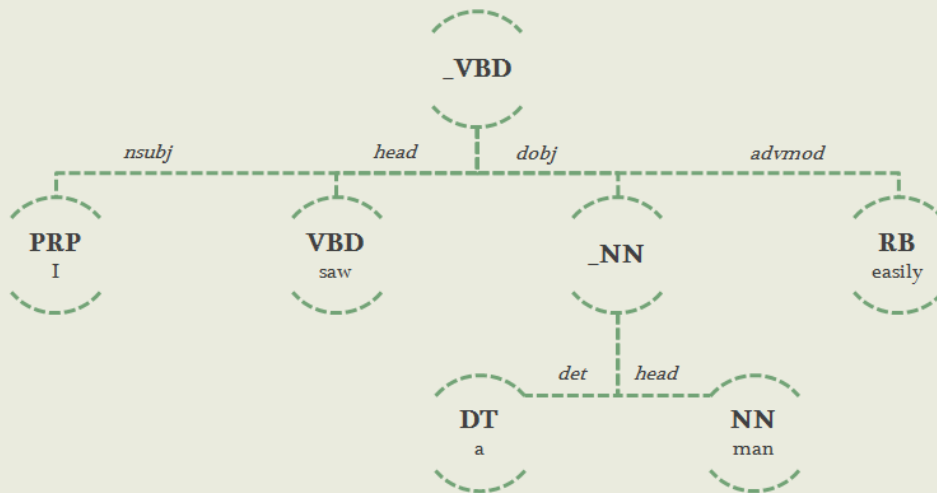
Bedingung:

2T = NN

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	

Conditioning Context III

13

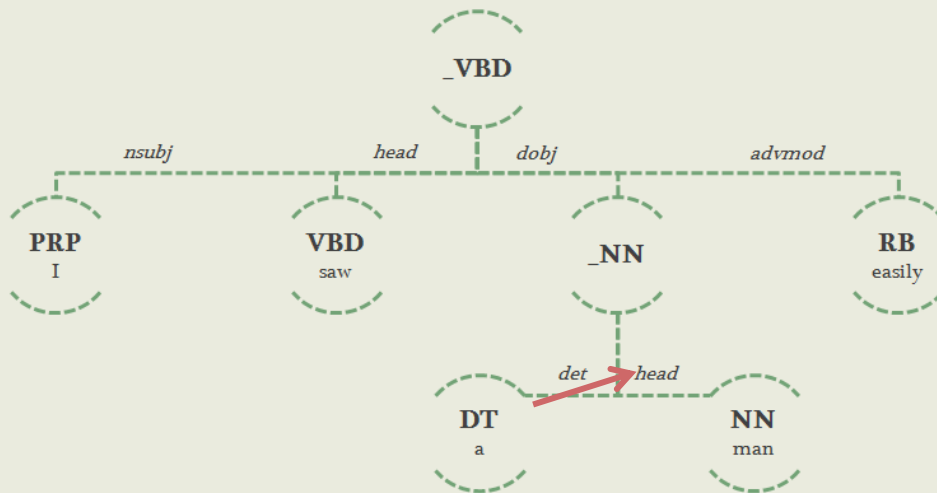


Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

13

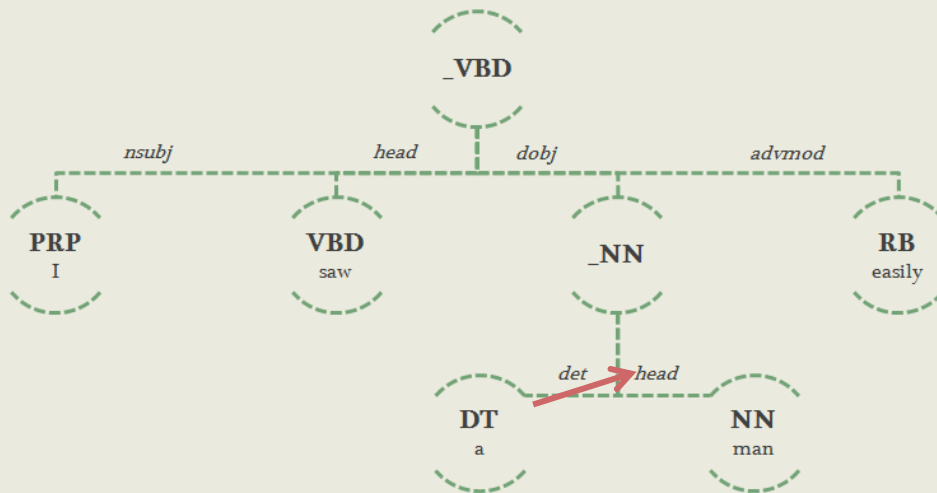


Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context III

13



Bedingung:

2L = head

Feature	Beschreibung
nT	POS-Tag des Knotens
nL	syntaktisches Label des Knotens
pT	POS-Tag des Mutterknotens
pL	syntaktisches Label des Mutterknotens
1T	POS-Tag des 1. Tochterknotens
1L	syntaktisches Label des 1. Tochterknotens
2T	POS-Tag des 2. Tochterknotens
2L	syntaktisches Label des 2. Tochterknotens
...	



Conditioning Context IV

14

Ein Conditioning Context ist eine Konjunktion aus Bedingungen, z.B.:

nT = _VBD

1T = PRP

1L = nsubj

3T = _NN

3L = dobj

Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

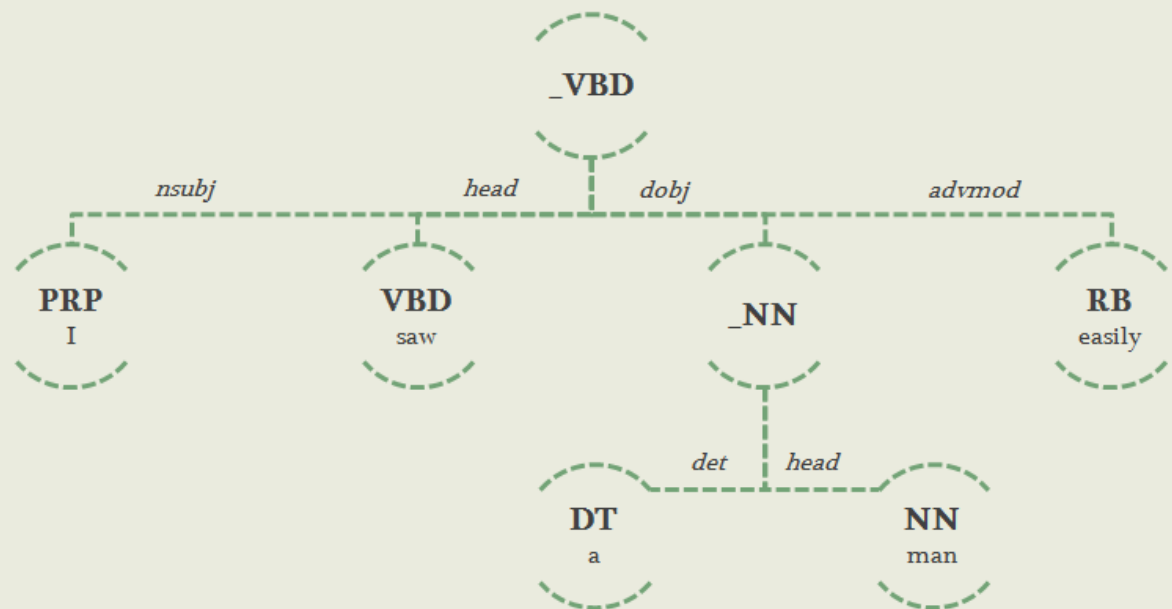
nT = _VBD

1T = PRP

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

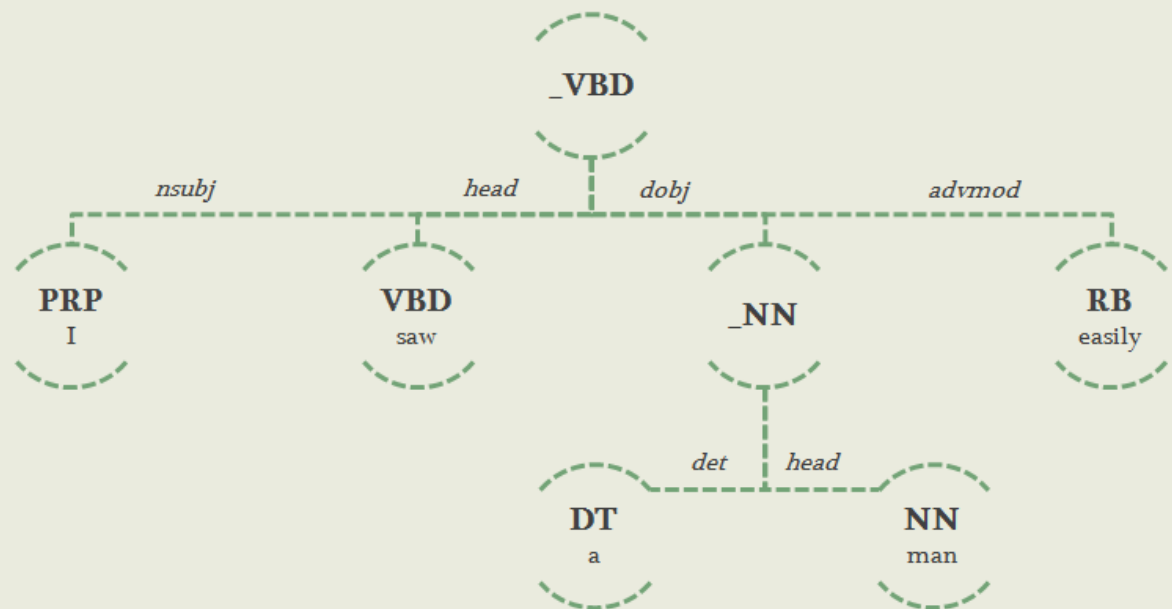
nT = _VBD

1T = PRP

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

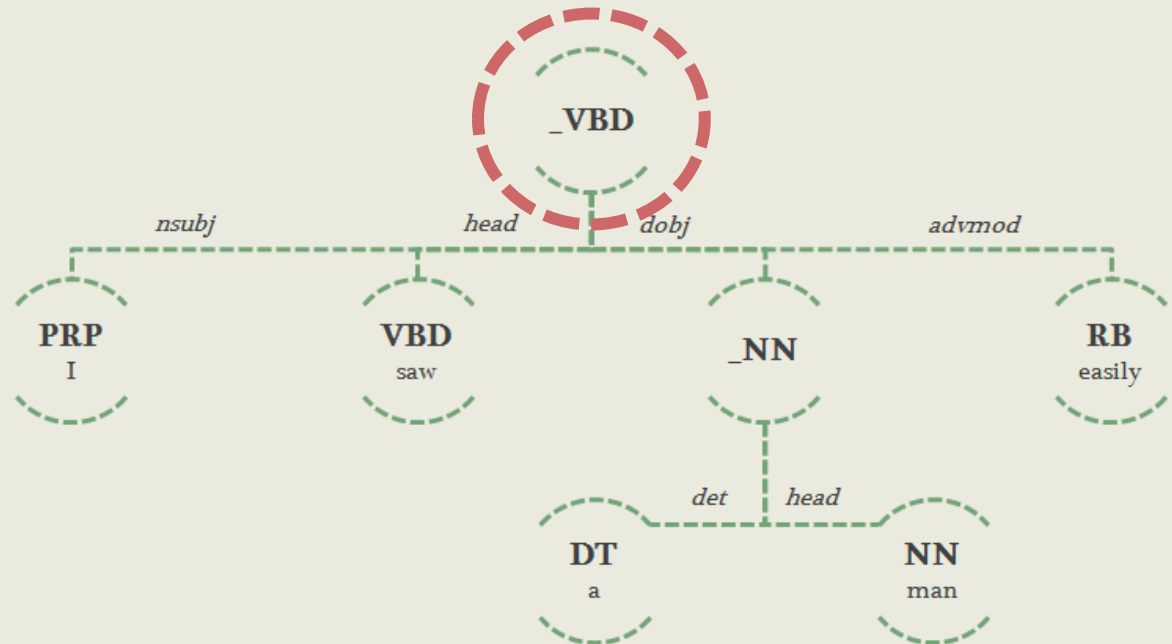
nT = _VBD

1T = PRP

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

nT = _VBD

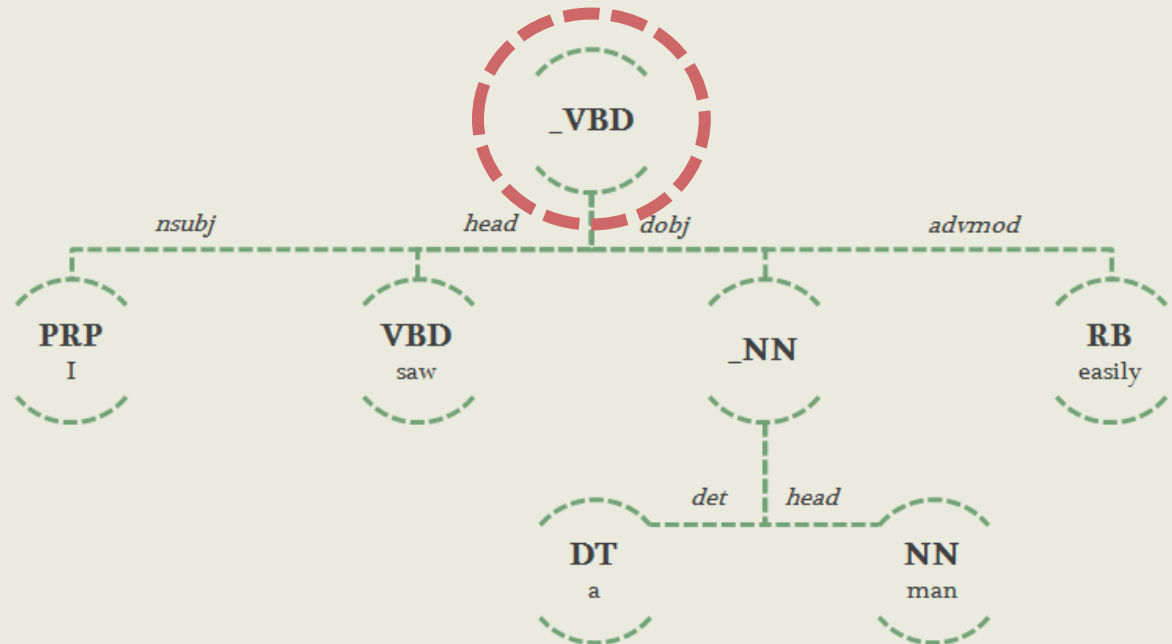


1T = PRP

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

nT = _VBD

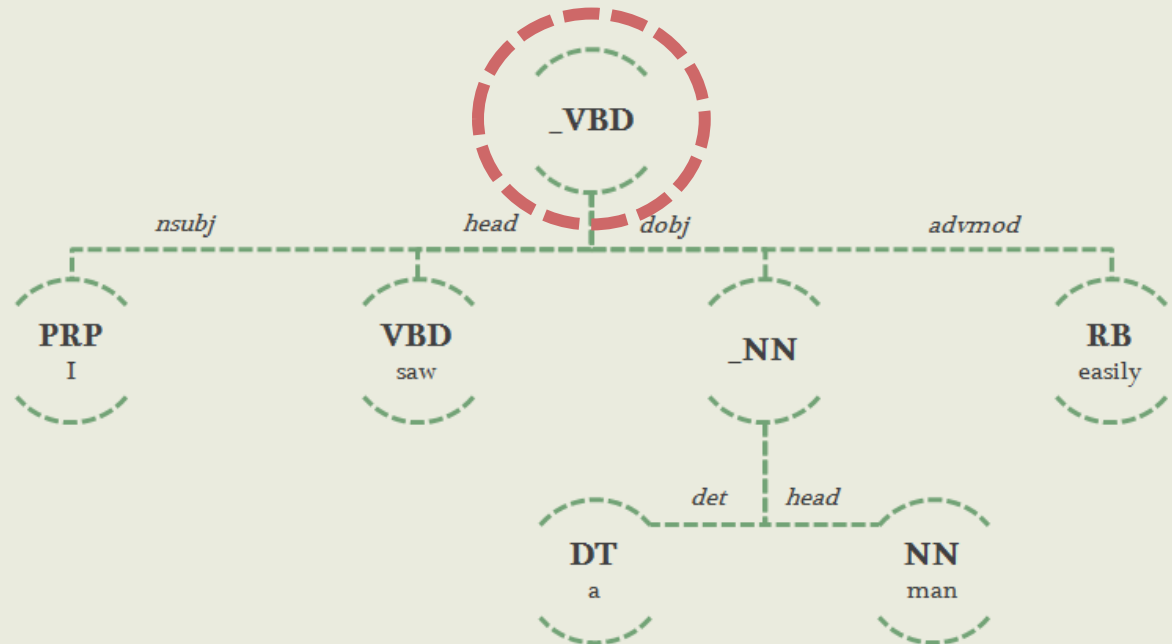


1T = PRP

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

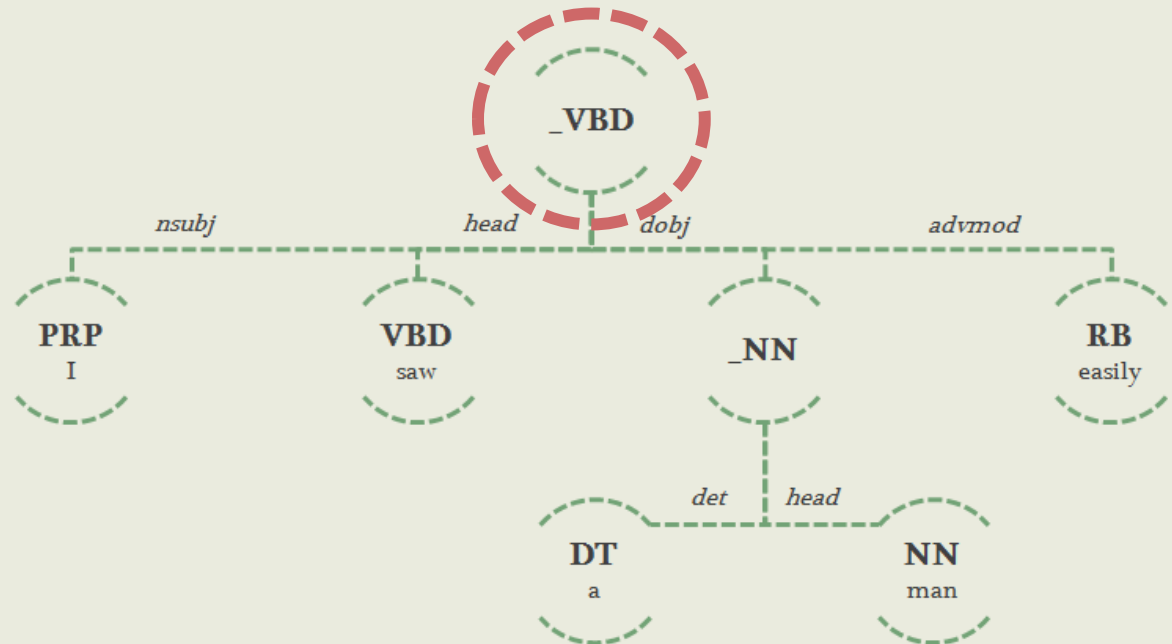
nT = _VBD ✓

1T = PRP ✓

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

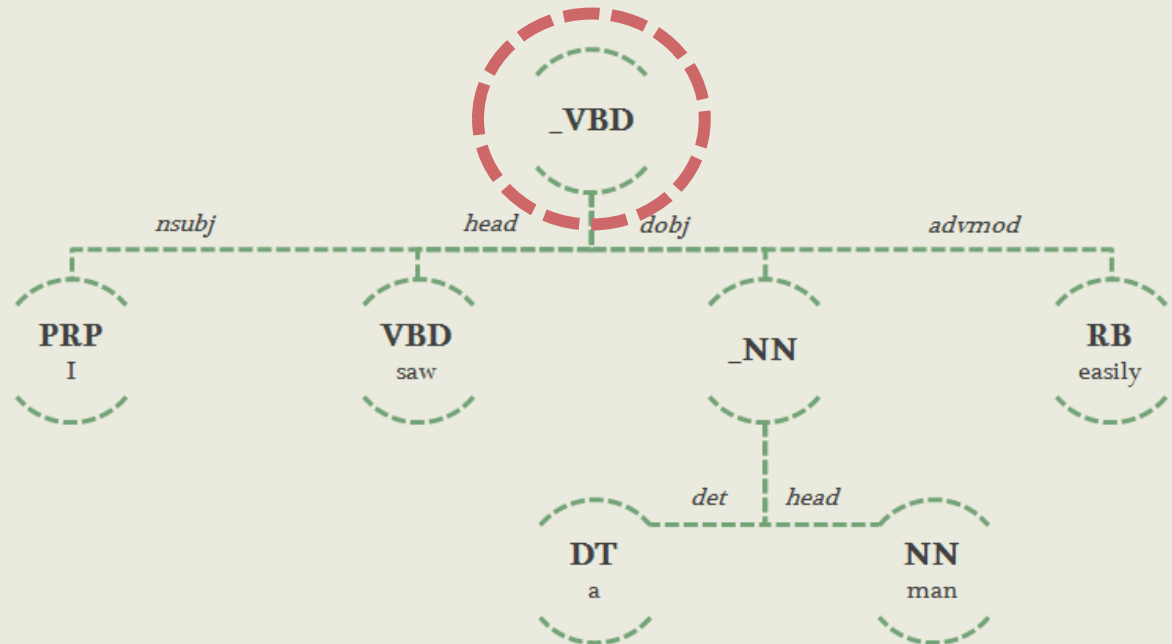
nT = _VBD ✓

1T = PRP ✓

1L = nsubj

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

nT = _VBD



1T = PRP

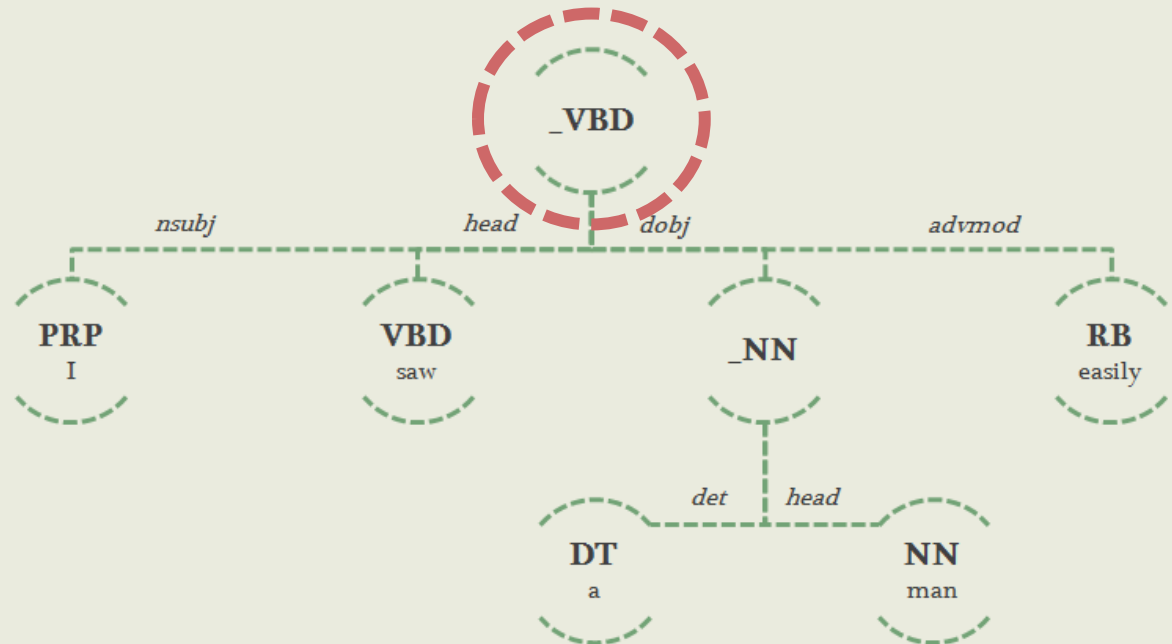


1L = nsubj



