Extended Tree Transducers in Natural Language Processing

Andreas Maletti

Institute for Natural Language Processing
Universität Stuttgart

Grenoble — May 28, 2015

Machine Translation

Original

Übersetzung (GOOGLE TRANSLATE)

- The addressees of this paper are students and students will be in the audience are.

Machine Translation

Original

- Die Adressaten dieses Vortrags sind Studierende und im Publikum werden sich Studierende befinden.

  (The addressees of this talk are students, and students will be in the audience.)

Übersetzung (Google Translate)

- The addressees of this paper are students and students will be in the audience are.

Machine Translation

Original

- Die Adressaten dieses Vortrags sind Studierende und im Publikum werden sich Studierende befinden.

  (The addressees of this talk are students, and students will be in the audience.)

Übersetzung (Google Translate)

- The addressees of this paper are students and students will be in the audience are.
- To scientific lecture, a public discussion follows on.

Machine Translation

Original

- Die Adressaten dieses Vortrags sind Studierende und im Publikum werden sich Studierende befinden.

  (The addressees of this talk are students, and students will be in the audience.)

- An den wissenschaftlichen Vortrag schließt sich eine öffentliche Diskussion an.

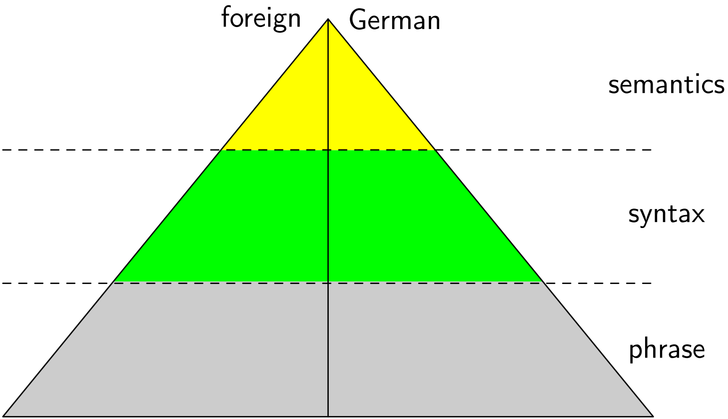
  (The scientific lecture is followed by a public discussion.)

Übersetzung (Google Translate)

- The addressees of this paper are students and students will be in the audience are.
- To scientific lecture, a public discussion follows on.

Machine Translation

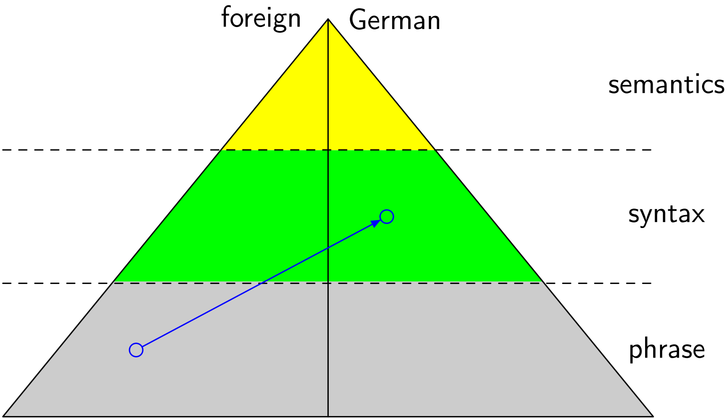
VAUQUOIS triangle:



Translation model:

Machine Translation

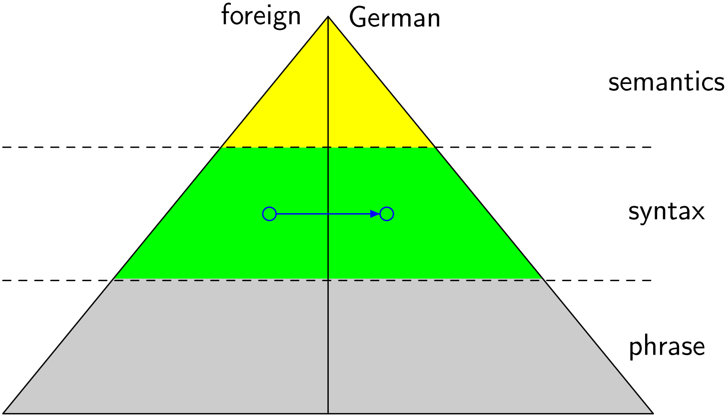
VAUQUOIS triangle:



Translation model: string-to-tree

Machine Translation

VAUQUOIS triangle:



Translation model: tree-to-tree

Machine Translation

Training data

- parallel corpus
- word alignments
- parse trees for the target sentences

Machine Translation

Training data

- parallel corpus
- word alignments
- parse trees for the target sentences

Parallel Corpus

linguistic resource containing example translations

(sentence level)

Machine Translation

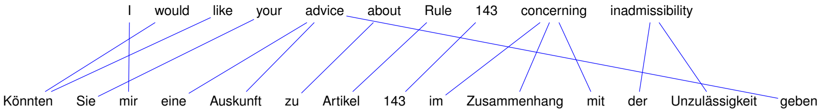
parallel corpus, word alignments, parse tree

I would like your advice about Rule 143 concerning inadmissibility

Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben

Machine Translation

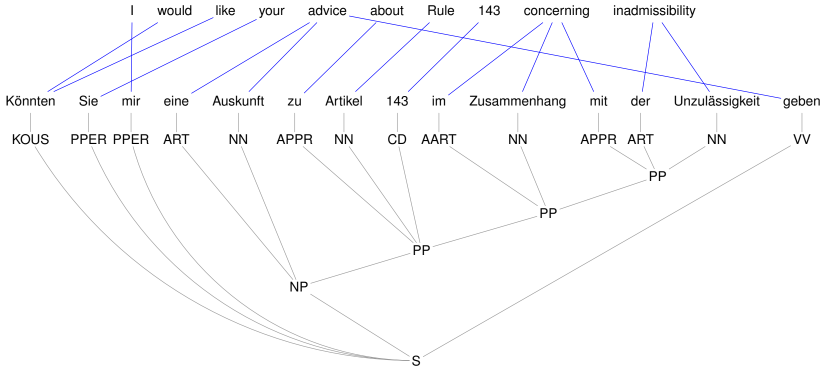
parallel corpus, word alignments, parse tree



via GIZA++ [OCH, NEY, 2003]

Machine Translation

parallel corpus, word alignments, parse tree



via BERKELEY parser [PETROV et al., 2006]

Extended Tree Transducer

Extended top-down tree transducer (STSG)

- variant of [M., Graehl, Hopkins, Knight, 2009]
- rules of the form NT $\rightarrow (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule

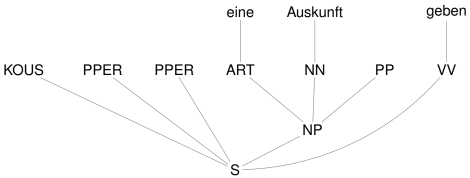
Extended Tree Transducer

Extended top-down tree transducer (STSG)

- variant of [M., GRAEHL, HOPKINS, KNIGHT, 2009]
- rules of the form NT $\rightarrow (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule

