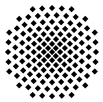
Machine Translation with Multi Bottom-up Tree Transducers

Andreas Maletti

Institute for Natural Language Processing
Universität Stuttgart, Germany

maletti@ims.uni-stuttgart.de

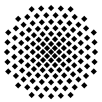
November 7, 2013



Machine translation

Translation

- Input:
  Official forecasts predicted just 3 percent, Bloomberg said.
- We (fast):
  Offiziellen Prognosen nur 3 Prozent prognostiziert,
  Bloomberg gesagt.
- Google (fast):
  Offizielle Prognosen vorhergesagt nur 3 Prozent,
  sagte Bloomberg.



Machine translation

Translation

- Input:
  Official forecasts predicted just 3 percent, Bloomberg said.
- We (fast):
  Offiziellen Prognosen nur 3 Prozent prognostiziert, Bloomberg gesagt.
- Google (fast):
  Offizielle Prognosen vorhergesagt nur 3 Prozent, sagte Bloomberg.

Machine translation

Translation

- Input:
  Official forecasts predicted just 3 percent, Bloomberg said.
- We (fast):
  Offiziellen Prognosen nur 3 Prozent prognostiziert, Bloomberg gesagt.
- Google (fast):
  Offizielle Prognosen vorhergesagt nur 3 Prozent, sagte Bloomberg.

Machine translation

Translation

- Input:
  The ECB wants to hold inflation to under two percent,
  or somewhere in that vicinity.
- We (fast):
  Die EZB die Inflation auf unter zwei Prozent
  oder irgendwo in die Nähe zu behalten will.
- Google (fast):
  Die EZB will die Inflation auf unter zwei Prozent zu halten,
  oder irgendwo in der Nähe.

Machine translation

Translation

- Input:
  The ECB wants to hold inflation to under two percent,
  or somewhere in that vicinity.
- We (fast):
  Die EZB die Inflation auf unter zwei Prozent
  oder irgendwo in die Nähe zu behalten will.
- Google (fast):
  Die EZB will die Inflation auf unter zwei Prozent zu halten,
  oder irgendwo in der Nähe.

Machine translation

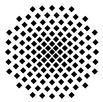
Translation

- Input:
  The ECB wants to hold inflation to under two percent,
  or somewhere in that vicinity.
- We (fast):
  Die EZB die Inflation auf unter zwei Prozent
  oder irgendwo in die Nähe zu behalten will.
- Google (fast):
  Die EZB will die Inflation auf unter zwei Prozent zu halten,
  oder irgendwo in der Nähe.

Machine translation

Translation

- Input:
  The proposal to remove Article 365 from the Code of Criminal Procedure, upon which the former Prime Minister was sentenced, was supported by 147 members of parliament.
- We (slow):
  Der Vorschlag, Artikel 365 aus der Code der kriminellen Geschäftsordnung, auf die der ehemalige Ministerpräsident verurteilt wurde zu entfernen, wurde von 147 Abgeordneten des Parlaments unterstützt.
- Google (fast):
  Der Vorschlag, Artikel 365 aus der Strafprozessordnung, auf dem der ehemalige Ministerpräsident wurde verurteilt zu entfernen, wurde von 147 Abgeordneten unterstützt.



Machine translation

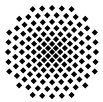
Translation

- Input:
  The proposal to remove Article 365 from the Code of Criminal Procedure, upon which the former Prime Minister was sentenced, was supported by 147 members of parliament.
- We (slow):
  Der Vorschlag, Artikel 365 aus der Code der kriminellen Geschäftsordnung, auf die der ehemalige Ministerpräsident verurteilt wurde zu entfernen, wurde von 147 Abgeordneten des Parlaments unterstützt.
- Google (fast):
  Der Vorschlag, Artikel 365 aus der Strafprozessordnung, auf dem der ehemalige Ministerpräsident wurde verurteilt zu entfernen, wurde von 147 Abgeordneten unterstützt.

Machine translation

Translation

- Input:
  The proposal to remove Article 365 from the Code of Criminal Procedure, upon which the former Prime Minister was sentenced, was supported by 147 members of parliament.
- We (slow):
  Der Vorschlag, Artikel 365 aus der Code der kriminellen Geschäftsordnung, auf die der ehemalige Ministerpräsident verurteilt wurde zu entfernen, wurde von 147 Abgeordneten des Parlaments unterstützt.
- Google (fast):
  Der Vorschlag, Artikel 365 aus der Strafprozessordnung, auf dem der ehemalige Ministerpräsident wurde verurteilt zu entfernen, wurde von 147 Abgeordneten unterstützt.



Machine translation

Translation

- Input:
  Their only excuse is that from their point of view, NATO raids were not controlled by an American general, but Allah, the greatness of whom they celebrated with each strike.
- We (slow):
  Ihre einzige Ausrede ist, dass aus ihrer Sicht, Schlachtzügen der NATO nicht von einem amerikanischen Allgemeinen, aber Allah, von denen die Größe sie mit jedem Streik feierte kontrolliert wurden.
- Google (fast):
  Ihre einzige Entschuldigung ist, dass aus ihrer Sicht, die NATO nicht Razzien wurden von einem amerikanischen General kontrolliert, aber Allah, die Größe, die sie feierten mit jedem Schlag.

Machine translation

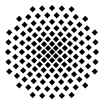
Translation

- Input:
  Their only excuse is that from their point of view, NATO raids were not controlled by an American general, but Allah, the greatness of whom they celebrated with each strike.
- We (slow):
  Ihre einzige Ausrede ist, dass aus ihrer Sicht, Schlachtzügen der NATO nicht von einem amerikanischen Allgemeinen, aber Allah, von denen die Größe sie mit jedem Streik feierte kontrolliert wurden.
- Google (fast):
  Ihre einzige Entschuldigung ist, dass aus ihrer Sicht, die NATO nicht Razzien wurden von einem amerikanischen General kontrolliert, aber Allah, die Größe, die sie feierten mit jedem Schlag.

Machine translation

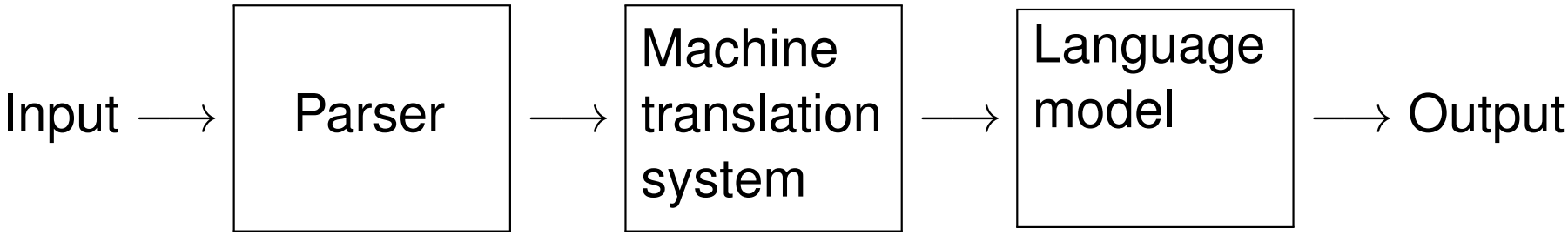
Translation

- Input:
  Their only excuse is that from their point of view, NATO raids were not controlled by an American general, but Allah, the greatness of whom they celebrated with each strike.
- We (slow):
  Ihre einzige Ausrede ist, dass aus ihrer Sicht, Schlachtzügen der NATO nicht von einem amerikanischen Allgemeinen, aber Allah, von denen die Größe sie mit jedem Streik feierte kontrolliert wurden.
- Google (fast):
  Ihre einzige Entschuldigung ist, dass aus ihrer Sicht, die NATO nicht Razzien wurden von einem amerikanischen General kontrolliert, aber Allah, die Größe, die sie feierten mit jedem Schlag.



Syntax-based machine translation

Syntax-based systems



What do we have?

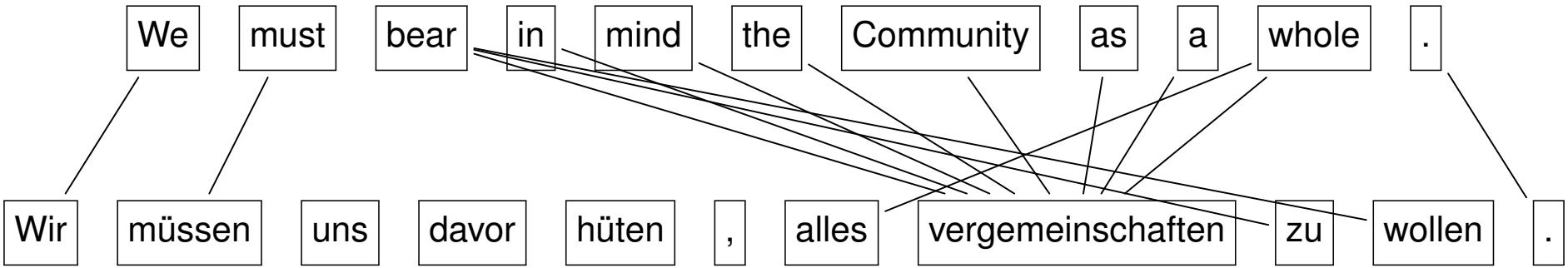
Input

- parallel text (English and German)
- here: EUROPARL

Example

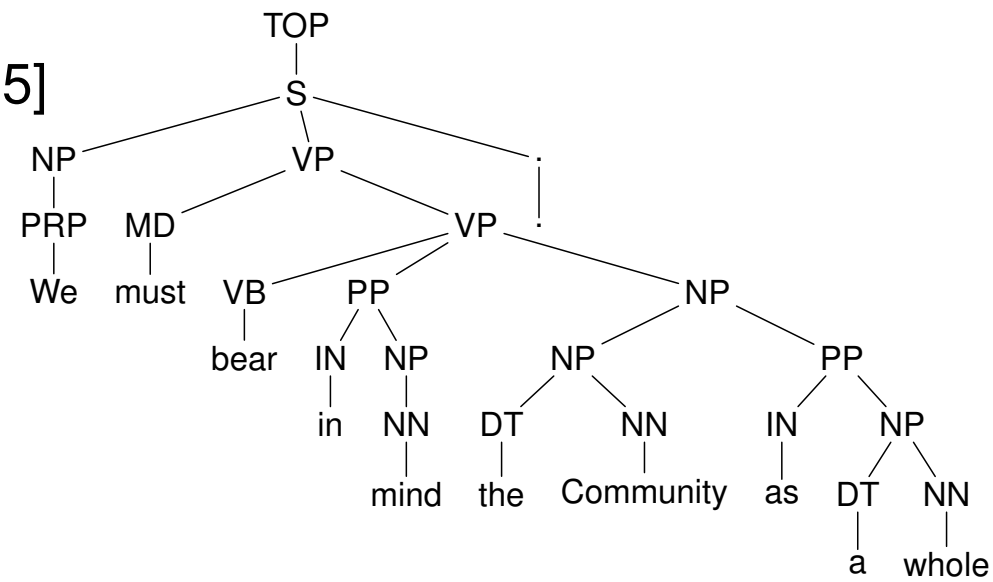
- “We must bear in mind the Community as a whole.”
- “Wir müssen uns davor hüten, alles vergemeinschaften zu wollen.”

Alignments by GIZA++ [OCH, NEY ’03]:

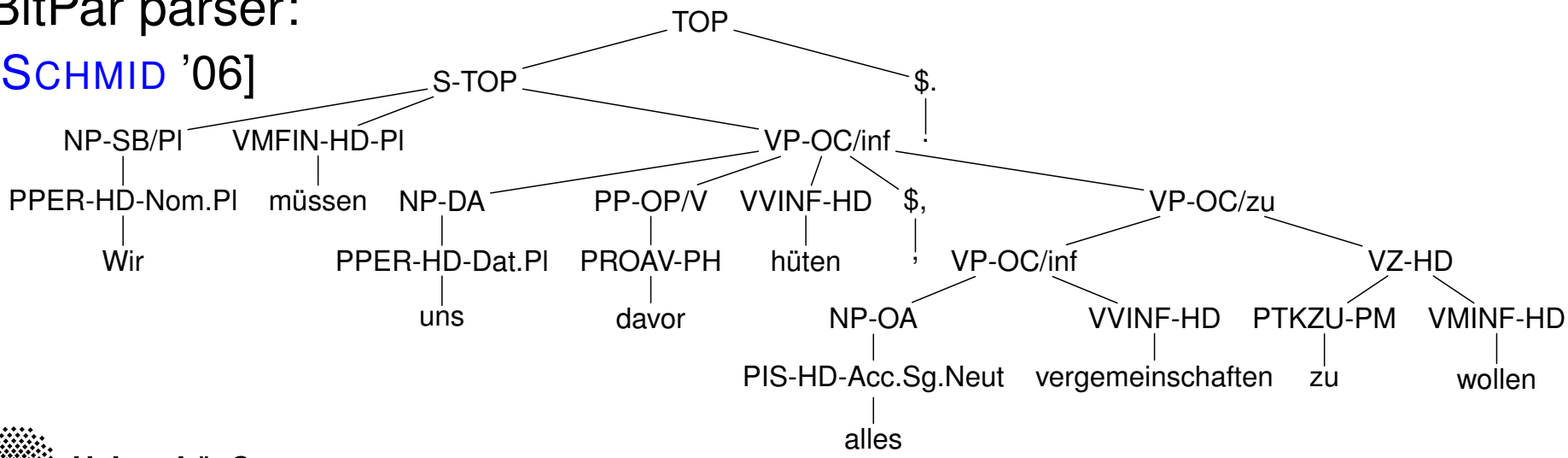


Parsing

CHARNIAK parser:
[CHARNIAK, JOHNSON '05]



BitPar parser:
[SCHMID '06]