3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

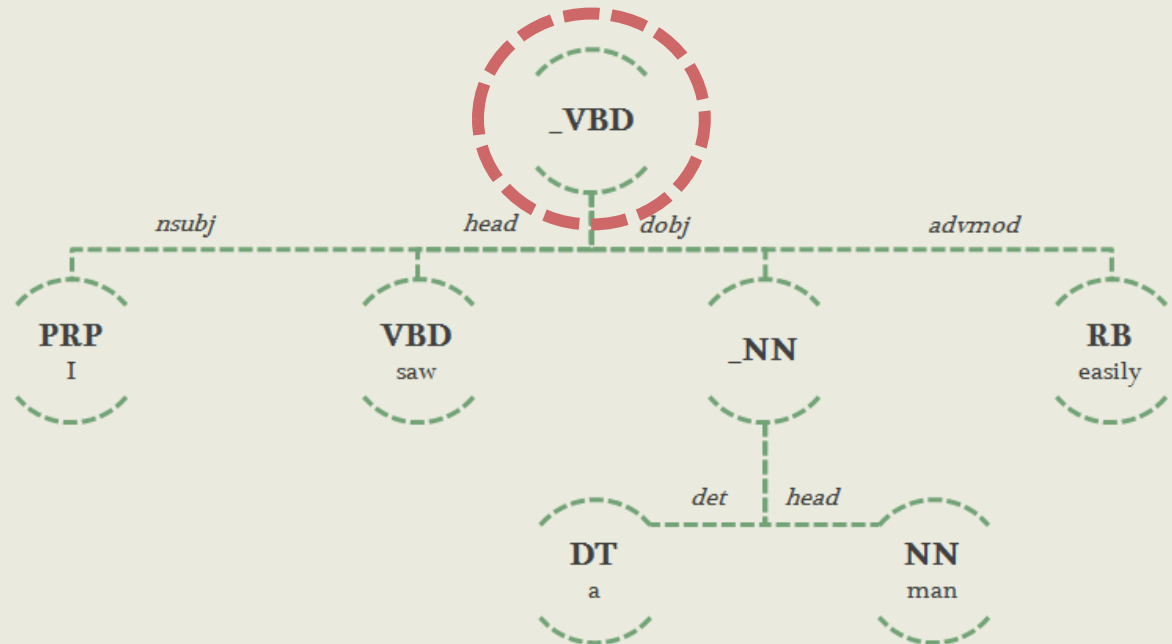
nT = _VBD ✓

1T = PRP ✓

1L = nsubj ✓

3T = _NN

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

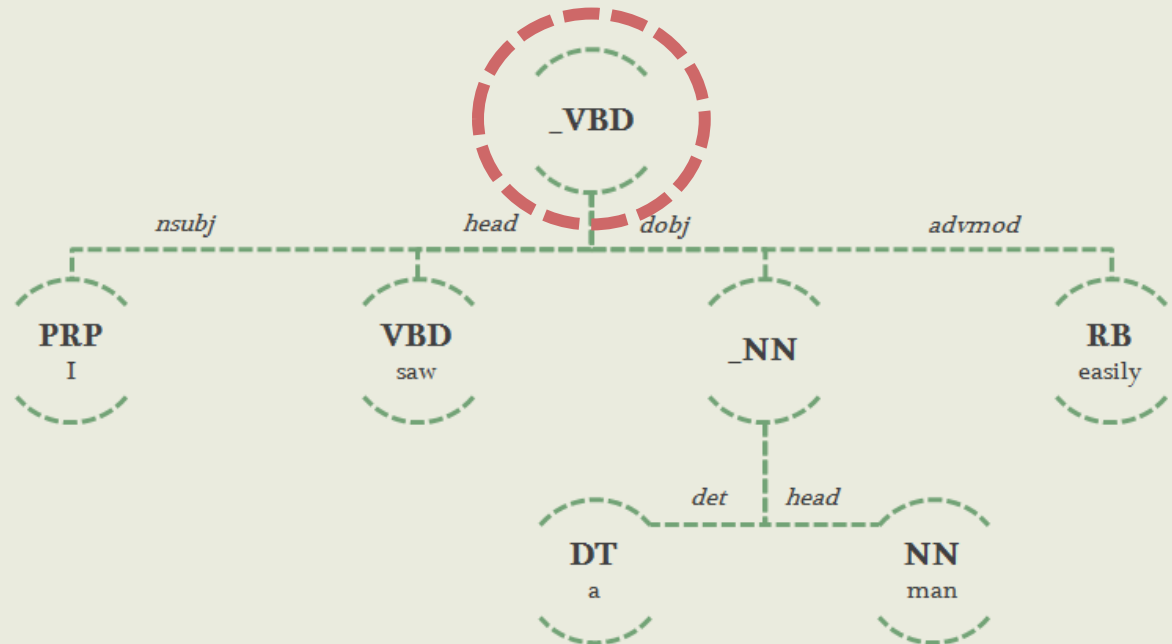
nT = _VBD ✓

1T = PRP ✓

1L = nsubj ✓

3T = _NN ✓

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

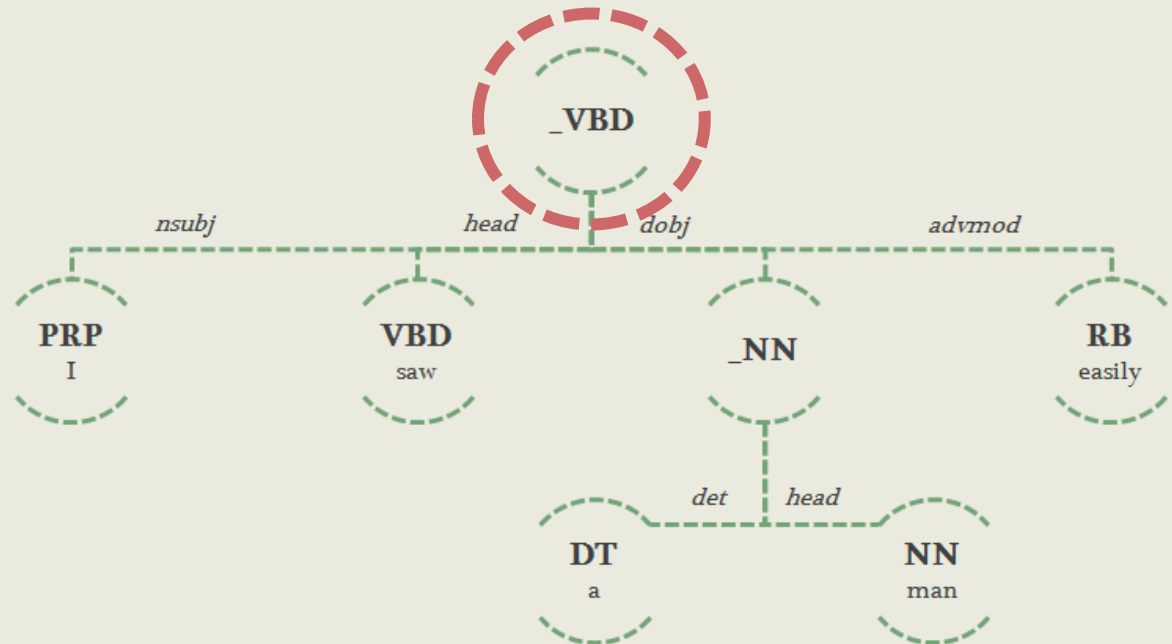
nT = _VBD ✓

1T = PRP ✓

1L = nsubj ✓

3T = _NN ✓

3L = dobj



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

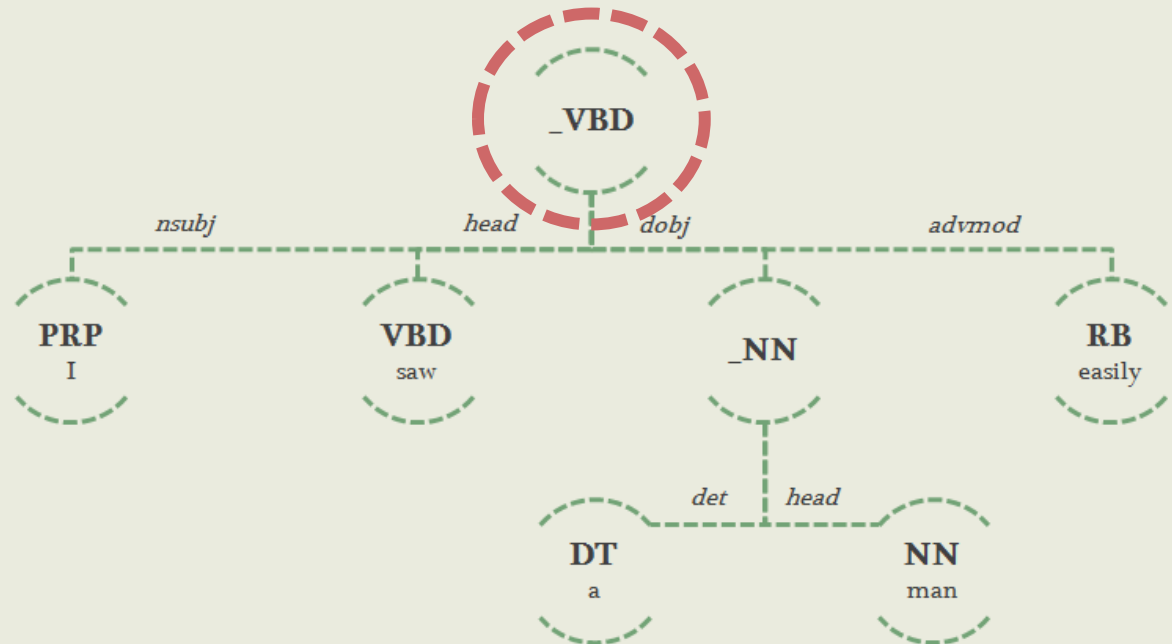
nT = _VBD ✓

1T = PRP ✓

1L = nsubj ✓

3T = _NN ✓

3L = dobj ✓



Conditioning Context V

15

Trifft ein Conditioning Context auf einen Knoten zu, wird eine Aktion durchgeführt.

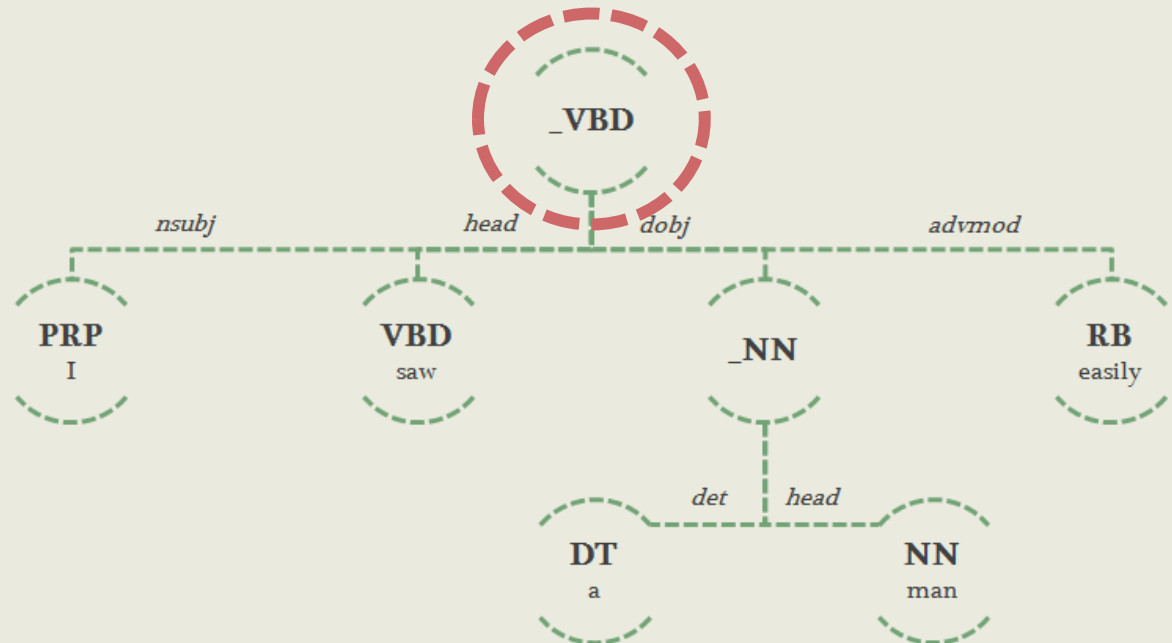
nT = _VBD ✓

1T = PRP ✓

1L = nsubj ✓

3T = _NN ✓

3L = dobj ✓



Regel kann durchgeführt werden.

Umordnungs-Metriken

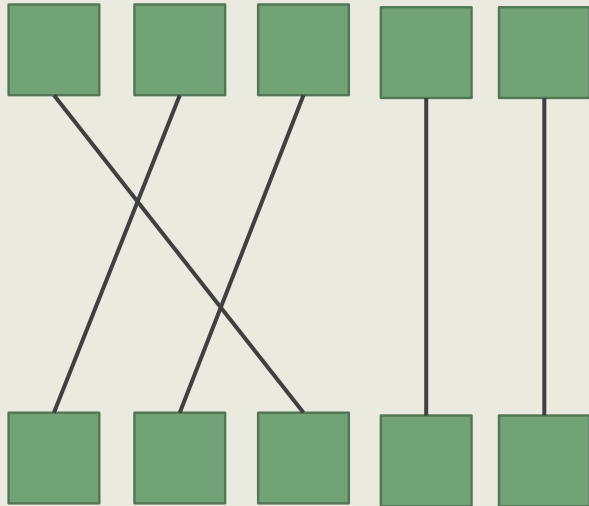
16

- *Crossing Score*: Anzahl der Kreuzungen von Alignments
- *Estimated BLEU Gain*: Unterschied des BLEU Scores mit und ohne die Anwendung einer Regel