PPER KOUS PPER advice PP

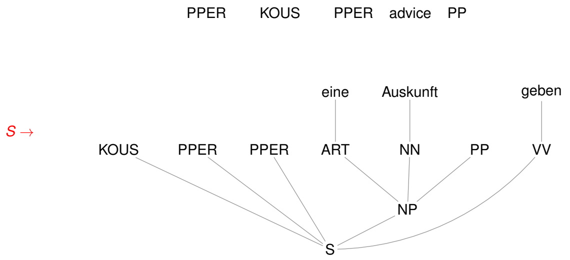
$S \rightarrow$



Extended Tree Transducer

Extended top-down tree transducer (STSG)

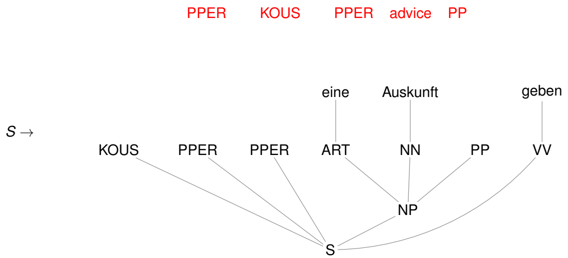
- variant of [M., GRAEHL, HOPKINS, KNIGHT, 2009]
- rules of the form NT $\to (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule



Extended Tree Transducer

Extended top-down tree transducer (STSG)

- variant of [M., GRAEHL, HOPKINS, KNIGHT, 2009]
- rules of the form NT $\rightarrow (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule



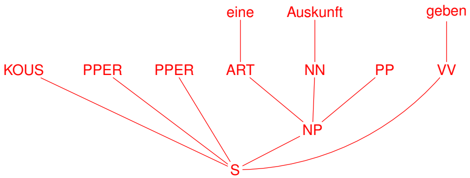
Extended Tree Transducer

Extended top-down tree transducer (STSG)

- variant of [M., GRAEHL, HOPKINS, KNIGHT, 2009]
- rules of the form NT $\rightarrow (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule

PPER KOUS PPER advice PP

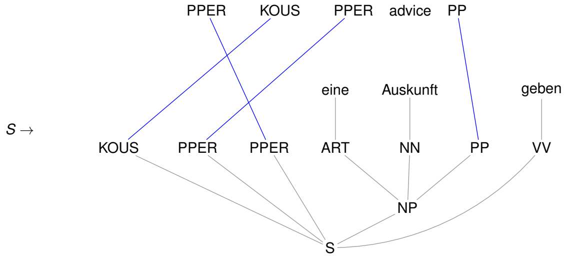
$S \rightarrow$



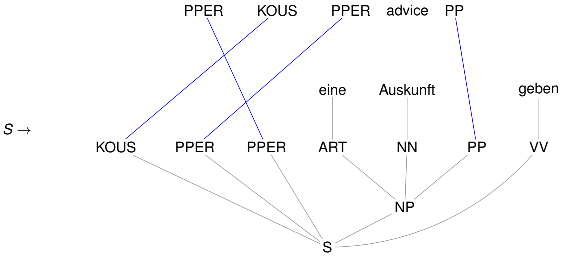
Extended Tree Transducer

Extended top-down tree transducer (STSG)

- variant of [M., GRAEHL, HOPKINS, KNIGHT, 2009]
- rules of the form NT $\rightarrow (r, r_1)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand side r_1 of regular tree grammar rule
- (bijective) synchronization of nonterminals



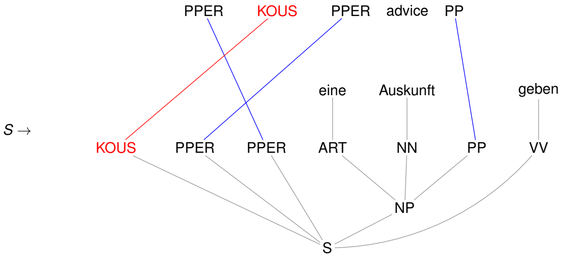
Extended Tree Transducer



Rule application

1. Selection of synchronous nonterminals

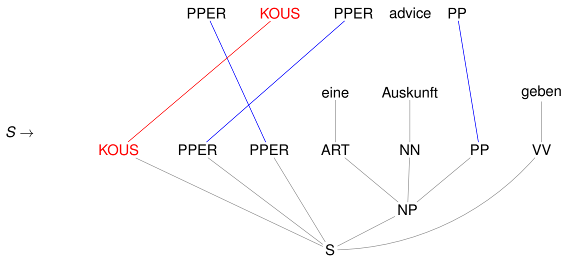
Extended Tree Transducer



Rule application

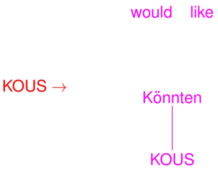
1. Selection of synchronous nonterminals