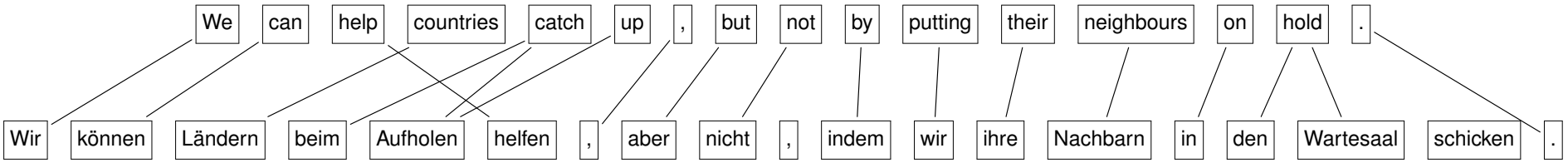


Better example

Example

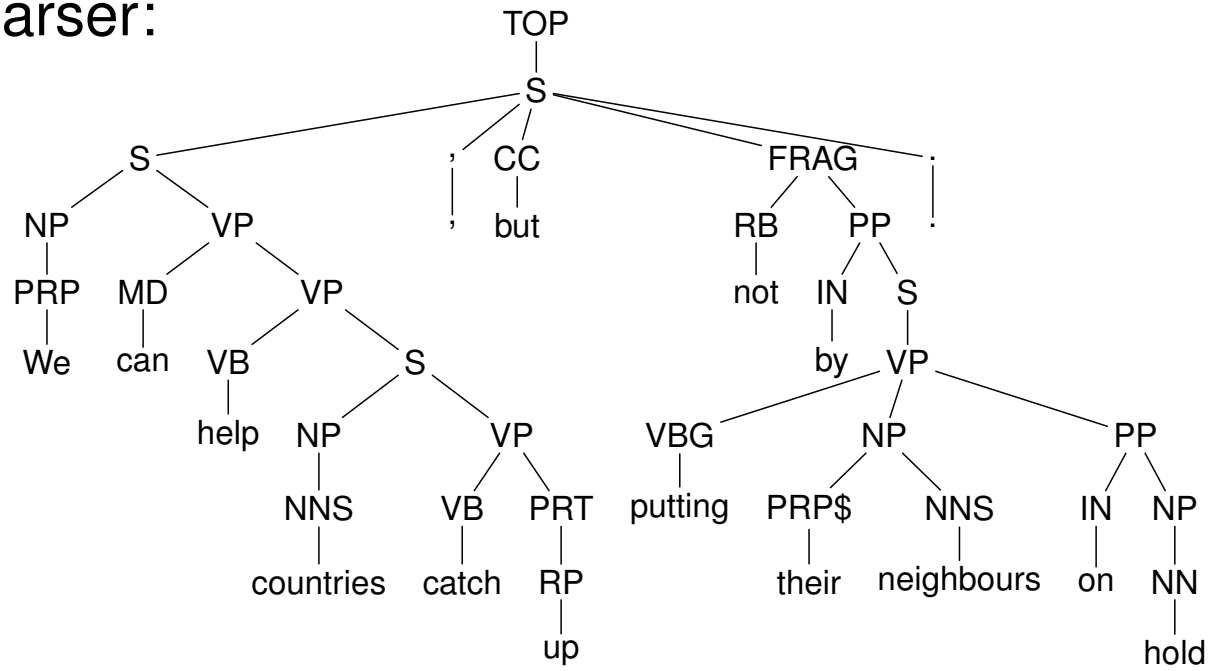
- “We can help countries catch up,
  but not by putting their neighbours on hold.”
- “Wir können Ländern beim Aufholen helfen,
  aber nicht, indem wir ihre Nachbarn in den Wartesaal schicken.”

Alignments by GIZA++:

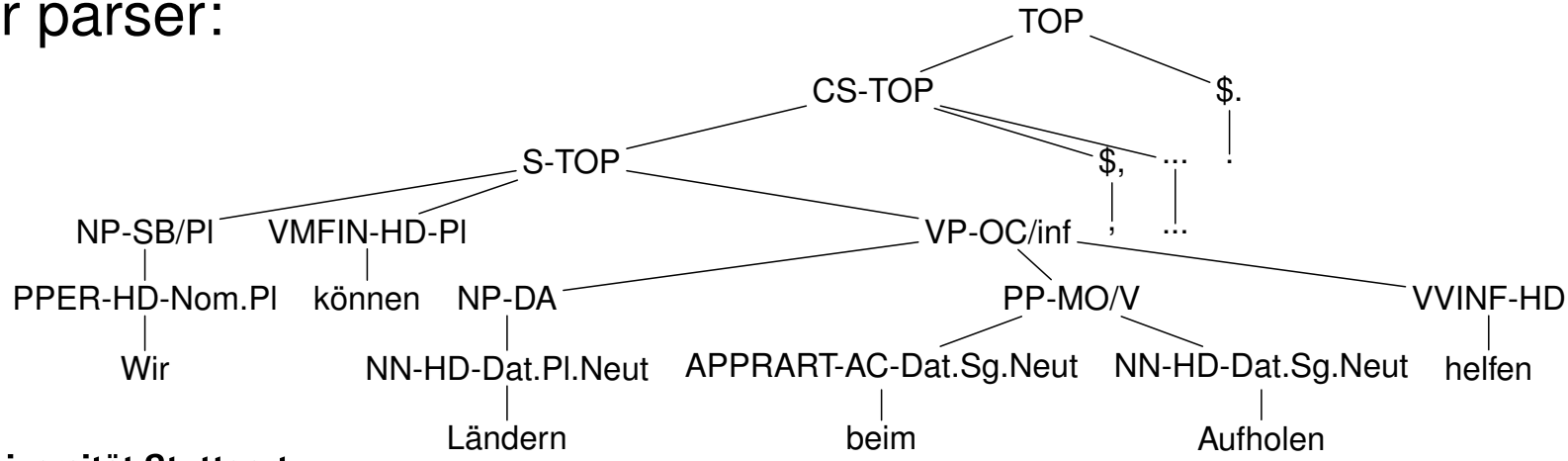


Better example

CHARNIAK parser:



BitPar parser:



Small example

Input

Yugoslav President Voislav signed for Serbia.

و تولى التوقيع عن صربيا الرئيس اليوغوسلافي فويسلاف

Transliteration: *w twlY AltwqyE En SrbyA Alr}ys Alywgwslafy fwyslAf.*

And then the matter was decided, and everything was put in place.

ف كان ان تم الحسم و وضعت الأمور في نصاب ها

Transliteration: *f kAn An tm AlHsm w wDEt Al>mwr fy nSAb hA.*

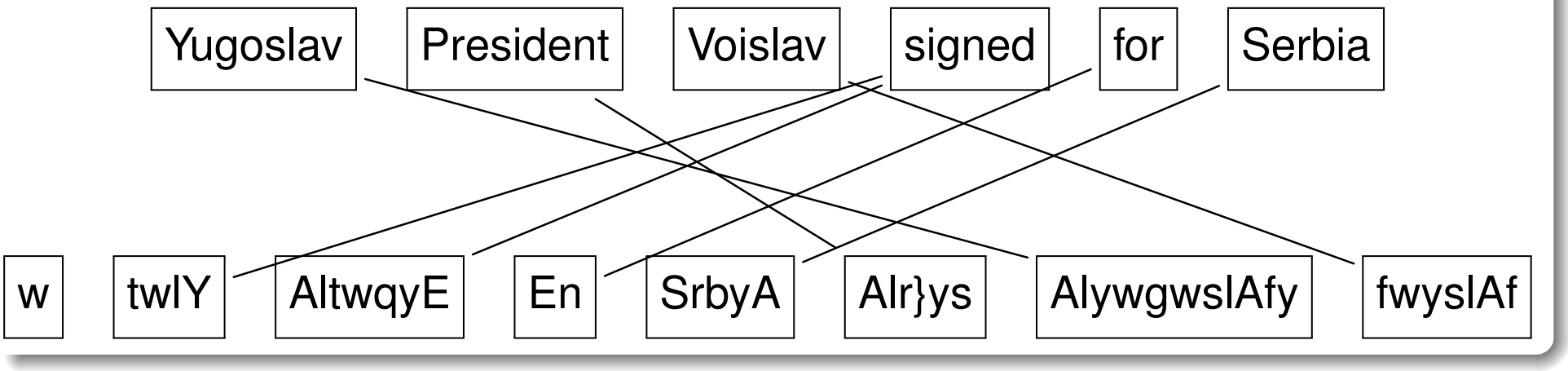
Below are the male and female winners in the different categories.

و هنا الأوائل و الأوليات في مختلف الفئات

Transliteration: *w hnA Al>wA}l w Al>wlyAt fy mxtlf Alf}At.*

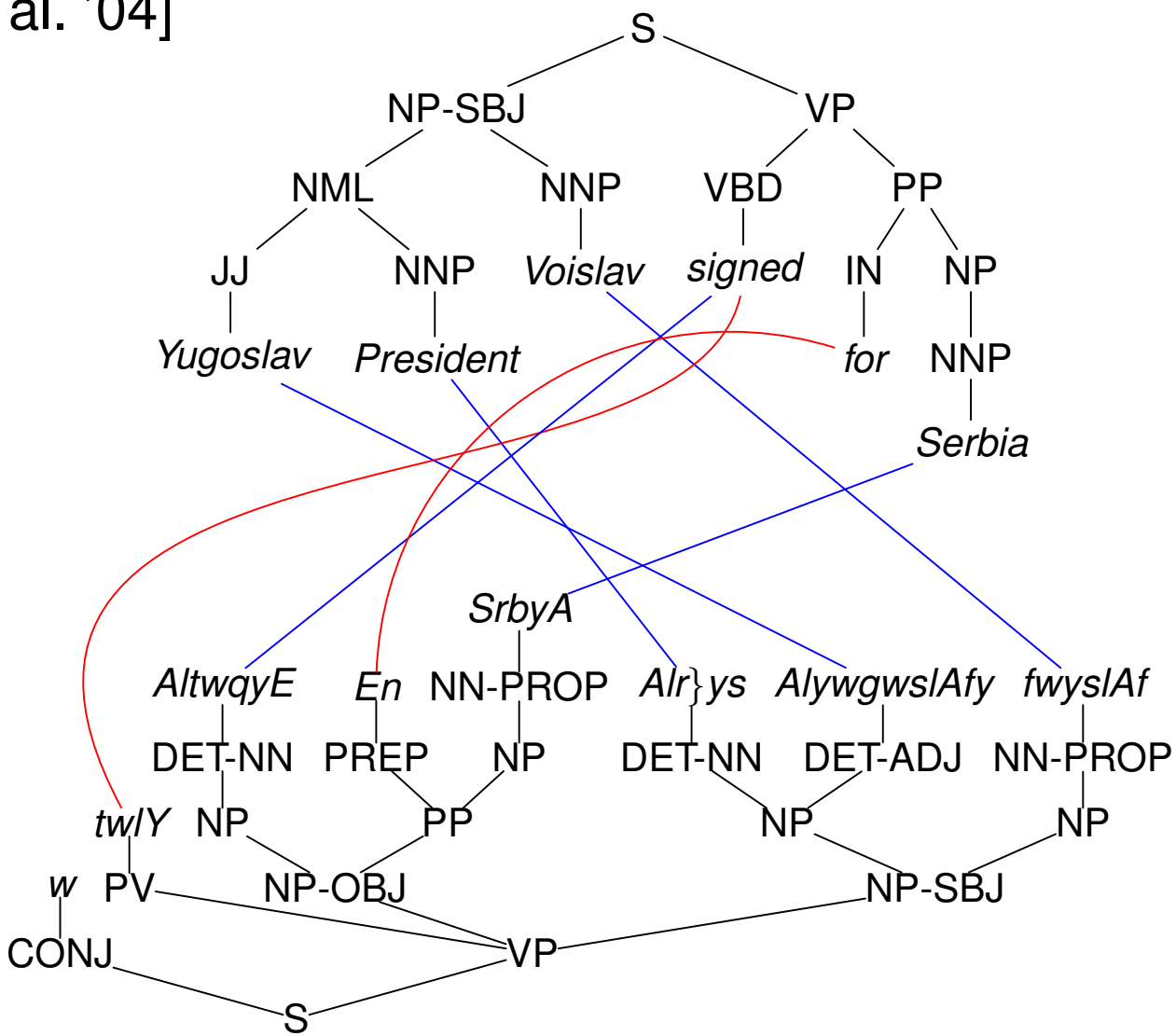
Small example

Alignment

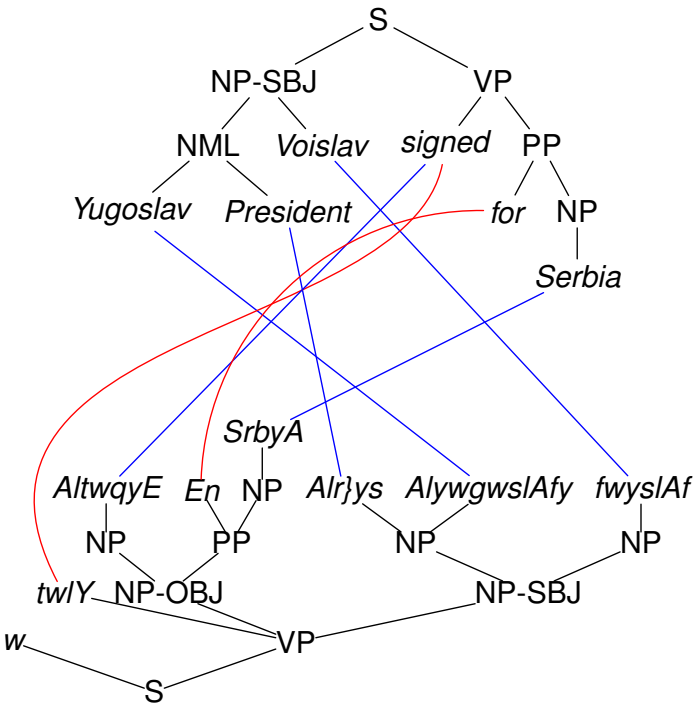


GHKM rule extraction

[GALLEY et al. '04]