Ziel: Nützlichkeit einer Regel erfassen, nicht des ganzen Systems

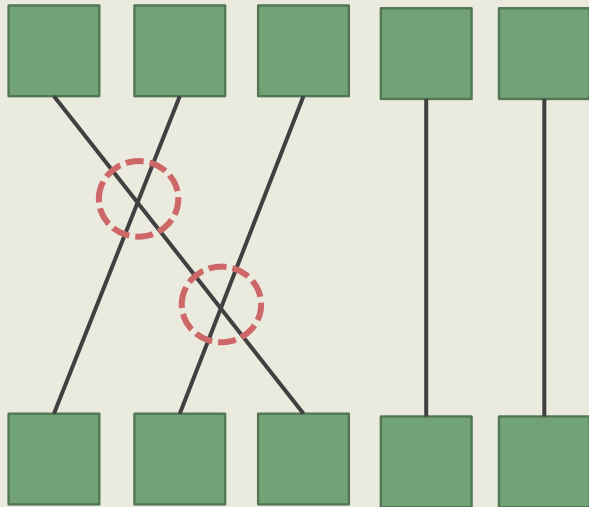
Crossing Score

17



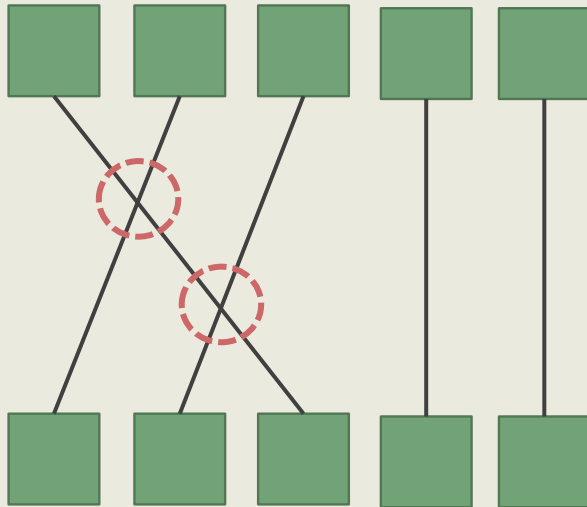
Crossing Score

17



Crossing Score

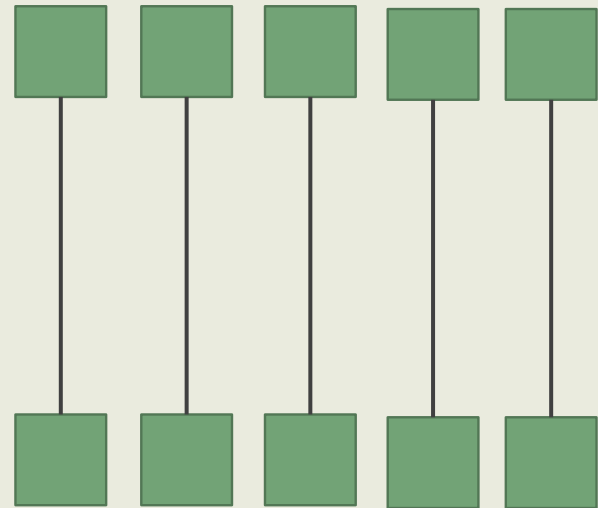
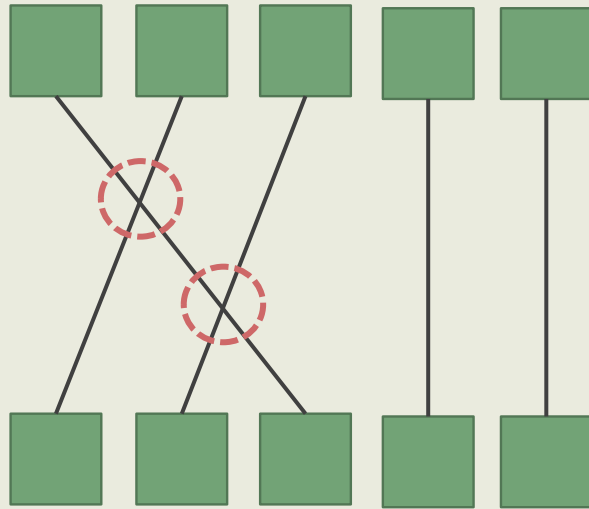
17



2 Crossings

Crossing Score

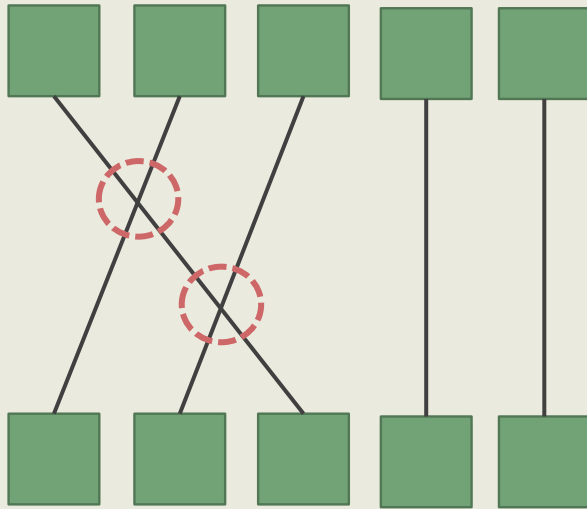
17



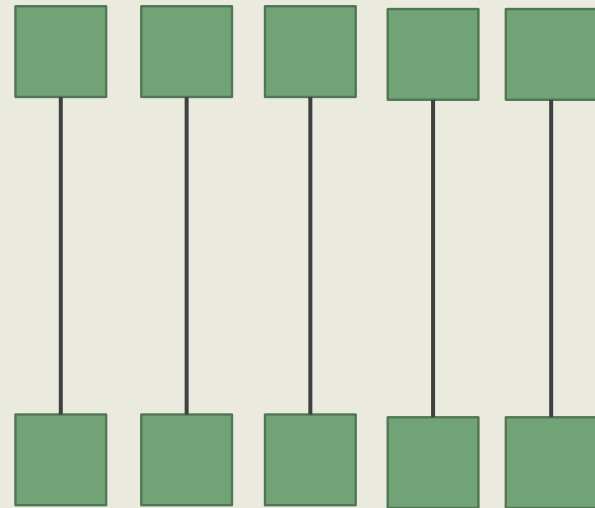
2 Crossings

Crossing Score

17



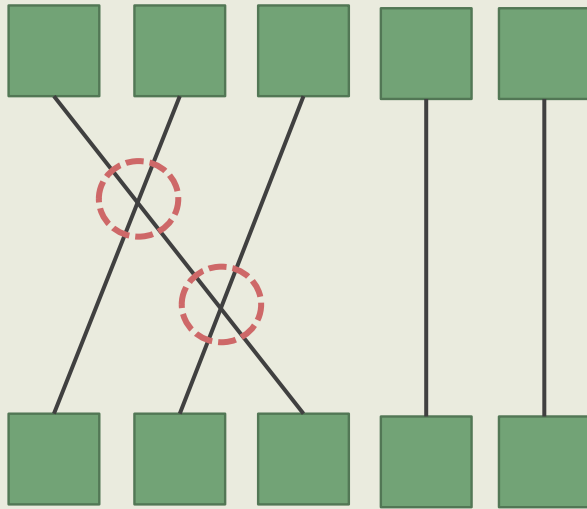
2 Crossings



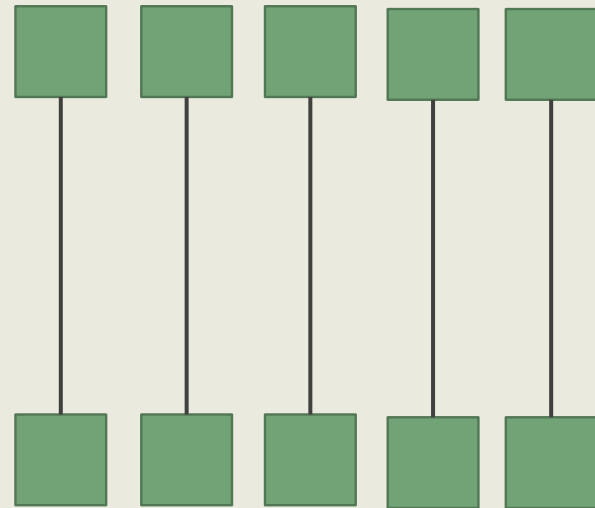
0 Crossings

Crossing Score

17



2 Crossings



0 Crossings

Optimal

Estimated BLEU Gain

18

- BLEU: Bilingual Evaluation Study
- Evaluiert Qualität einer maschinellen Übersetzung
- *Estimated BLEU Gain*
Was man nach Anwendung einer Regel bzw. einer Sequenz an Regeln an BLEU „dazugewinnen“ würde

Agenda

19

Das Problem

Der Lösungsansatz

Der Algorithmus

Evaluation

Vorbereitungen

20

- Input: Wort-alignierte Satzpaare (nach IBM 1)
- Alignments mit einer Wahrscheinlichkeit unter 0,25 entfernen
- Anschließend Satzpaare mit nur sehr wenigen Alignments löschen

ZIEL: Optimierung der Alignment Links

Funktionsweise I

21

```
input: A set of aligned sentence pairs
base = <empty sequence>
for several iterations do
    candidate_rules=GenerateAllCandidateRules(input,base)
    base.append(MinCost(candidate_rules))
end for
```


Funktionsweise I

21

```
input: A set of aligned sentence pairs
base = <empty sequence>
for several iterations do
    candidate_rules=GenerateAllCandidateRules(input,base)
    base.append(MinCost(candidate_rules))
end for
```

Wortalignierte Satzpaare werden eingelesen

Funktionsweise I

21

```
input: A set of aligned sentence pairs
base = <empty sequence>
for several iterations do
    candidate_rules=GenerateAllCandidateRules(input,base)
    base.append(MinCost(candidate_rules))
end for
```

„Speicher“ für die besten Regeln

Funktionsweise I

21

input: A set of aligned sentence pairs

base = <empty sequence>

for several iterations **do**

 candidate_rules=GenerateAllCandidateRules(input,base)

 base.append(MinCost(candidate_rules))

end for

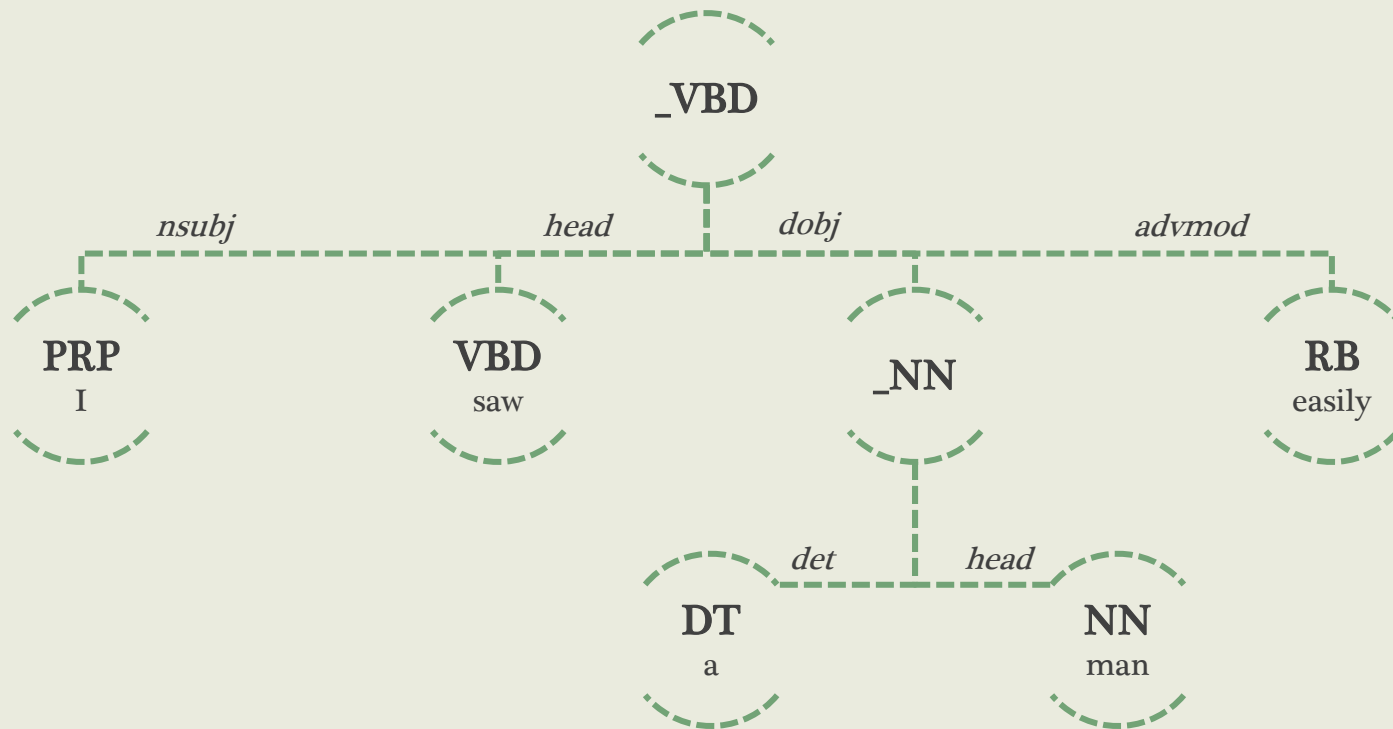
Ermittlung der besten Regel

Nach jeder Iteration wird die Sequenz um eine Regel erweitert

Funktionsweise II

22

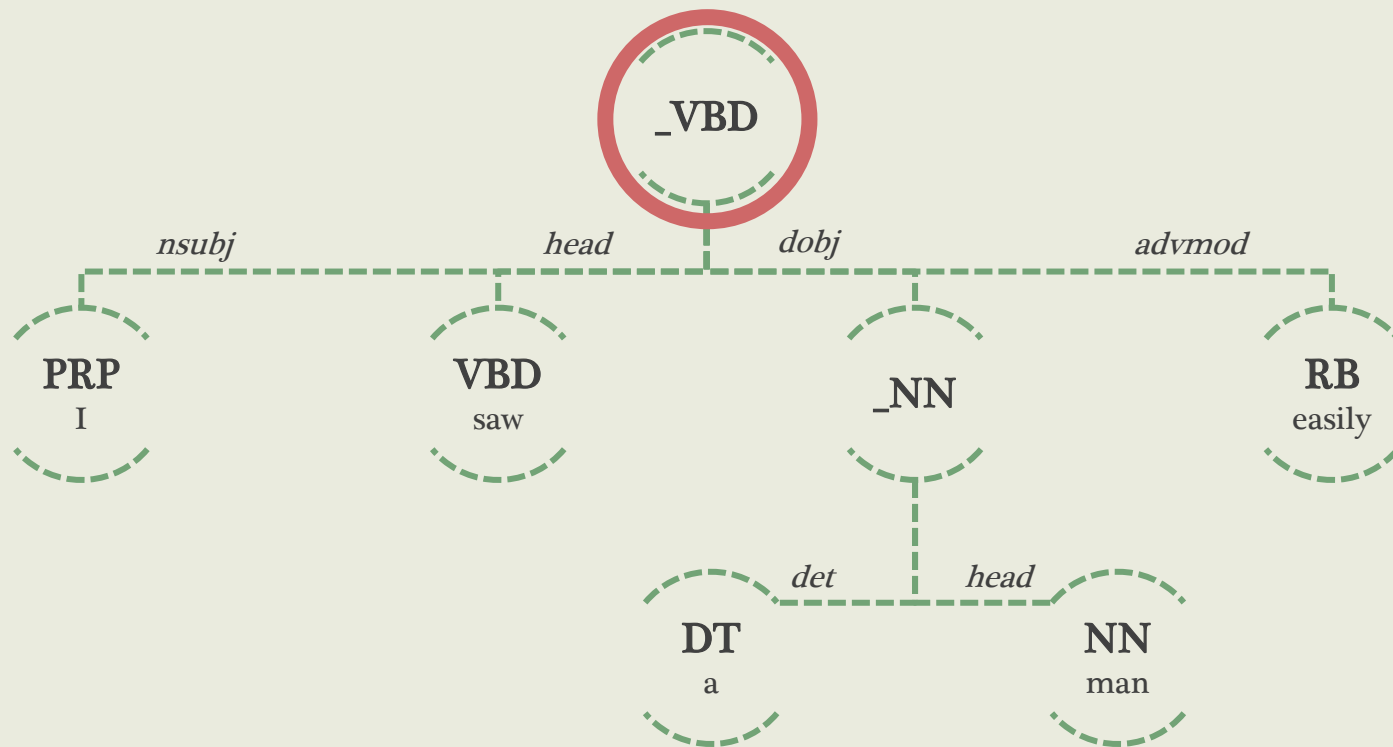
Durchgehen der Knoten in *Sliding Windows*



Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

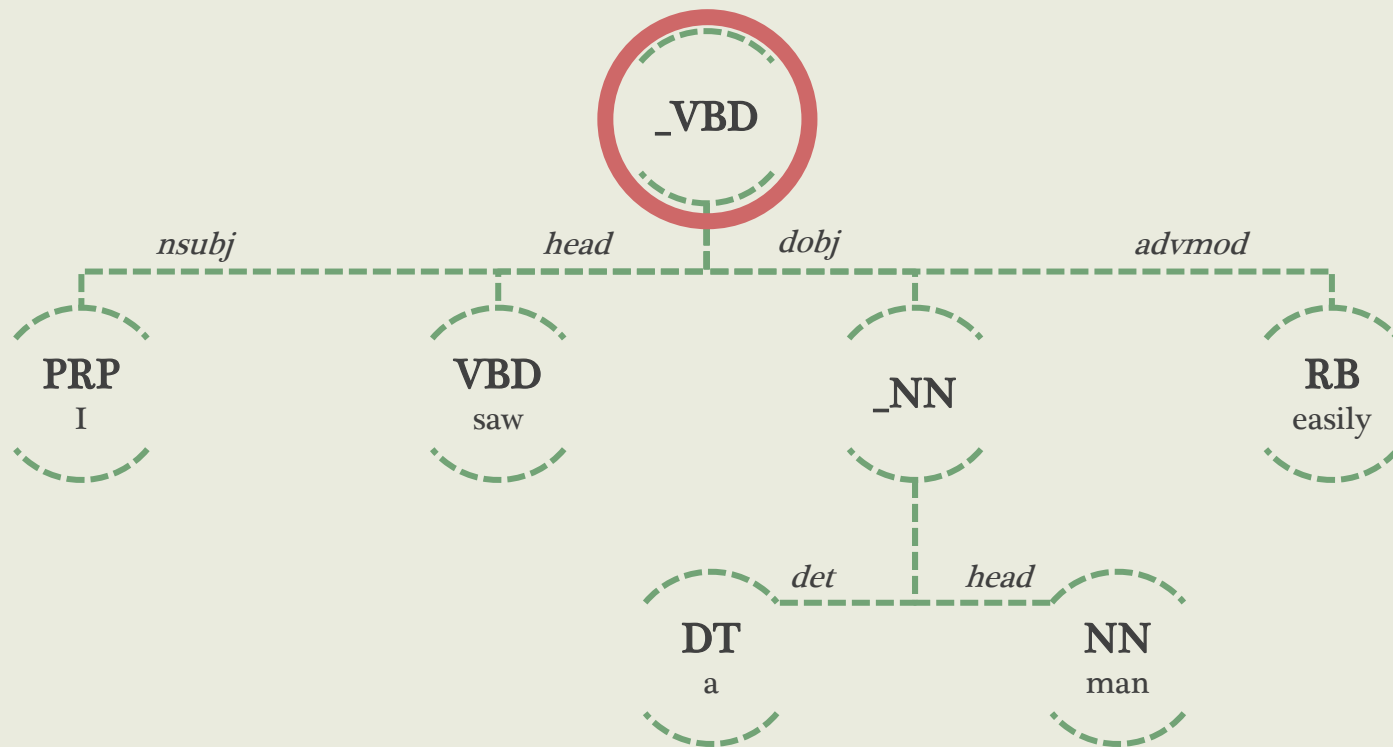


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

3 Töchter = 2 Fenster

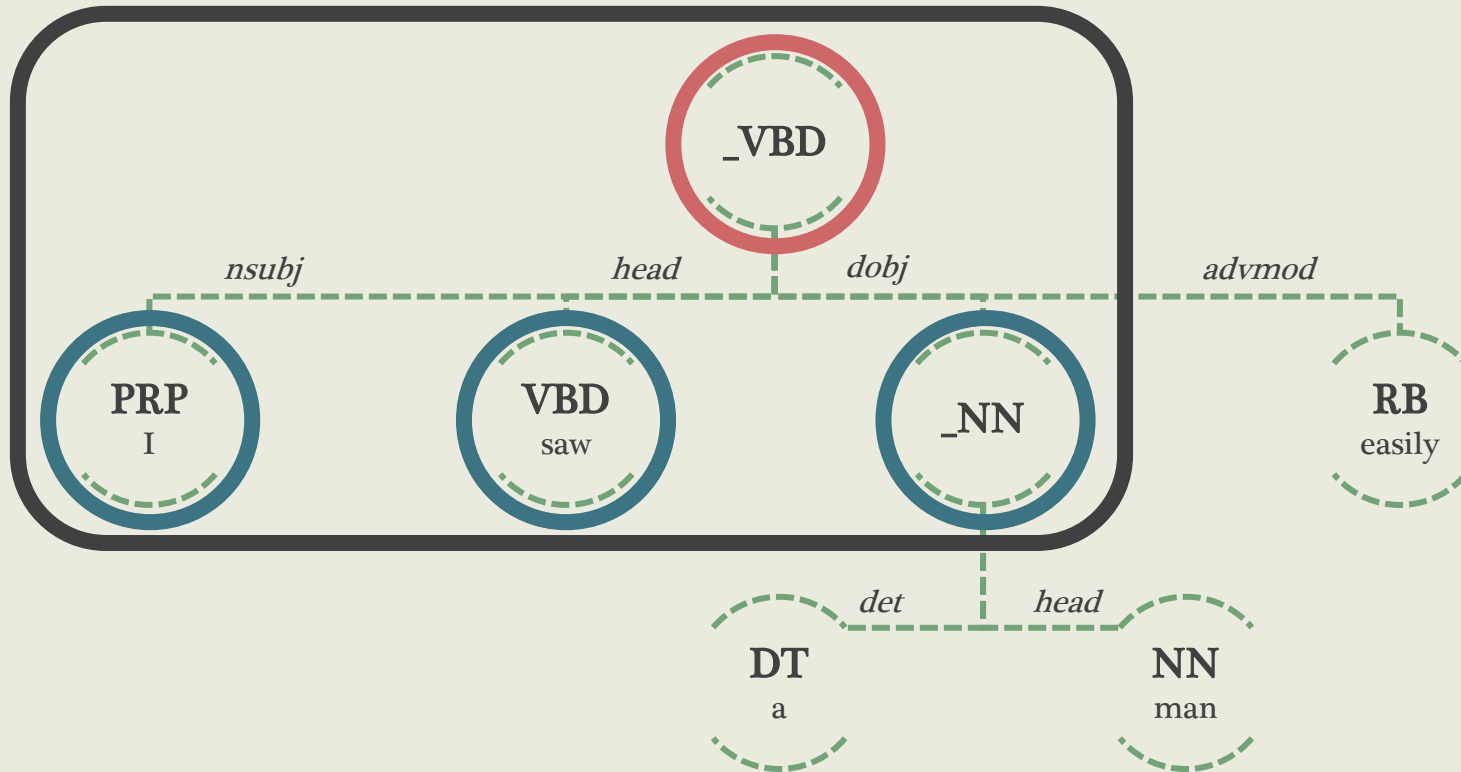


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

3 Töchter = 2 Fenster

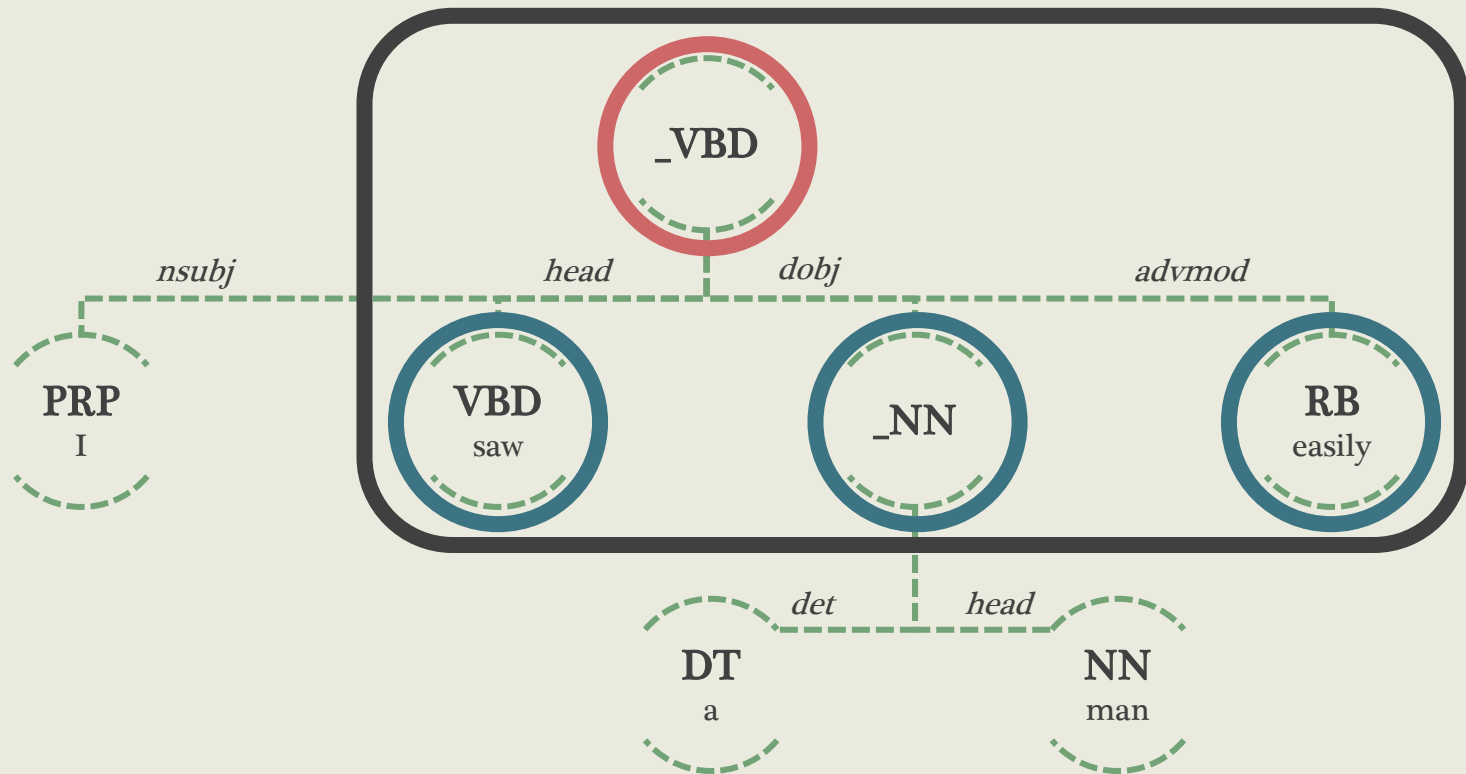


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

3 Töchter = 2 Fenster

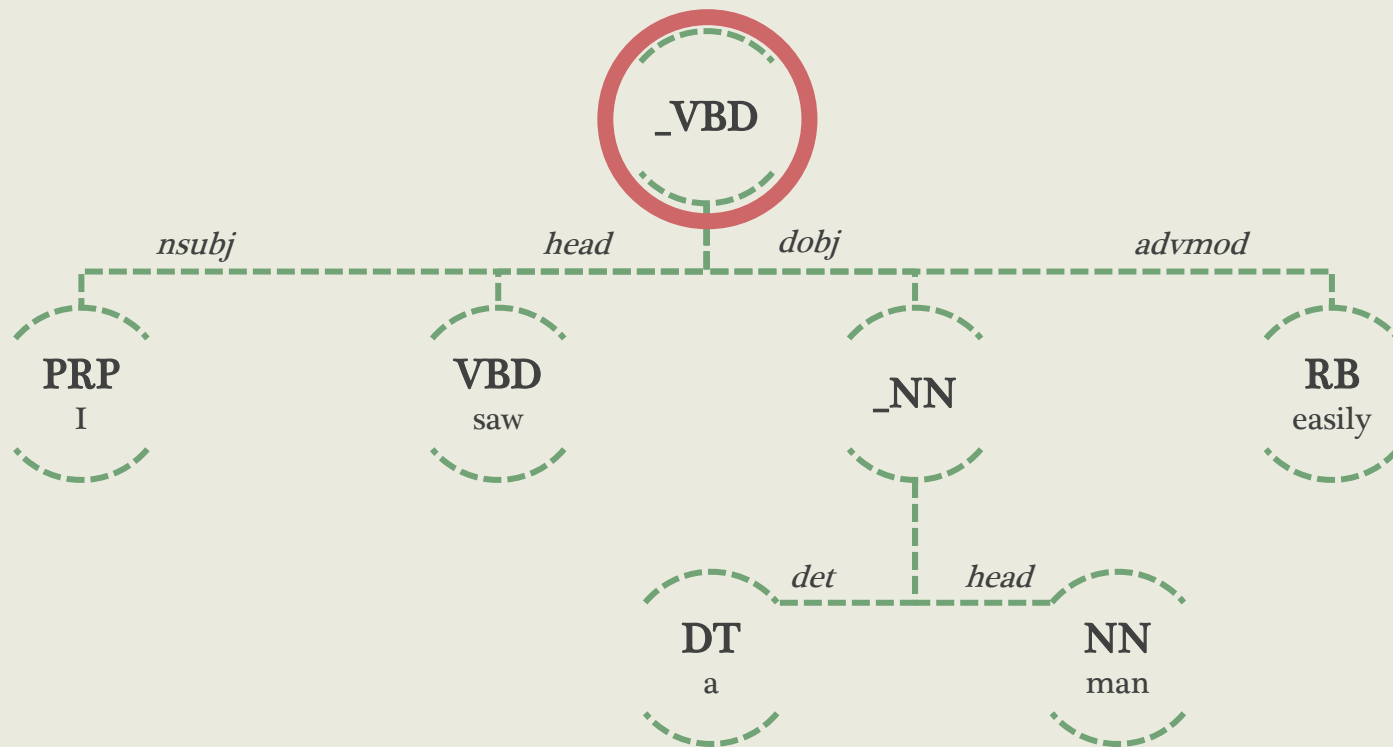


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

2 Töchter = 3 Fenster

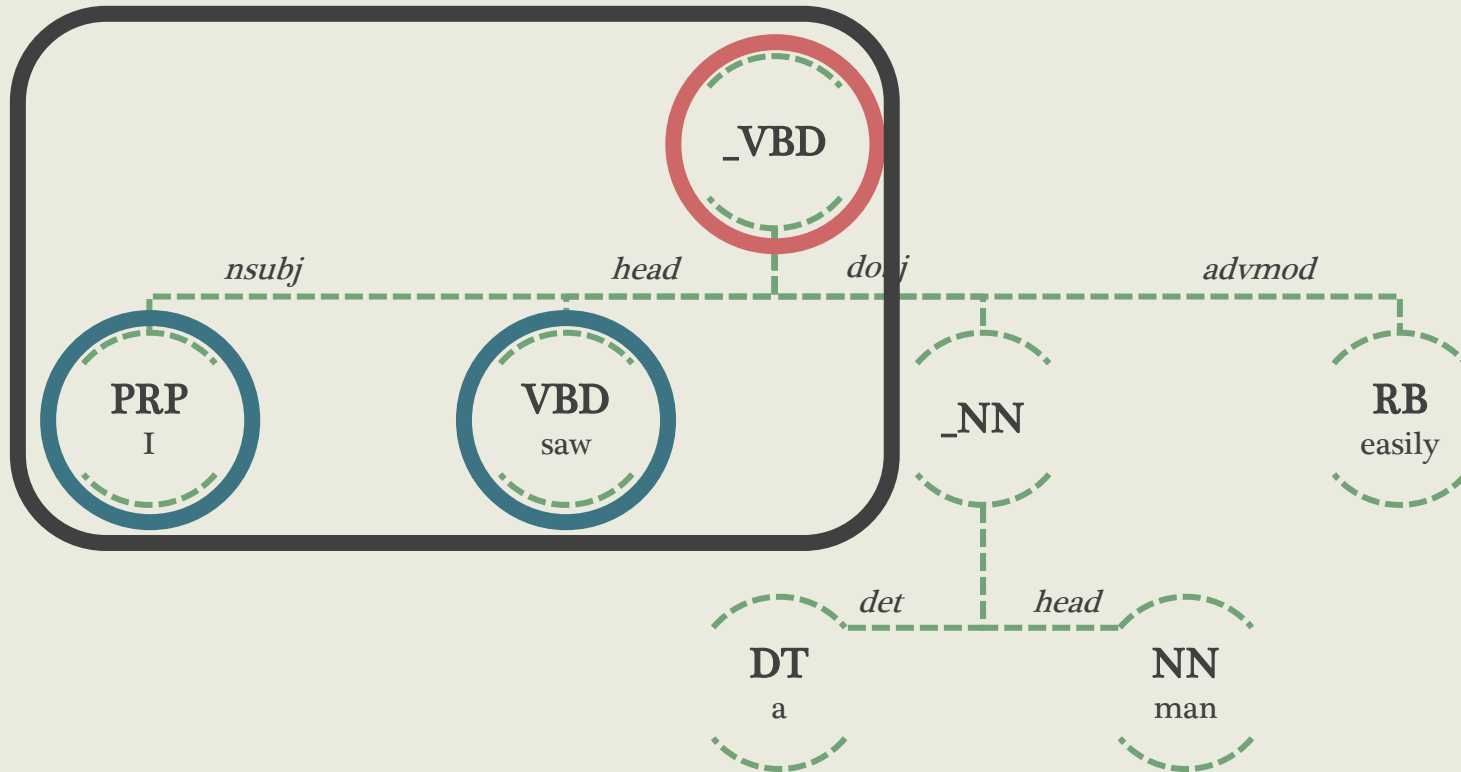


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

2 Töchter = 3 Fenster

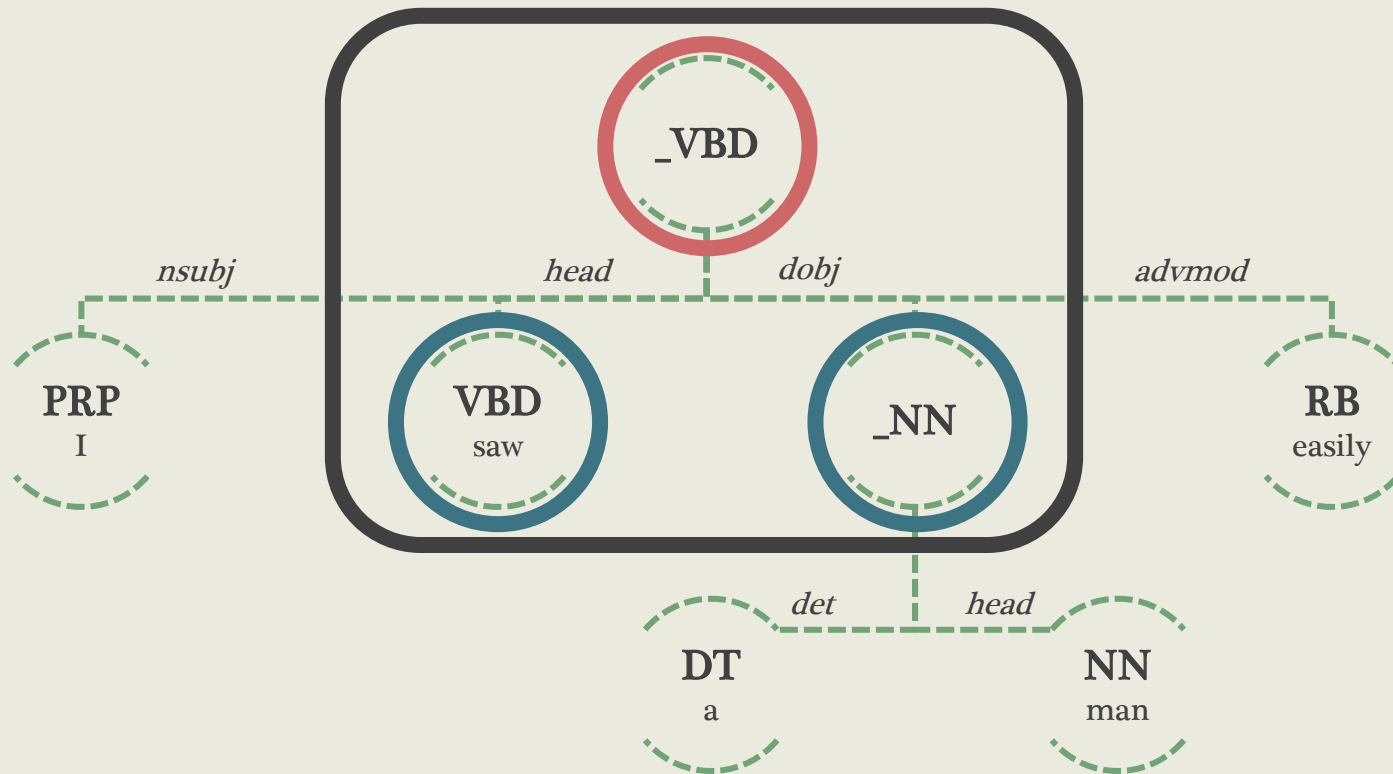


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

2 Töchter = 3 Fenster

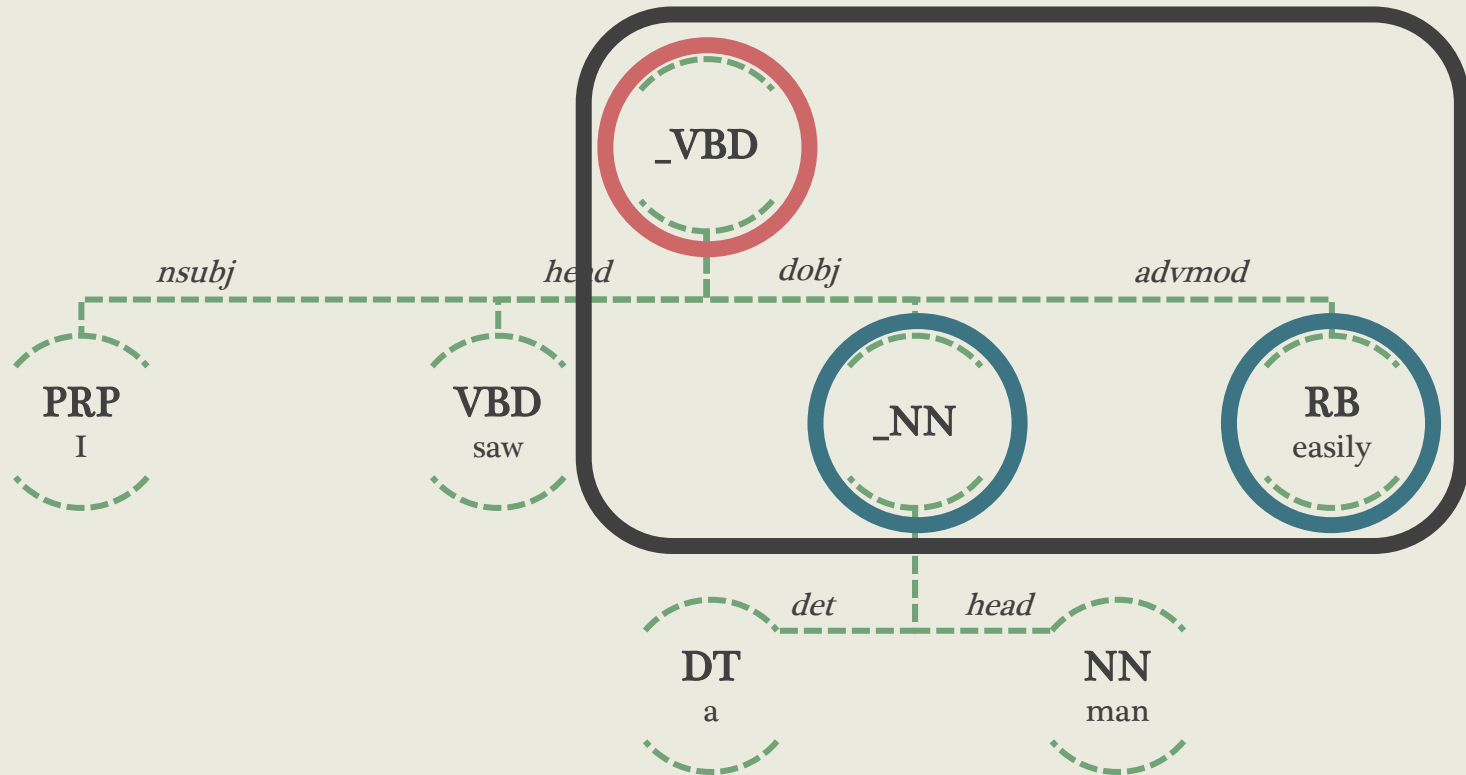


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

2 Töchter = 3 Fenster

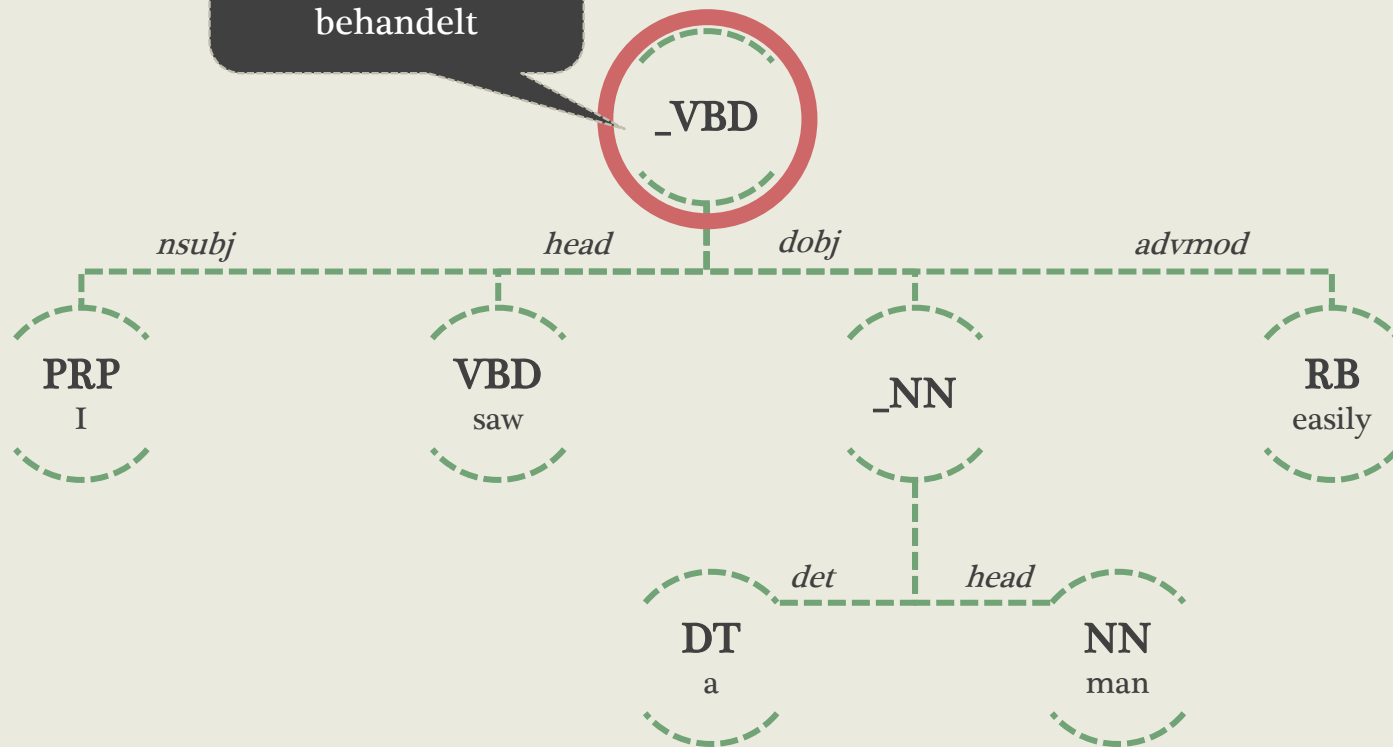


Funktionsweise II

22

Durchgehen der Knoten in *Sliding Windows*

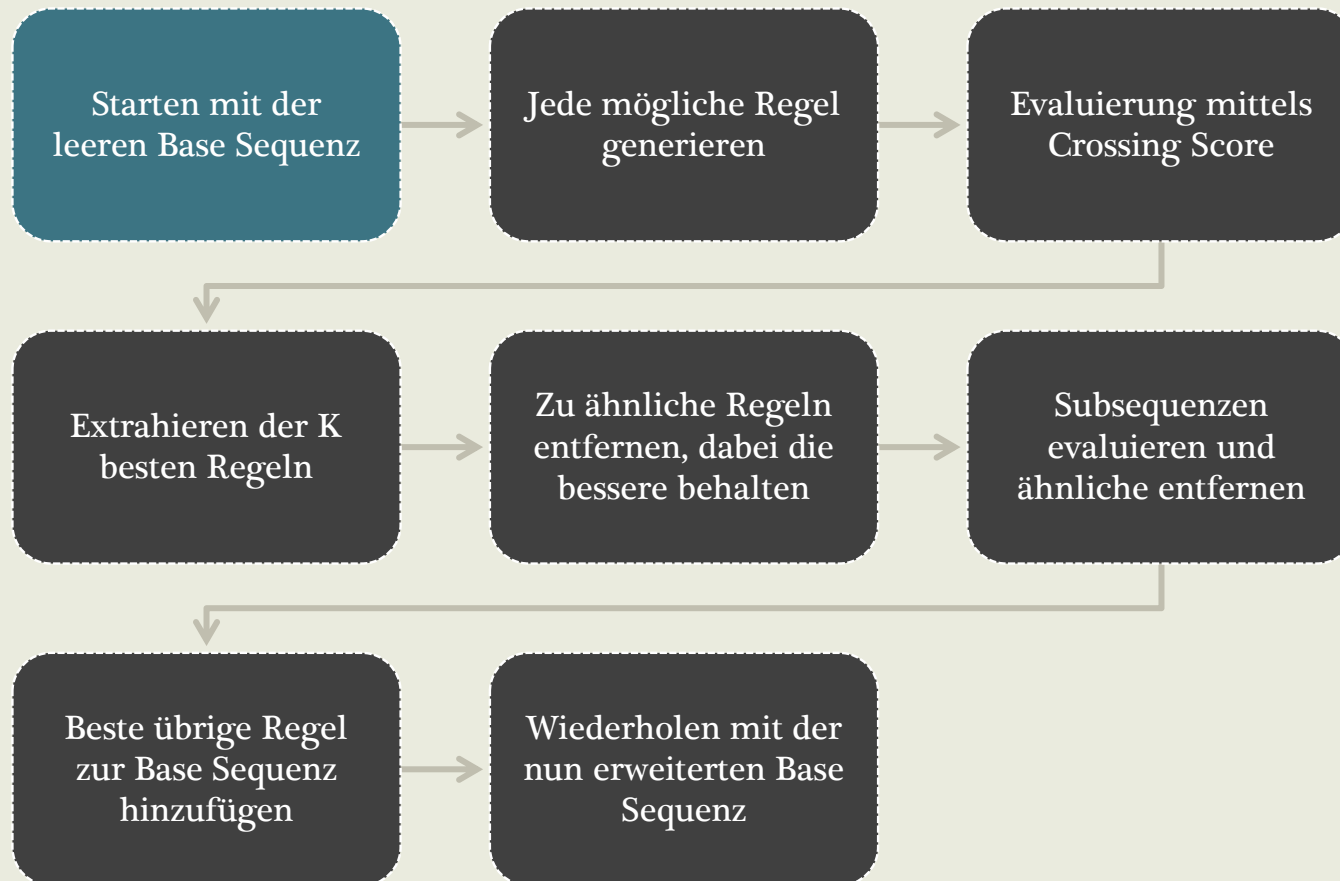
Wird wie zwei bzw.
drei Knoten
behandelt