Extended Tree Transducer

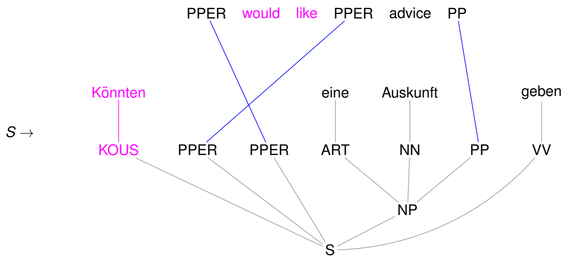


Rule application

1. Selection of synchronous nonterminals
2. Selection of suitable rule

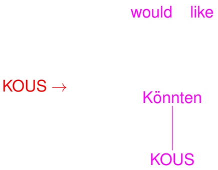


Extended Tree Transducer

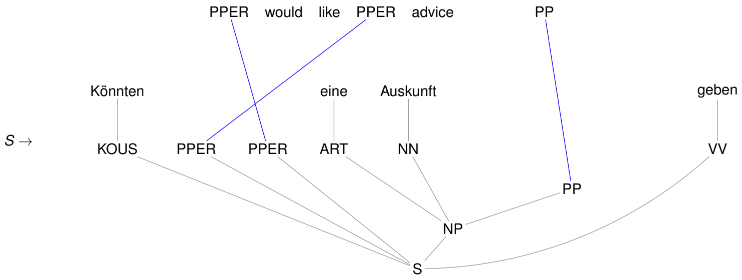


Rule application

1. Selection of synchronous nonterminals
2. Selection of suitable rule
3. Replacement on both sides



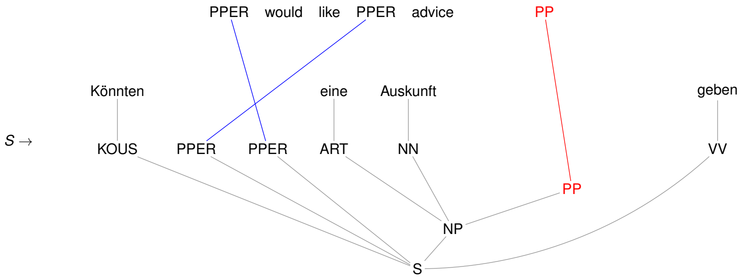
Extended Tree Transducer



Rule application

1. synchronous nonterminals

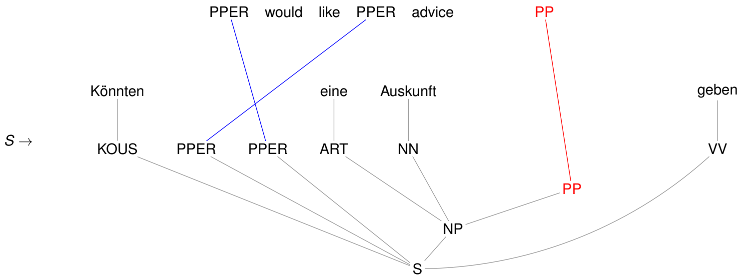
Extended Tree Transducer



Rule application

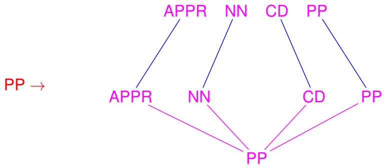
1. synchronous nonterminals

Extended Tree Transducer

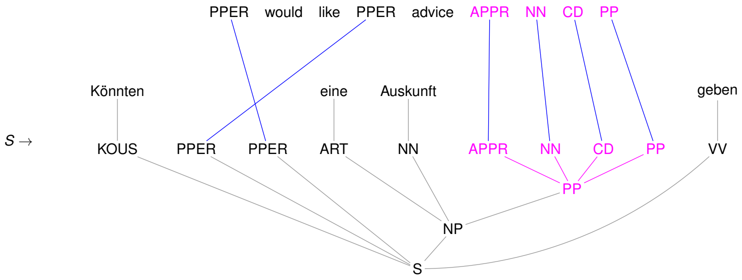


Rule application

1. synchronous nonterminals
2. suitable rule

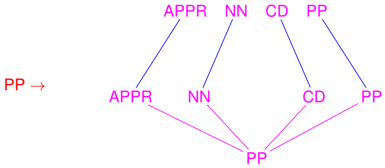


Extended Tree Transducer



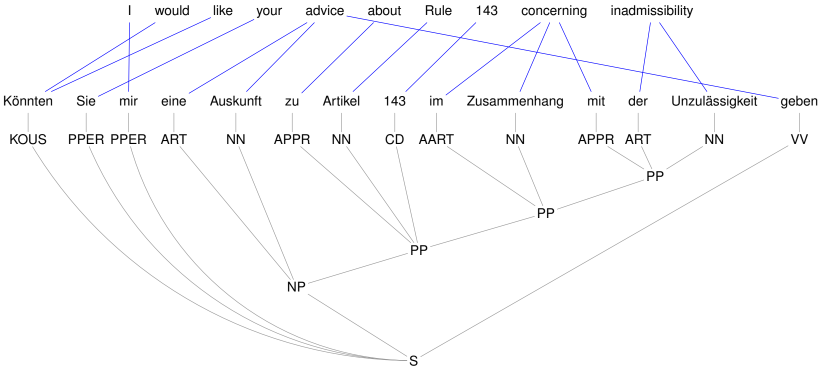
Rule application

1. synchronous nonterminals
2. suitable rule
3. replacement



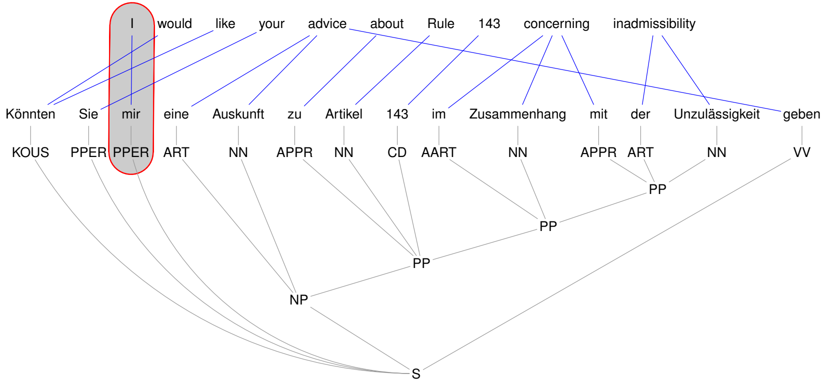
Rule extraction

following [Galley, Hopkins, Knight, Marcu, 2004]



Rule extraction

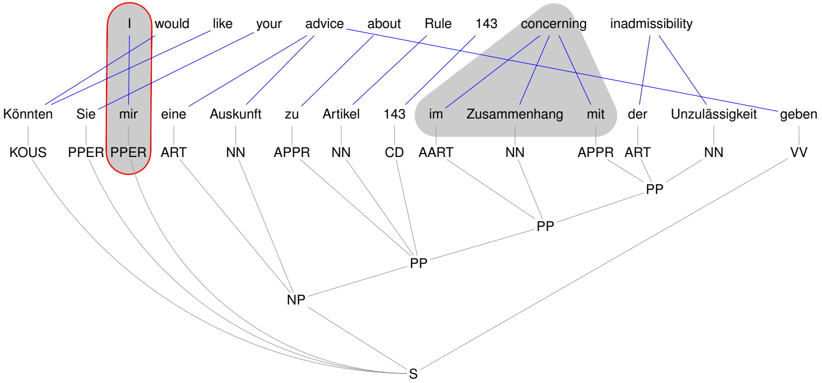
following [Galley, Hopkins, Knight, Marcu, 2004]



extractable rules marked in red

Rule extraction

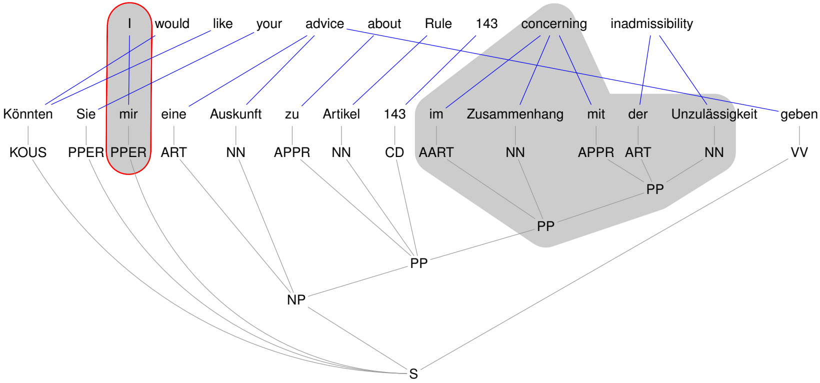
following [GALLEY, HOPKINS, KNIGHT, MARCU, 2004]



extractable rules marked in red

Rule extraction

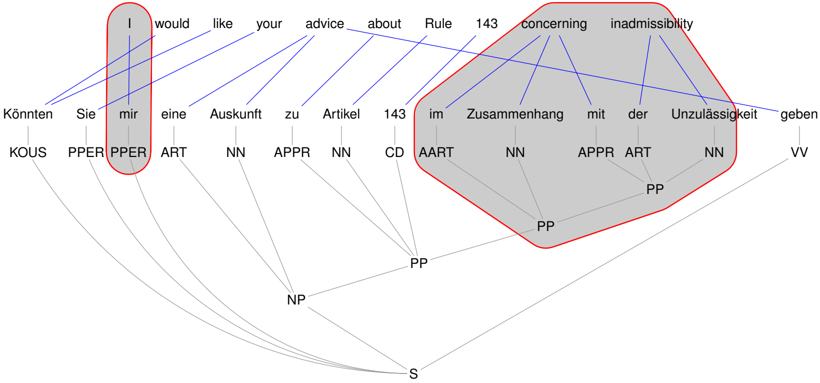
following [Galley, Hopkins, Knight, Marcu, 2004]



