Expressive Power of Syntax-based Machine Translation Formalisms

Andreas Maletti

Institute for Natural Language Processing
Universität Stuttgart

Stuttgart — December 8, 2015

Machine Translation

Review Translation [by Google Translate]

1. The room it is not narrowly was a simple, bathtub was also attached.
2. Wi-fi, TV and I was available.
3. Church looked When morning awake open the curtain.
4. When looking at often, wives, went out and is invited to try to go [...].
5. But was a little cold, morning walks was good.

Machine Translation

Review Translation [by Google Translate]

1. The room it is not narrowly was a simple, bathtub was also attached.
2. Wi-fi, TV and I was available.
3. Church looked When morning awake open the curtain.
4. When looking at often, wives, went out and is invited to try to go [...].
5. But was a little cold, morning walks was good.

Original [Japanese — © tripadvisor]

1. 部屋もシンプルでしたが狭くなく、バスタブもついていました。
2. Wi-fi、テレビも利用出来ました。
3. 朝起きてカーテンを開けると教会が見えました。
4. しばし眺めていると、妻たちは、[...]るから行こうとさそわれ出かけました。
5. ちょっと寒かったけれど、朝の散策はグッドでしたよ。

Machine Translation

Speech transcript


Okay, what is your name?


Abdul.


And your last name?


Al Farran.

Text taken from [Jones, Shen, Herzog: Machine translation for government applications. *Lincoln Laboratory Journal* 18(1), 2009]
Icons by Rion (Wikipedia) and Ivlichev Viktor Petrovich (`http://www.aha-soft.com/`)

Machine Translation

Speech transcript


Okay, what is your name?


Abdul.


And your last name?


Al Farran.

Speech translation [undisclosed system; most likely phrase-based]


Okay, what's your name?


milk a mechanic and I am here I mean yes

Text taken from [Jones, Shen, Herzog: Machine translation for government applications. *Lincoln Laboratory Journal* 18(1), 2009]
Icons by Rion (Wikipedia) and Ivlichev Viktor Petrovich (`http://www.aha-soft.com/`)

Machine Translation

Speech transcript


Okay, what is your name?


Abdul.


And your last name?


Al Farran.

Speech translation [undisclosed system; most likely phrase-based]


Okay, what's your name?


milk a mechanic and I am here I mean yes


What is your last name?


every two weeks my son's name is ismail

Text taken from [Jones, Shen, Herzog: Machine translation for government applications. *Lincoln Laboratory Journal* 18(1), 2009]
Icons by Rion (Wikipedia) and Ivlichev Viktor Petrovich (`http://www.aha-soft.com/`)

Short History

Timeline

1960 – Dark age

- rule-based systems (e.g., SYSTRAN)
- Chomskyan approach (perfect translation, poor coverage)

1991 – Reformation

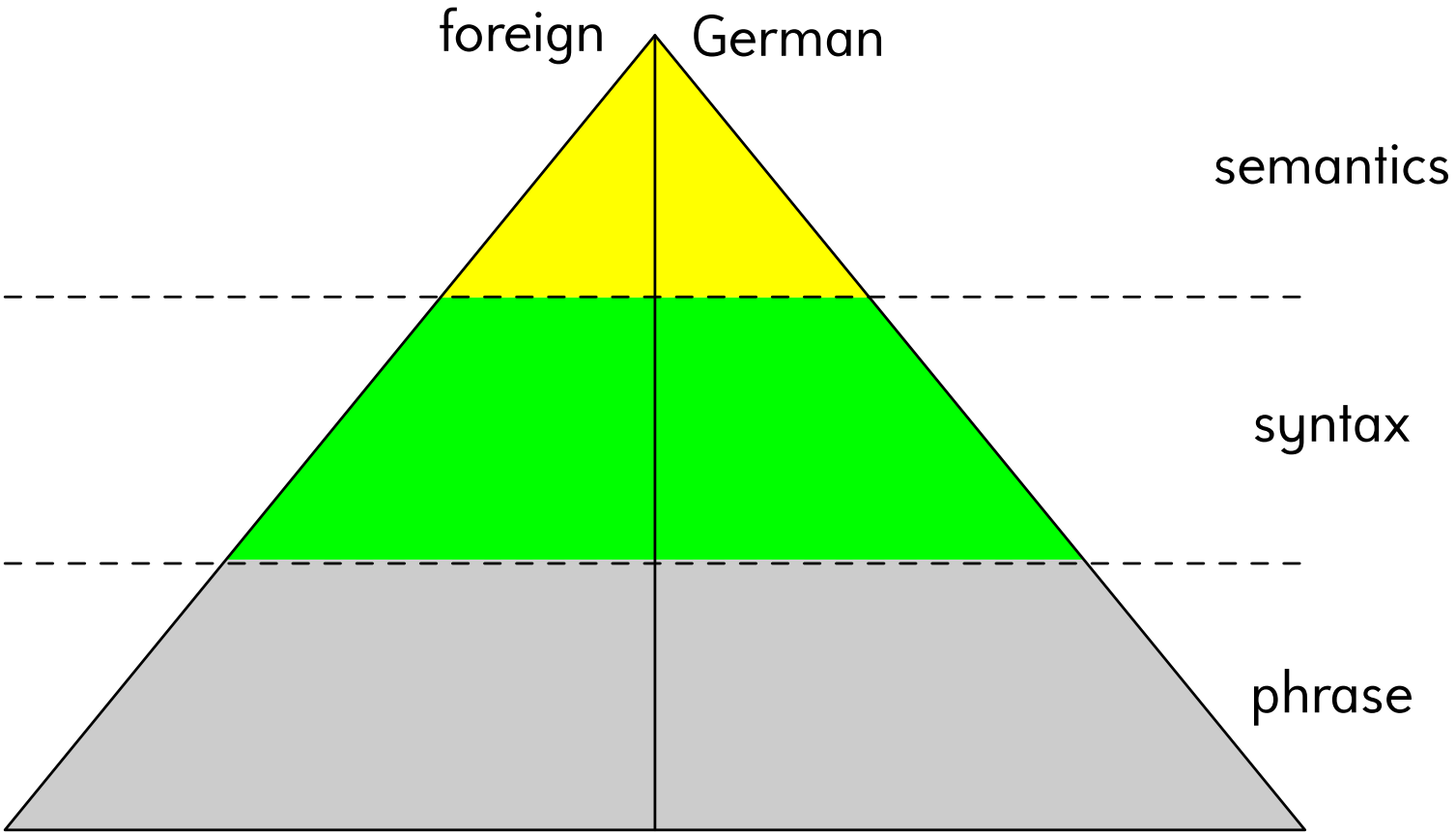
- phrase-based and syntax-based systems
- statistical approach (cheap, automatically trained)

2015 – Potential future

- semantics-based systems (e.g., FrameNet-based)
- semi-supervised, statistical approach
- basic understanding of (translated) text

Machine Translation

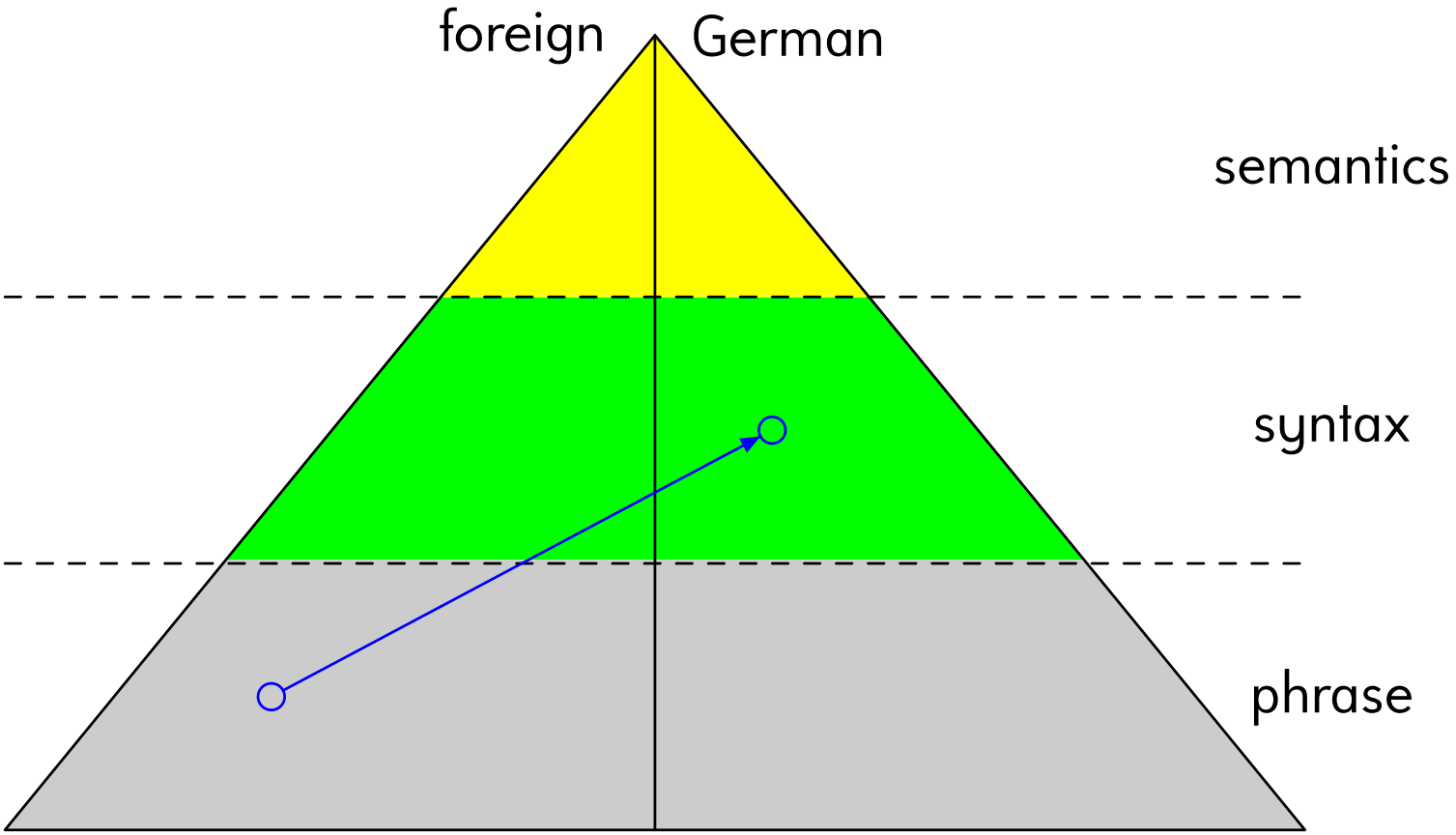
Vauquois triangle:



Translation model:

Machine Translation

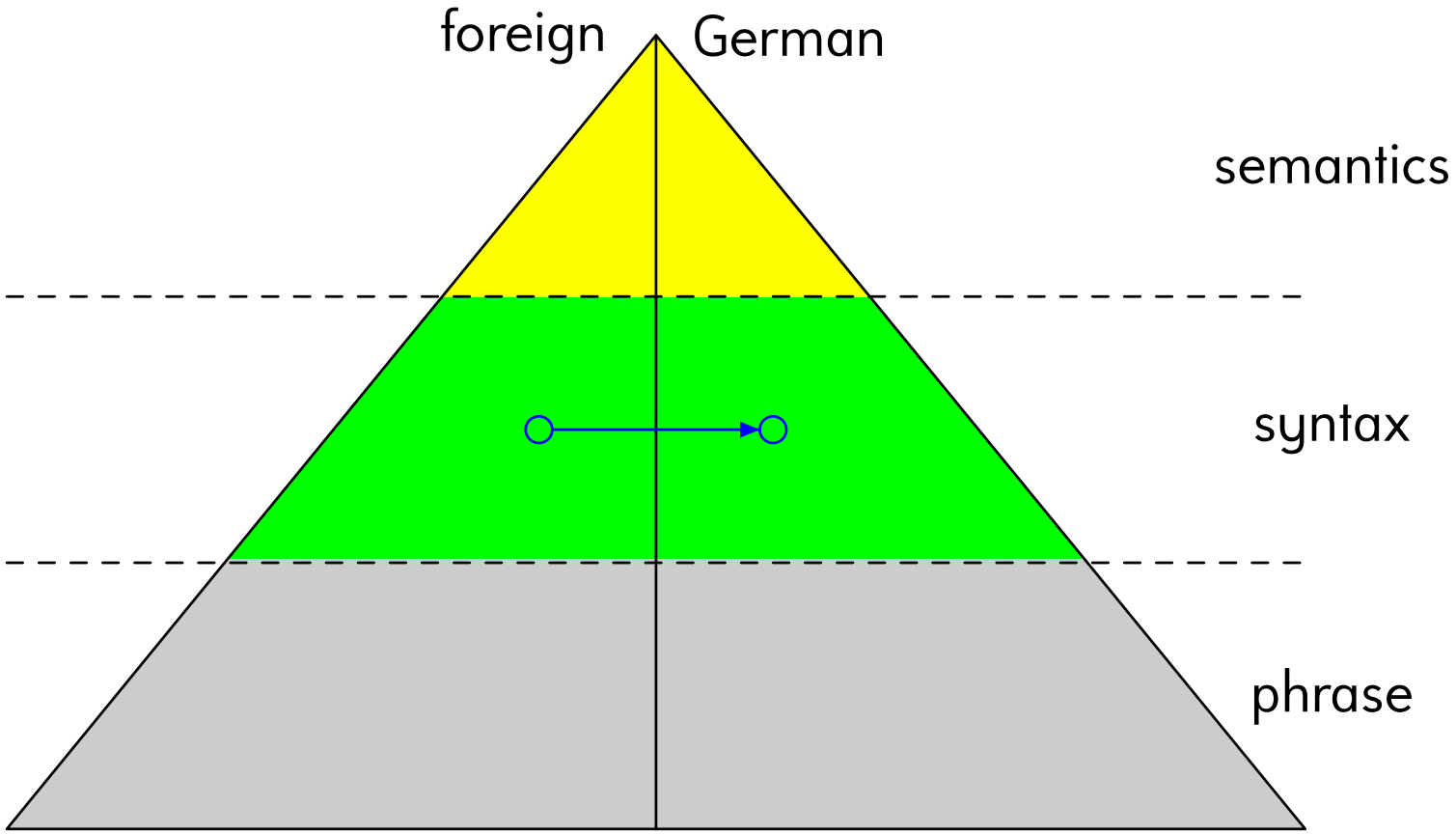
Vauquois triangle:



Translation model: string-to-tree

Machine Translation

Vauquois triangle:



Translation model: tree-to-tree

Machine Translation

Training data

- parallel corpus
- word alignments
- parse trees

Machine Translation

Training data

- parallel corpus
- word alignments
- parse trees

Parallel Corpus

linguistic resource containing (sentence-by-sentence) example translations

Machine Translation

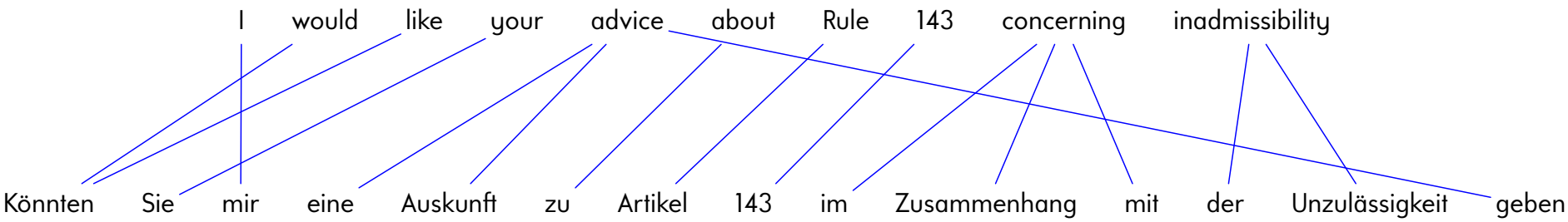
parallel corpus, word alignments, parse tree

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

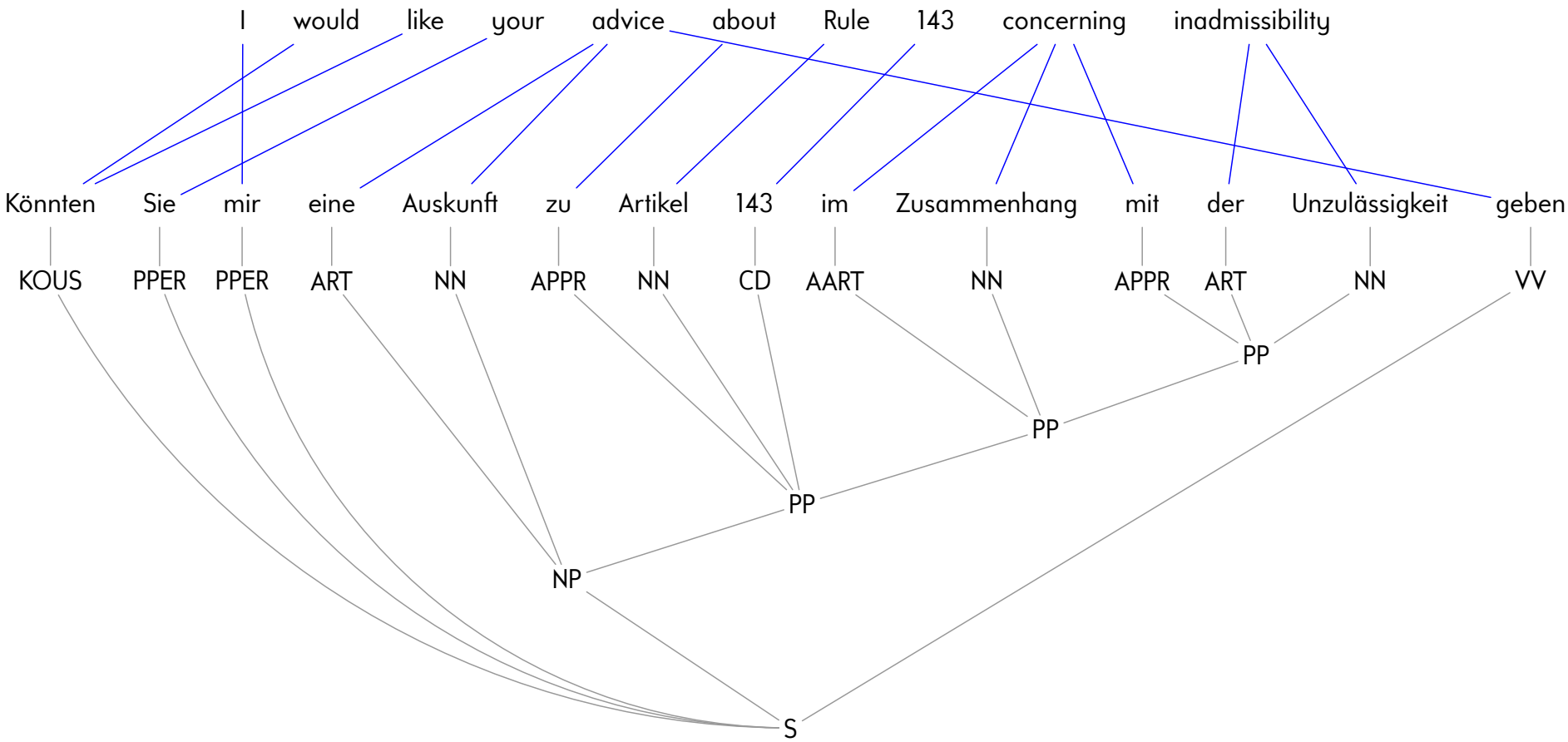
parallel corpus, word alignments, parse tree



via GIZA++ [Och, Ney: A systematic comparison of various statistical alignment models. *Computational Linguistics* 29(1), 2003]

Machine Translation

parallel corpus, word alignments, parse tree

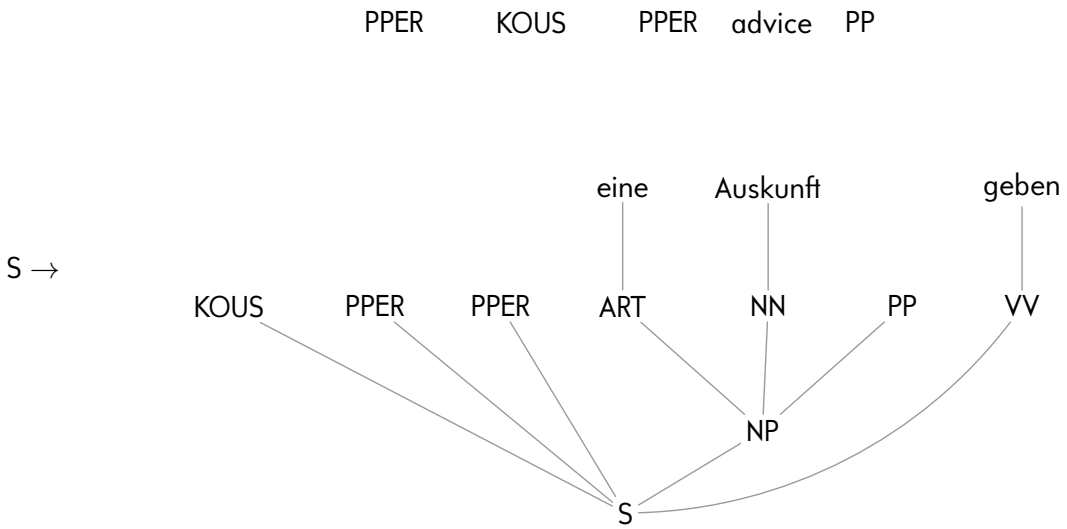


via Berkeley parser [Petrov, Barrett, Thibaux, Klein: Learning accurate, compact, and interpretable tree annotation. *Proc. ACL*, 2006]

Synchronous Grammars

Synchronous tree substitution grammar: productions $N \rightarrow (r, r_1)$

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand side r_1 of tree substitution grammar production

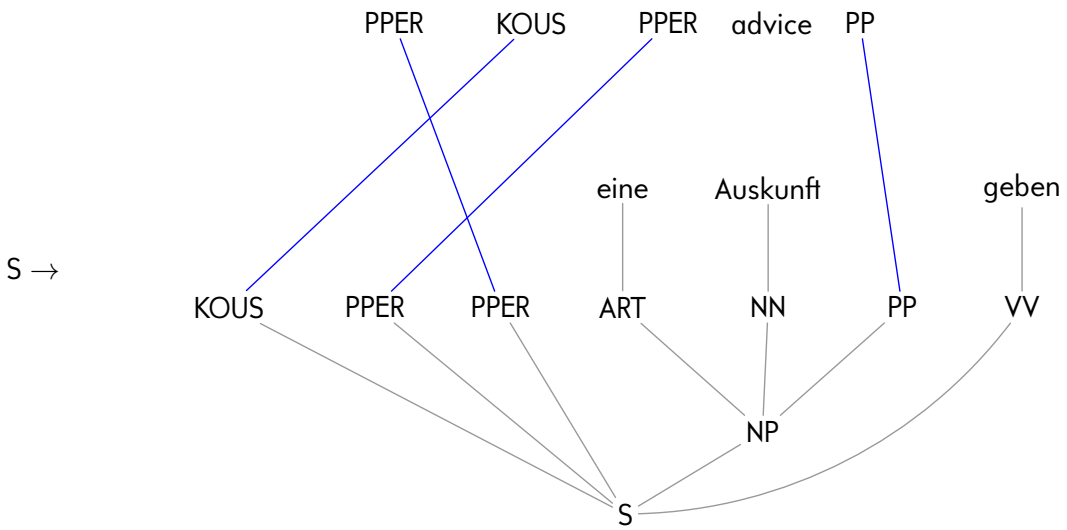


variant of [M., Graehl, Hopkins, Knight: The power of extended top-down tree transducers. *SIAM Journal on Computing* 39(2), 2009]