Variante 1

23

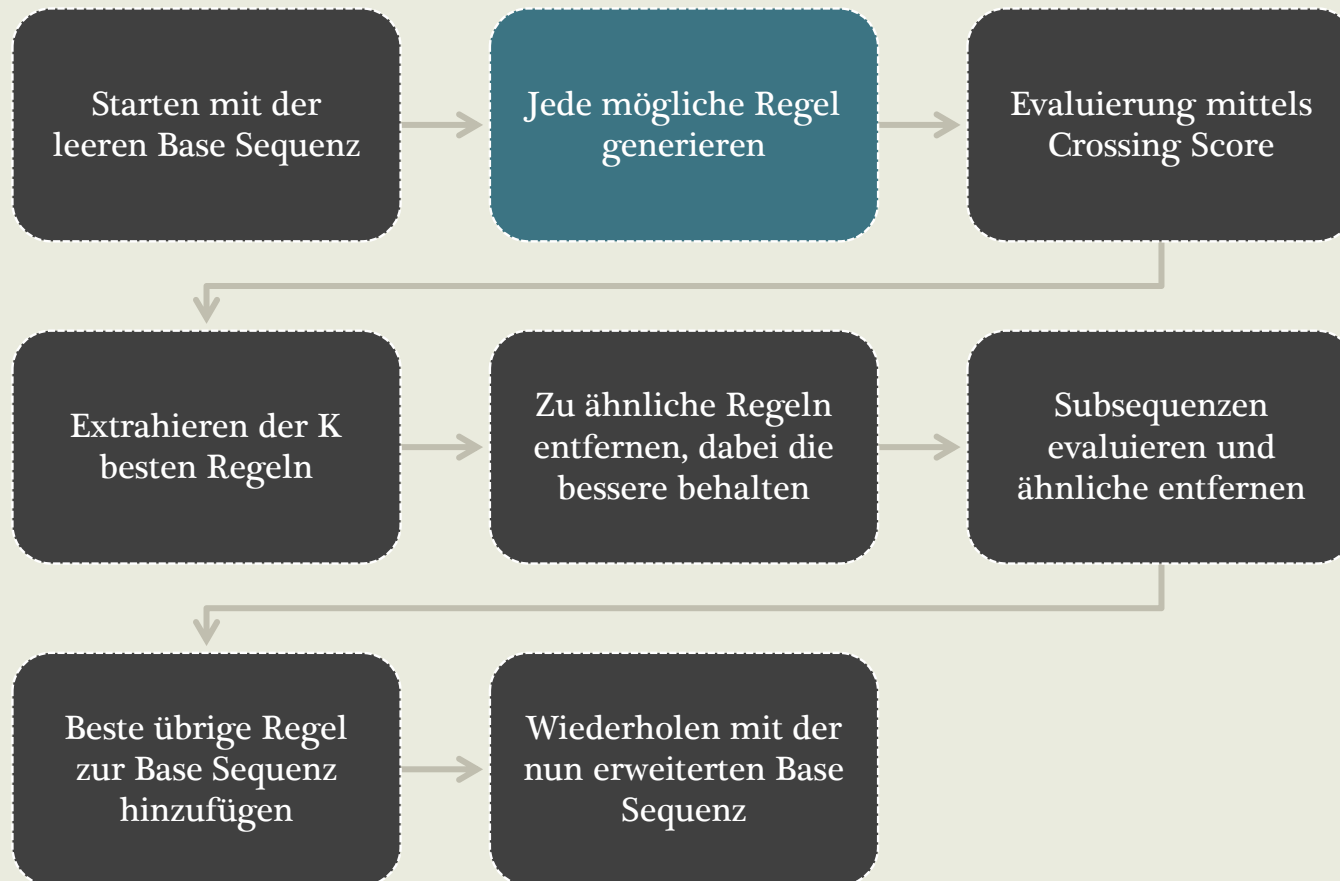
Optimieren des *Crossing Scores*



Variante 1

23

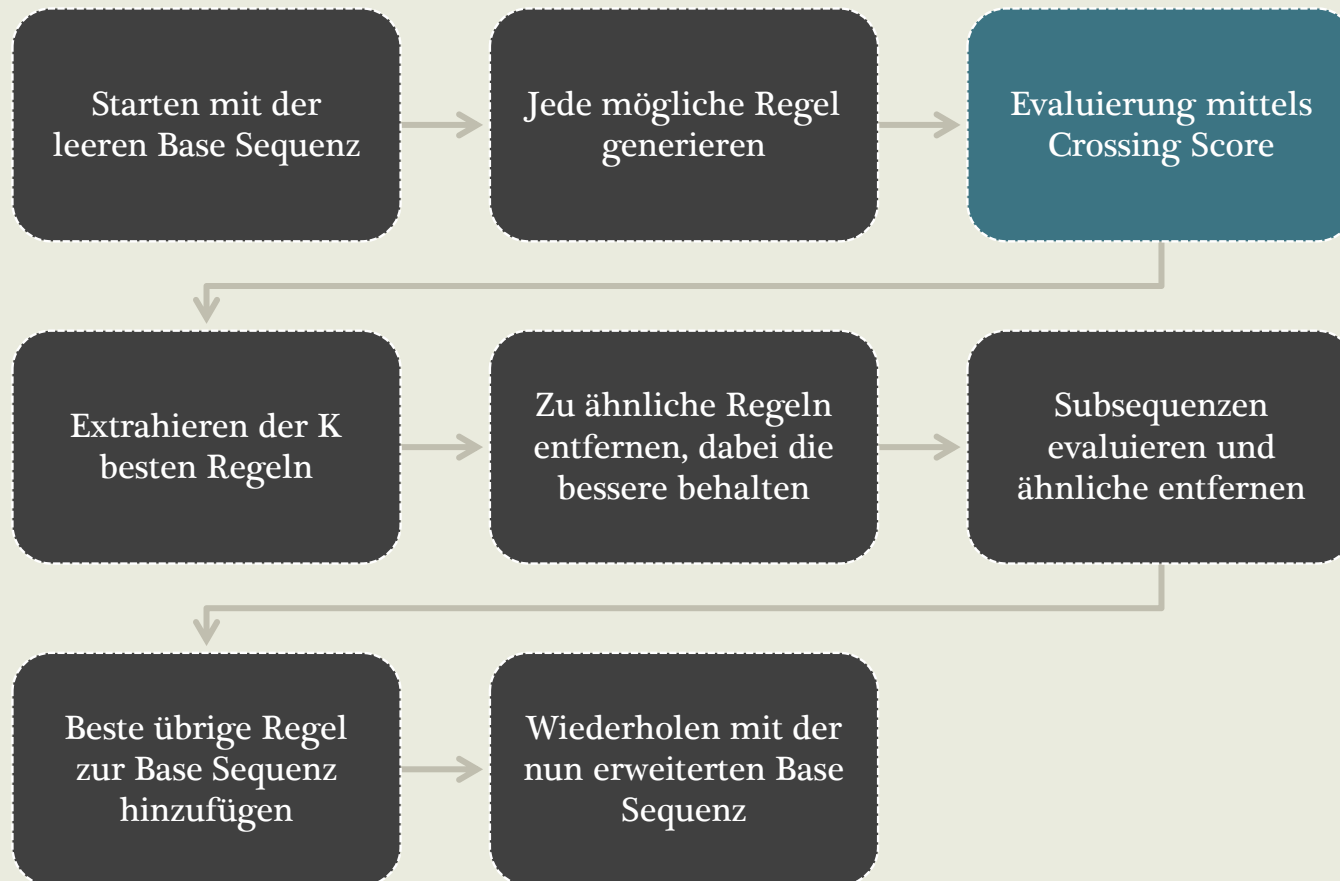
Optimieren des *Crossing Scores*



Variante 1

23

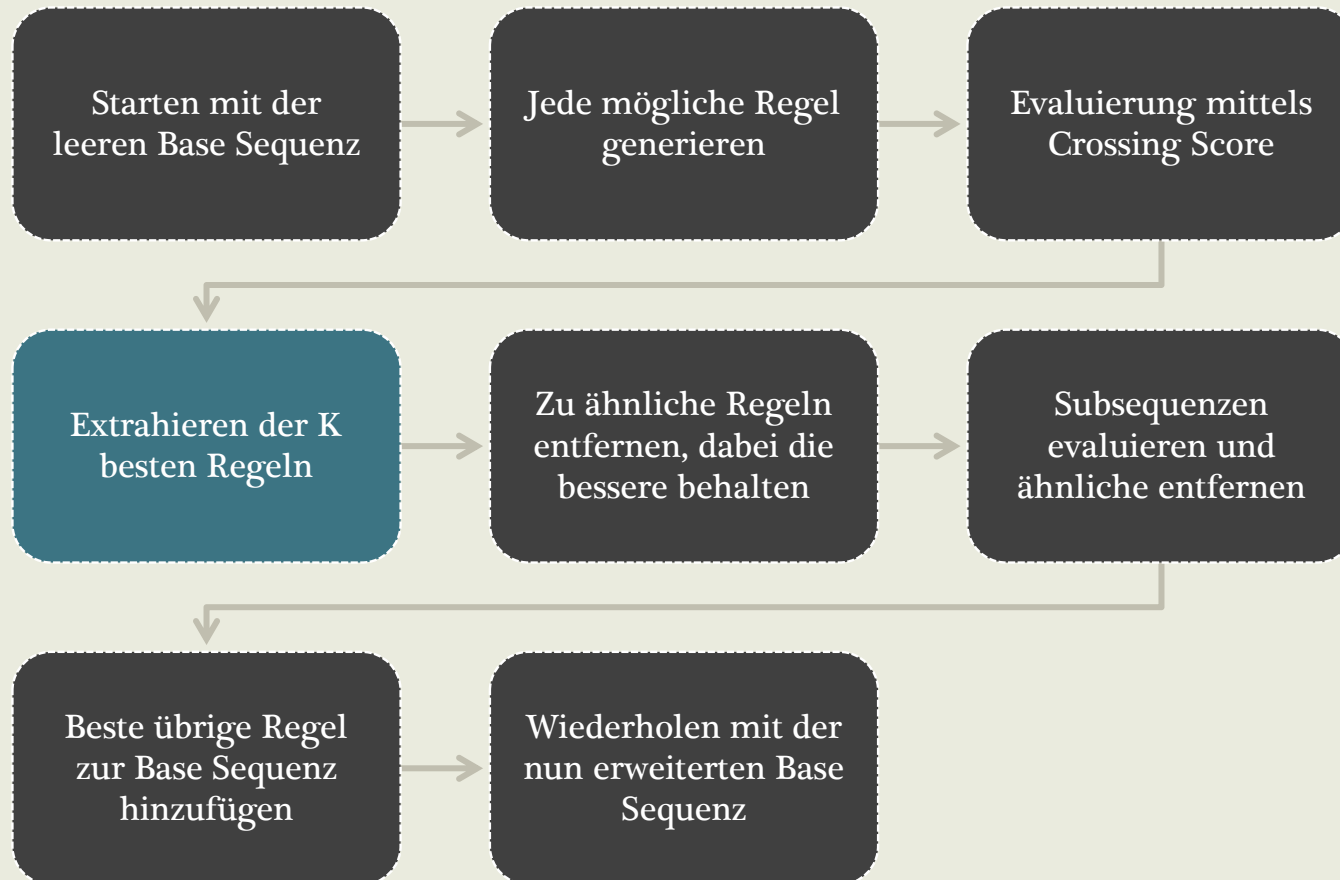
Optimieren des *Crossing Scores*



Variante 1

23

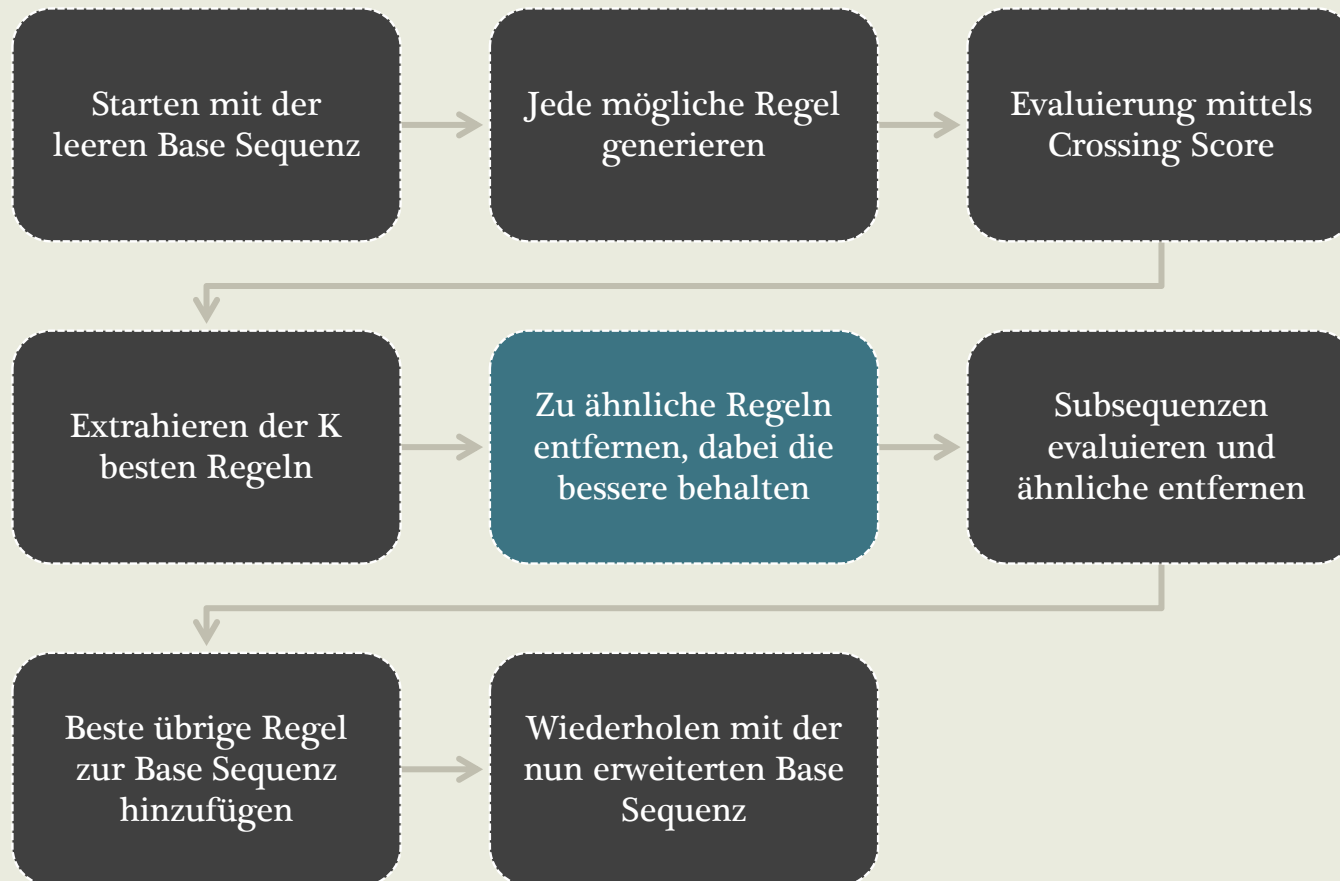
Optimieren des *Crossing Scores*



Variante 1

23

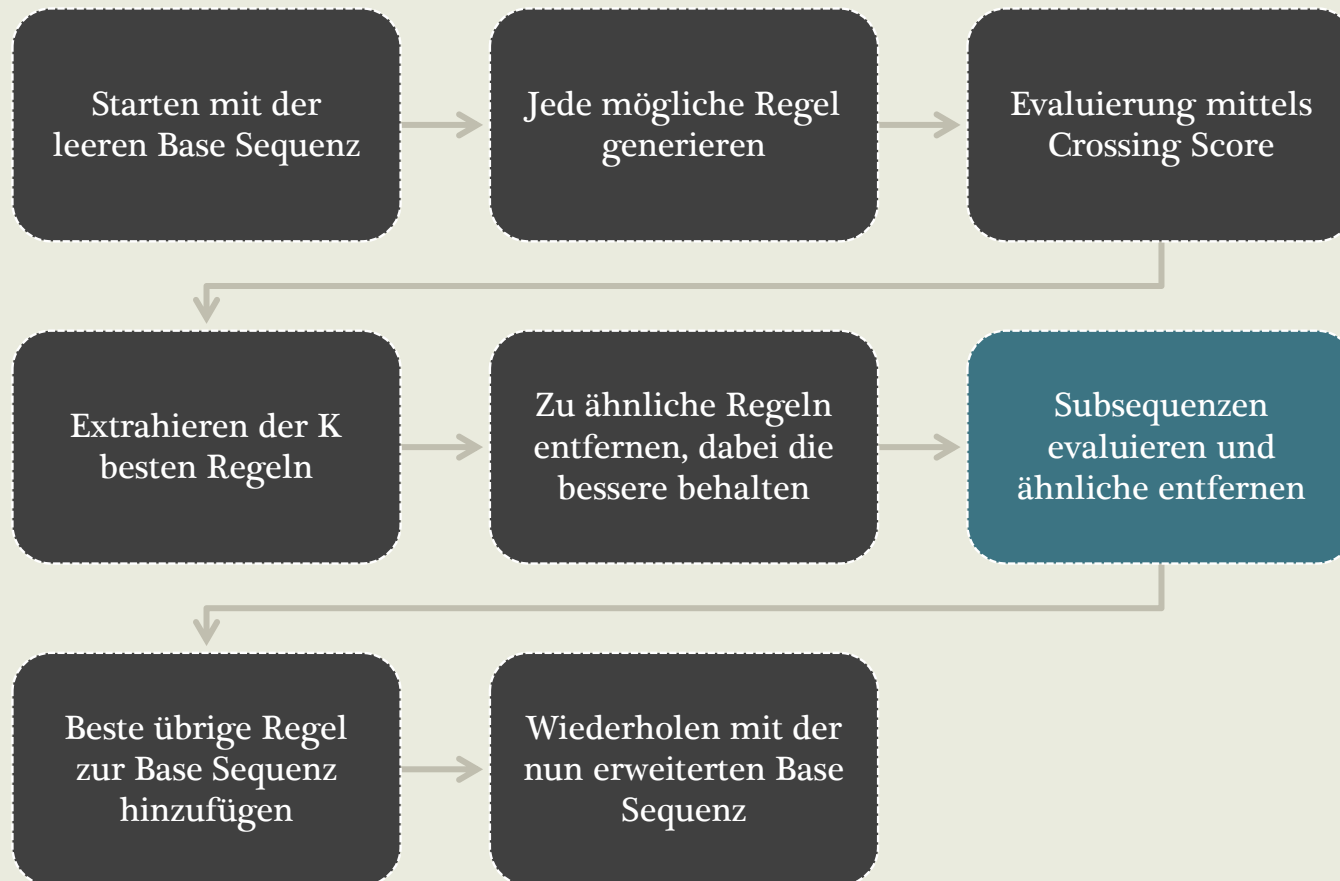
Optimieren des *Crossing Scores*



Variante 1

23

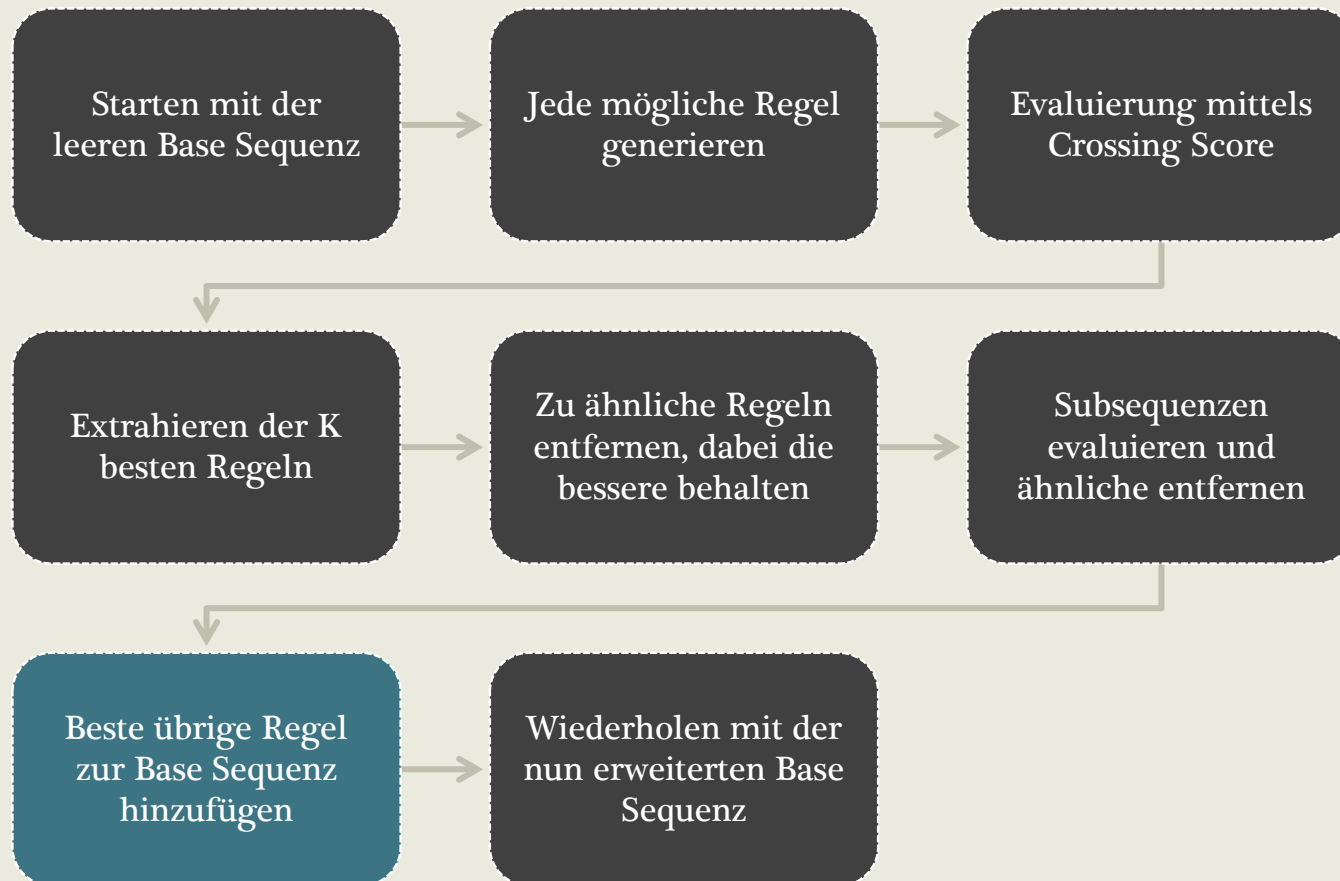
Optimieren des *Crossing Scores*



Variante 1

23

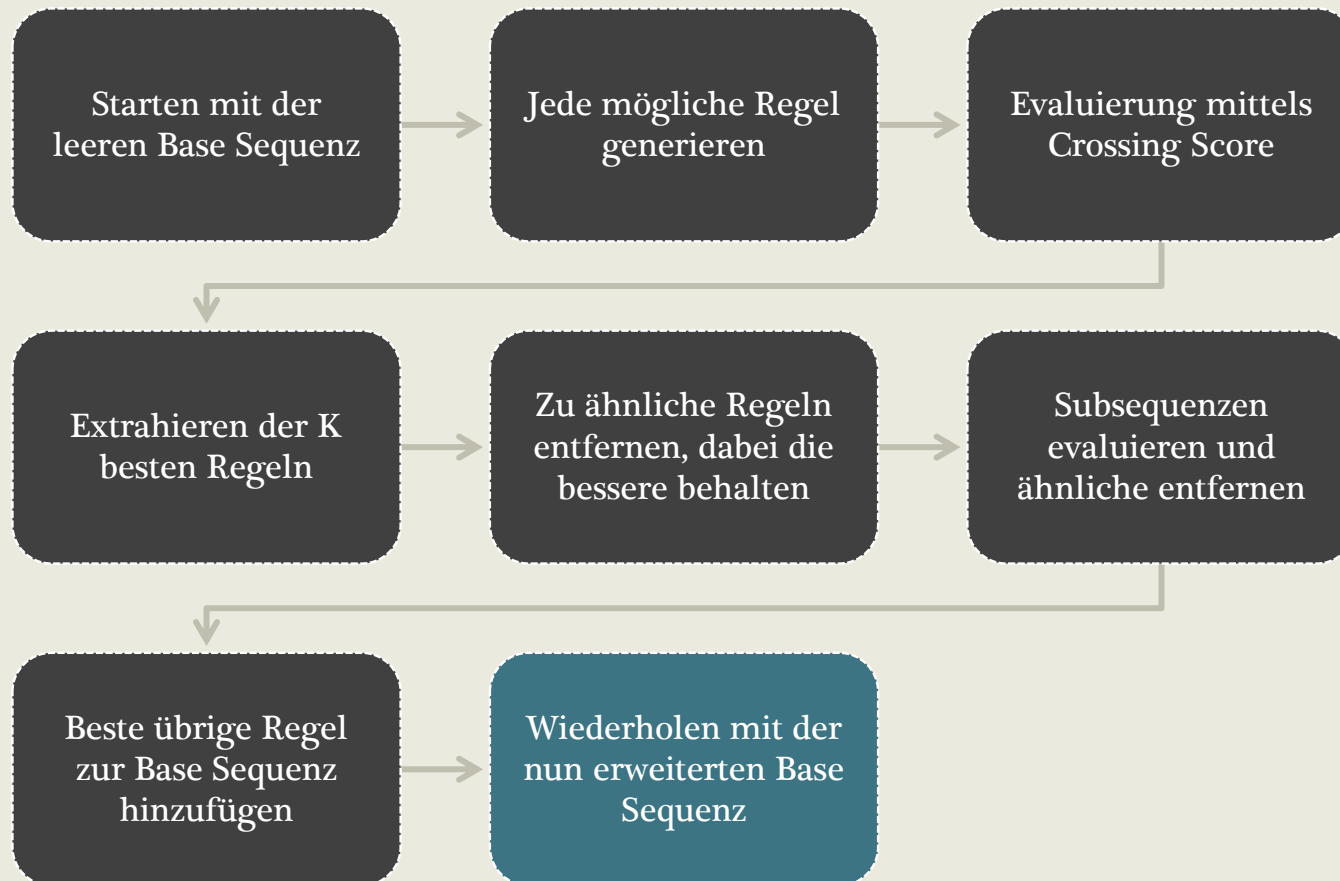
Optimieren des *Crossing Scores*



Variante 1

23

Optimieren des *Crossing Scores*



Variante 1 - Beispiel

24

- $B = \{ \}$

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$
- R3 und R5 zu ähnlich $\rightarrow \{R3, R2, R12\}$

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$
- R3 und R5 zu ähnlich $\rightarrow \{R3, R2, R12\}$
- Vergleich der Sequenzen auf Ähnlichkeit:
R3, R2, R12, R3 – R2, R2 – R12, R3 – R2 – R12

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$
- R3 und R5 zu ähnlich $\rightarrow \{R3, R2, R12\}$
- Vergleich der Sequenzen auf Ähnlichkeit:
R3, R2, R12, R3 – R2, R2 – R12, R3 – R2 – R12
- Übrig bleiben R3 und R2 – R12

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$
- R3 und R5 zu ähnlich $\rightarrow \{R3, R2, R12\}$
- Vergleich der Sequenzen auf Ähnlichkeit:
R3, R2, R12, R3 – R2, R2 – R12, R3 – R2 – R12
- Übrig bleiben R3 und R2 – R12
- R2 – R12 hat besseren Score

Variante 1 - Beispiel

24

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- K beste = $\{R3, R5, R2, R12\}$
- R3 und R5 zu ähnlich $\rightarrow \{R3, R2, R12\}$
- Vergleich der Sequenzen auf Ähnlichkeit:
R3, R2, R12, R3 – R2, R2 – R12, R3 – R2 – R12
- Übrig bleiben R3 und R2 – R12
- R2 – R12 hat besseren Score
- $B = \{R2 - R12\}$

Variante 2

25

Optimieren des *Estimated BLEU Gain*



Variante 2

25

Optimieren des *Estimated BLEU Gain*



Variante 2

25

Optimieren des *Estimated BLEU Gain*



Variante 2

25

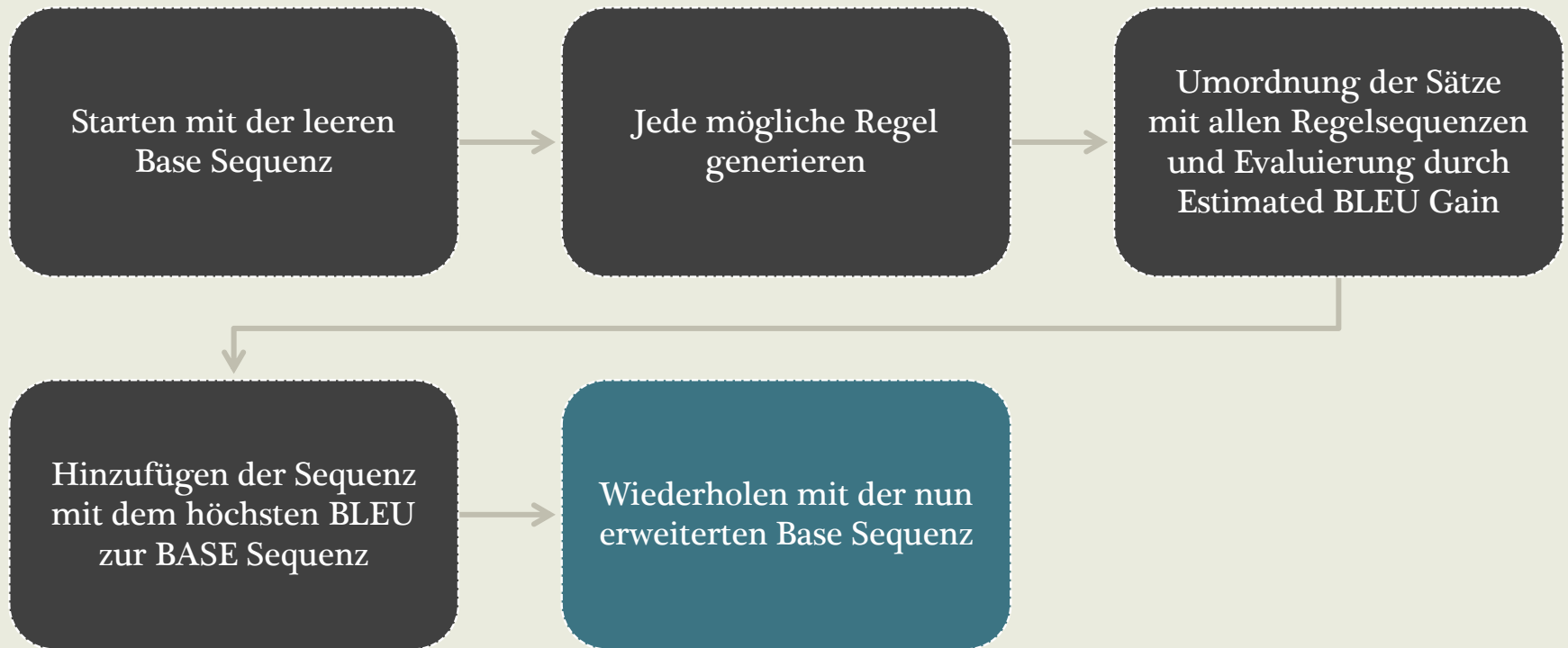
Optimieren des *Estimated BLEU Gain*



Variante 2

25

Optimieren des *Estimated BLEU Gain*



Variante 2 - Beispiel

26

- $B = \{ \}$

Variante 2 - Beispiel

26

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$

Variante 2 - Beispiel

26

- $B = \{ \}$
- Regeln = $\{R_1, R_2, R_3, R_4, R_5, R_6, \dots, R_n\}$
- Regelsequenzen bilden und zur Umordnung anwenden, z.B.:
 $R_1 - R_2 - R_3$ oder $R_3 - R_7 - R_1$

Variante 2 - Beispiel

26

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regelsequenzen bilden und zur Umordnung anwenden, z.B.:
R1 – R2 – R3 oder R3 – R7 – R1
- BLEU nach jeder Umordnung berechnen

Variante 2 - Beispiel

26

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regelsequenzen bilden und zur Umordnung anwenden, z.B.:
R1 – R2 – R3 oder R3 – R7 – R1
- BLEU nach jeder Umordnung berechnen
- Sequenz, die höchsten BLEU erzeugt: R1 – R2 – R3