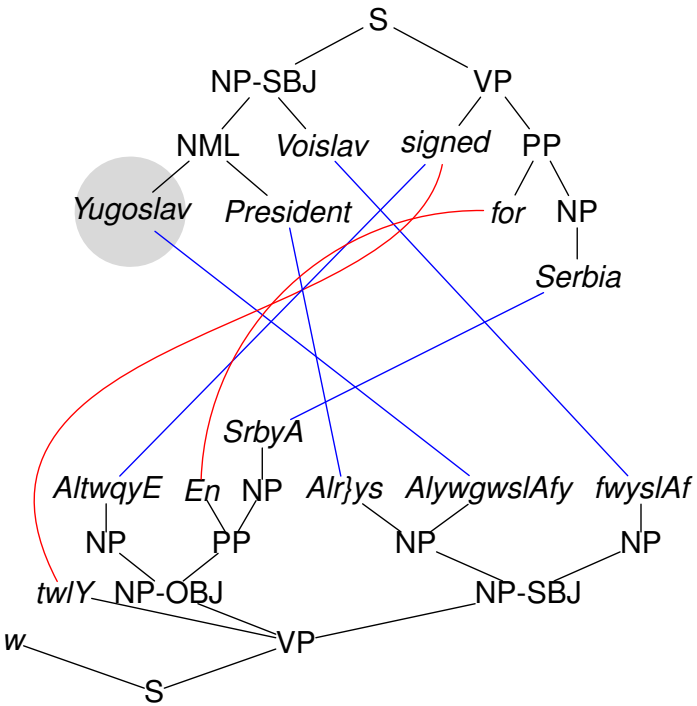


GHKM rule extraction



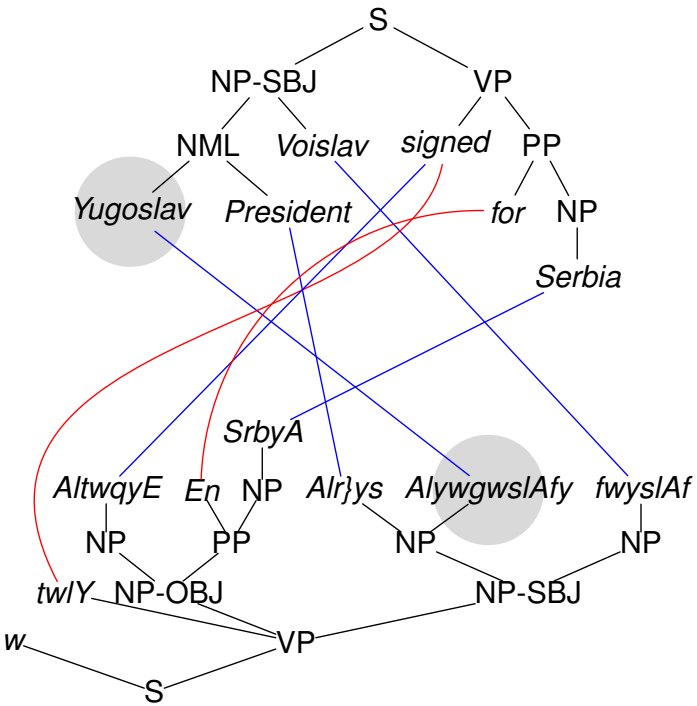
- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

GHKM rule extraction



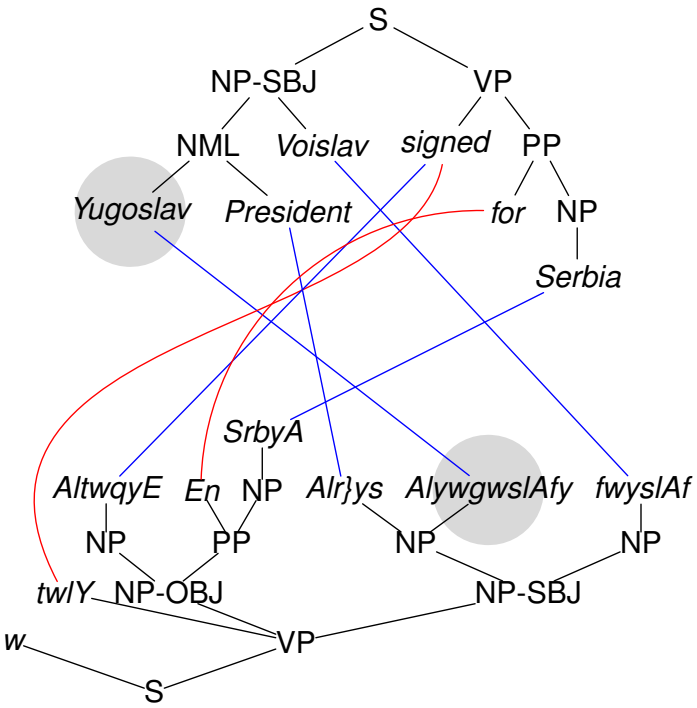
- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

GHKM rule extraction



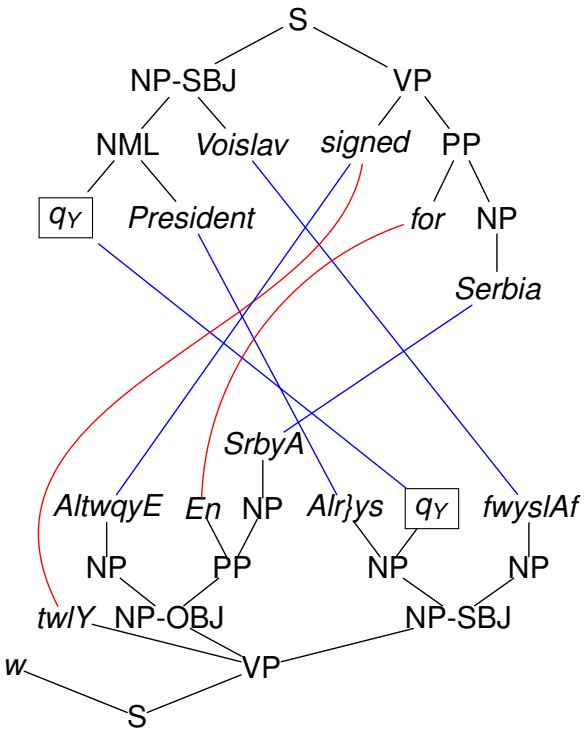
- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

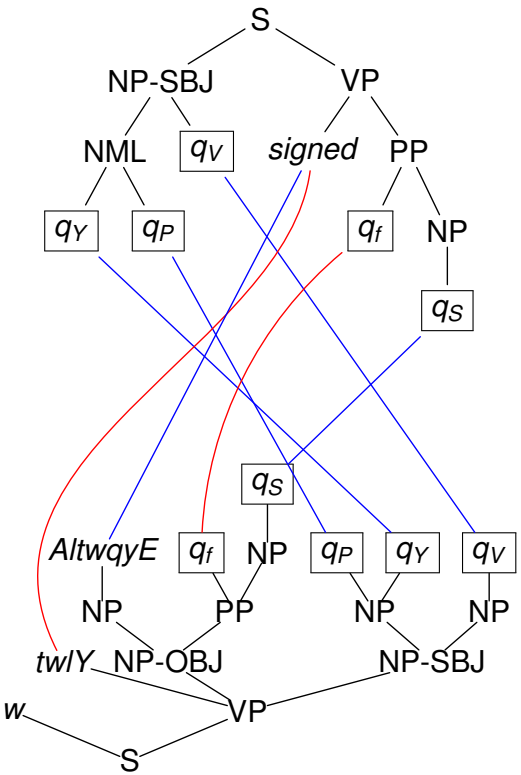
GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

  $Yugoslav \frac{q_Y}{} AlywgwslAfy$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy}$$

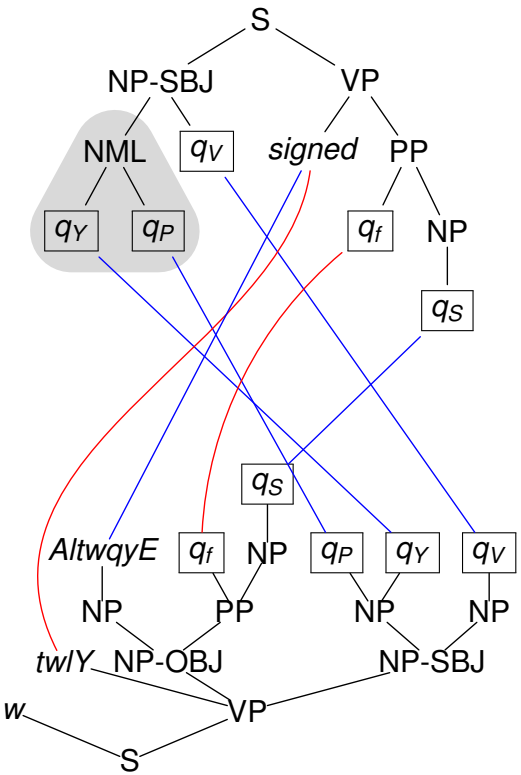
$$\textit{President} \xrightarrow{q_P} \textit{Alr\}ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf}$$

$$\textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy}$$

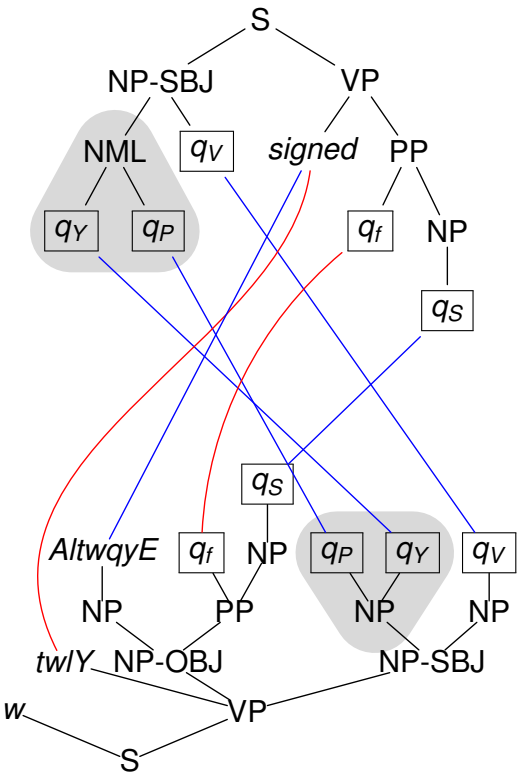
$$\textit{President} \xrightarrow{q_P} \textit{Alr\}ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf}$$

$$\textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

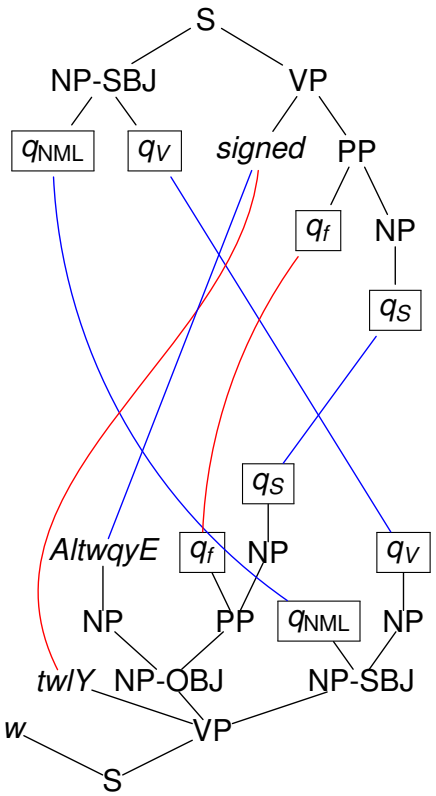
GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\mathrm{NML}(q_Y, q_P) \overset{q_{\mathrm{NML}}}{\underline{\quad}} \mathrm{NP}(q_P, q_Y)$$

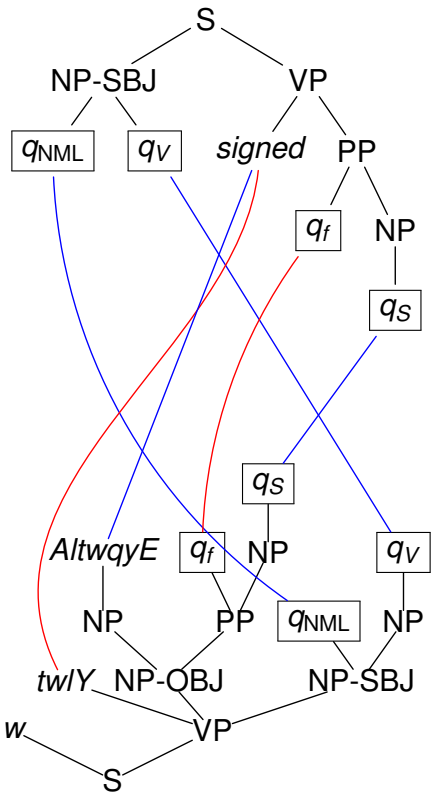
GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\mathrm{NML}(q_Y, q_P) \overset{q_{\mathrm{NML}}}{\underline{\quad}} \mathrm{NP}(q_P, q_Y)$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

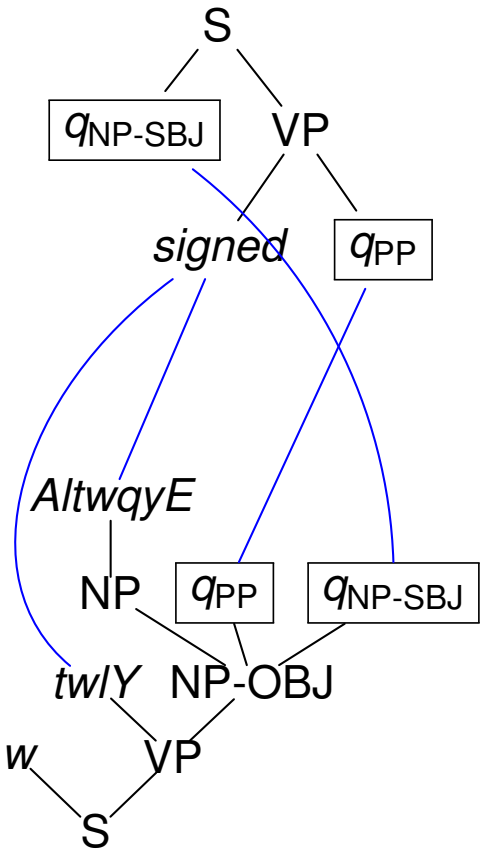
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

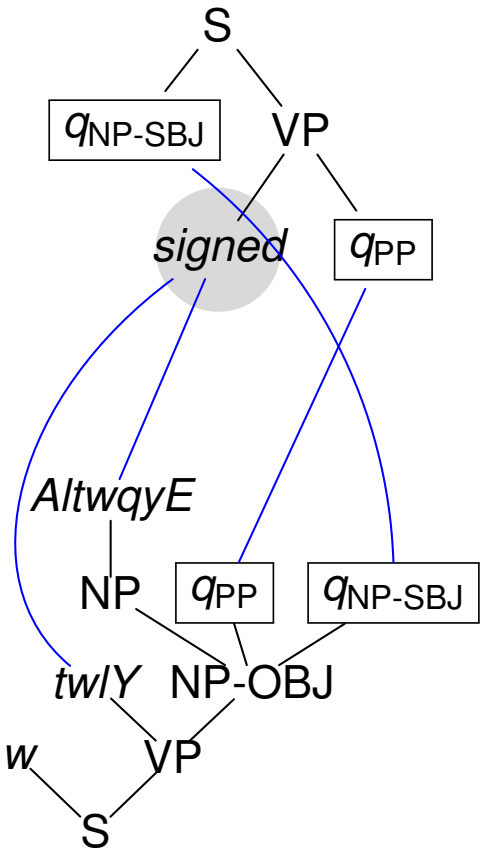
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\quad}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\quad}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\quad}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