extractable rules marked in red

Rule extraction

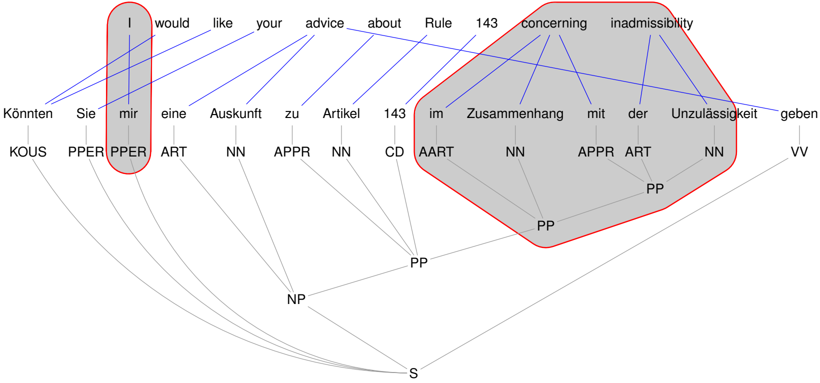
following [Galley, Hopkins, Knight, Marcu, 2004]



extractable rules marked in red

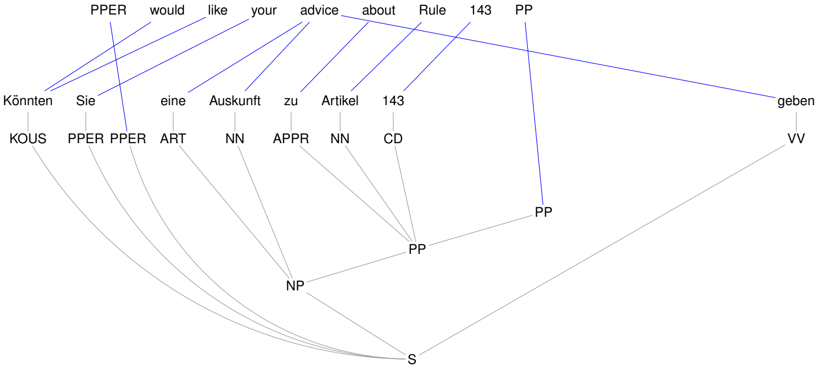
Rule extraction

Removal of extractable rule:



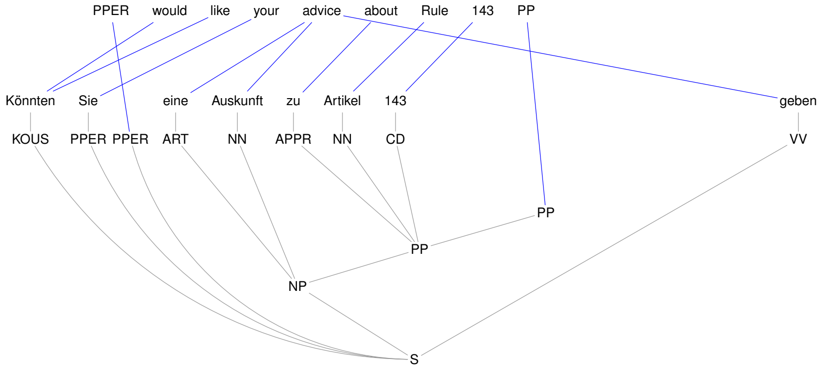
Rule extraction

Removal of extractable rule:



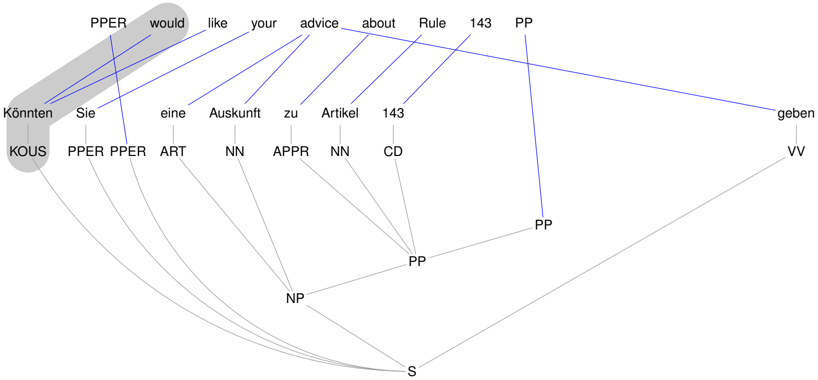
Rule extraction

Repeated rule extraction:



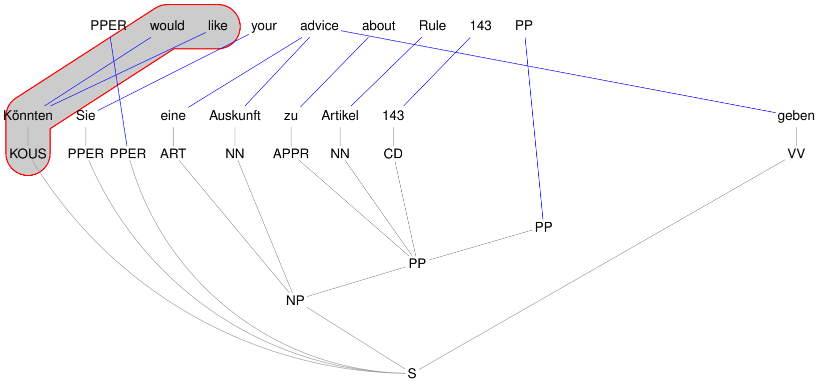
Rule extraction

Repeated rule extraction:



Rule extraction

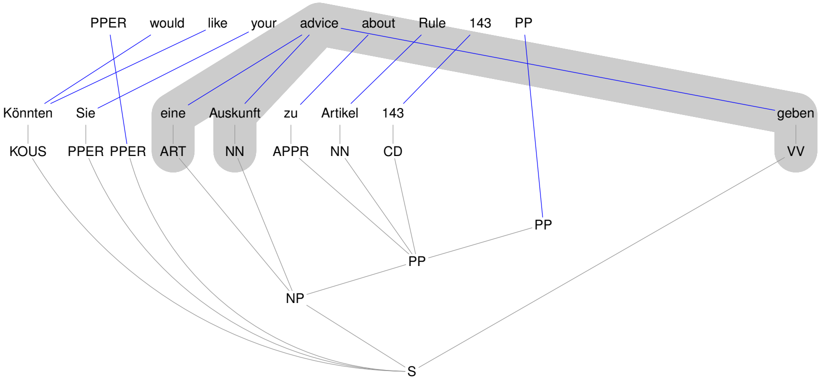
Repeated rule extraction:



extractable rules marked in red

Rule extraction

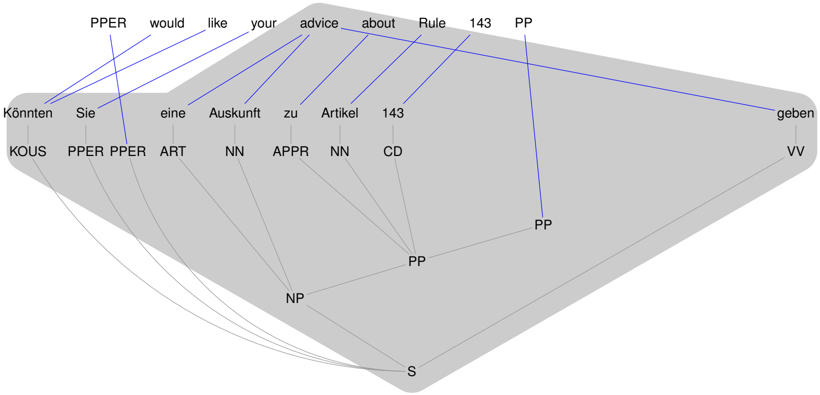
Repeated rule extraction:



extractable rules marked in red

Rule extraction

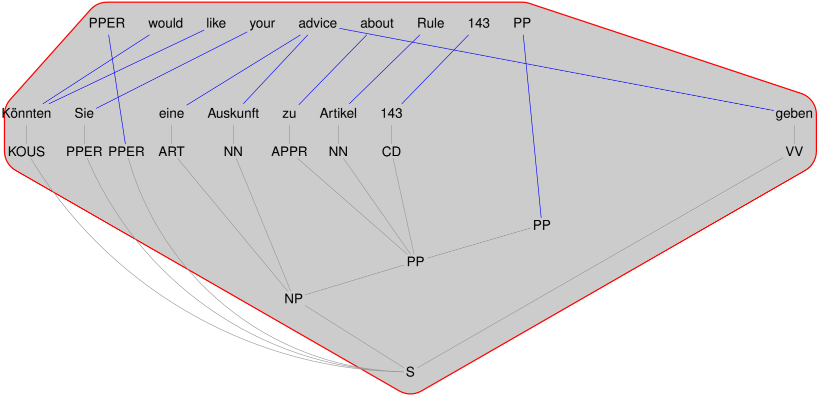
Repeated rule extraction:



extractable rules marked in red

Rule extraction

Repeated rule extraction:



extractable rules marked in red

Extended Tree Transducer

Advantages

- very simple
- implemented in MOSES [KOEHN et al., 2007]
- “context-free”

Extended Tree Transducer

Advantages

- very simple
- implemented in MOSES [KOEHN et al., 2007]
- “context-free”

Disadvantages

- problems with discontinuities
- composition and binarization not possible [M. et al., 2009] and [ZHANG et al., 2006]
- “context-free”

Extended Tree Transducer

Remarks

- synchronization breaks almost all existing constructions
  (e.g., the normalization construction)

$\rightarrow$ the basic grammar model very important

Extended Tree Transducer

Remarks

- synchronization breaks almost all existing constructions
  (e.g., the normalization construction)

→ the basic grammar model very important

- tree-to-tree models use trees on both sides

Extended Tree Transducer

Major (tree-to-tree) models

1. linear top-down tree transducer (with look-ahead)
   - input-side: tree automaton
   - output-side: regular tree grammar
   - synchronization: mapping output NT to input NT

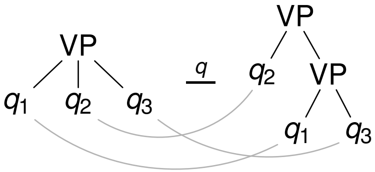
Extended Tree Transducer

Major (tree-to-tree) models

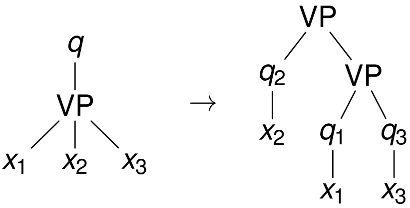
1. linear top-down tree transducer (with look-ahead)
   - input-side: tree automaton
   - output-side: regular tree grammar
   - synchronization: mapping output NT to input NT
2. linear extended top-down tree transducer (w. look-ahead)
   - input-side: regular tree grammar
   - output-side: regular tree grammar
   - synchronization: mapping output NT to input NT

Extended Tree Transducer

Synchronous grammar rule:



"Classical" top-down tree transducer rule:



Extended Tree Transducer

Syntactic restrictions

- nondeleting if synchronization bijective (in all rules)
- strict if r_1 not a nonterminal (for all rules $q \to (r, r_1)$)
- ε-free if r not a nonterminal (for all rules $q \to (r, r_1)$)

Composition (COMP)

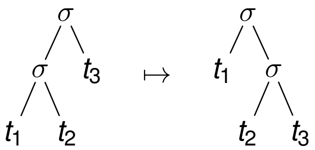
executing transformations $\tau \subseteq T_\Sigma \times T_\Delta$ and $\tau' \subseteq T_\Delta \times T_\Gamma$ one after the other:

$$\tau \,;\, \tau' = \{(s, u) \mid \exists t \in T_\Delta \colon (s, t) \in \tau, (t, u) \in \tau'\}$$

Extended Tree Transducer

Rotations (Rot)

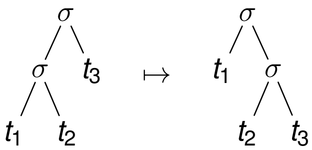
$$\{\langle \sigma(\sigma(t_1, t_2), t_3),\ \sigma(t_1, \sigma(t_2, t_3))\rangle \mid t_1, t_2, t_3 \in T_\Sigma\}$$



Extended Tree Transducer

Rotations (ROT)

$$\{\langle \sigma(\sigma(t_1, t_2), t_3),\ \sigma(t_1, \sigma(t_2, t_3))\rangle \mid t_1, t_2, t_3 \in T_\Sigma\}$$



Preservation of regularity (PRES)

Given $\tau \subseteq T_\Sigma \times T_\Delta$ and $L \subseteq T_\Sigma$ regular, is $\tau(L)$ regular?

$$\tau(L) = \{u \mid \exists t \in L\colon (t, u) \in \tau\}$$

Extended Tree Transducer

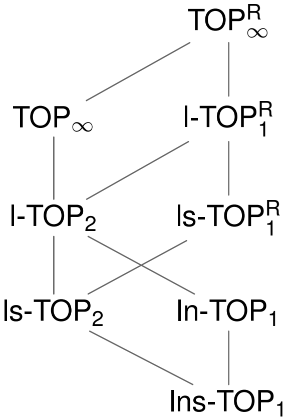
Notation

- (X)TOP = class of tree transformations computable by (extended) top-down tree transducers
- $(\mathrm{X})\mathrm{TOP}^{\mathrm{R}}$ = class of ... transducers with regular look-ahead
- x-$(\mathrm{X})\mathrm{TOP}^{(\mathrm{R})}$ = class of ... transducers with properties x

Example

ln-TOP = class of tree transformations computable by linear and nondeleting top-down tree transducers

Top-down Tree Transducer



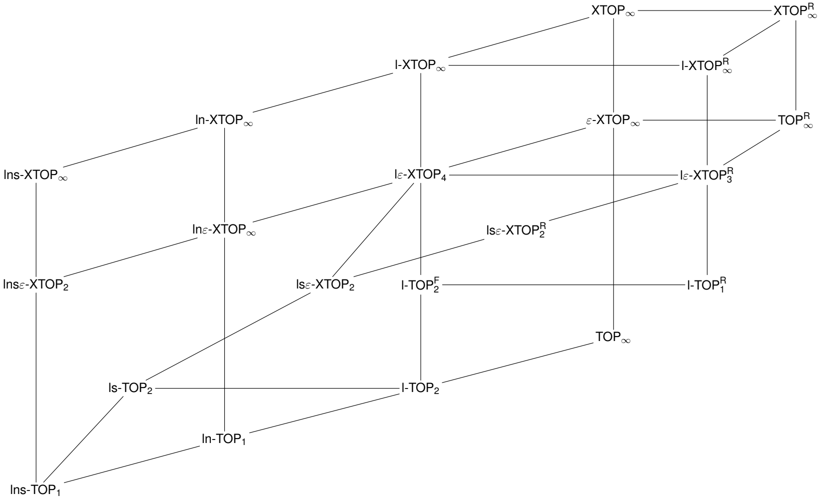
composition closure indicated in subscript