Synchronous Grammars

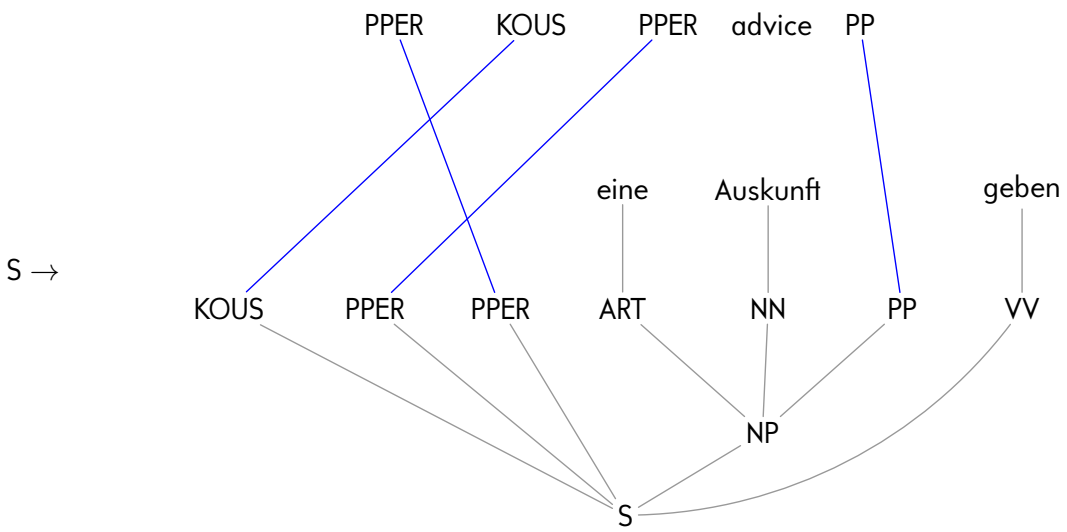
Synchronous tree substitution grammar: productions $N \rightarrow (r, r_1)$

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand side r_1 of tree substitution grammar production
- (bijective) synchronization of nonterminals



variant of [M., Graehl, Hopkins, Knight: The power of extended top-down tree transducers. *SIAM Journal on Computing* 39(2), 2009]