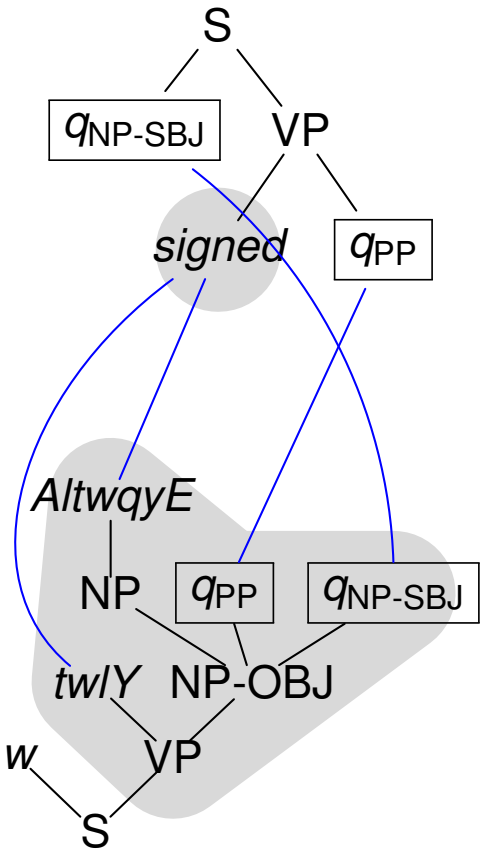
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\quad}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\quad}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\quad}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

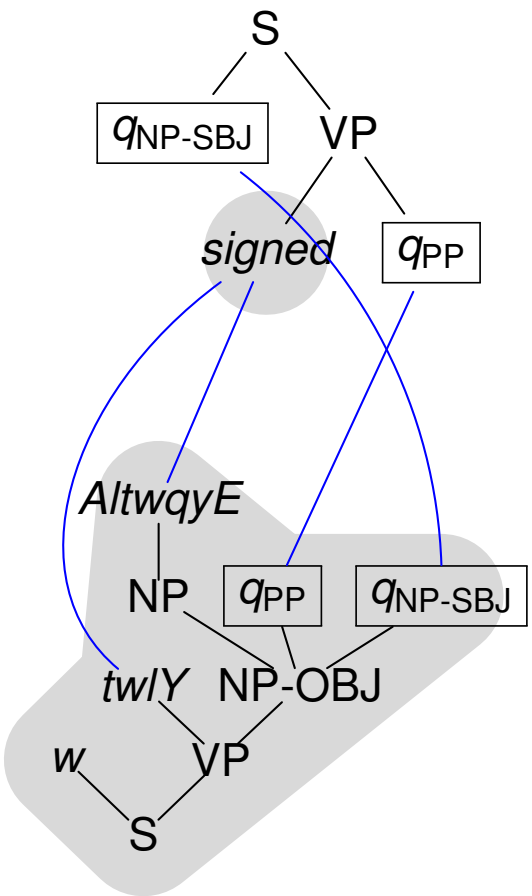
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\quad}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\quad}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\quad}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

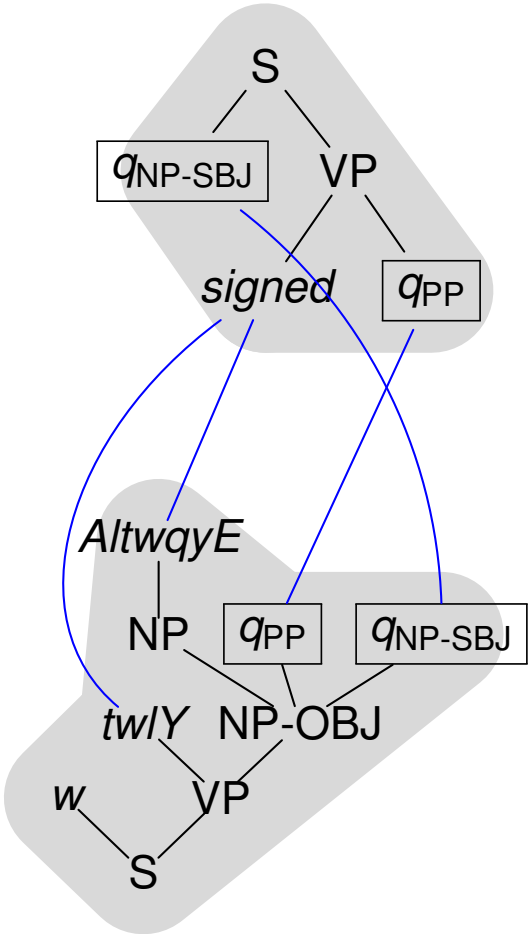
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\quad}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\quad}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\quad}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

GHKM rule extraction



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Extended top-down tree transducer

Advantages

- ✓ simple and natural model
- ✓ easy to train (from linguistic resources) [GRAEHL et al. '08]
- ✓ symmetric

Implementation

- TIBURON [MAY, KNIGHT '06]

Extended top-down tree transducer

Advantages

- ✓ simple and natural model
- ✓ easy to train (from linguistic resources) [GRAEHL et al. '08]
- ✓ symmetric

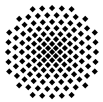
Implementation

- TIBURON [MAY, KNIGHT '06]

Extended top-down tree transducer

Disadvantages (also of STSG)

- ✗ no discontinuities
- ✗ not binarizable
  [AHO, ULLMAN '72; ZHANG et al. '06]
- ✗ inefficient input/output restriction
  [M., SATTA '10]
- ✗ not composable
  [ARNOLD, DAUCHET '82]



Our model — Syntax

Definition

Multi bottom-up tree transducer (MBOT)
system (Q, Σ, F, R)

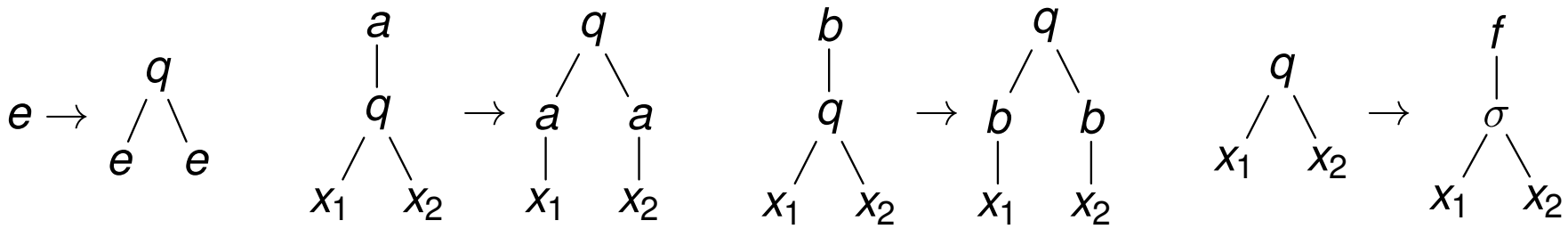
- Q *ranked* alphabet (states)
- Σ ranked alphabet (input/output symbols)
- $F \subseteq Q_1$ (final states)
- R finite set of rules $\ell \to r$ (rules)
  - linear $\ell \in T_\Sigma(Q(X))$ and $r \in Q(T_\Sigma(\mathrm{var}(\ell)))$
  - $\mathrm{var}(r) = \mathrm{var}(\ell)$

Our model — Syntax

Example

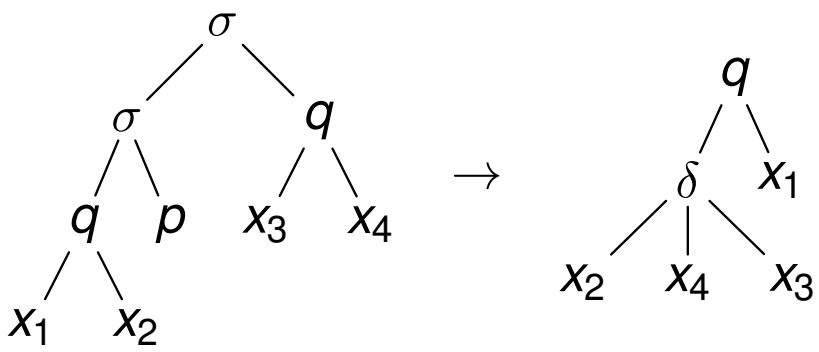
MBOT $(Q, \Sigma, \{f\}, R)$

- $Q = \{q^{(2)}, f^{(1)}\}$
- $\Sigma = \{\sigma^{(2)}, a^{(1)}, b^{(1)}, e^{(0)}\}$
- R contains:

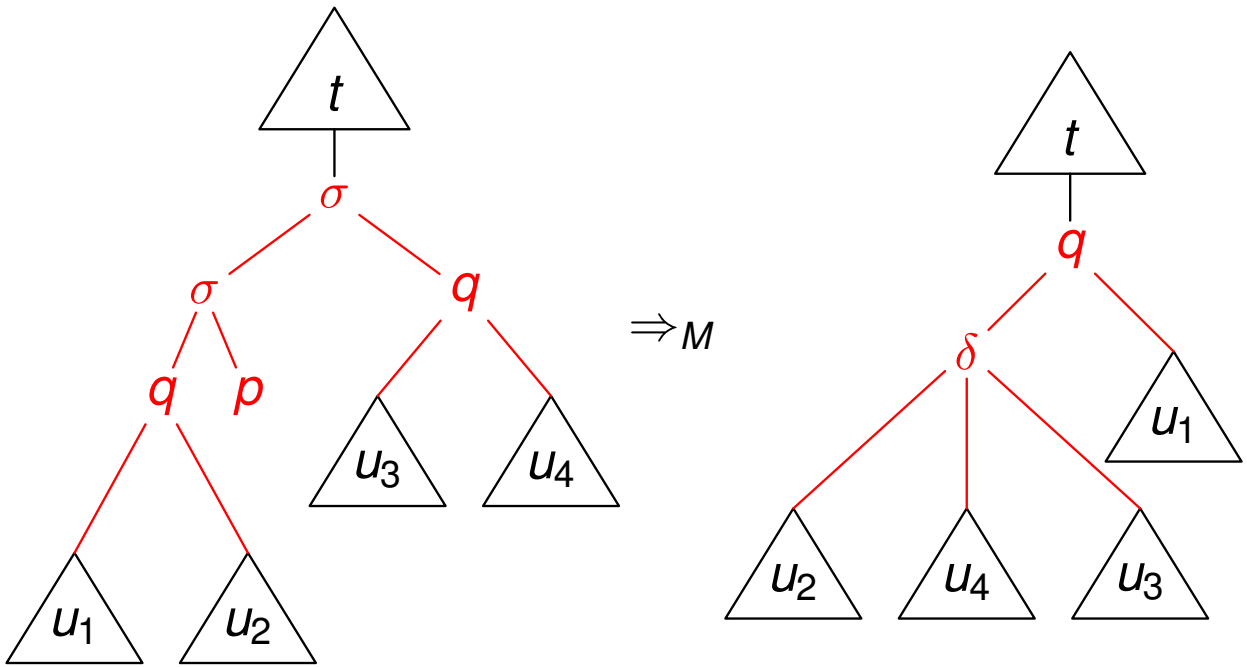


Our model — Semantics

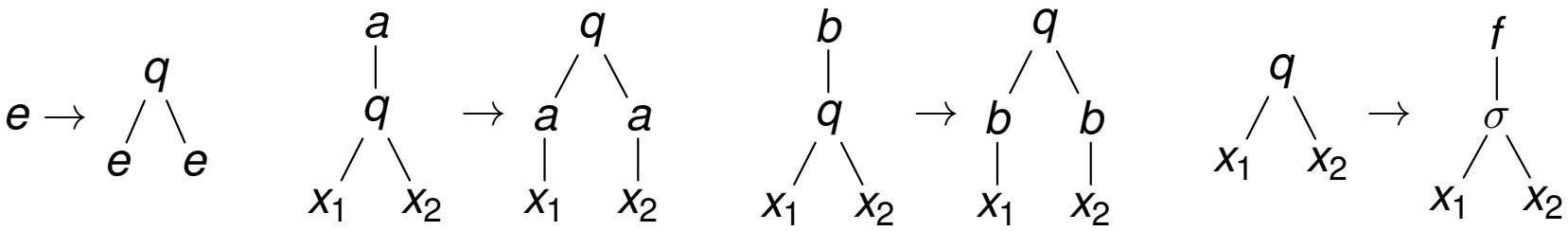
Rule:



Derivation:

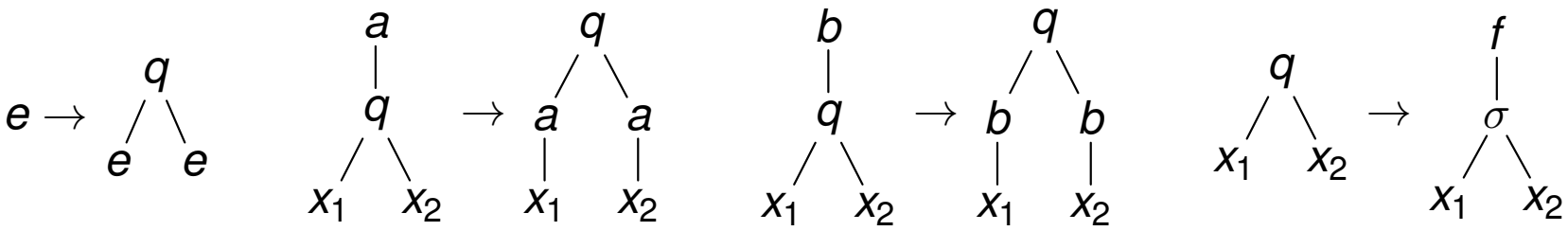


Our model — Semantics

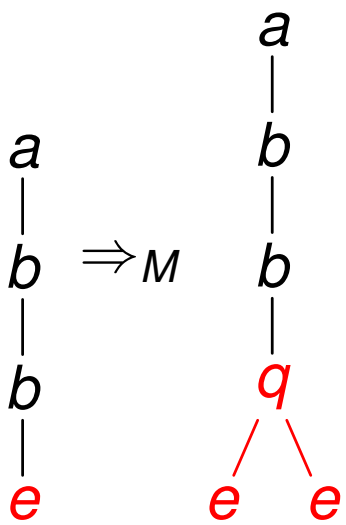


Example (Derivation)