Top-down Tree Transducer

Model \ Criterion	**Rot**	**Sym**	**Pres**	**Pres^{-1}**	**Comp**
ln-TOP	✗	✗	✓	✓	✓
ln-TOP	✗	✗	✓	✓	✓
ls-TOP	✗	✗	✓	✓	✗$_2$
l-TOP	✗	✗	✓	✓	✗$_2$
ls-TOP^R	✗	✗	✓	✓	✓
l-TOP^R	✗	✗	✓	✓	✓
TOP	✓	✗	✗	✓	✗$_\infty$
TOP^R	✓	✗	✗	✓	✗$_\infty$

(Sym = symmetric)

Extended Top-down Tree Transducer

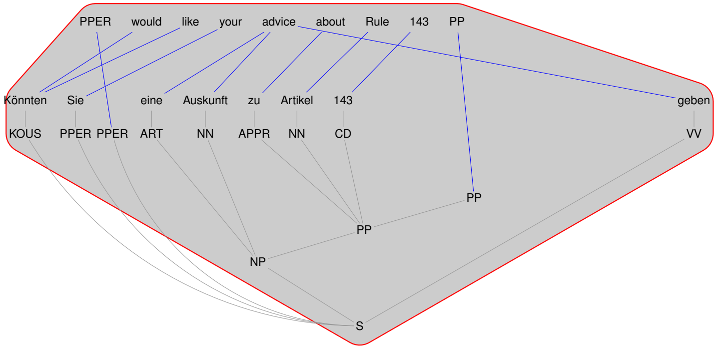


composition closure indicated in subscript

Extended Top-down Tree Transducer

Model \ Criterion	**Rot**	**Sym**	**Pres**	**Pres^{-1}**	**Comp**
ln-TOP	✗	✗	✓	✓	✓
l-TOP	✗	✗	✓	✓	✗$_2$
l-TOP^R	✗	✗	✓	✓	✓
TOP^R	✓	✗	✗	✓	✗$_\infty$
lsε-XTOP	✓	✓	✓	✓	✗$_2$
ls-XTOP	✓	✗	✓	✓	✗$_\infty$
lsε-XTOP$^{(R)}$	✓	✗	✓	✓	✗$_2$
lε-XTOP	✓	✗	✓	✓	✗$_4$
lε-XTOP^R	✓	✗	✓	✓	✗$_3$
(s)l-XTOP$^{(R)}$	✓	✗	✓	✓	✗$_\infty$
XTOP	✓	✗	✗	✓	✗$_\infty$
XTOP^R	✓	✗	✗	✓	✗$_\infty$

Rule extraction



- very specific rule
- every rule for “advice” contains sentence structure
- (syntax “in the way”)

Extended Tree Transducer

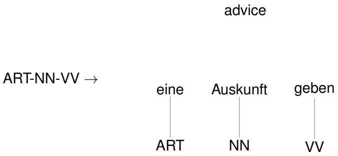
Extended Multi Bottom-up Tree Transducer (MBOT)

- variant of [M., 2010]
- rules of the form NT $\to \big(r, \langle r_1, \dots, r_n \rangle\big)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand sides $r_1, \dots, r_n$ of regular tree grammar rule

Extended Tree Transducer

Extended Multi Bottom-up Tree Transducer (MBOT)

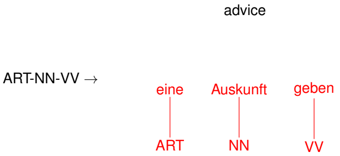
- variant of [M., 2010]
- rules of the form NT $\rightarrow \big(r, \langle r_1, \ldots, r_n \rangle\big)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand sides $r_1, \ldots, r_n$ of regular tree grammar rule



Extended Tree Transducer

Extended Multi Bottom-up Tree Transducer (MBOT)

- variant of [M., 2010]
- rules of the form NT $\rightarrow \big(r, \langle r_1, \ldots, r_n \rangle\big)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand sides $r_1, \ldots, r_n$ of regular tree grammar rule



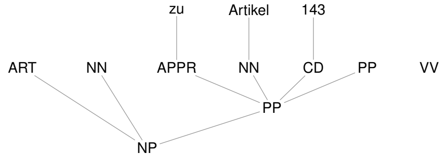
Extended Tree Transducer

Extended Multi Bottom-up Tree Transducer (MBOT)

- variant of [M., 2010]
- rules of the form NT $\rightarrow \big(r, \langle r_1, \ldots, r_n \rangle\big)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand sides $r_1, \ldots, r_n$ of regular tree grammar rule

ART-NN-VV about Rule 143 PP

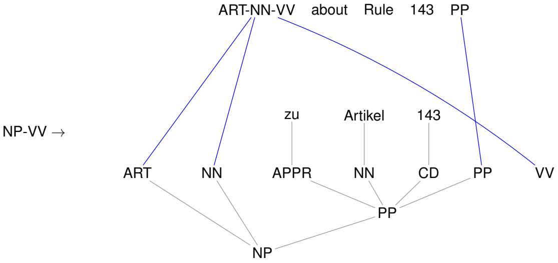
NP-VV $\rightarrow$



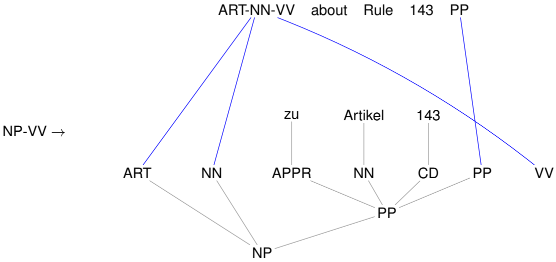
Extended Tree Transducer

Extended Multi Bottom-up Tree Transducer (MBOT)

- variant of [M., 2010]
- rules of the form NT $\rightarrow \big(r, \langle r_1, \ldots, r_n \rangle\big)$
  - nonterminal NT
  - right-hand side r of context-free grammar rule
  - right-hand sides $r_1, \ldots, r_n$ of regular tree grammar rule
- synchronization via map NT $r_1, \ldots, r_n$ to NT r



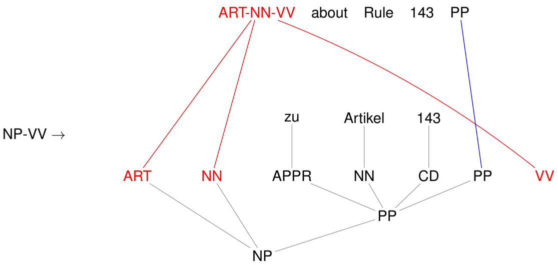
Extended Multi Bottom-up Tree Transducer



Rule application

1. synchronous nonterminals

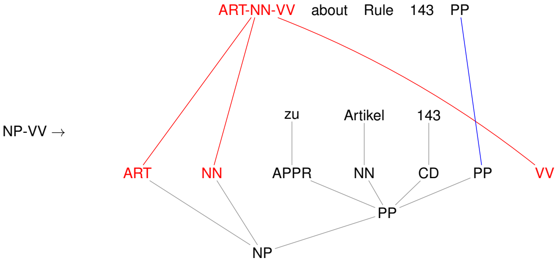
Extended Multi Bottom-up Tree Transducer