Variante 2 - Beispiel

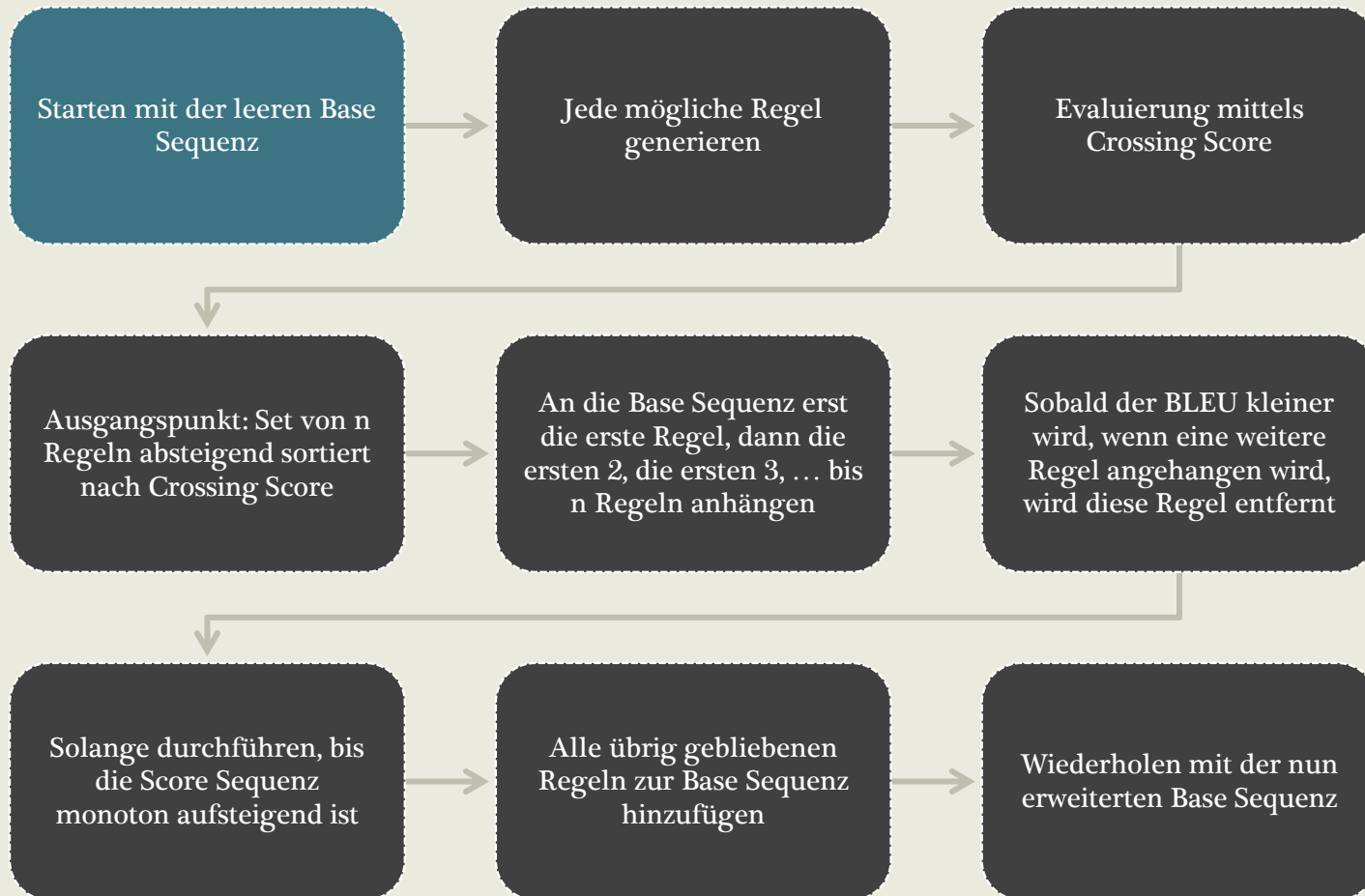
26

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regelsequenzen bilden und zur Umordnung anwenden, z.B.:
R1 – R2 – R3 oder R3 – R7 – R1
- BLEU nach jeder Umordnung berechnen
- Sequenz, die höchsten BLEU erzeugt: R1 – R2 – R3
- $B = \{R1 - R2 - R3\}$

Variante 3

27

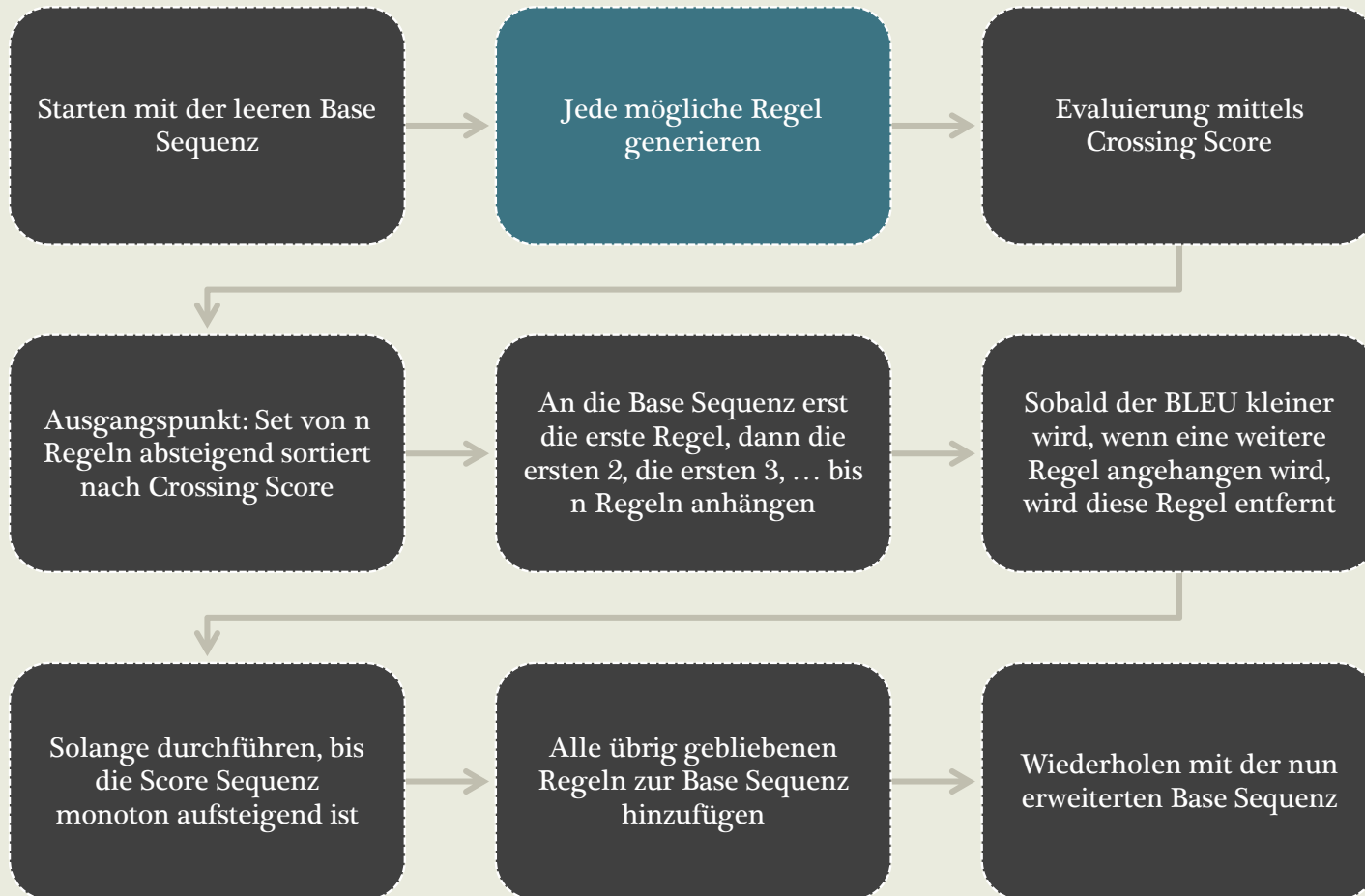
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

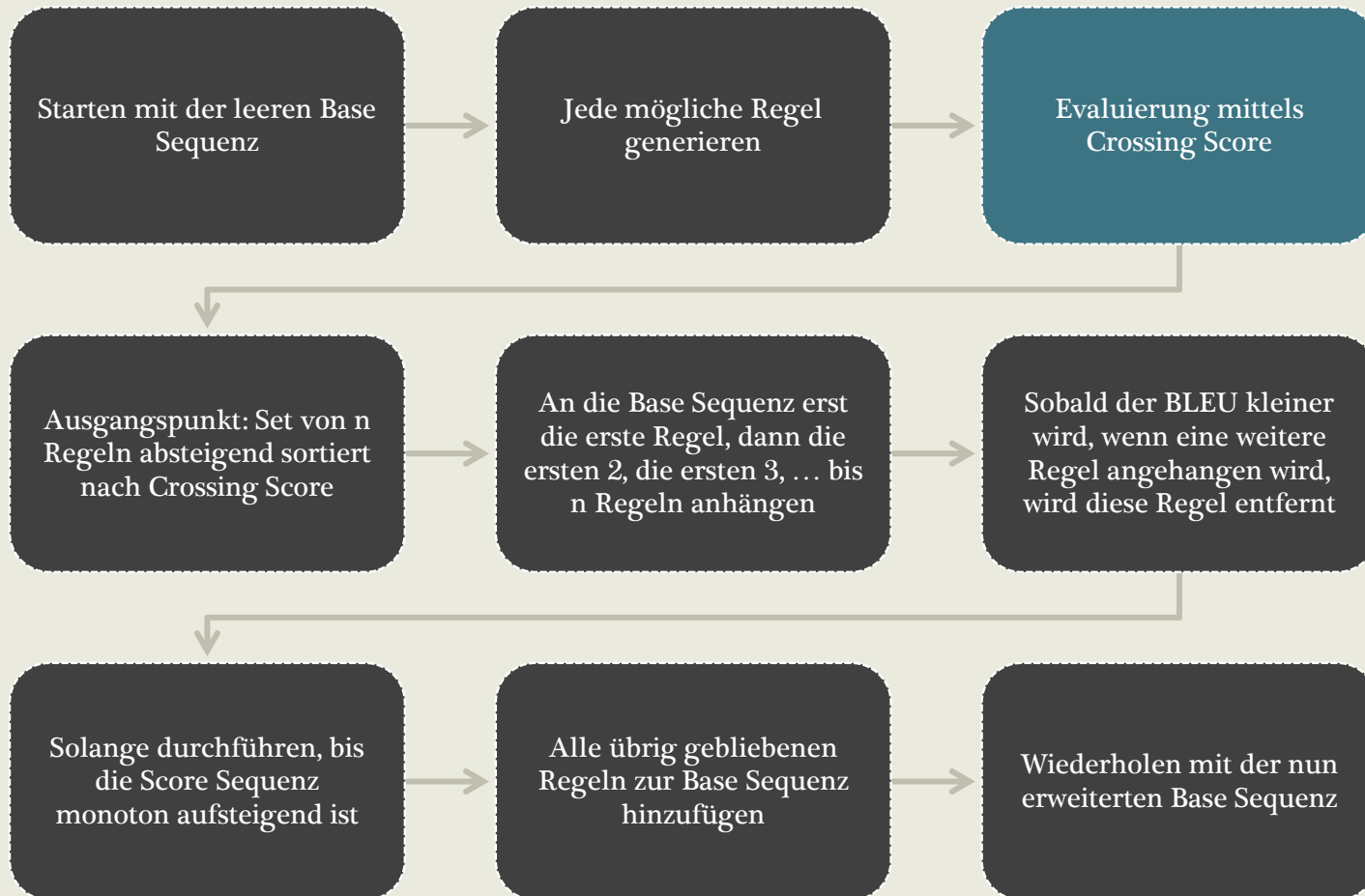
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

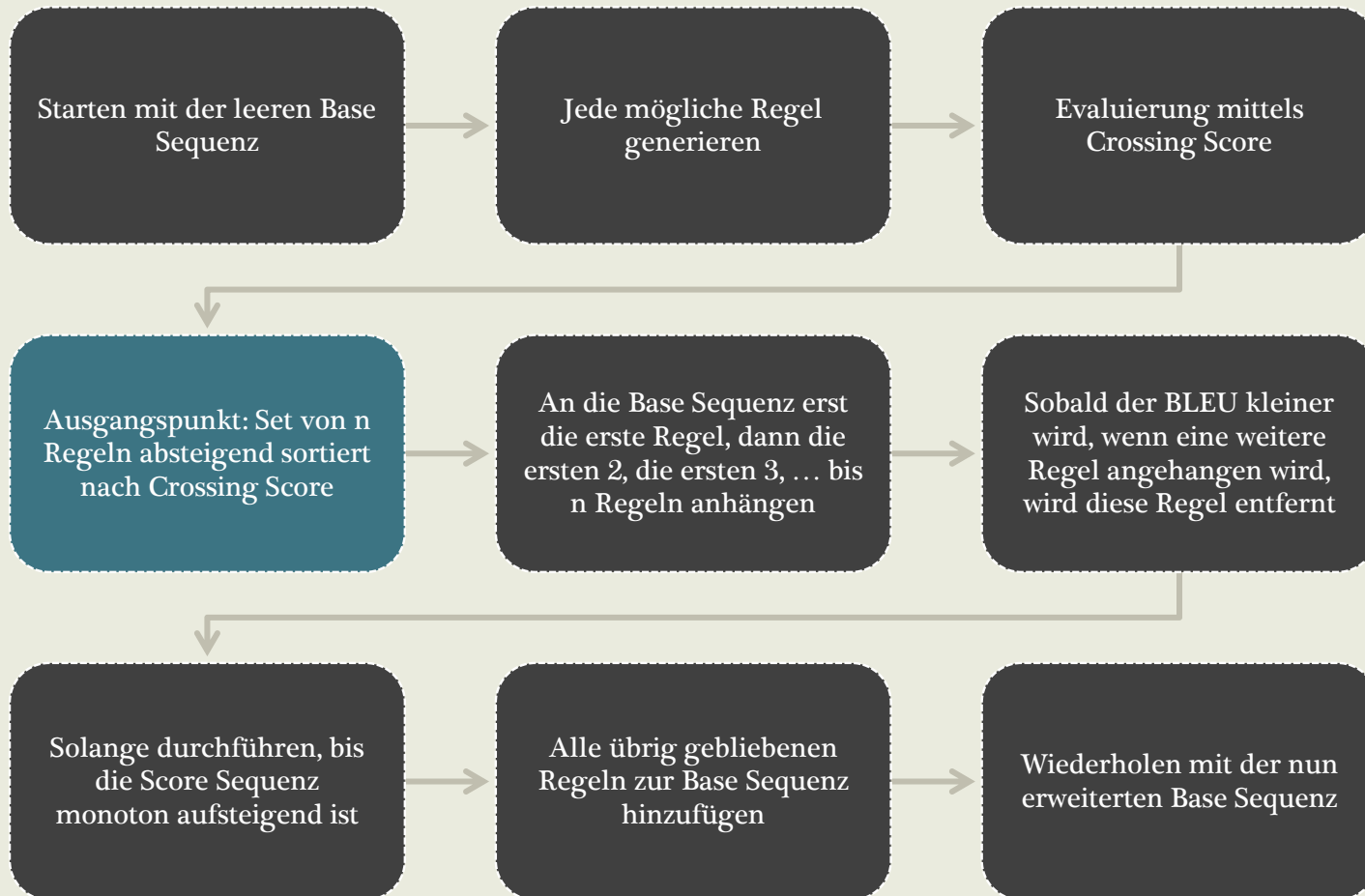
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

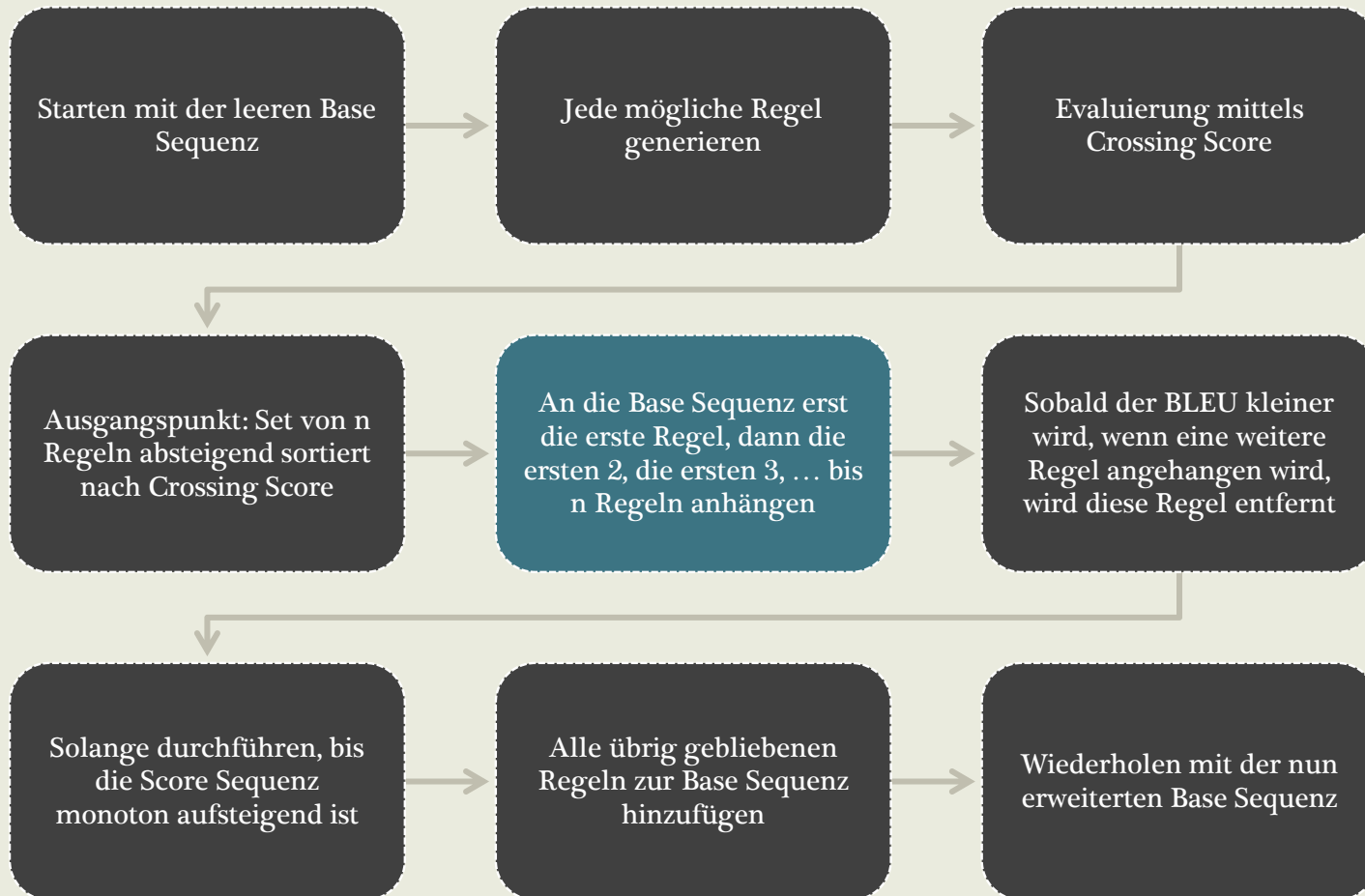
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