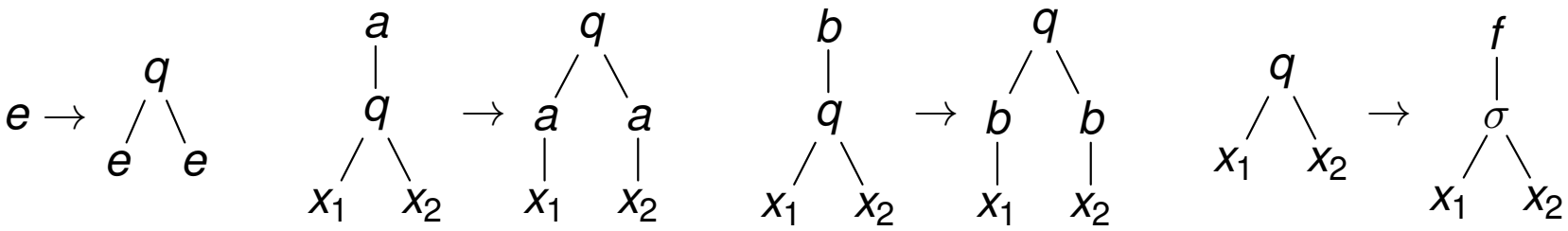
Our model — Semantics



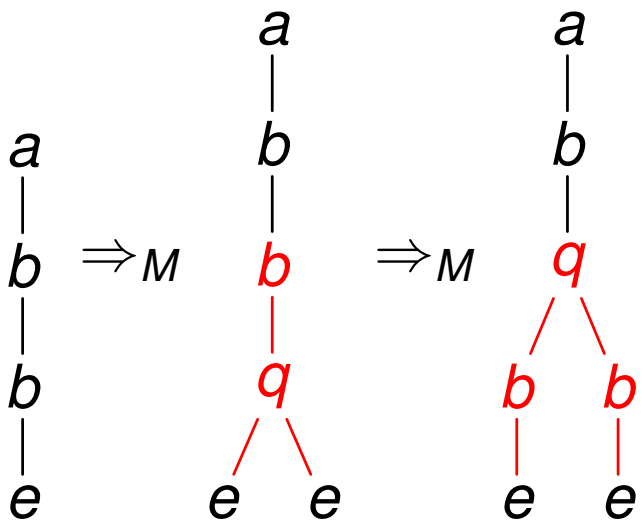
Example (Derivation)



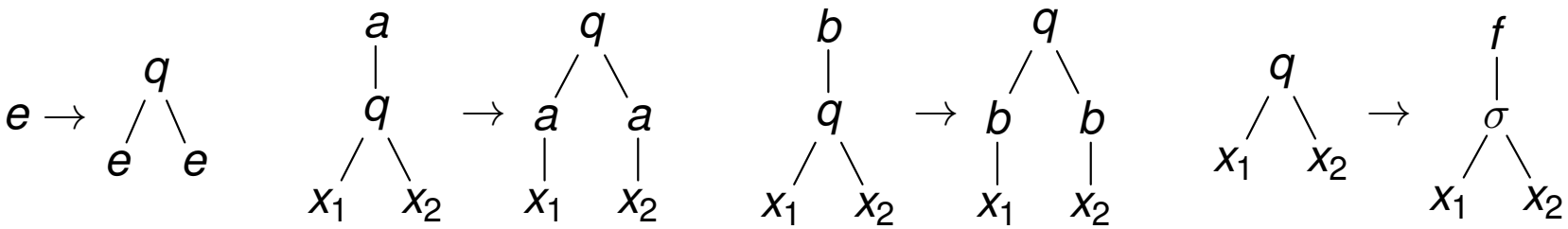
Our model — Semantics



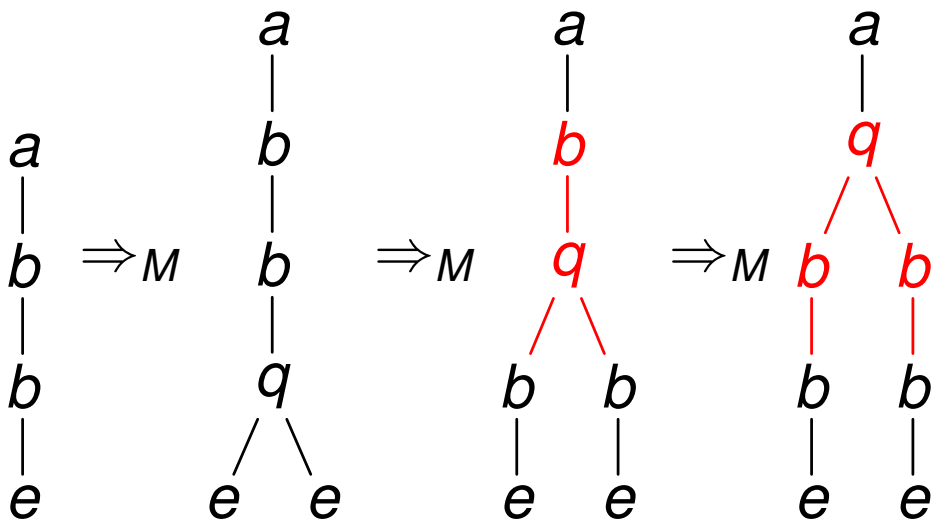
Example (Derivation)



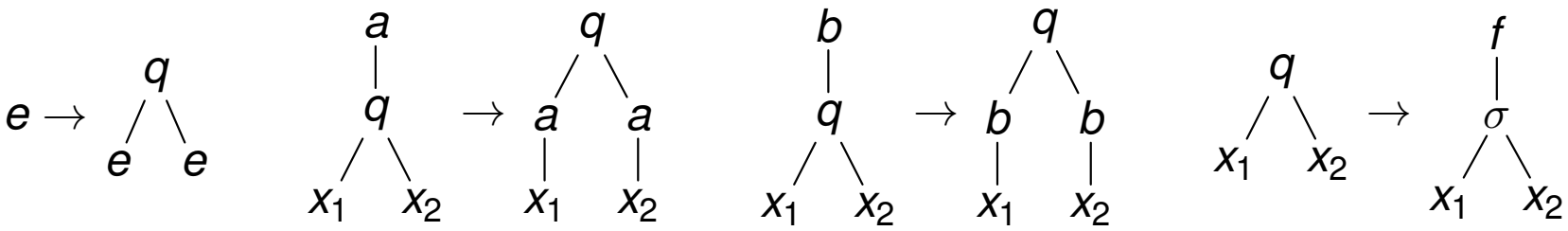
Our model — Semantics



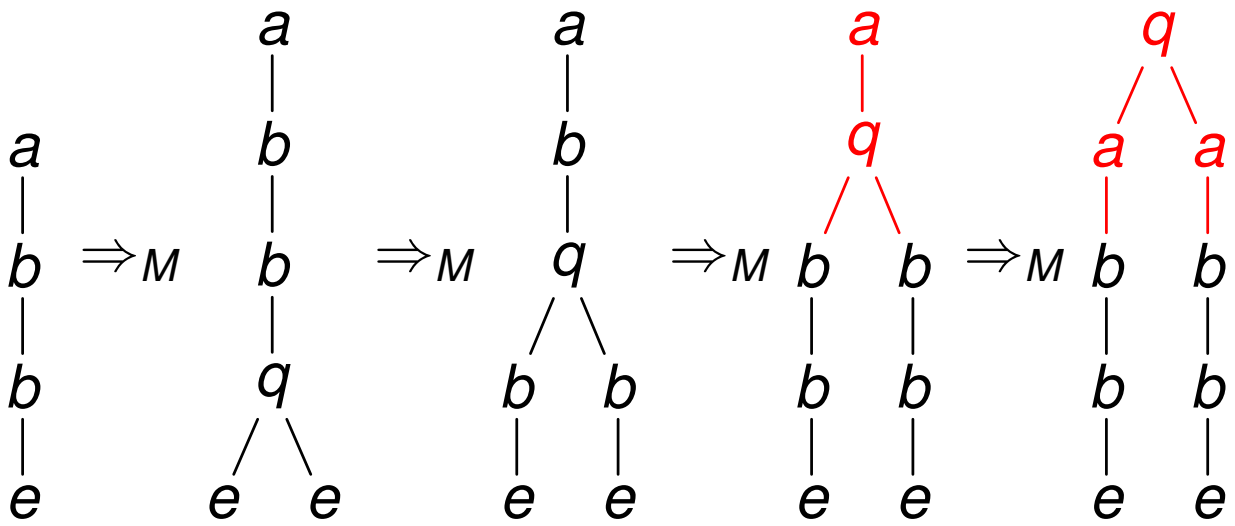
Example (Derivation)



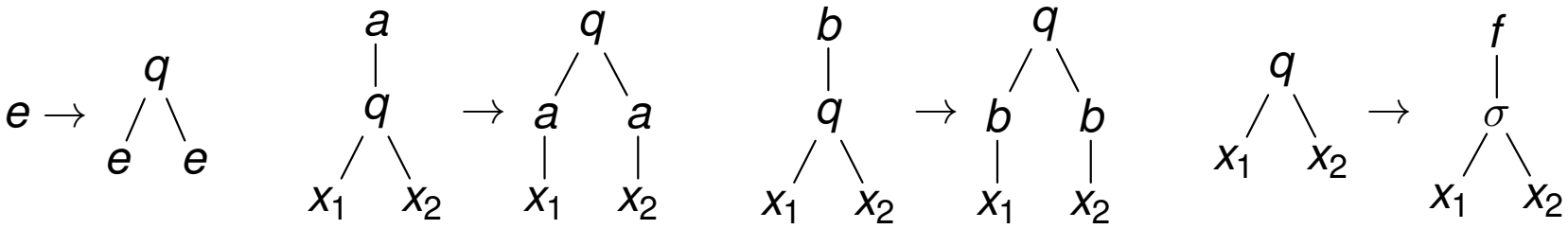
Our model — Semantics



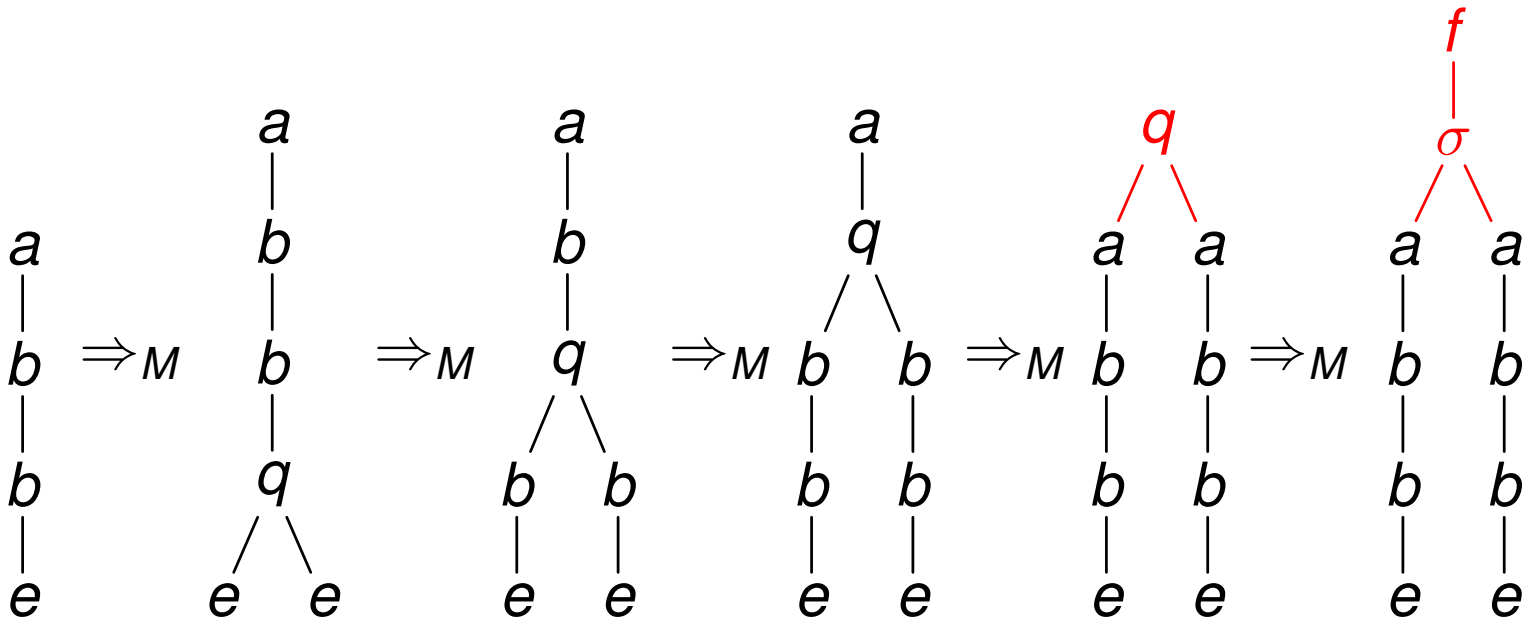
Example (Derivation)



Our model — Semantics



Example (Derivation)



Our model — Semantics

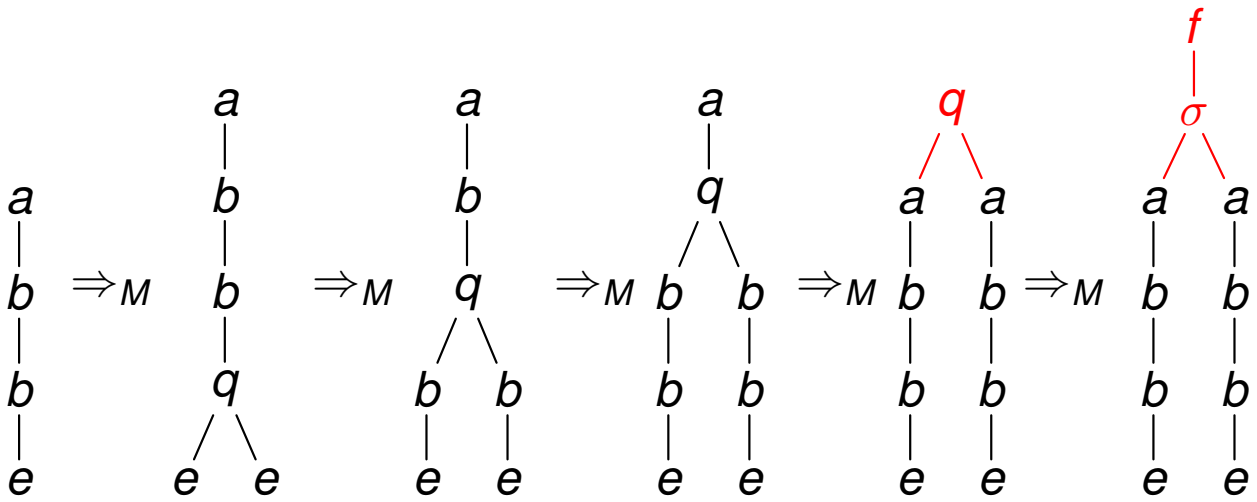
Definition

MBOT $M = (Q, \Sigma, F, R)$

$$\tau_M = \{(t, u) \in T_\Sigma \times T_\Sigma \mid \exists q \in F : t \Rightarrow_M^* q(u)\}$$

Discontinuities

Example (Derivation)