Rule application

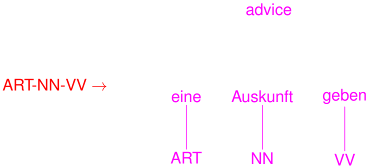
1. synchronous nonterminals

Extended Multi Bottom-up Tree Transducer

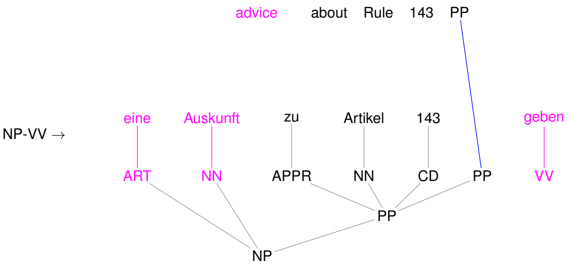


Rule application

1. synchronous nonterminals
2. suitable rule

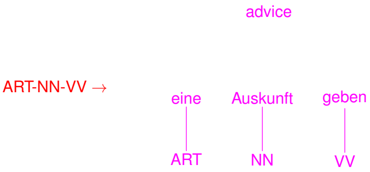


Extended Multi Bottom-up Tree Transducer



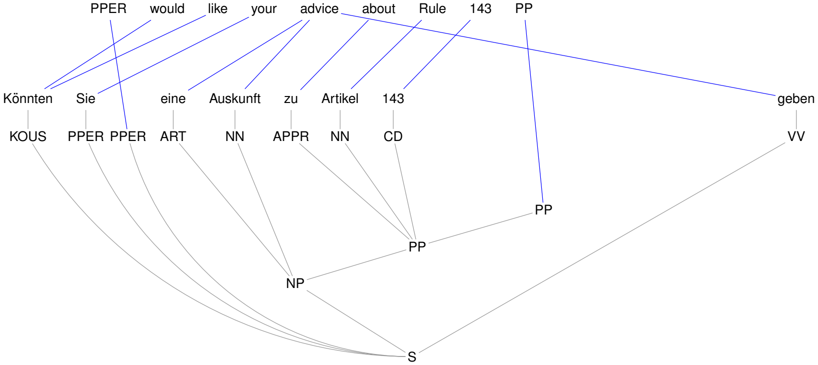
Rule application

1. synchronous nonterminals
2. suitable rule
3. replacement



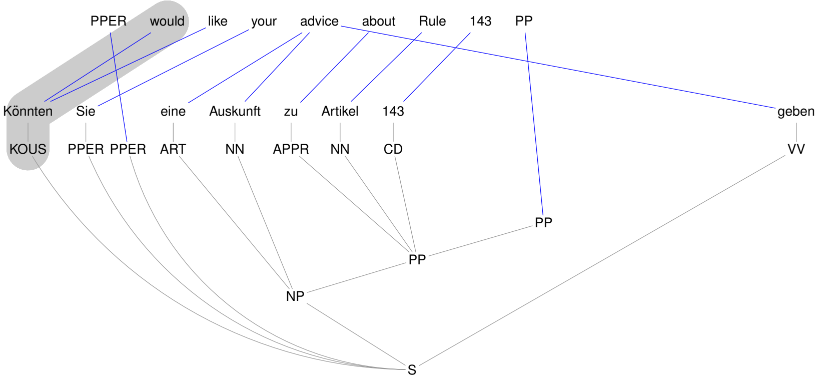
Rule extraction

following [M., 2011]



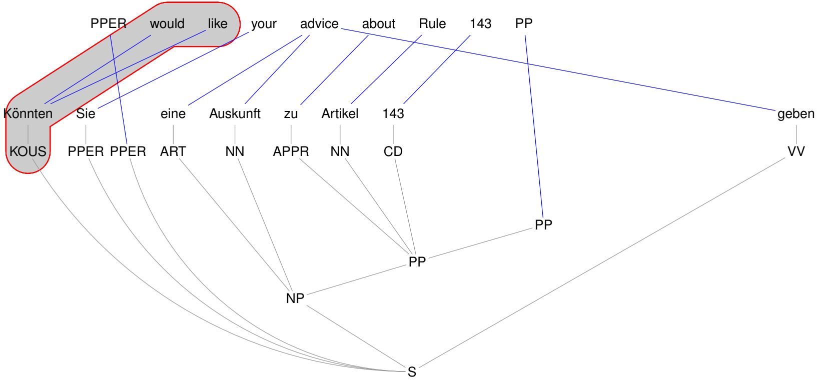
Rule extraction

following [M., 2011]



Rule extraction

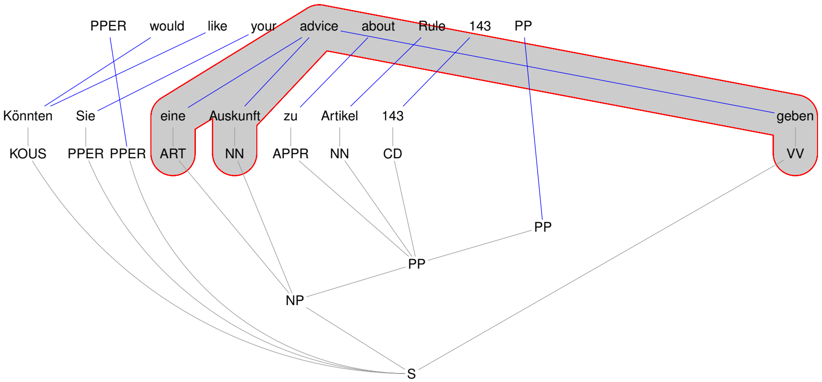
following [M., 2011]



extractable rules marked in red

Rule extraction

following [M., 2011]



extractable rules marked in red

Extended Multi Bottom-up Tree Transducer

- complicated discontinuities
- also available in MOSES [BRAUNE et al., 2013]
- binarizable, composable

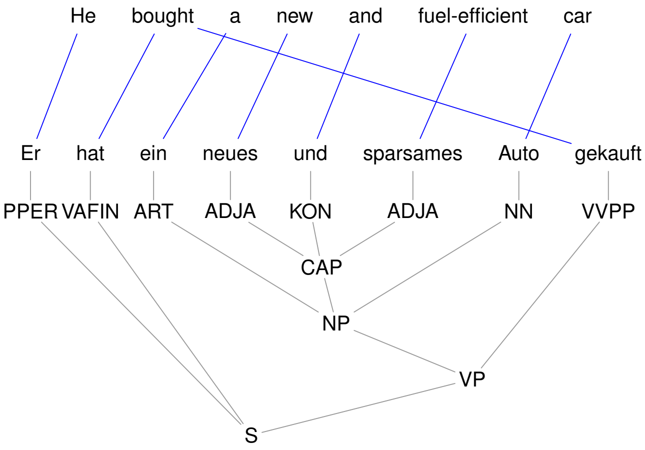
Extended Multi Bottom-up Tree Transducer

- complicated discontinuities
- also available in MOSES [BRAUNE et al., 2013]
- binarizable, composable

Disadvantages

- output not regular (as tree language)
- not symmetric (input context-free; output not)

Discontinuity



Extended Multi Bottom-up Tree Transducer

Theorem [Engelfriet et al., 2009]

$$\text{l-XTOP}^{\text{R}} = \text{l-XBOT}$$

Proof.
Standard construction trading input-deletion for output-deletion
see l-TOP $\subseteq$ l-BOT by [Engelfriet '75] □

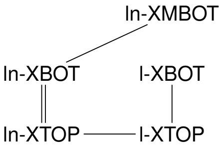
Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l-XTOP}^{\text{R}} = \text{l-XBOT}$$

Proof.

Standard construction trading input-deletion for output-deletion
see l-TOP $\subseteq$ l-BOT by [ENGELFRIET '75] □



Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l-XMBOT} = \text{ln-XMBOT}$$

Proof.