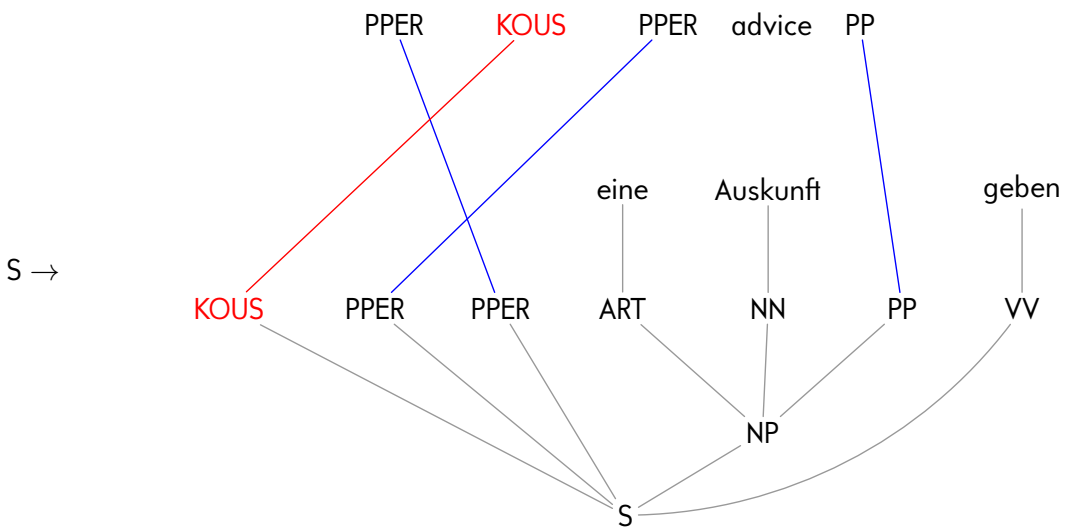
Synchronous Grammars



Production application

1. Selection of synchronous nonterminals

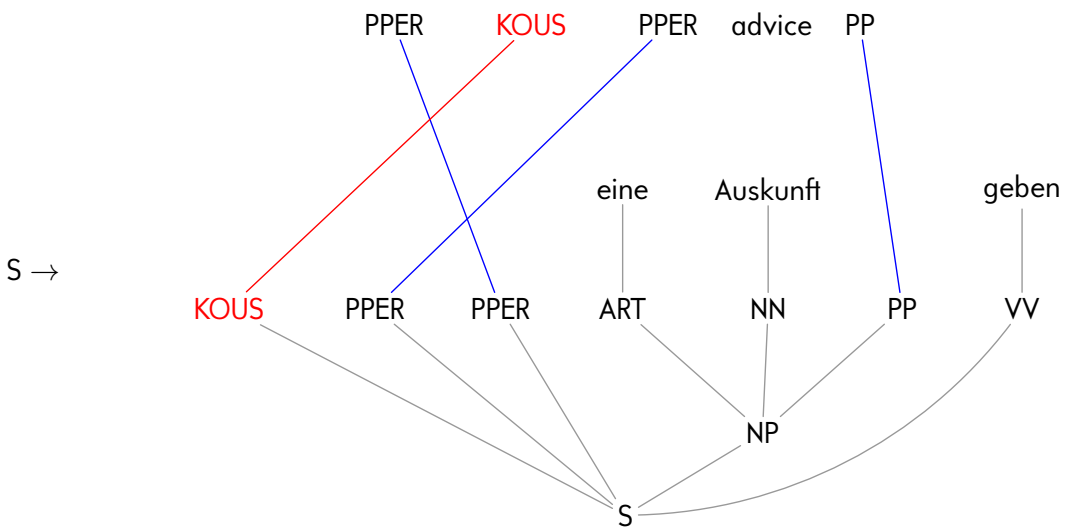
Synchronous Grammars



Production application

1. Selection of synchronous nonterminals

Synchronous Grammars



Production application

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

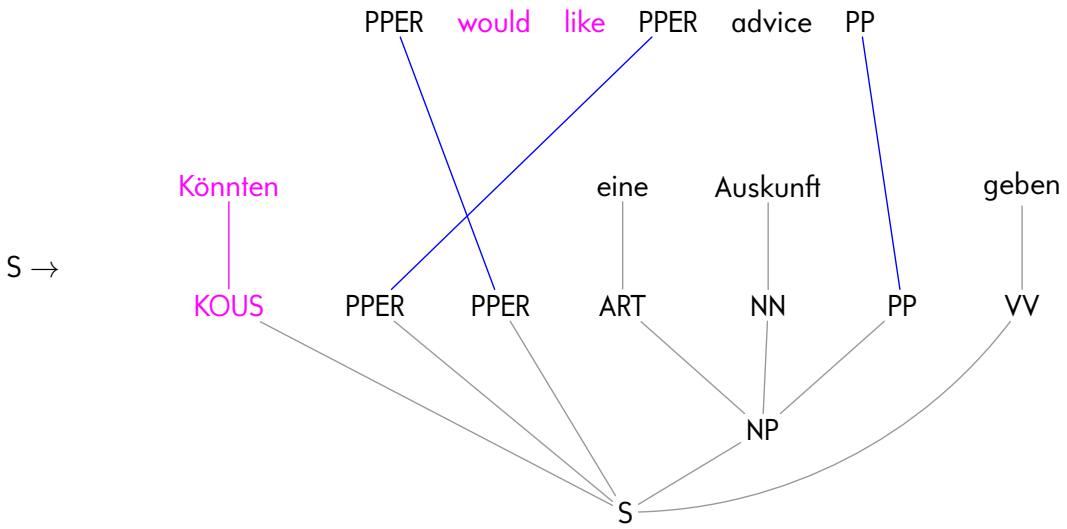
would like

KOUS →

Könnten

KOUS

Synchronous Grammars



Production application

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

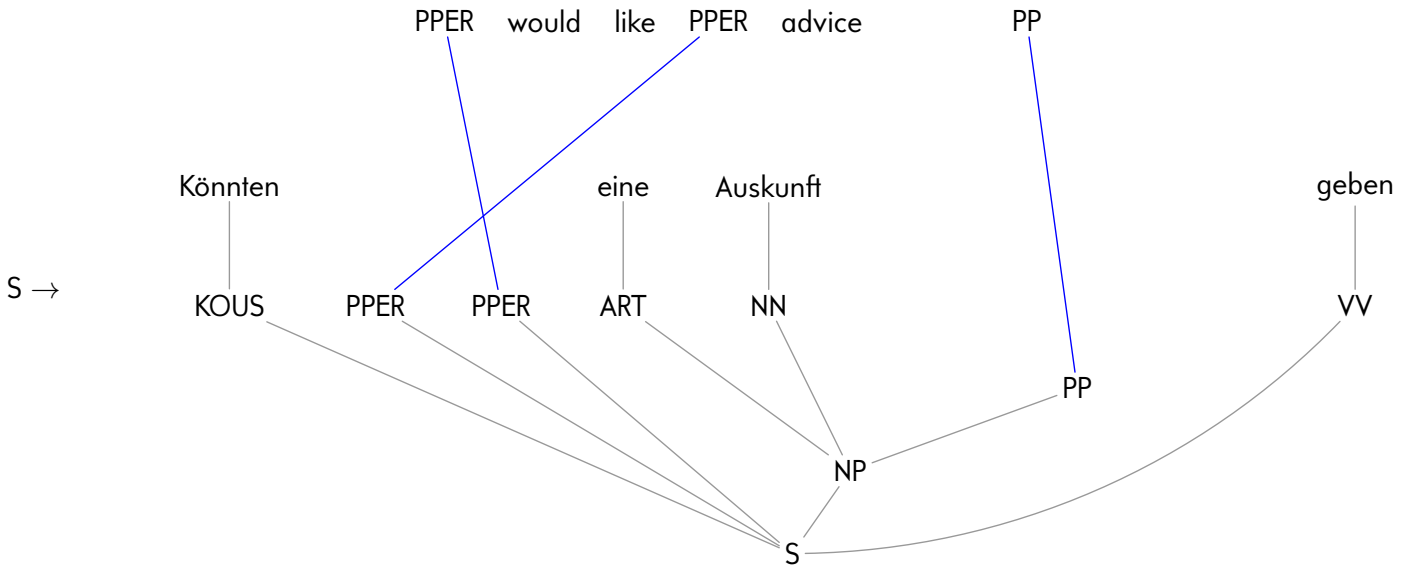
would like

KOUS →

Könnten

KOUS

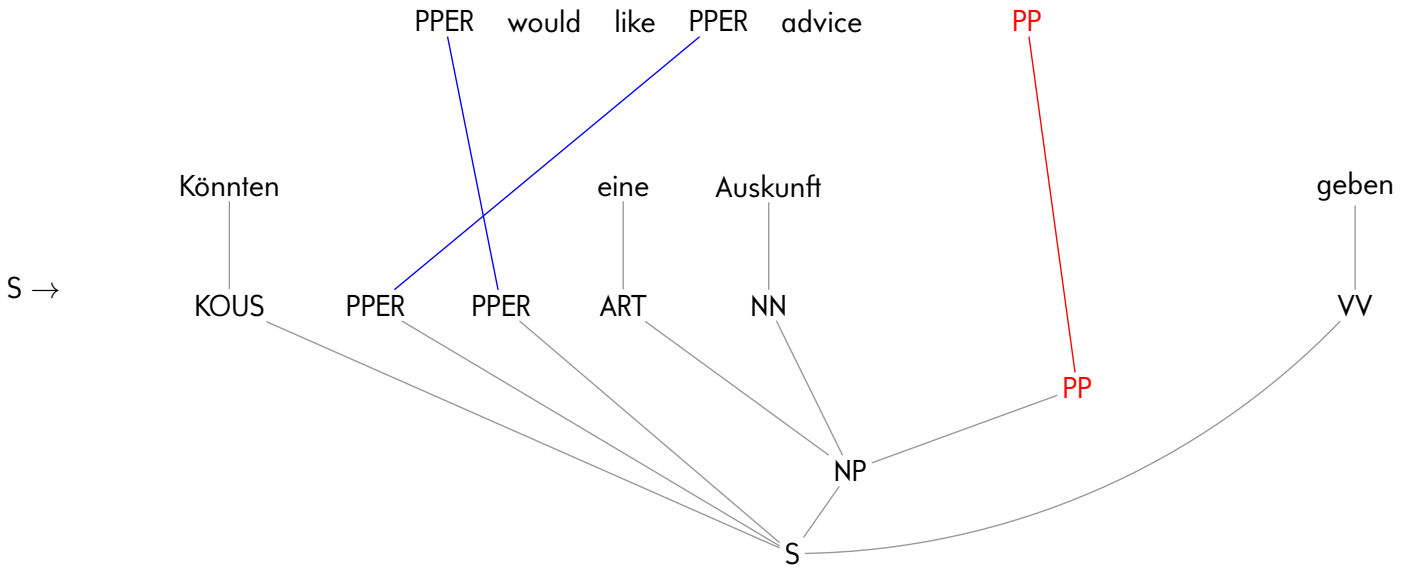
Synchronous Grammars



Production application

1. synchronous nonterminals

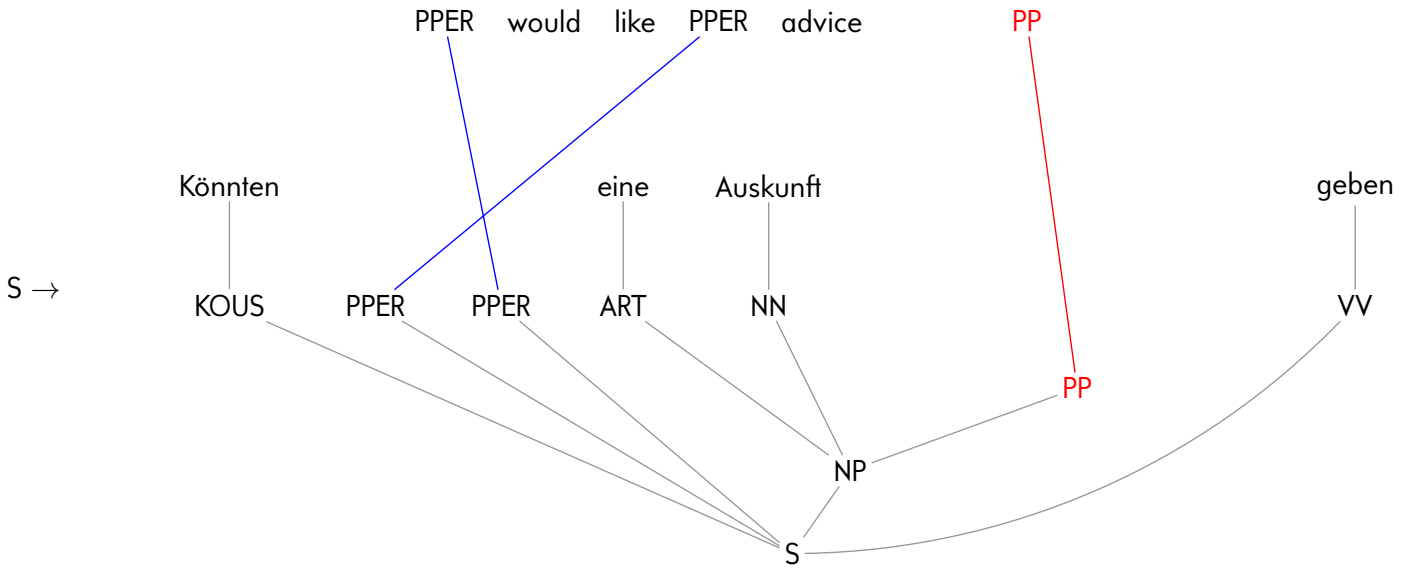
Synchronous Grammars



Production application

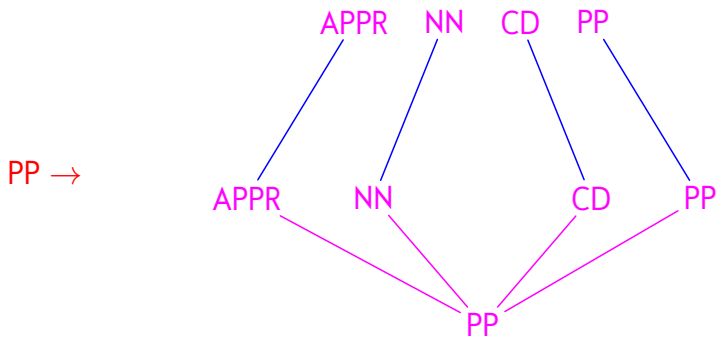
1. synchronous nonterminals