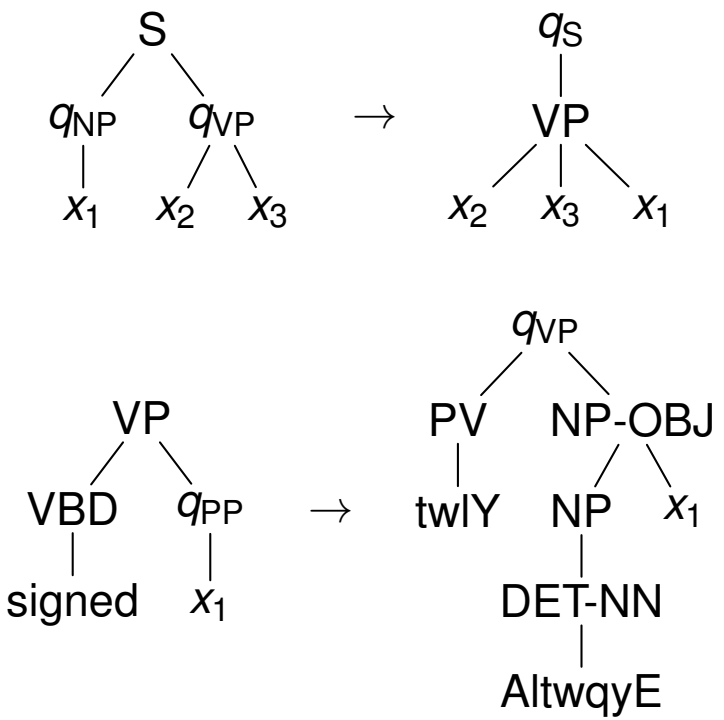


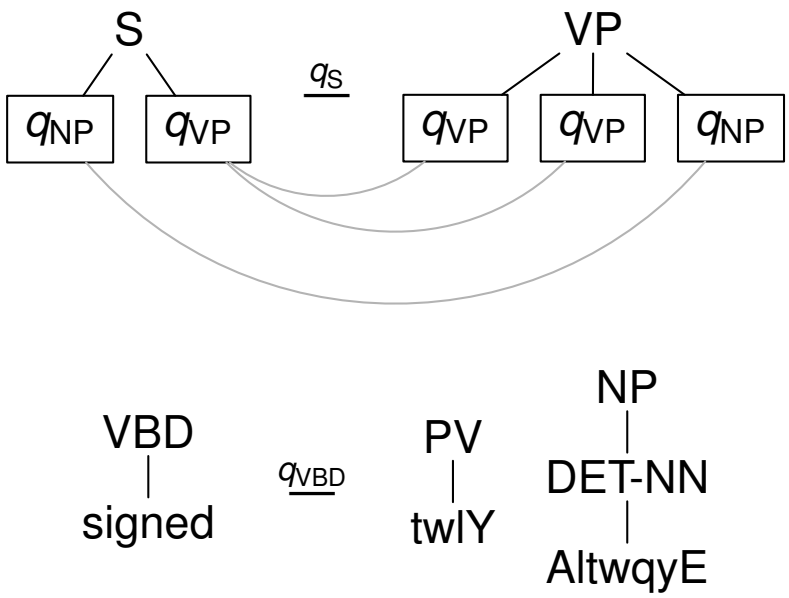
Discontinuities

✗ state covers 1 input subtree → no input discontinuities

✓ state covers several output subtrees → output discontinuities

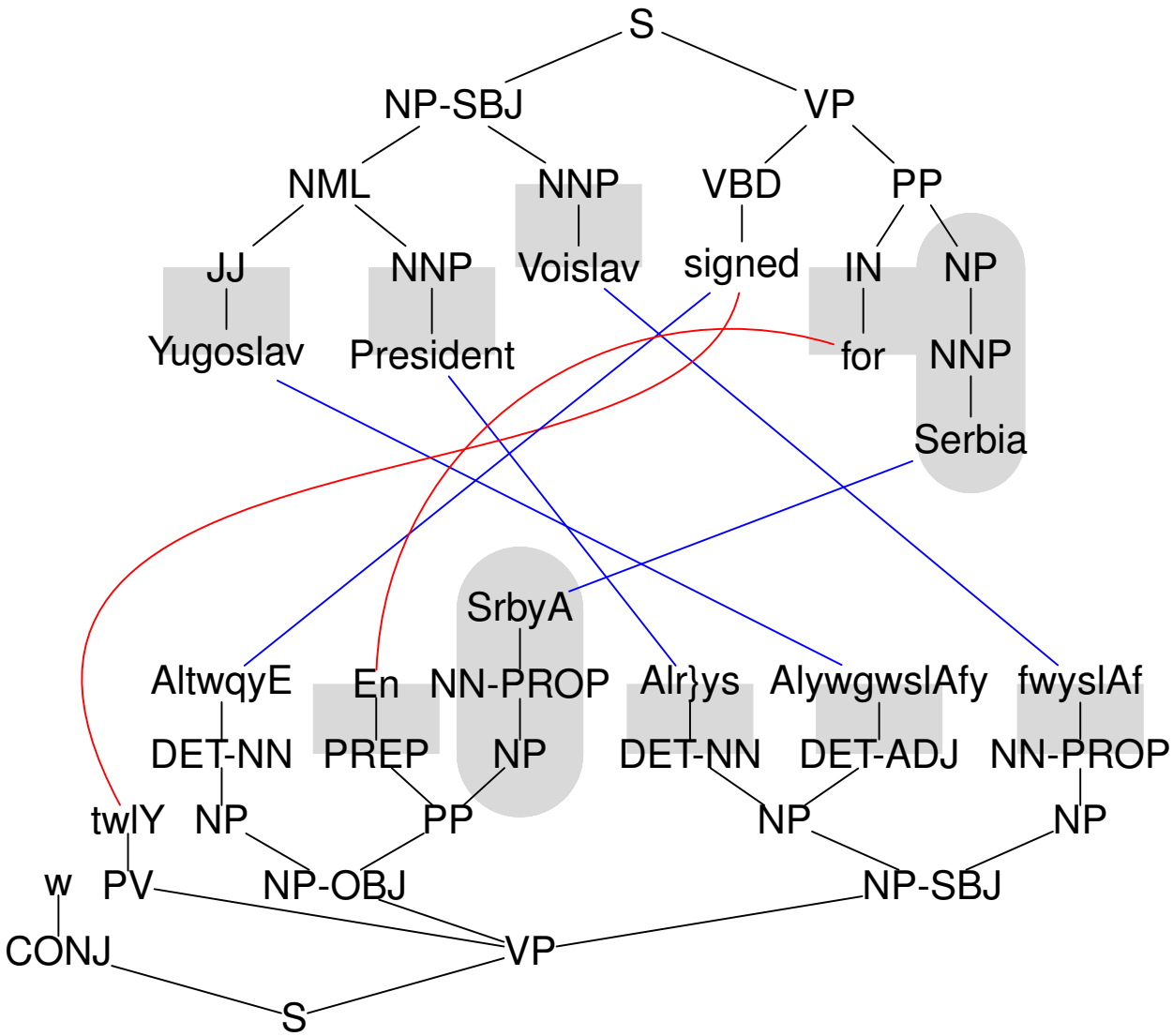
A top-down variant



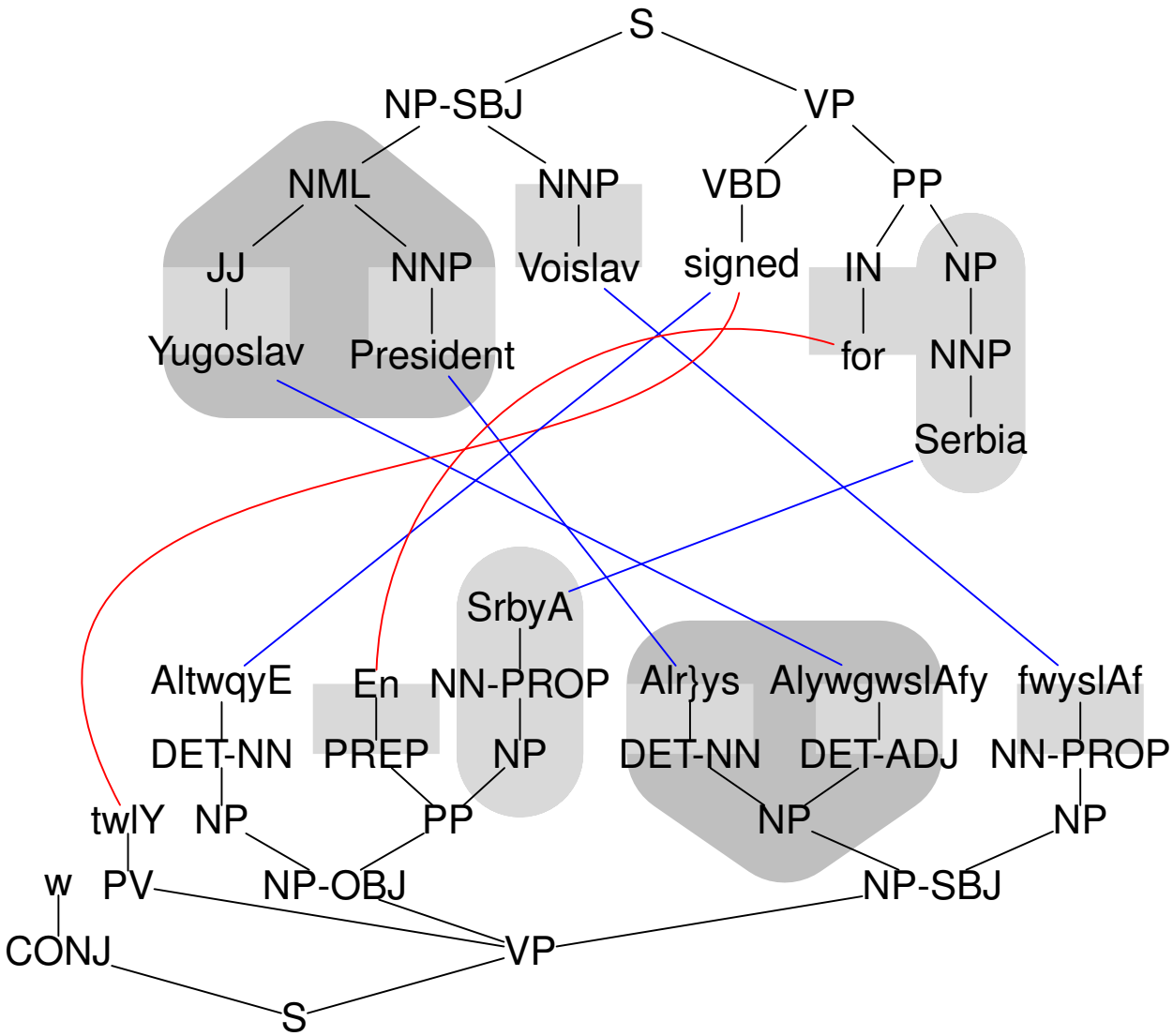


[M. '11]