- guess subtrees that will be deleted
- process them using look-ahead □

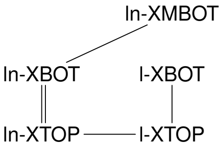
Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l-XMBOT} = \text{ln-XMBOT}$$

Proof.

- guess subtrees that will be deleted
- process them using look-ahead □



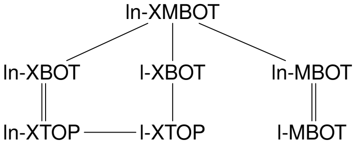
Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l-XMBOT} = \text{ln-XMBOT}$$

Proof.

- guess subtrees that will be deleted
- process them using look-ahead □



Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l}\varepsilon\text{-XMBOT} = \text{l-MBOT}$$

Proof.

- decompose large left-hand sides using “multi”-states
- attach finite effect of ε-rules □

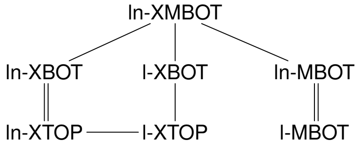
Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l}\varepsilon\text{-XMBOT} = \text{l-MBOT}$$

Proof.

- decompose large left-hand sides using “multi”-states
- attach finite effect of ε-rules □



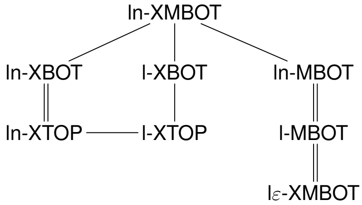
Extended Multi Bottom-up Tree Transducer

Theorem [ENGELFRIET et al., 2009]

$$\text{l}\varepsilon\text{-XMBOT} = \text{l-MBOT}$$

Proof.

- decompose large left-hand sides using “multi”-states
- attach finite effect of ε-rules □



Extended Multi Bottom-up Tree Transducer

Theorem [M., 2014]

$$\text{ln-MBOT} \not\subseteq \left(\text{ln-XTOP}^{\text{R}}\right)^*$$

Extended Multi Bottom-up Tree Transducer

Theorem [M., 2014]

$$\text{ln-MBOT} \not\subseteq (\text{ln-XTOP}^{\text{R}})^*$$

Theorem [Gildea, 2012]

$$\text{yd}_{\text{out}}(\text{ln-MBOT}) = \text{LCFRS}$$

Summary

Model \ Criterion	ROT	SYM	PRES	PRES^{-1}	COMP
ln-TOP	✗	✗	✓	✓	✓
l-TOP	✗	✗	✓	✓	$✗_2$
l-TOP^{R}	✗	✗	✓	✓	✓
TOP^{R}	✓	✗	✗	✓	$✗_\infty$
lnsε-XTOP	✓	✓	✓	✓	$✗_2$
lns-XTOP	✓	✗	✓	✓	$✗_\infty$
lsε-$\text{XTOP}^{\text{(R)}}$	✓	✗	✓	✓	$✗_2$
lε-XTOP	✓	✗	✓	✓	$✗_4$
lε-XTOP^{R}	✓	✗	✓	✓	$✗_3$
(s)l-$\text{XTOP}^{\text{(R)}}$	✓	✗	✓	✓	$✗_\infty$
$\text{XTOP}^{\text{(R)}}$	✓	✗	✗	✓	$✗_\infty$
l(n)-XMBOT	✓	✗	✗	✓	✓
XMBOT	✓	✗	✗	✓	$✗_\infty$
reg.-preserving l-XMBOT	✓	✗	✓	✓	✓
invertable l-XMBOT	✓	✓	✓	✓	✓

Evaluation

Task	System	BLEU
English → German	STSG	15.22
	MBOT	15.90
	phrase-based	16.73
	hierarchical	16.95
	GHKM	17.10
English → Arabic	STSG	48.32
	MBOT	49.10
	phrase-based	50.27
	hierarchical	51.71
	GHKM	46.66
English → Chinese	STSG	17.69
	MBOT	18.35
	phrase-based	18.09
	hierarchical	18.49
	GHKM	18.12

from [Seemann, Braune, M., 2015]

Literature

Selected references

- ARNOLD, DAUCHET: *Morphismes et Bimorphismes d'Arbres*
  Theoret. Comput. Sci. 20, 1982
- ENGELFRIET: *Bottom-up and Top-down Tree Transformations — A Comparison.* Math. Systems Theory 9, 1975
- ENGELFRIET, MANETH: *Macro Tree Translations of Linear Size Increase are MSO Definable.* SIAM J. Comput. 32, 2003
- ENGELFRIET, LILIN, $\sim$: *Extended Multi Bottom-up Tree Transducers — Composition and Decomposition.* Acta Inf. 46, 2009
- ROUNDS: *Mappings and Grammars on Trees*
  Math. Systems Theory 4, 1970
- THATCHER: *Generalized^2 Sequential Machine Maps*
  J. Comput. System Sci. 4, 1970

Current Research

Decoding

- input regular tree language
- extended CYK algorithm for translation
  (parse the input; translation develops)

Current Research

Decoding

- input regular tree language
- extended CYK algorithm for translation
  (parse the input; translation develops)

Observations

- phrase-based system makes no search errors
  [CHANG, COLLINS, 2011]

Current Research

Decoding

- input regular tree language
- extended CYK algorithm for translation
  (parse the input; translation develops)

Observations

- phrase-based system makes no search errors
  [Chang, Collins, 2011]
- STSG and MBOT <u>do</u>
  - heuristics (??? BLEU)
  - exact decoding with syntax forest (+2–3 BLEU)

Current Research

Rule extraction

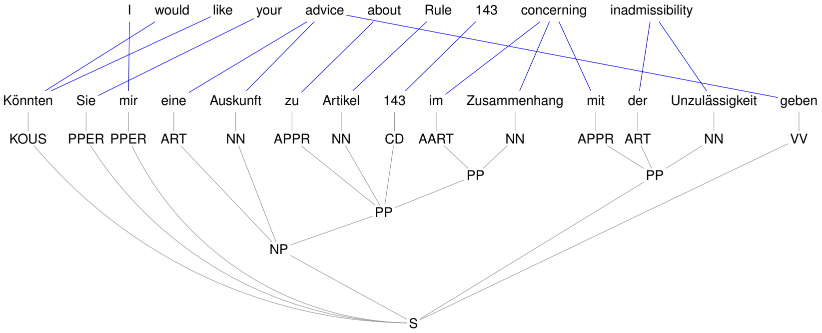
- too many extractable rules
  - which restrictions? [SEEMANN, BRAUNE, M., 2015]
  - efficient representation (maybe symbolic)

Current Research

Rule extraction

- too many extractable rules
  - which restrictions? [SEEMANN, BRAUNE, M., 2015]
  - efficient representation (maybe symbolic)
- only best syntax tree
  - rule extraction with syntax forest (ambitious)

Current Research



Current Research

Rule extraction

- too many extractable rules
  - which restrictions? [SEEMANN, BRAUNE, M., 2015]
  - efficient representation (maybe symbolic)
- only best syntax tree
  - rule extraction with syntax forest (ambitious)

Translation models

- only word-based systems for word alignment
  - efficient restrictions of modern systems
  - unsupervised learning

Current Research

Rule extraction

- too many extractable rules
  - which restrictions? [SEEMANN, BRAUNE, M., 2015]
  - efficient representation (maybe symbolic)
- only best syntax tree
  - rule extraction with syntax forest (ambitious)

Translation models

- only word-based systems for word alignment
  - efficient restrictions of modern systems
  - unsupervised learning
- models for semantics-based translation
  - graph-based models