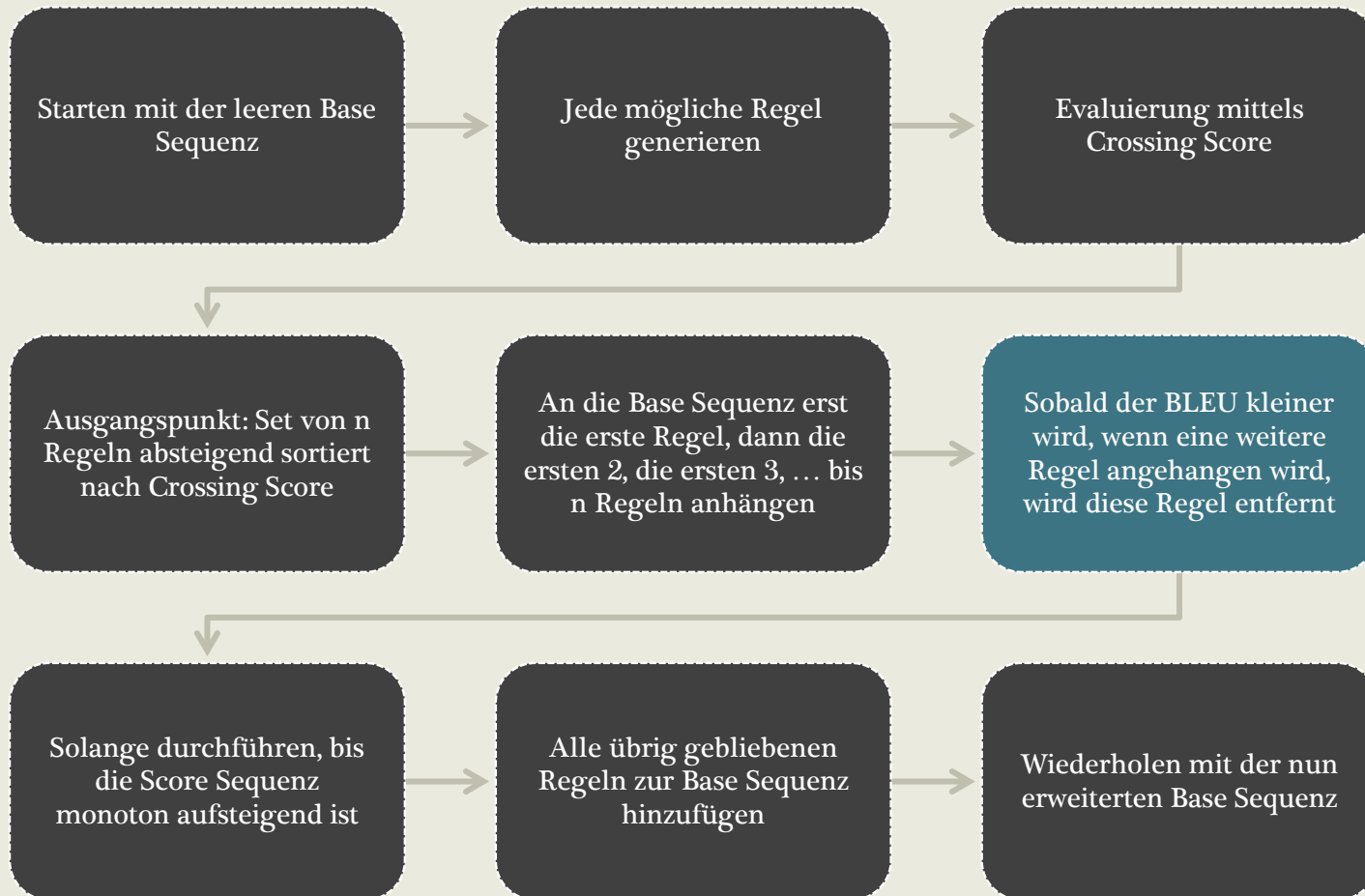
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

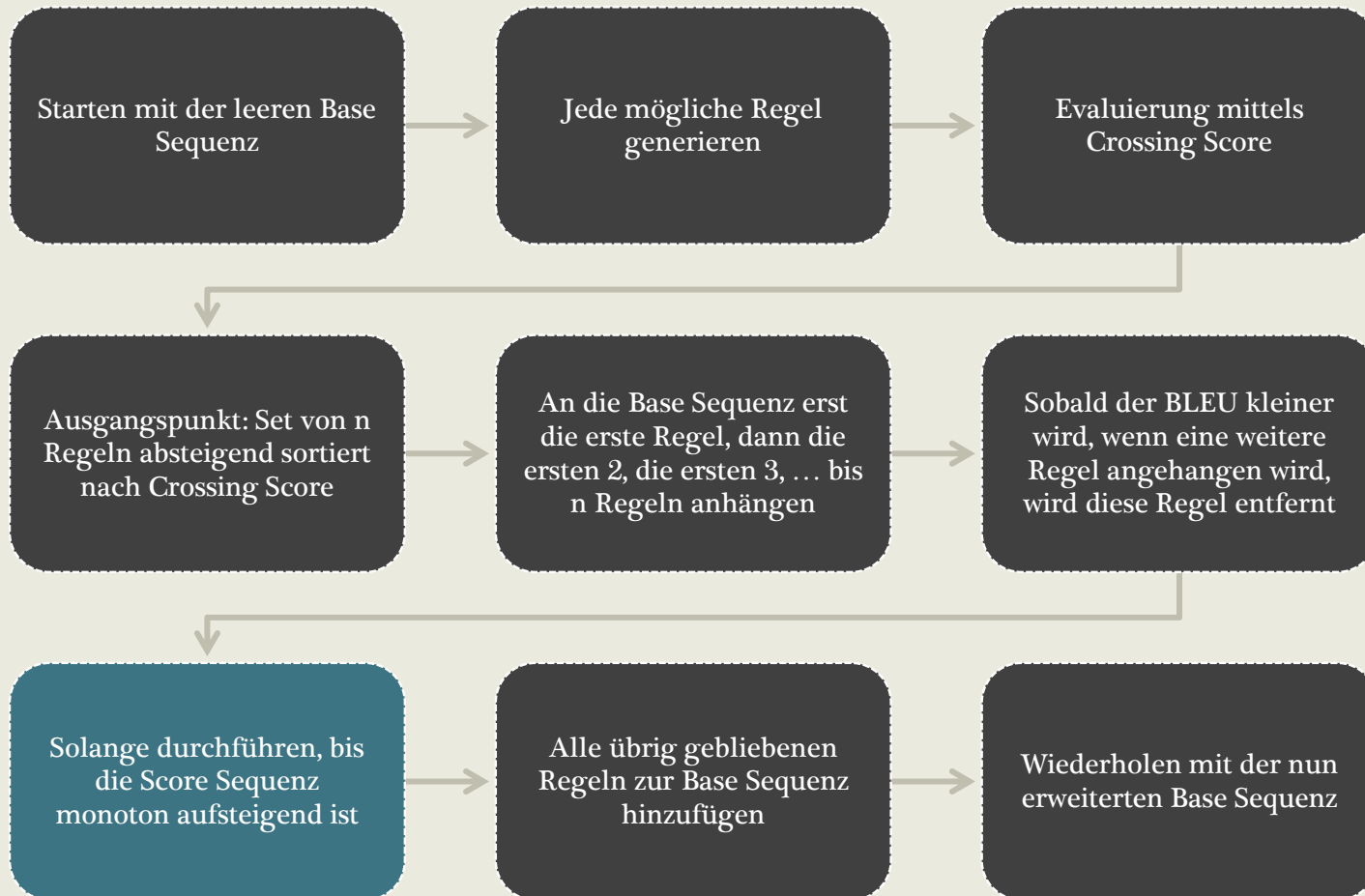
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

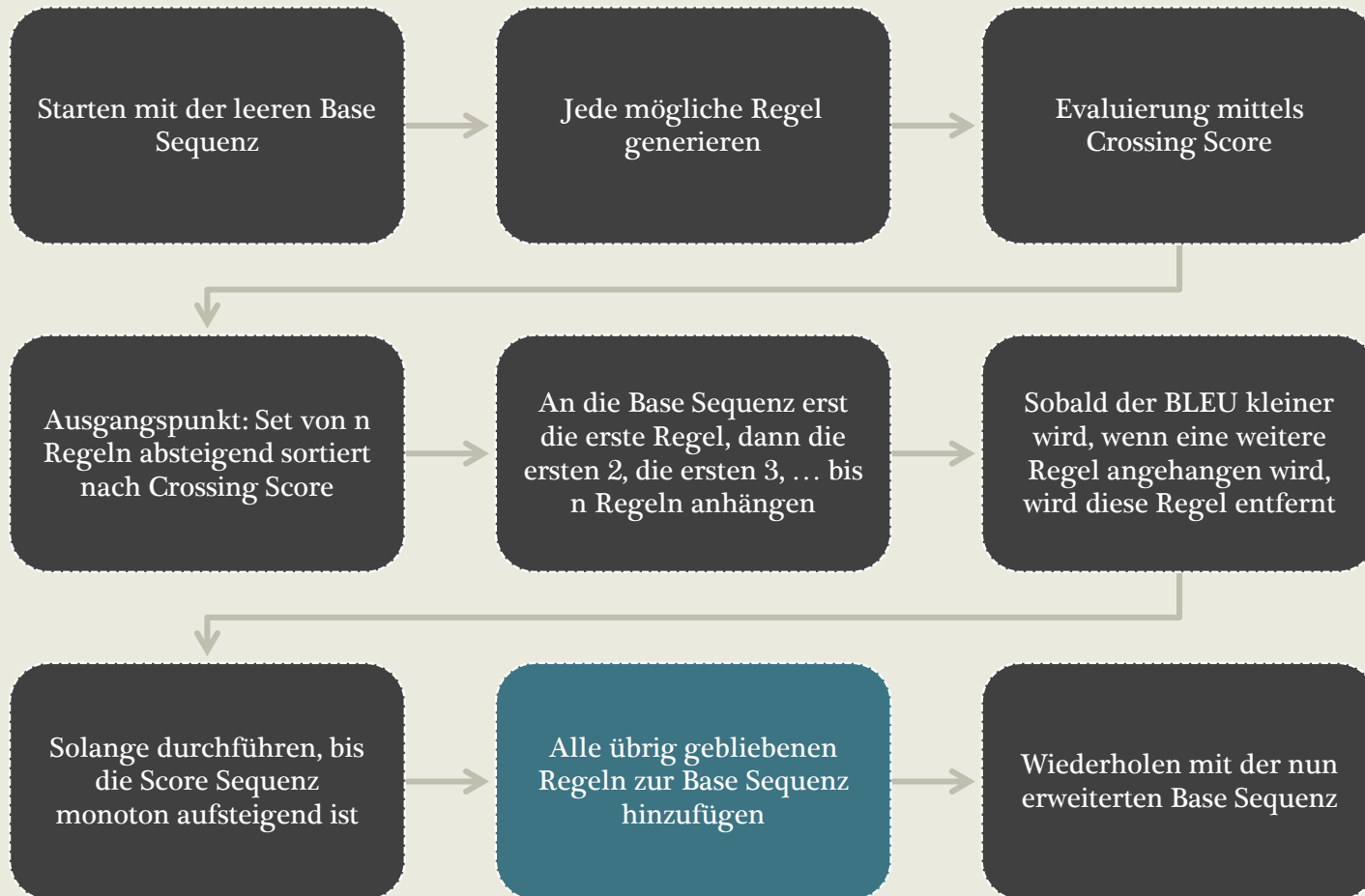
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

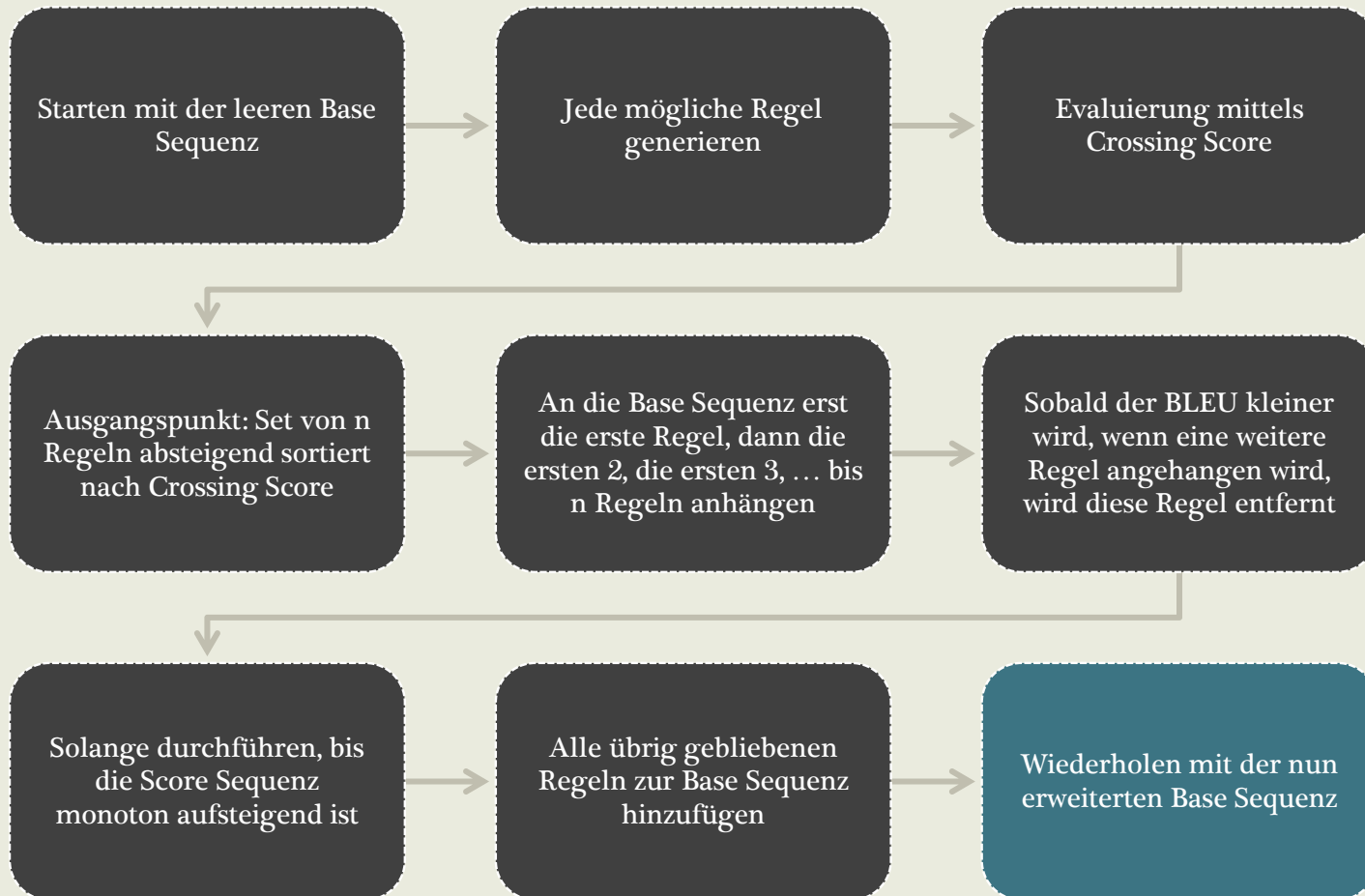
Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3

27

Optimieren des *sequentiellen Estimated BLEU Gain*



Variante 3 - Beispiel

28

- $B = \{ \}$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R1\} \rightarrow BLEU = 4,2$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R1\} \rightarrow BLEU = 4,2$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R1\} \rightarrow BLEU = 4,2$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R15\} \rightarrow BLEU = 4,6$

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R1\} \rightarrow BLEU = 4,2$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R15\} \rightarrow BLEU = 4,6$
- ...

Variante 3 - Beispiel

28

- $B = \{ \}$
- Regeln = $\{R1, R2, R3, R4, R5, R6, \dots, Rn\}$
- Regeln absteigend nach Crossing Score sortiert:
 $\{R2, R7, R1, R15, R4, \dots\}$
- $B = \{R2\} \rightarrow BLEU = 4$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R1\} \rightarrow BLEU = 4,2$
- $B = \{R2 - R7\} \rightarrow BLEU = 4,5$
- $B = \{R2 - R7 - R15\} \rightarrow BLEU = 4,6$
- ...
- $B = \{R2 - R7 - R15 - R4\}$

Agenda

29

Das Problem

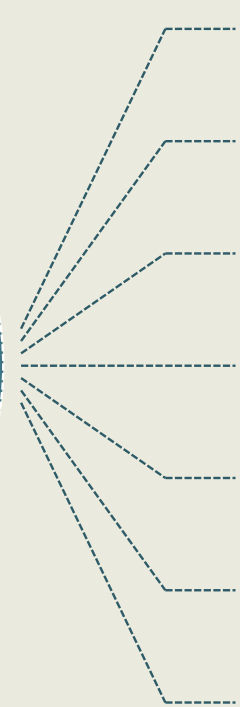
Der Lösungsansatz

Der Algorithmus

Evaluation

Trainingsdaten I

30



Deutsch

Japanisch

Koreanisch

Hindi

Walisisch

Russisch

Tschechisch

Trainingsdaten II

31

- **Standard-Datensätze**
 - Englisch – Deutsch
 - Deutsch – Englisch
- **Unveröffentlichte Datensätze**
 - Alle Sprachen

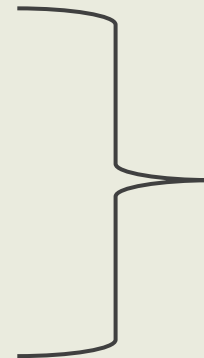
Unveröffentlichte Datensätze

32

- Nachrichtenartikel von *WikiNews* (996 Sätze)
- Zufällige Sätze aus dem Web (9000 Sätze)

- 3 Sets:

<i>Set</i>	<i>Wiki</i>	<i>Web</i>
<i>dev1</i>	486	3000
<i>dev2</i>	0	1000
<i>test</i>	510	5000



Für alle
Sprachen
gleich

Vorgehen

33

Für alle Testdaten
Durchlaufen des
Algorithmus' in allen
drei Varianten



Reordering Rule
Sequenz
(40-50 Regeln)

Durchlauf OHNE
Umordnung im
Dekodierer



Umordnung im
Dekodierer ist nicht
notwendig

Ergebnisse I

34

- Varianten 1 und 3 meistens besser als Variante 2
- Jedoch alle nützlich

Standard-Datensatz

<i>Sprache</i>	<i>Base %BLEU</i>	<i>Var. 1 %BLEU</i>	<i>Var. 2 %BLEU</i>	<i>Var. 3 %BLEU</i>
EN → DE	16,09	16,30	16,35	16,40
DE → EN	21,00	22,45	22,13	22,05

Ergebnisse II

35

Interne Datensätze

<i>Sprache</i>	<i>Google %BLEU</i>	<i>Base %BLEU</i>	<i>Var.1 Gain</i>	<i>Var. 2 Gain</i>	<i>Var. 3 Gain</i>
Tschechisch	16,68	15,35	-0,08	0,13	0,19
Deutsch	20,34	18,65	0,47	0,30	0,39
Hindi	19,15	16,85	2,25	2,08	0,15
Japanisch	30,74	25,91	3,05	2,60	3,05
Koreanisch	27,99	23,61	3,34	3,77	4,16
Russisch	16,80	15,33	0,08	0,10	0,10
Walisisch	27,38	25,48	1,25	0,77	1,43

Ergebnisse III: Top Regeln

36

<i>Sprache</i>	<i>Kontext</i>	<i>Reihenfolge</i>	<i>Beispiel</i>
Hindi	1L:head 3L:none	2,1,3	I see him → I him see
Japanisch, Koreanisch	2L:prep	2,1	eat with a spoon → eat a spoon with
Deutsch	1T:VBN 2L:prep	2,1	struck with a ball → with a ball struck
Russisch, Tschechisch	1L:nn 2L:head	2,1	a building entrance → a entrance building
Walisisch	1L:amod 2L:head	2,1	Blue ball → ball blue

Diskussion

37

- Generieren von Regeln ohne Hilfe von Linguisten oder Muttersprachlern möglich
- Nur Englisch und Deutsch als Quellsprachen
- Auch auf andere Sprachen anwendbar → Trainingsdaten

Vielen Dank!

Referenzen

- Genzel, D. (2010). Automatically Learning Source-side Reordering Rules for Large Scale Machine Translation. In C.-R. Huang & D. Jurafsky (Eds.), *COLING 2010. The 23rd International Conference on Computational Linguistics : proceedings of the main conference, August 23-27, 2010, Beijing, China* (pp. 376–384). Beijing: Tsinghua University Press. 📄