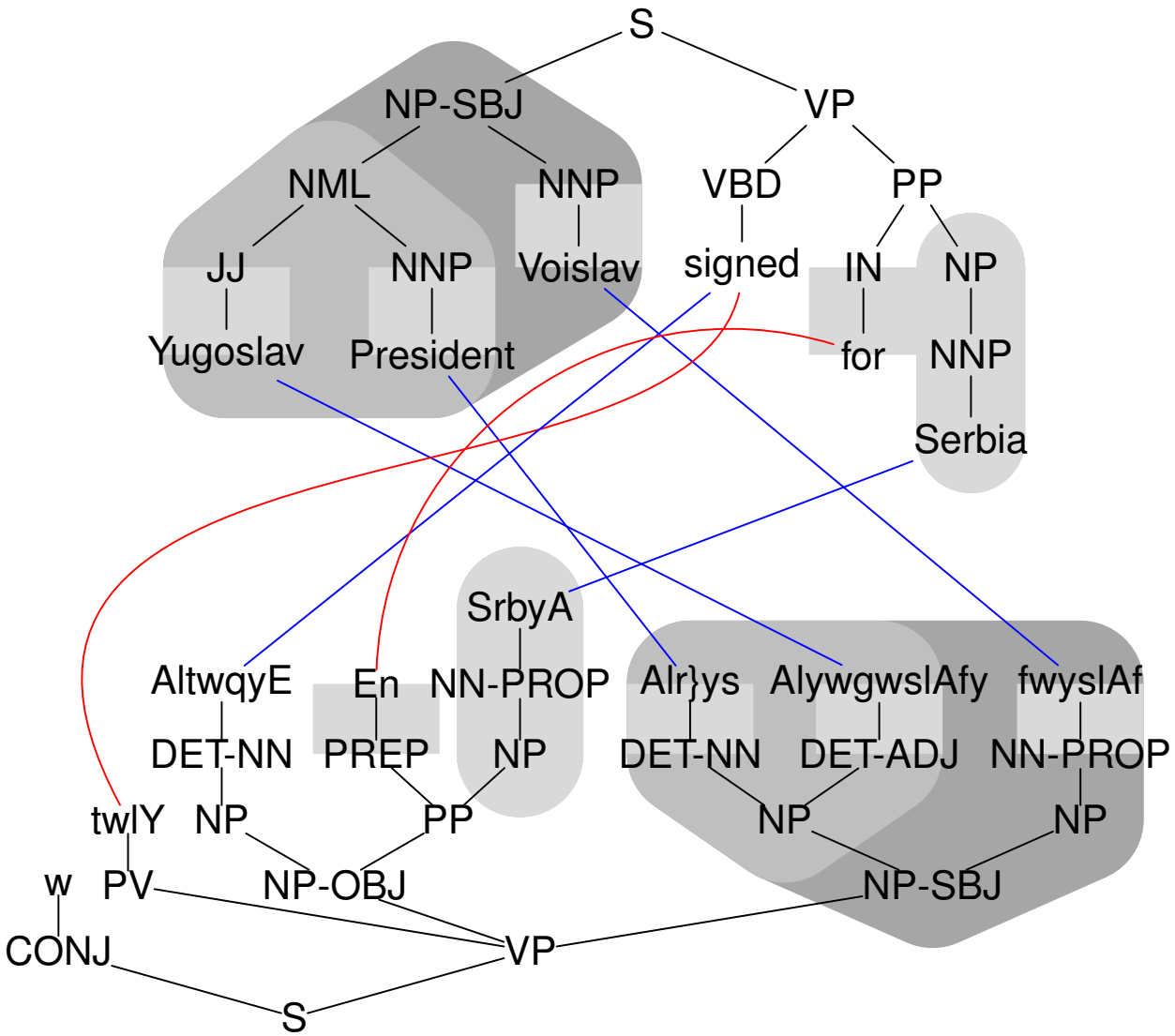
Rule extraction



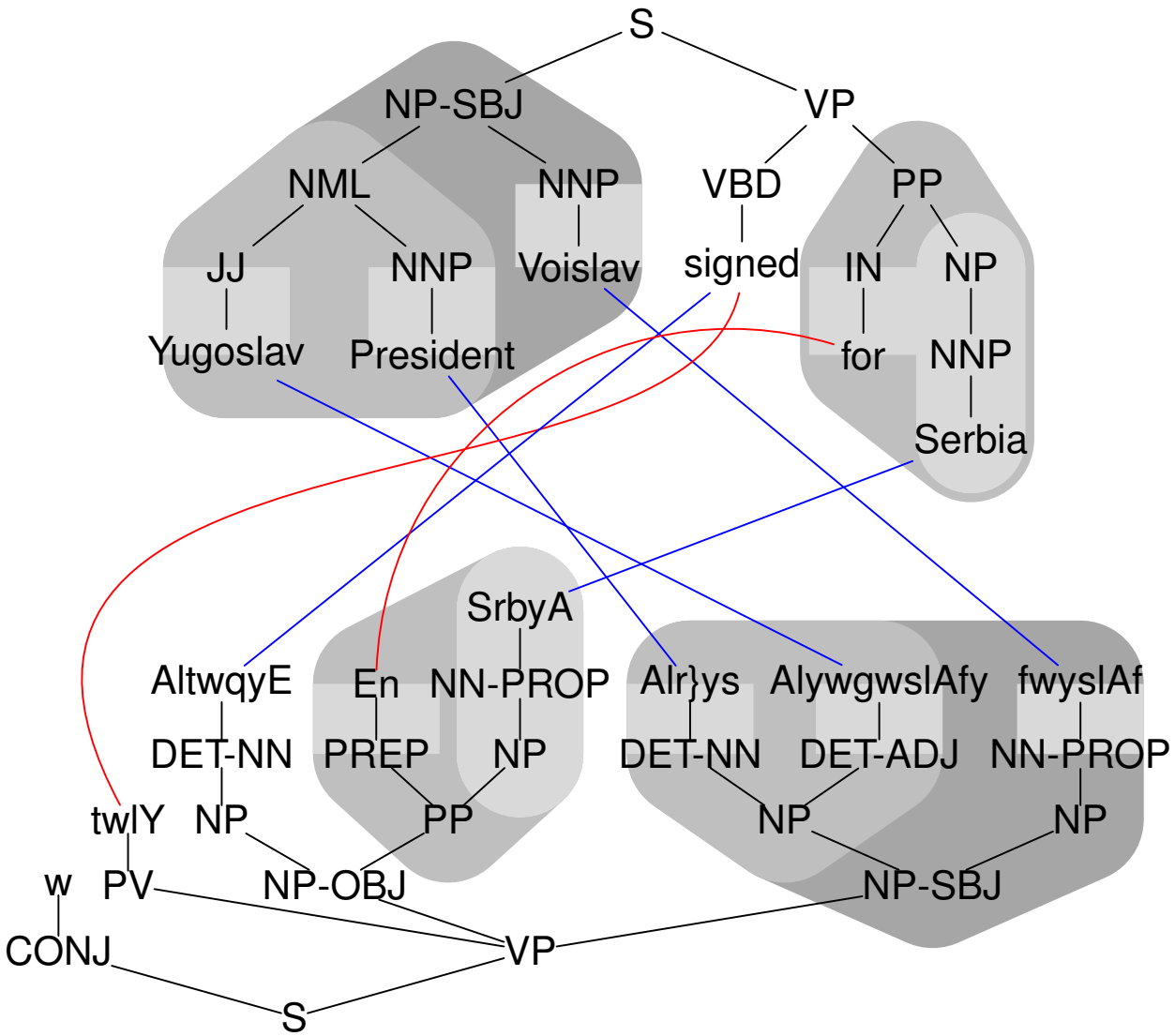
Rule extraction



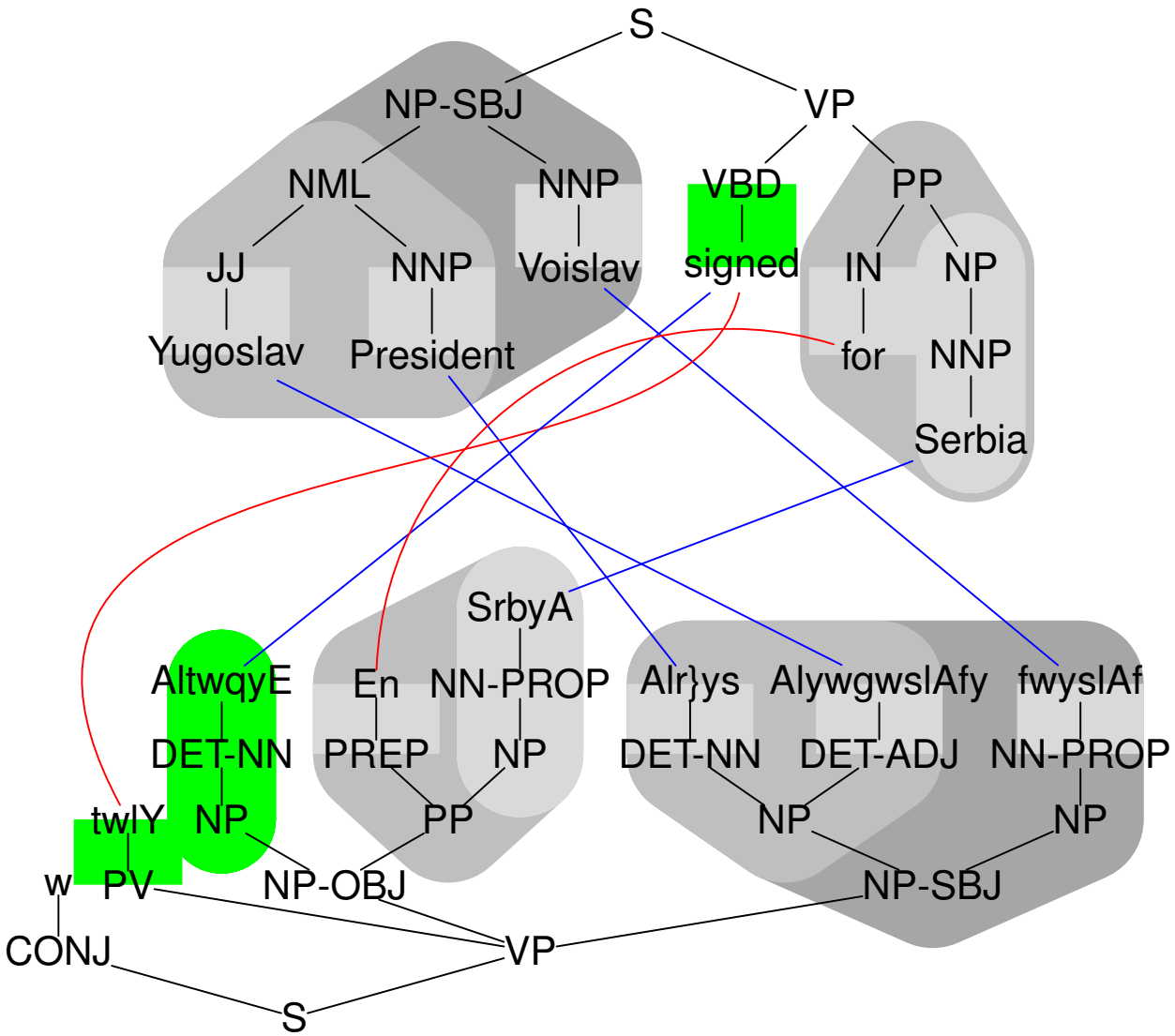
Rule extraction



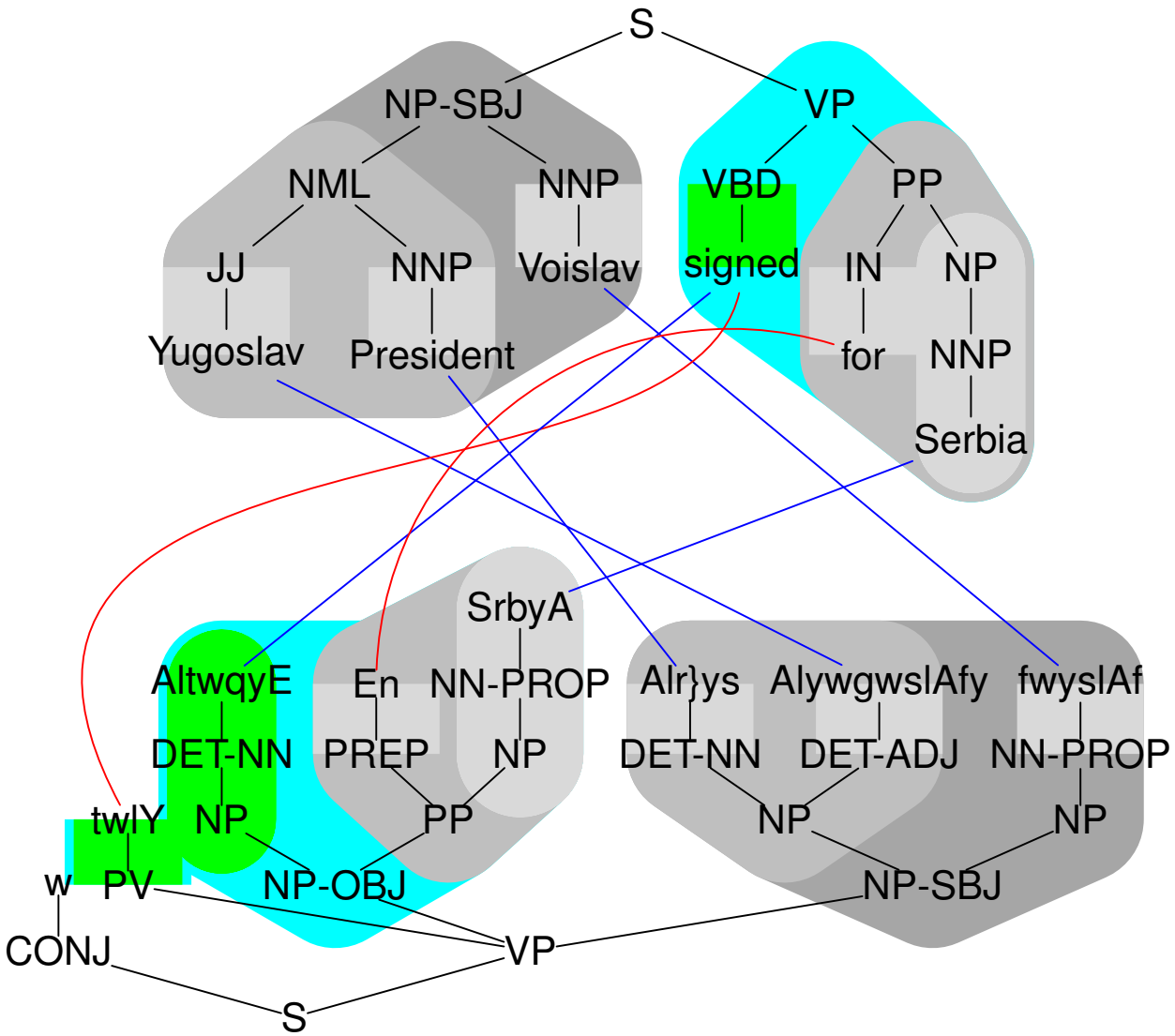
Rule extraction



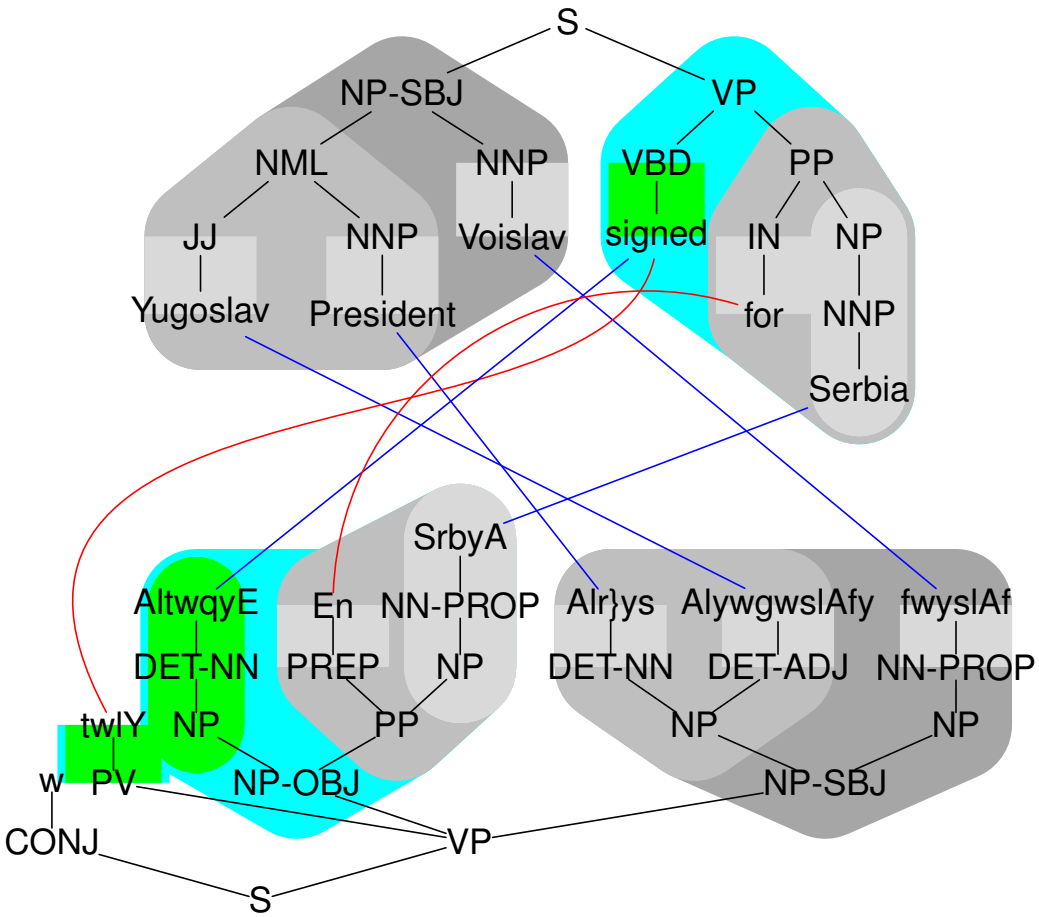
Rule extraction

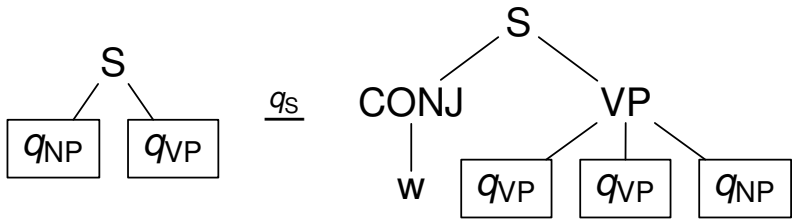


Rule extraction

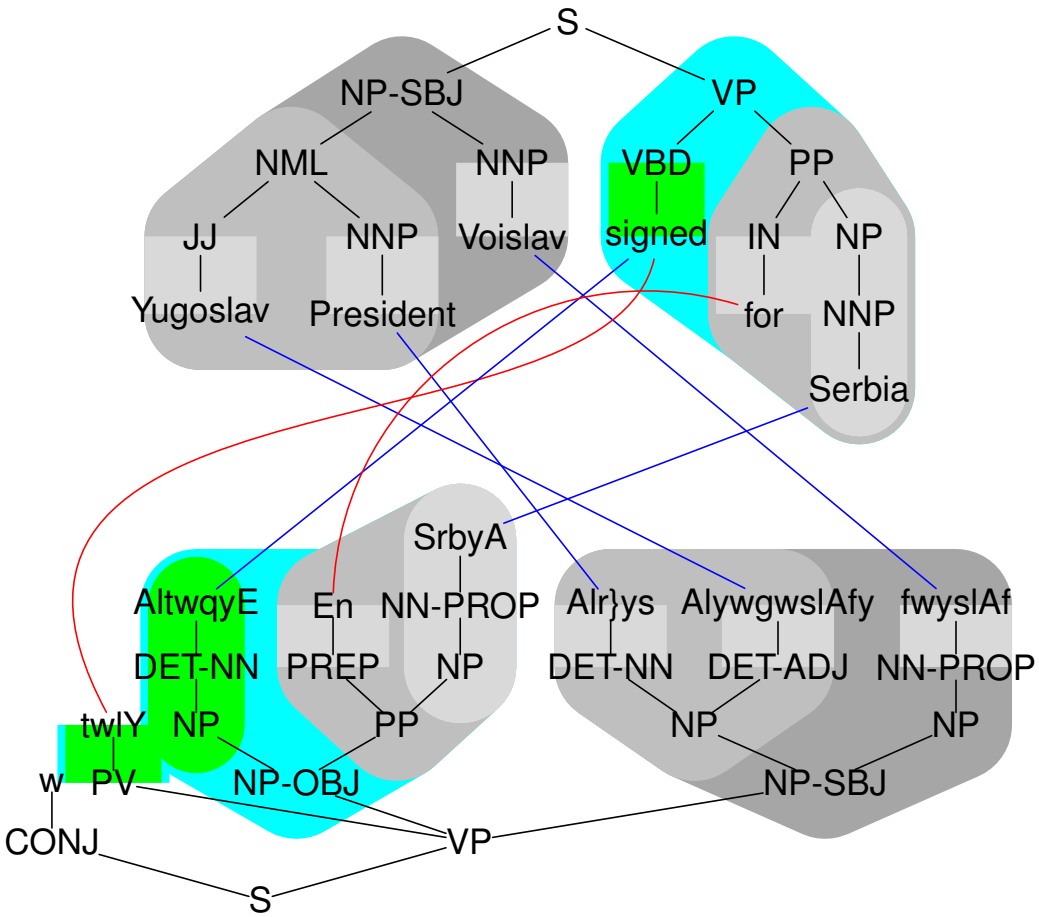


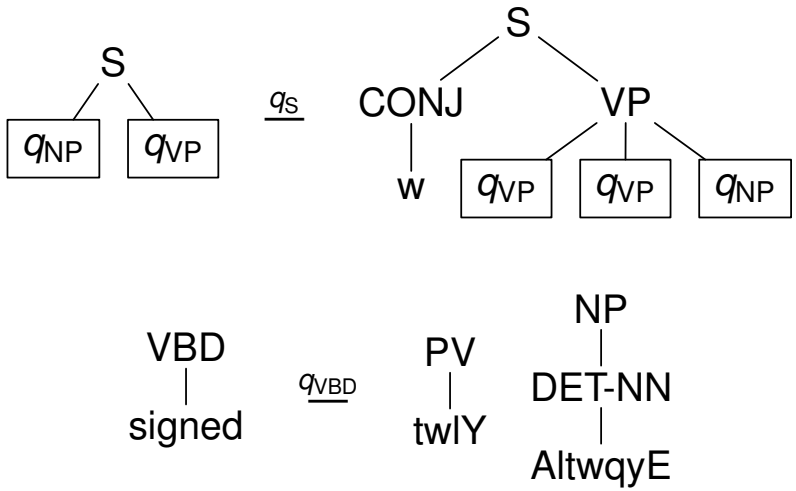
Rule extraction





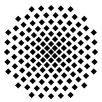
Rule extraction





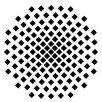
Theoretical summary

- ✓ generalize XTOP
- ✓ discontinuities
- ✓ binarizable
- ✓ efficient input/output restriction
- ✓ efficiently trainable
- ✓ closed under composition



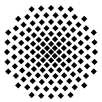
Theoretical summary

- ✓ generalize XTOP
- ✓ discontinuities
- ✓ binarizable
- ✓ efficient input/output restriction
- ✓ efficiently trainable
- ✓ closed under composition
- ✓ preserve regularity backward



Theoretical summary

- ✓ generalize XTOP
- ✓ discontinuities
- ✓ binarizable
- ✓ efficient input/output restriction
- ✓ efficiently trainable
- ✓ closed under composition
- ✓ preserve regularity backward
- ✗ preserve regularity forward
- ✗ symmetric



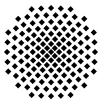
MBOT in machine translation

Moses [KOEHN et al. '07]

- framework for statistical MT
- implementations for many standard tasks
  (alignment, lexical scores, language model, BLEU scoring)
- supports syntax-based MT

We added

- MBOT rule support
- MBOT chart decoder
- adjusted language model calls



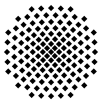
MBOT in machine translation

Moses [KOEHN et al. '07]

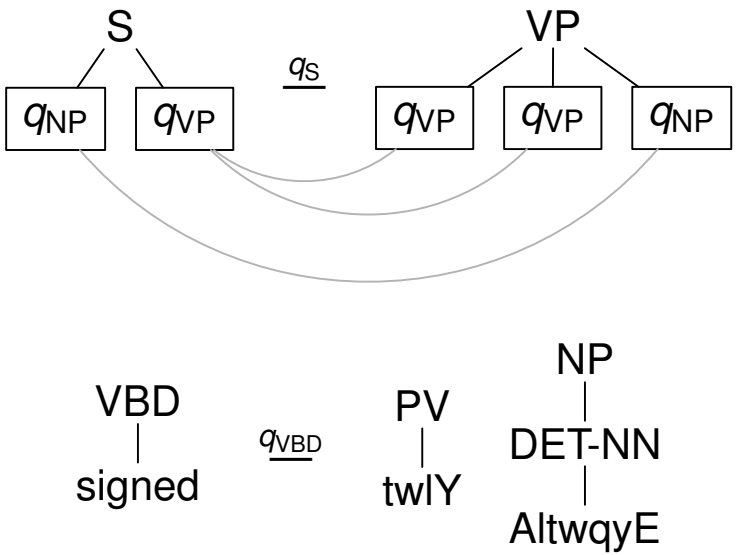
- framework for statistical MT
- implementations for many standard tasks (alignment, lexical scores, language model, BLEU scoring)
- supports syntax-based MT

We added

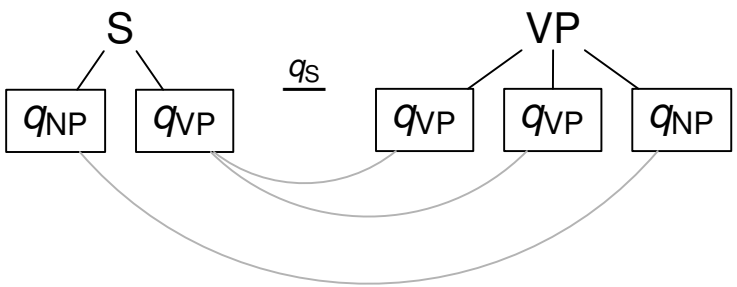
- MBOT rule support
- MBOT chart decoder
- adjusted language model calls

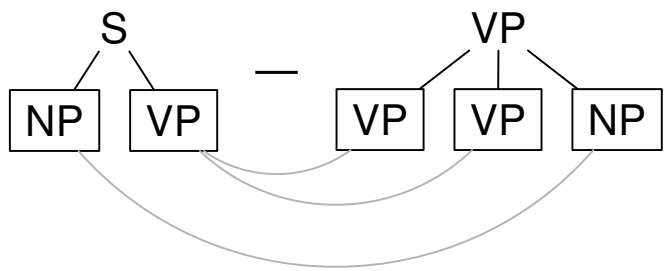


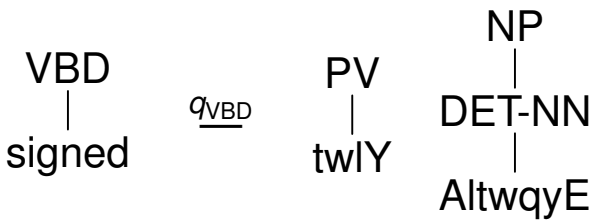
MBOT rule encoding

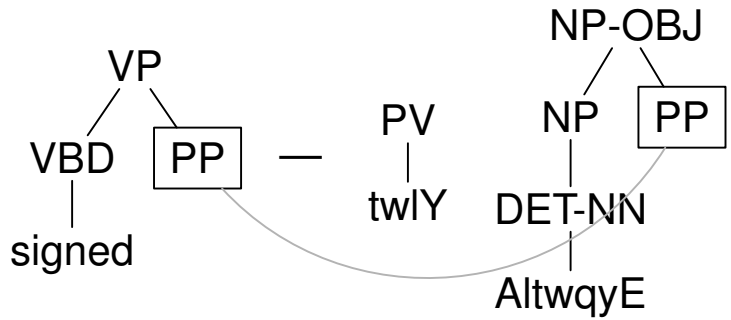


MBOT rule encoding

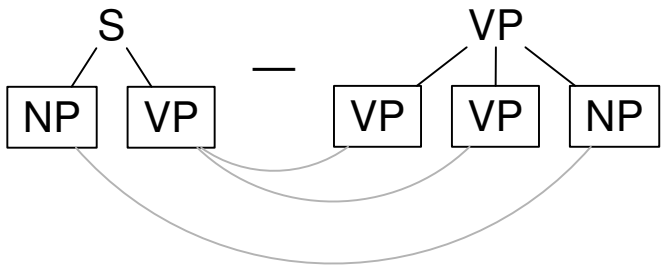






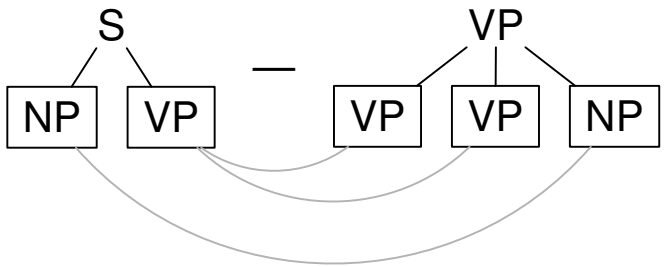


MBOT rule encoding

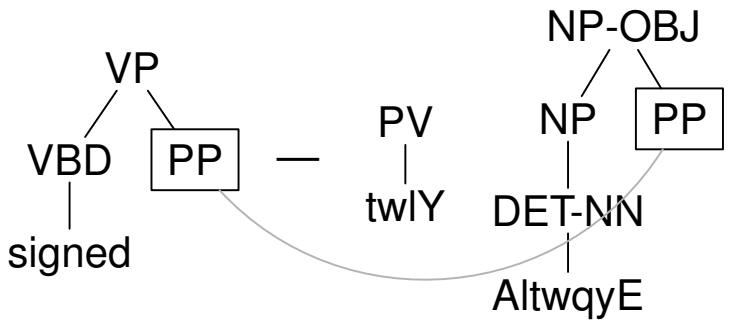



```