Synchronous Grammars

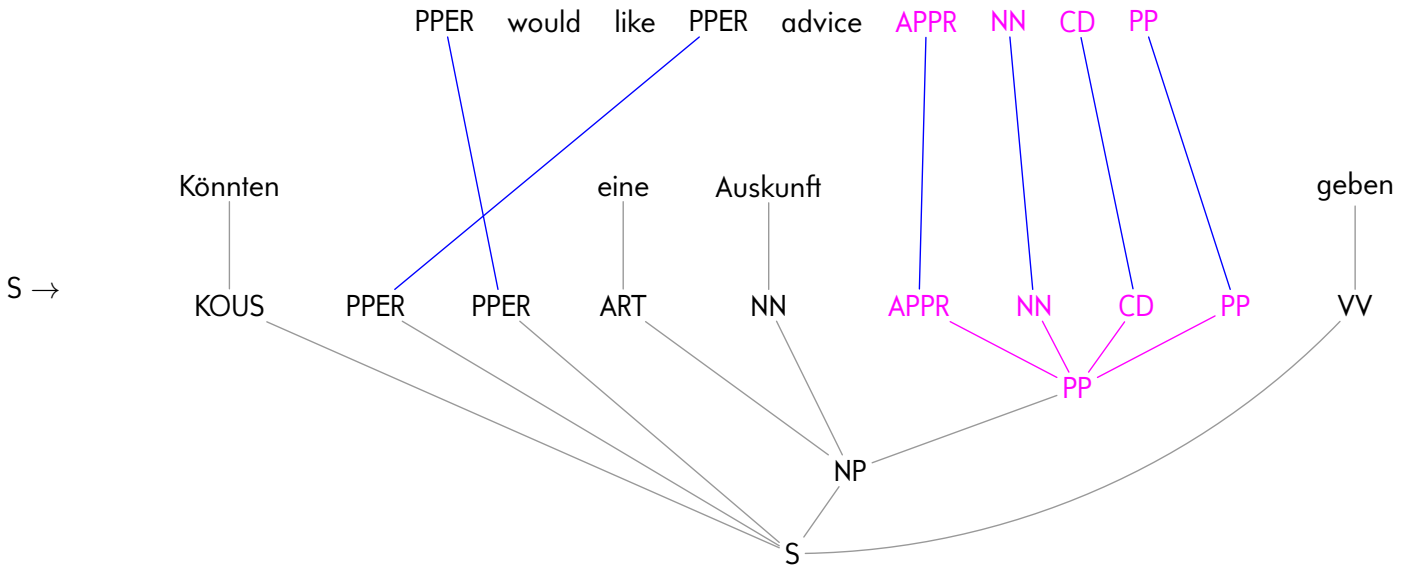


Production application

1. synchronous nonterminals
2. suitable production

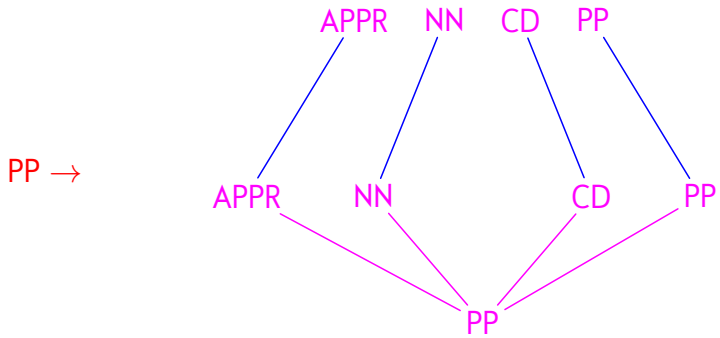


Synchronous Grammars

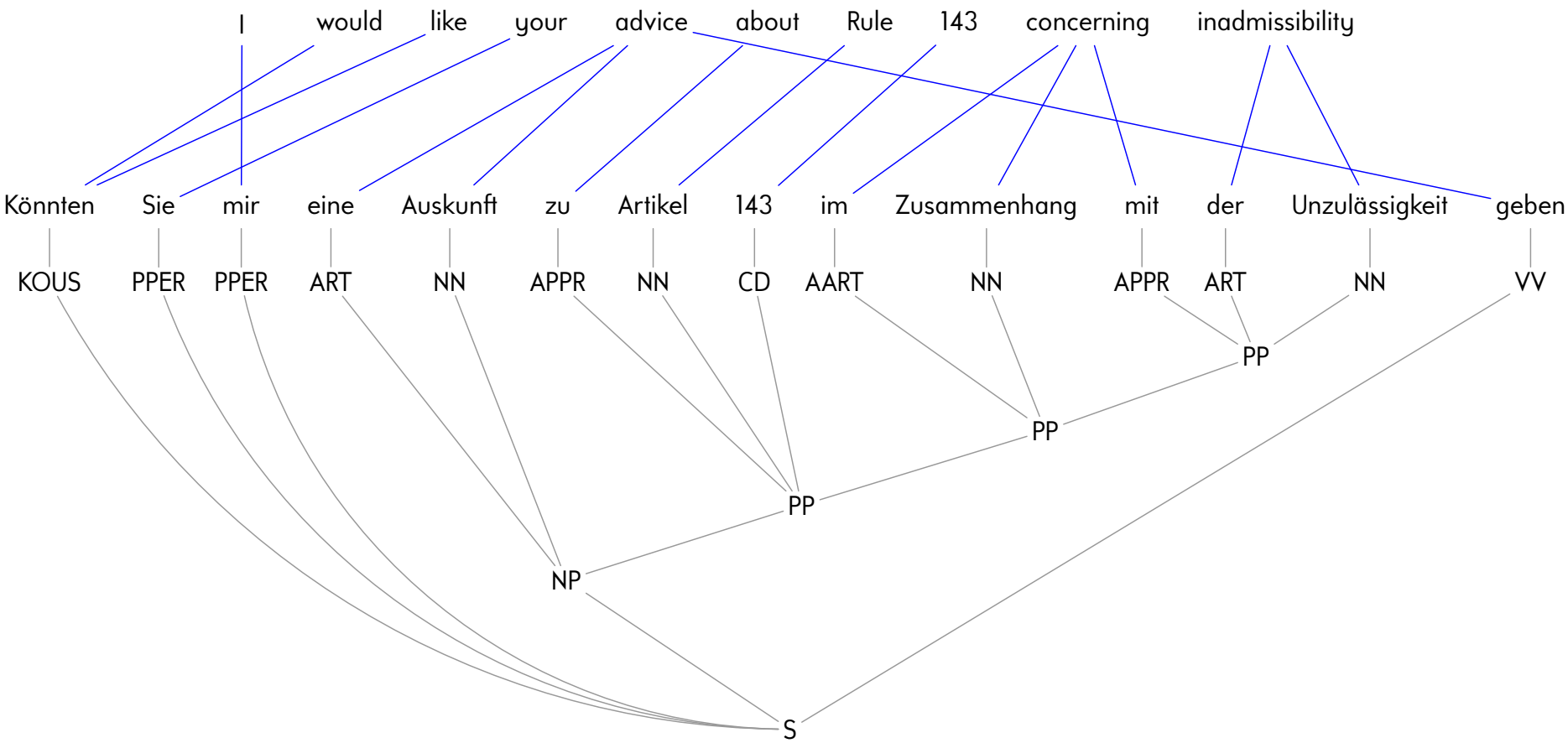


Production application

1. synchronous nonterminals
2. suitable production
3. replacement



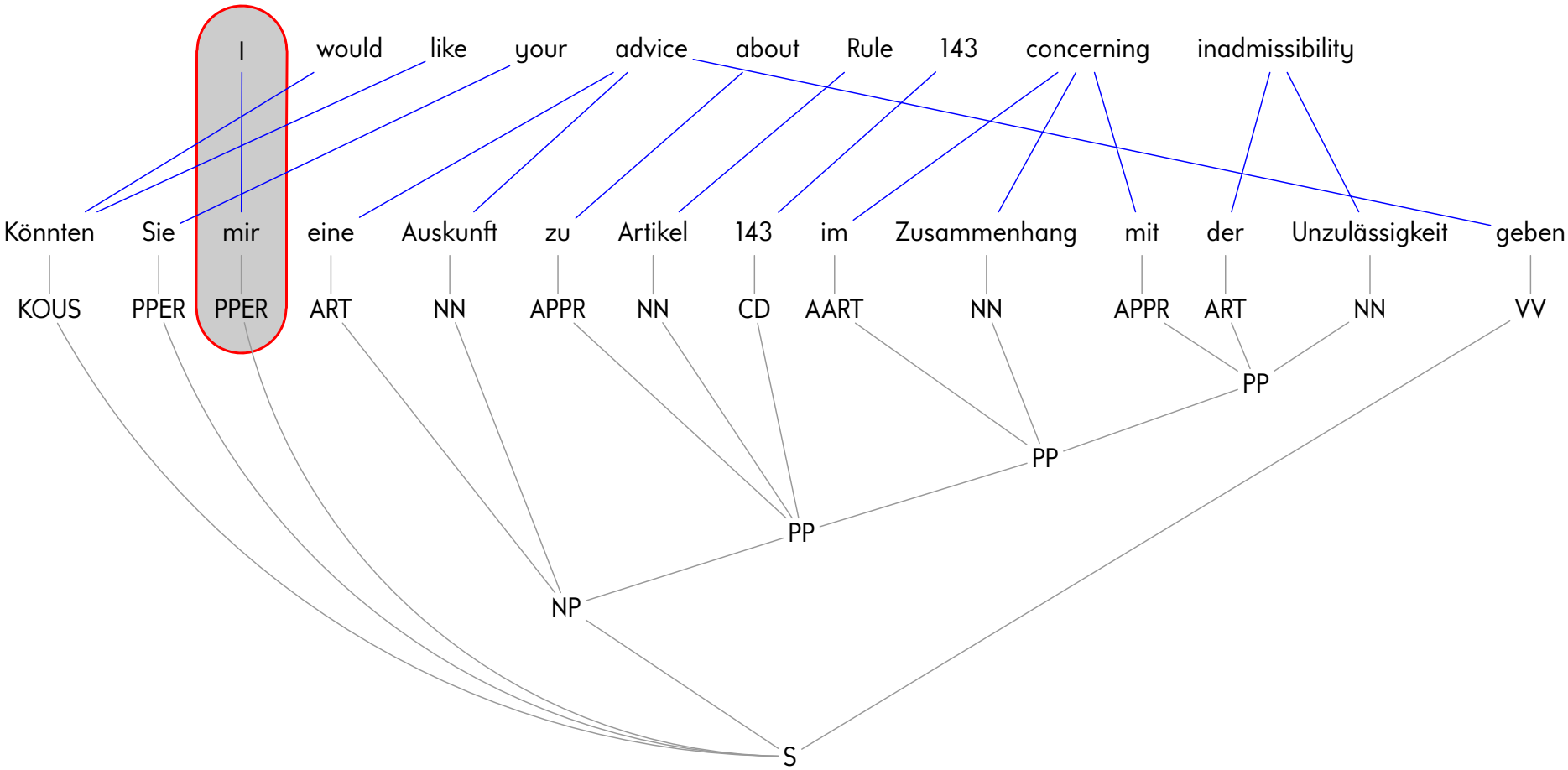
Production Extraction



following [Galley, Hopkins, Knight, Marcu: What's in a translation rule? *Proc. NAACL*, 2004]

Production Extraction

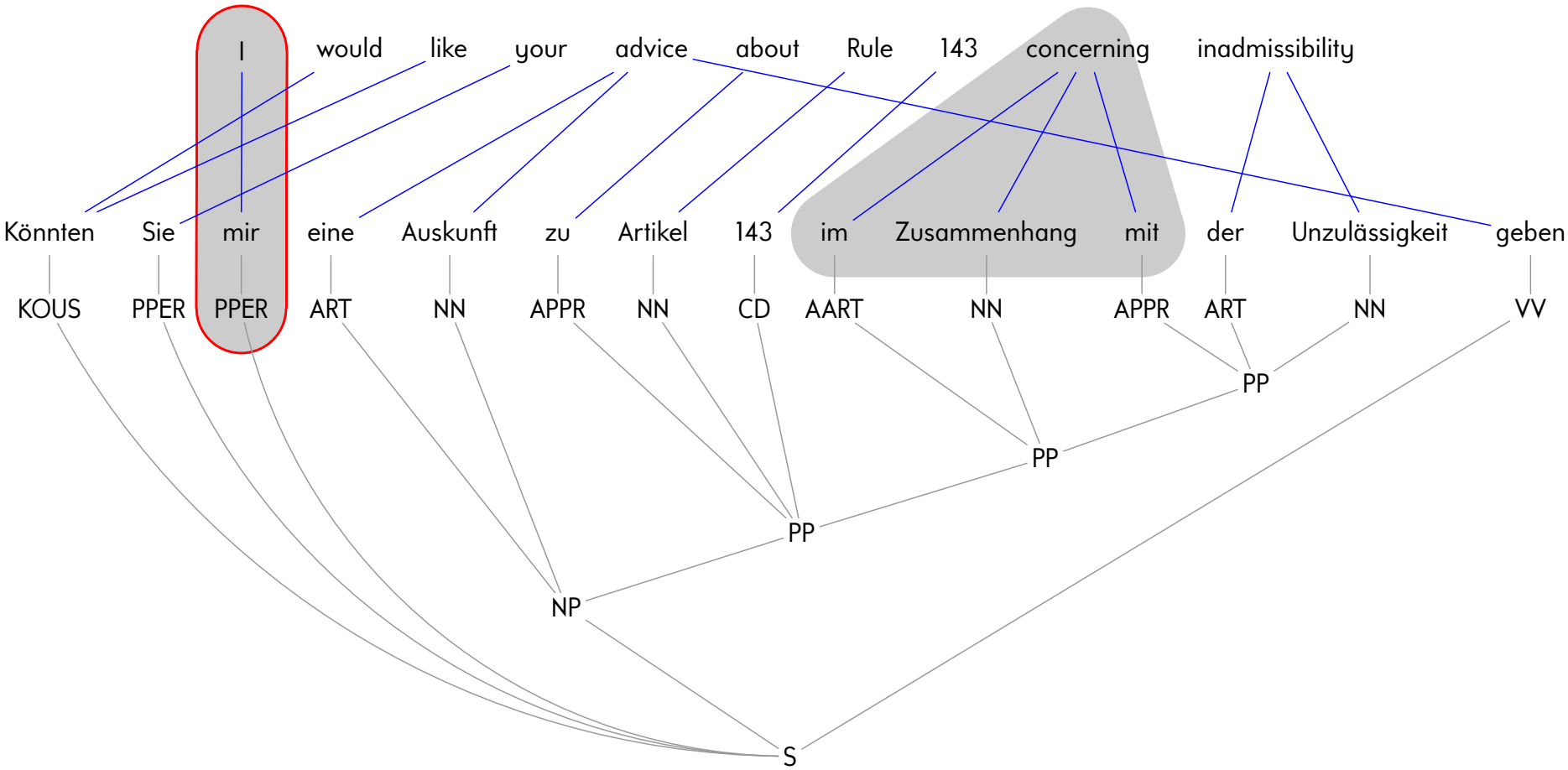
(extractable productions marked in red)



following [Galley, Hopkins, Knight, Marcu: What's in a translation rule? *Proc. NAACL*, 2004]

Production Extraction

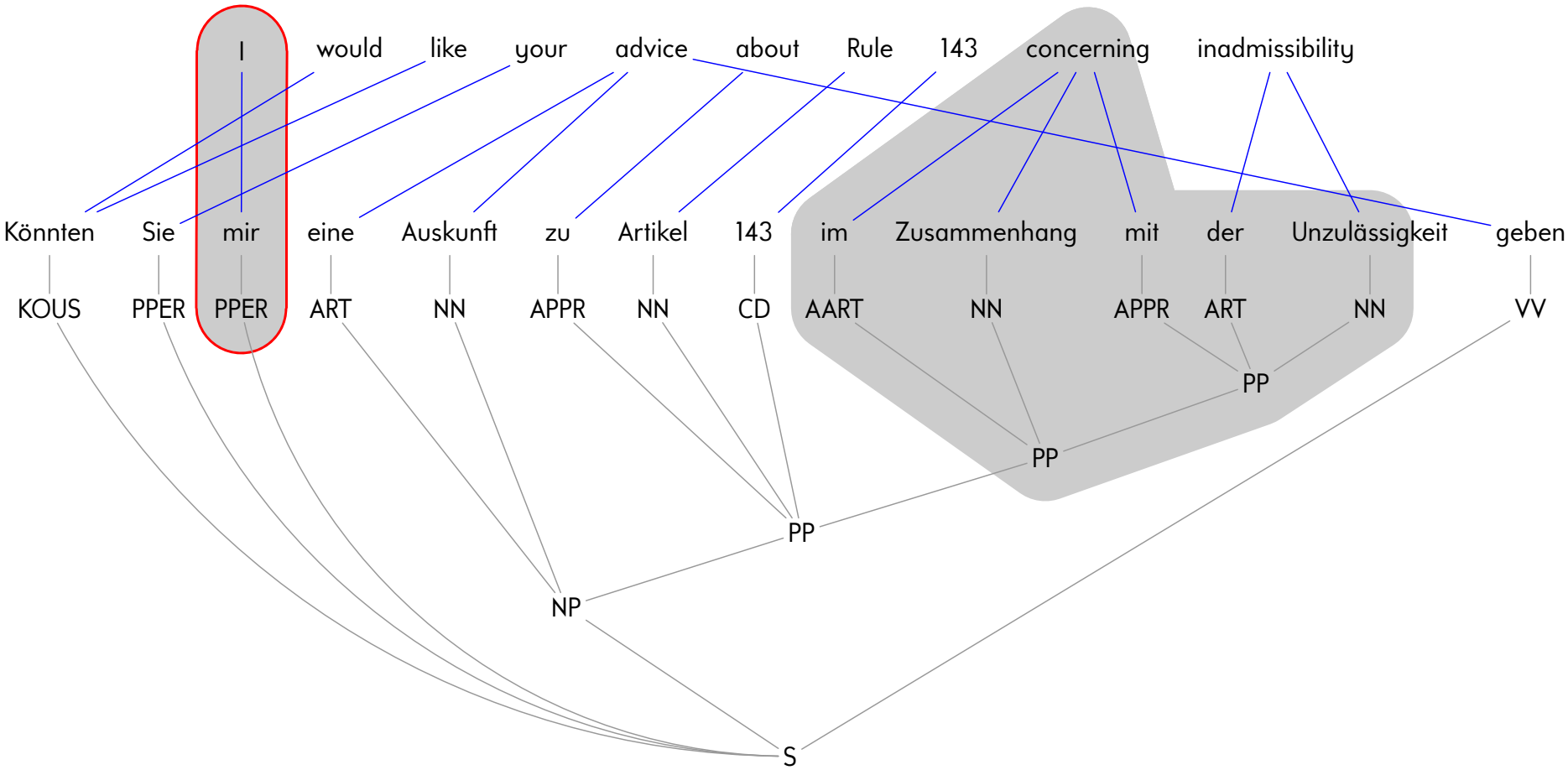
(extractable productions marked in red)



following [Galley, Hopkins, Knight, Marcu: What's in a translation rule? *Proc. NAACL*, 2004]