S(NP,VP) ||| VP(VP,VP,NP) ||| S ||| VP ||| 0-2 1-0 1-1 ||| ...
```

MBOT rule encoding




```
S(NP,VP) ||| VP(VP,VP,NP) ||| S ||| VP ||| 0-2 1-0 1-1 ||| ...
```




```
VP(VBD(signed),PP) ||| PV(twlY) || NP-OBJ(NP(DET-NN(AltwqyE)),PP) |||

VP ||| PV NP-OBJ ||| || 0-0 ||| ...
```

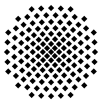
MBOT decoder

Fabienne Braune

- CYK-like chart parser
- only forward application
- supports all standard features
- integrated cube pruning with language model

Notes

- fast
- generated the examples in Motivation
- described in the ACL article you read



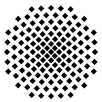
MBOT decoder

Fabienne Braune

- CYK-like chart parser
- only forward application
- supports all standard features
- integrated cube pruning with language model

Notes

- fast
- generated the examples in Motivation
- described in the ACL article you read



MBOT external tools

NINA SEEMANN

- rule extraction
- input/output restriction
- EM training
- conversion tools, pipeline scripts, . . .

Notes

- in PYTHON (not inside MOSES)
- computationally quite expensive
- variants for reduced POS-tags

MBOT external tools

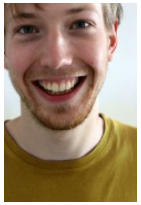
NINA SEEMANN

- rule extraction
- input/output restriction
- EM training
- conversion tools, pipeline scripts, . . .

Notes

- in PYTHON (not inside MOSES)
- computationally quite expensive
- variants for reduced POS-tags

Exact MBOT decoder



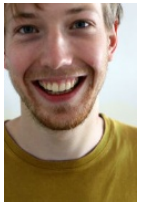
DANIEL QUERNHEIM

- performs no cheats (no pruning, no flattening, etc.)
- utilizes parse forests
- uses theoretical properties (1-symbol normal form)
- language model scoring after exact decoding

Notes

- in PYTHON (not inside MOSES)
- computationally expensive → slow
- but can handle data of WMT 2013

Exact MBOT decoder



DANIEL QUERNHEIM

- performs no cheats (no pruning, no flattening, etc.)
- utilizes parse forests
- uses theoretical properties (1-symbol normal form)
- language model scoring after exact decoding

Notes

- in PYTHON (not inside MOSES)
- computationally expensive $\rightarrow$ slow
- but can handle data of WMT 2013

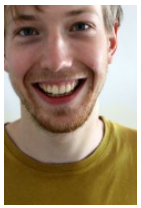
Posters and demos

Nina Seemann

- Poster: MBOT implementation in Moses

Fabienne Braune

- Demo: MBOT implementation in Moses



Daniel Quernheim

- Demo: Exact MBOT decoding