Production Extraction

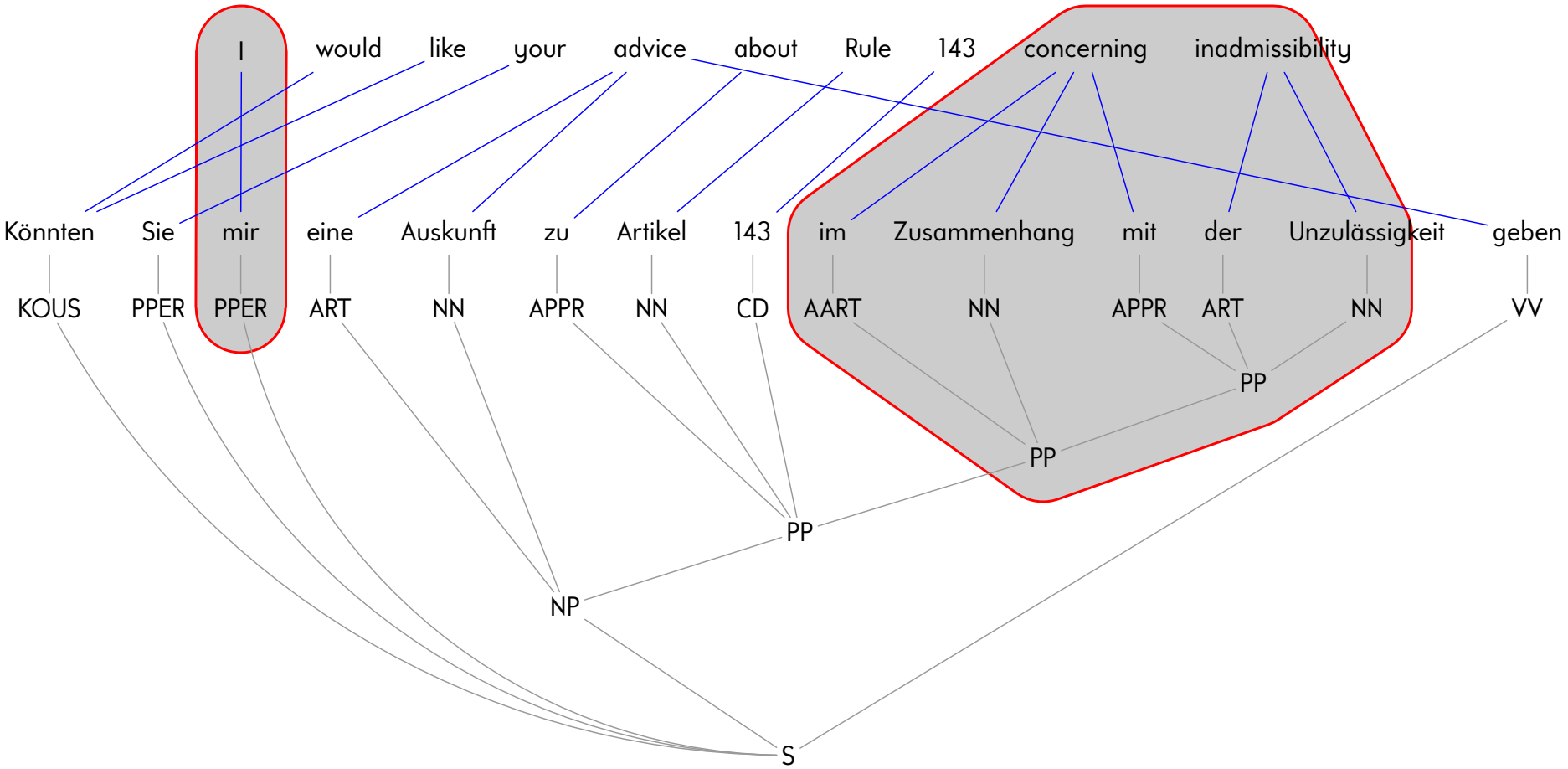
(extractable productions marked in red)



following [Galley, Hopkins, Knight, Marcu: What's in a translation rule? *Proc. NAACL*, 2004]

Production Extraction

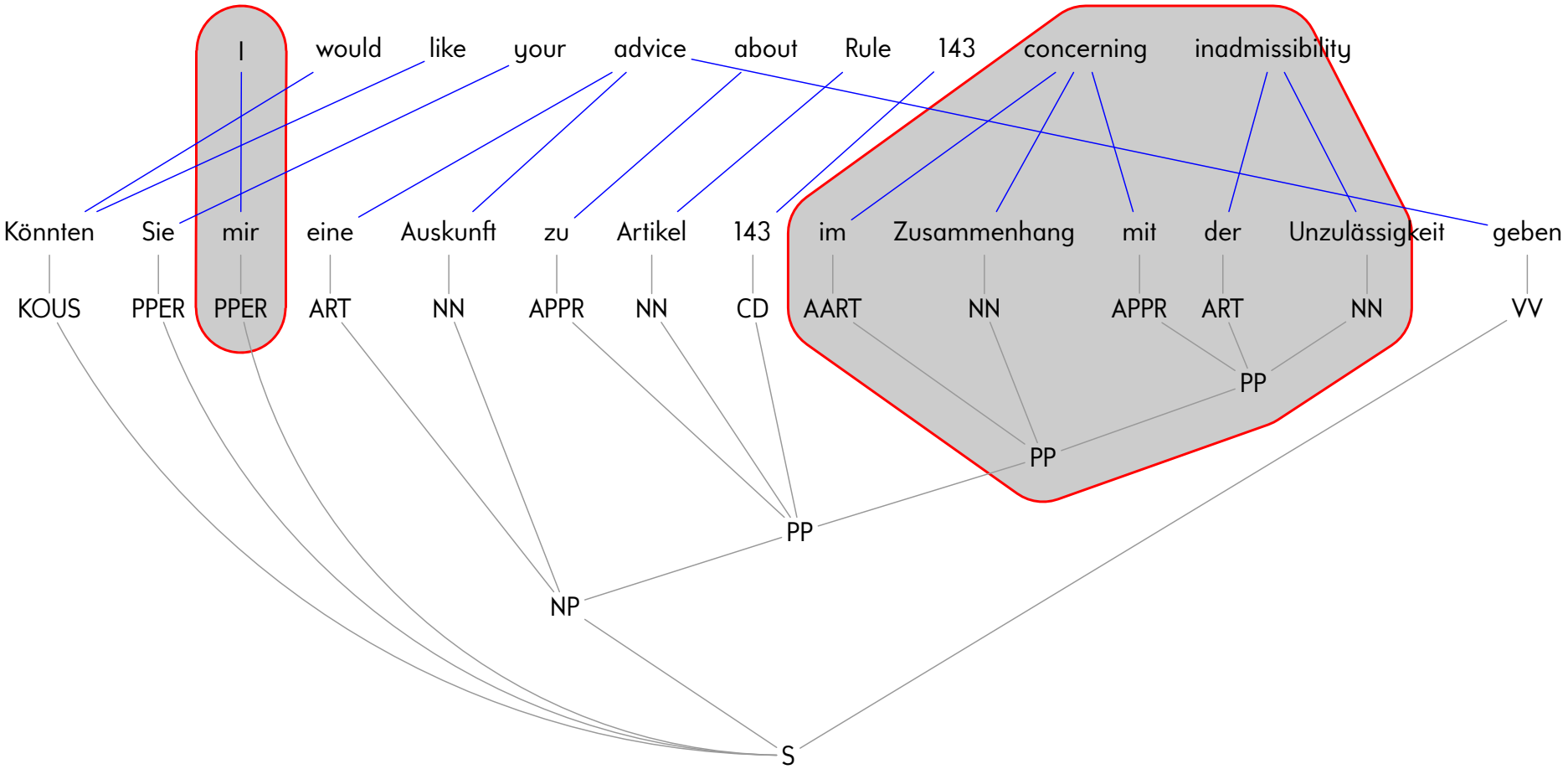
(extractable productions marked in red)



following [Galley, Hopkins, Knight, Marcu: What's in a translation rule? *Proc. NAACL*, 2004]

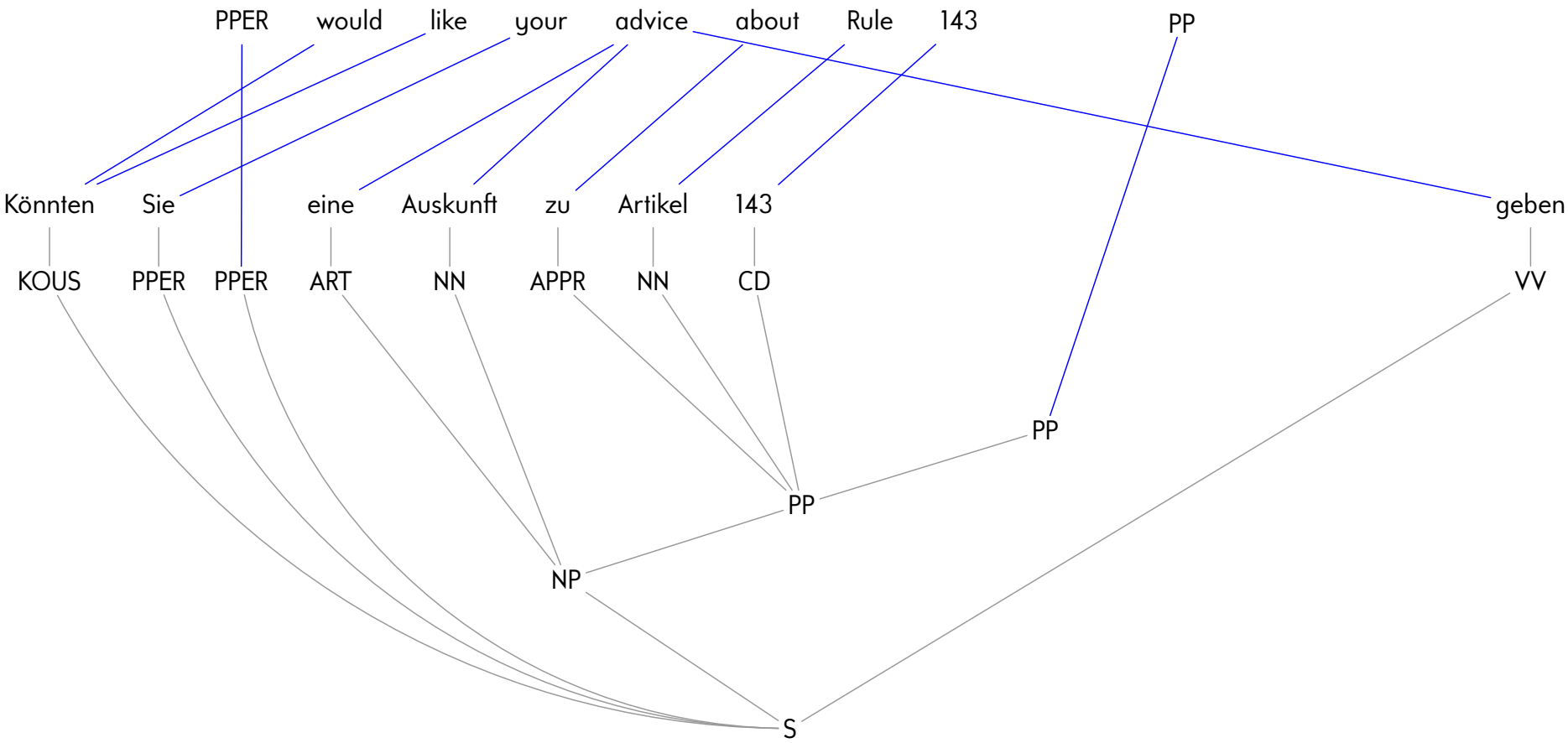
Production Extraction

Removal of extractable production:



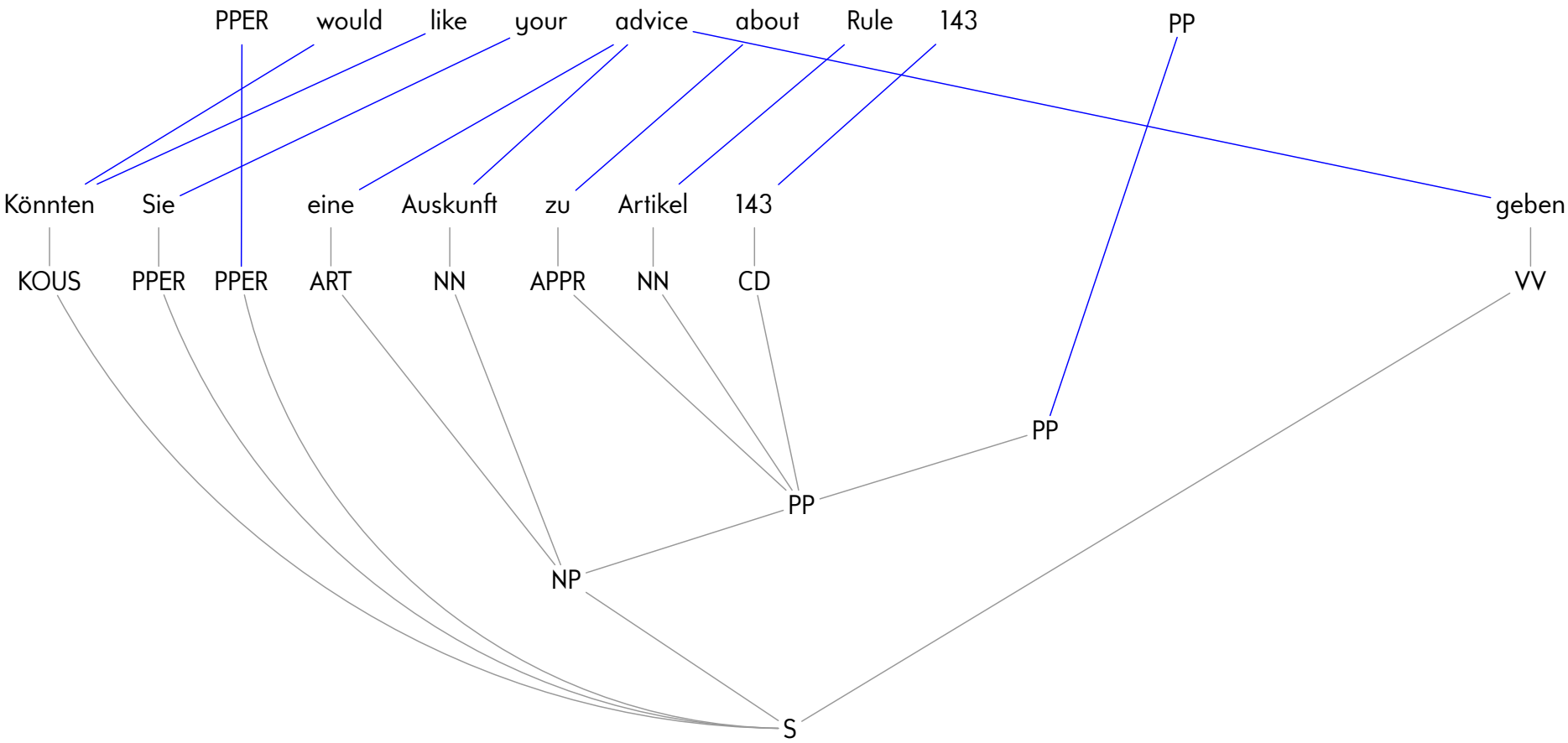
Production Extraction

Removal of extractable production:



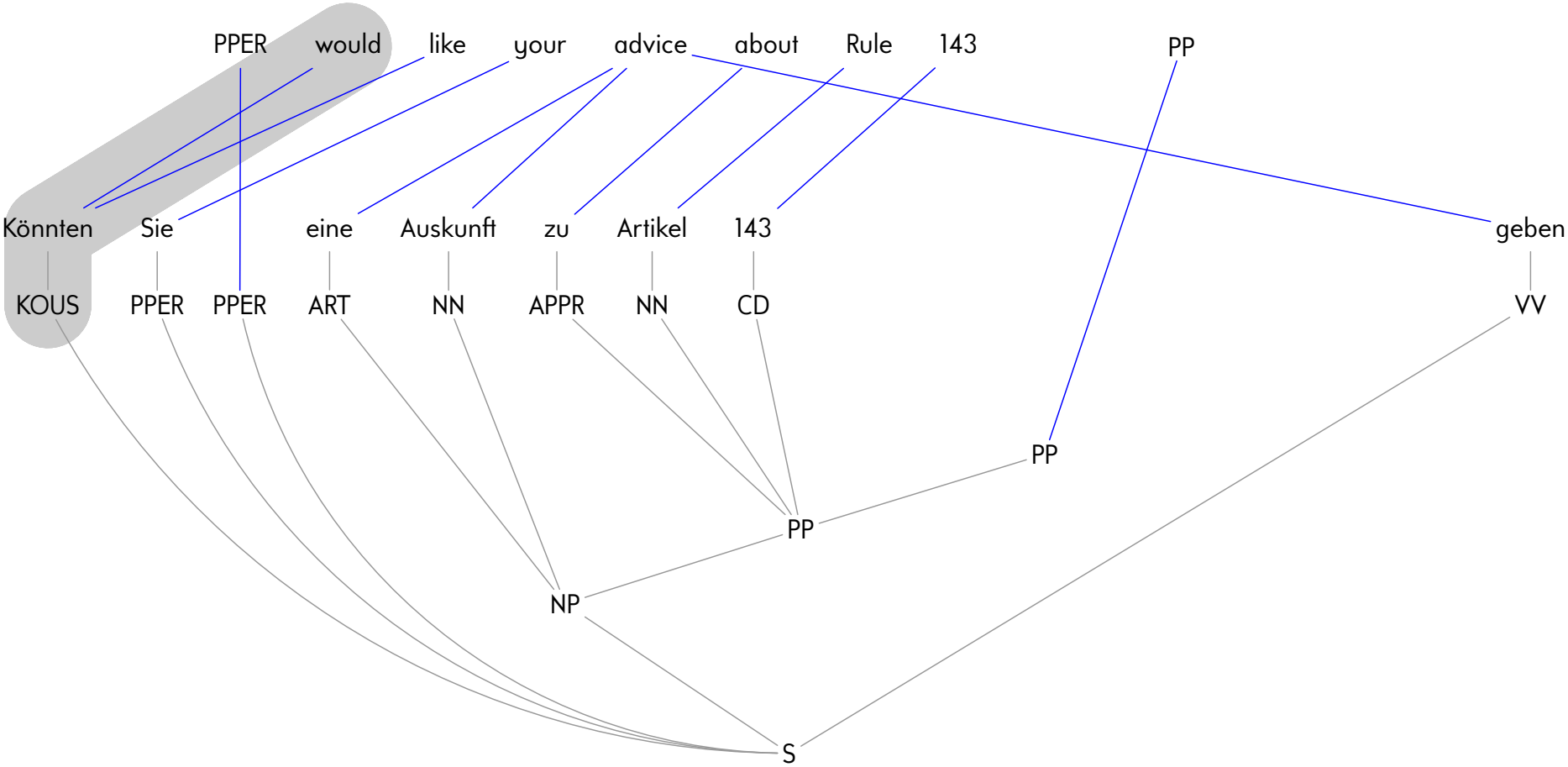
Production Extraction

Repeated production extraction:



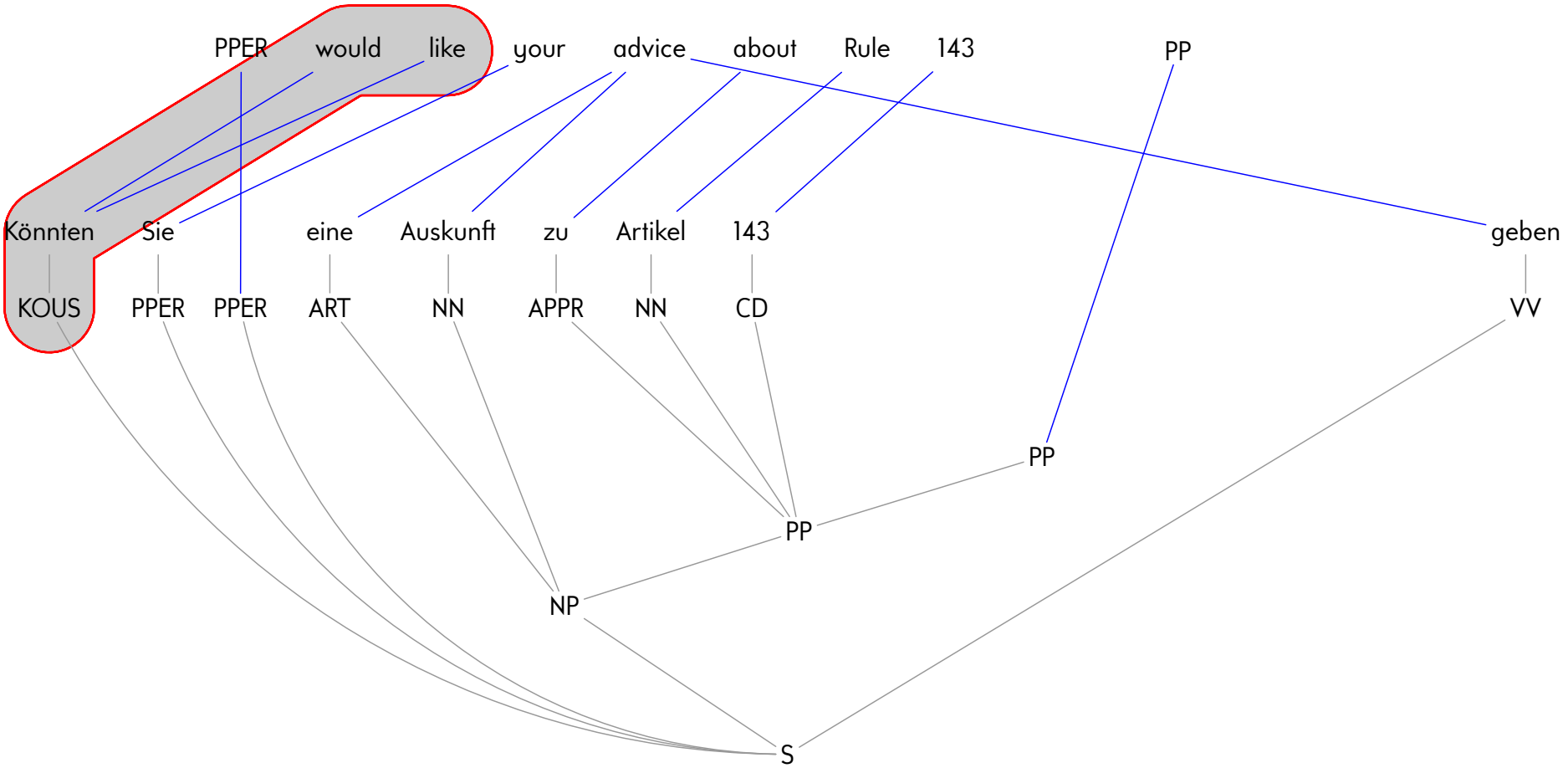
Production Extraction

Repeated production extraction:



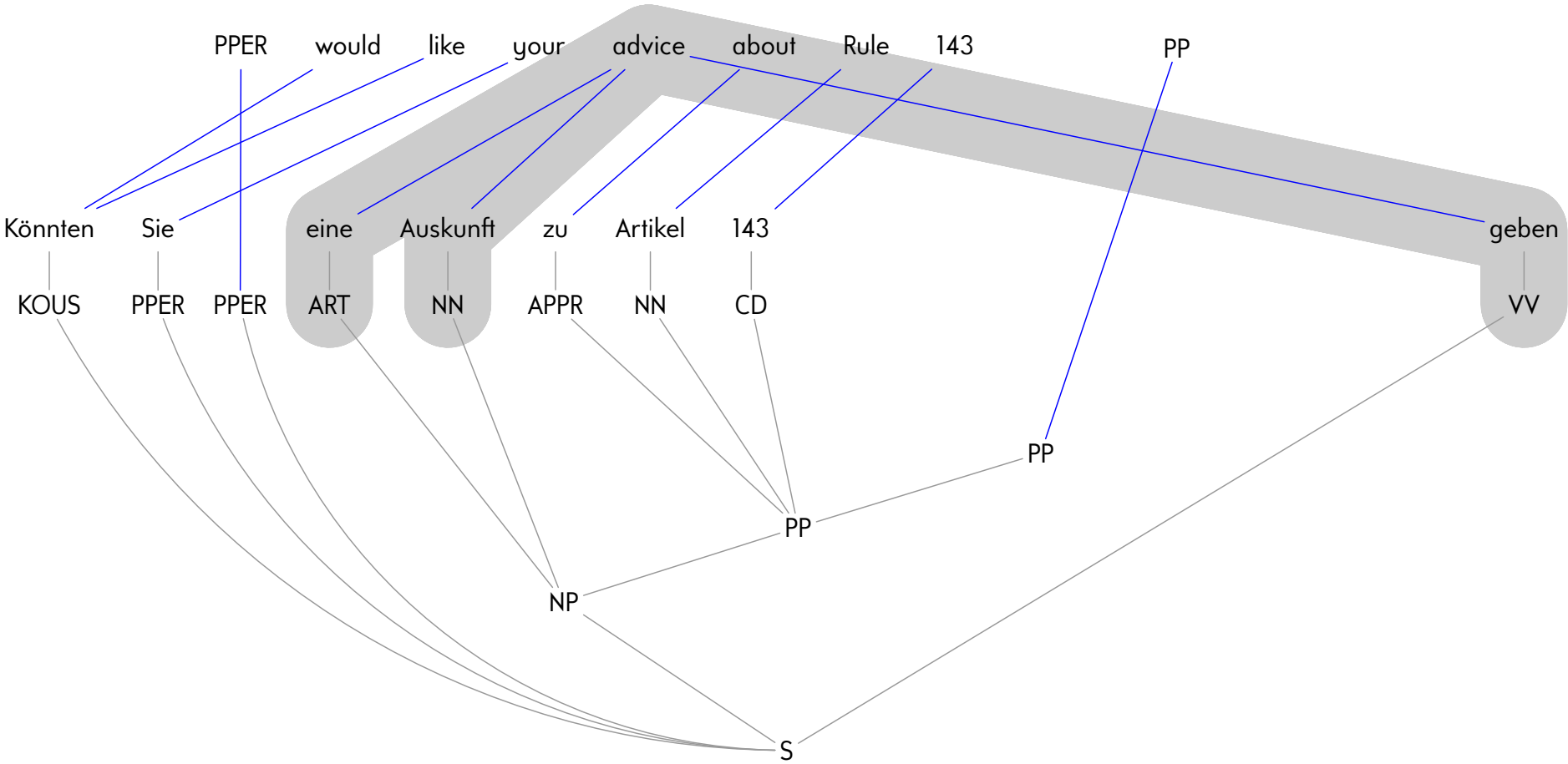
Production Extraction

Repeated production extraction: (extractable productions marked in red)



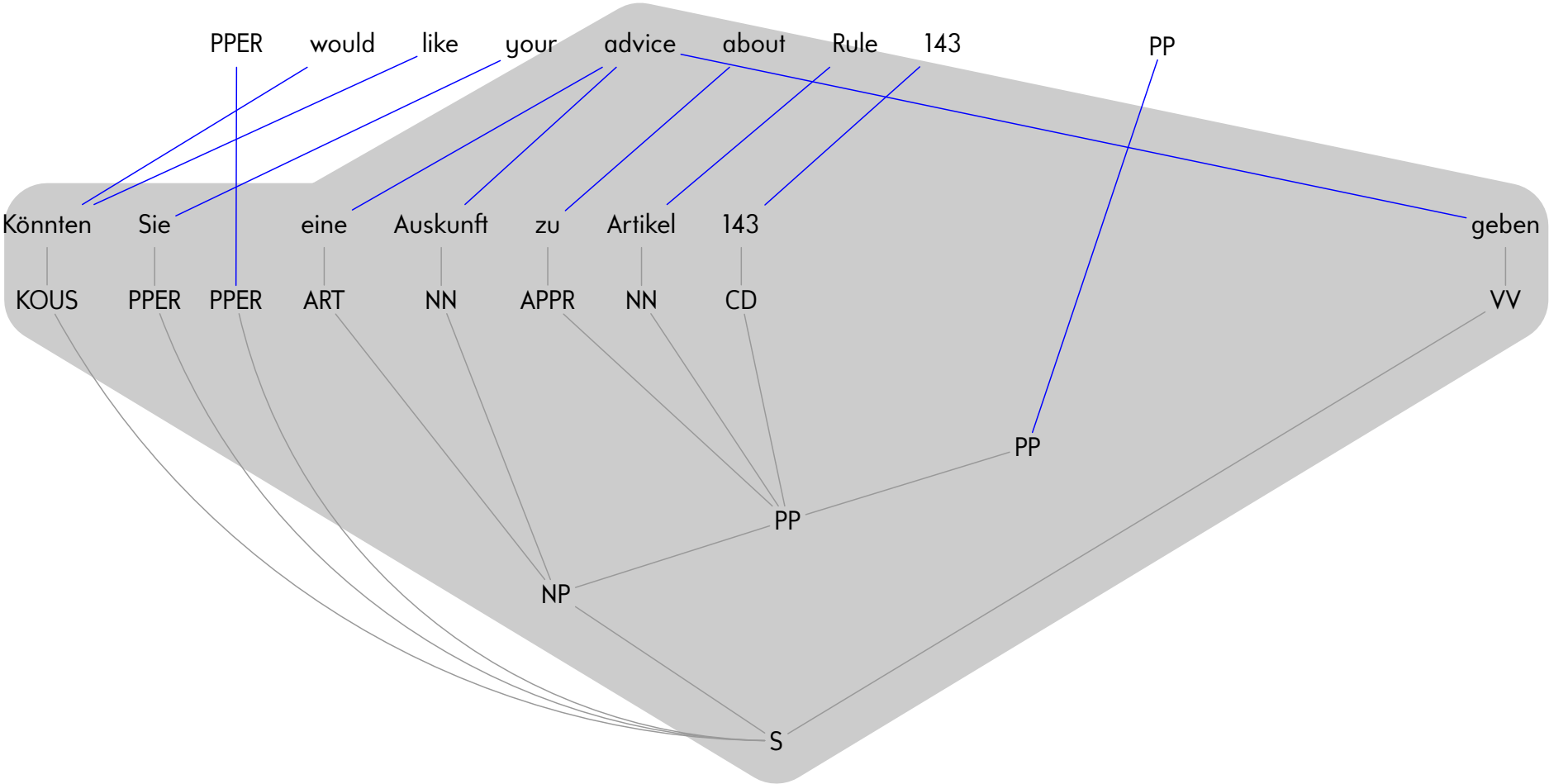
Production Extraction

Repeated production extraction: (extractable productions marked in red)



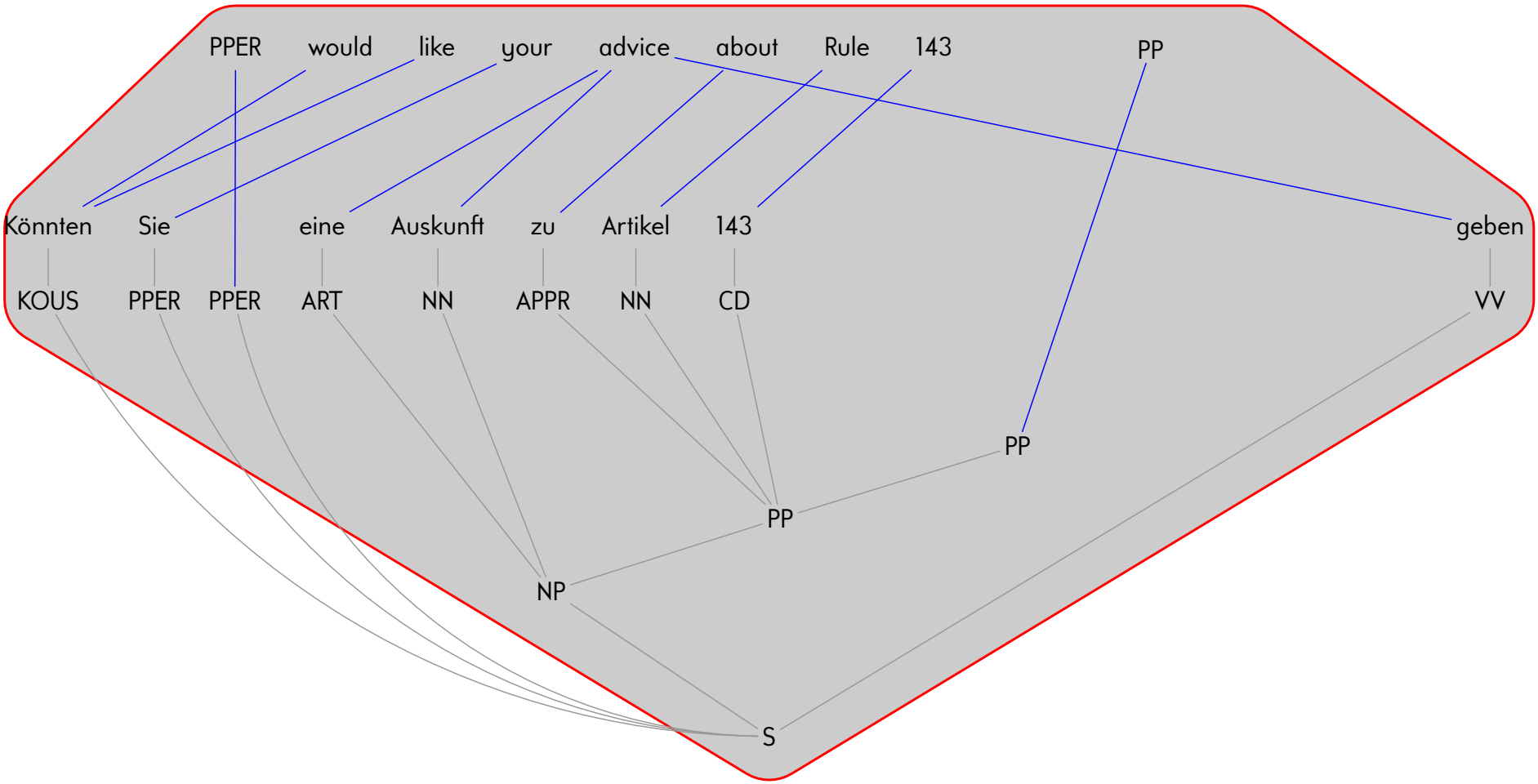
Production Extraction

Repeated production extraction: (extractable productions marked in red)



Production Extraction

Repeated production extraction: (extractable productions marked in red)



Synchronous Tree Substitution Grammars

Advantages

- very simple
- implemented in framework ‘Moses’
  [Koehn et al.: Moses — Open source toolkit for statistical machine translation. *Proc. ACL*, 2007]
- “context-free”

Synchronous Tree Substitution Grammars

Advantages

- very simple
- implemented in framework ‘Moses’
  [Koehn et al.: Moses — Open source toolkit for statistical machine translation. *Proc. ACL*, 2007]
- “context-free”

Disadvantages

- problems with discontinuities
- composition and binarization not possible
  [M., Graehl, Hopkins, Knight: The power of extended top-down tree transducers. *SIAM Journal on Computing* 39(2), 2009]
  [Zhang, Huang, Gildea, Knight: Synchronous Binarization for Machine Translation. *Proc. NAACL*, 2006]
- “context-free”

Evaluation

English → German translation task: (higher BLEU is better)

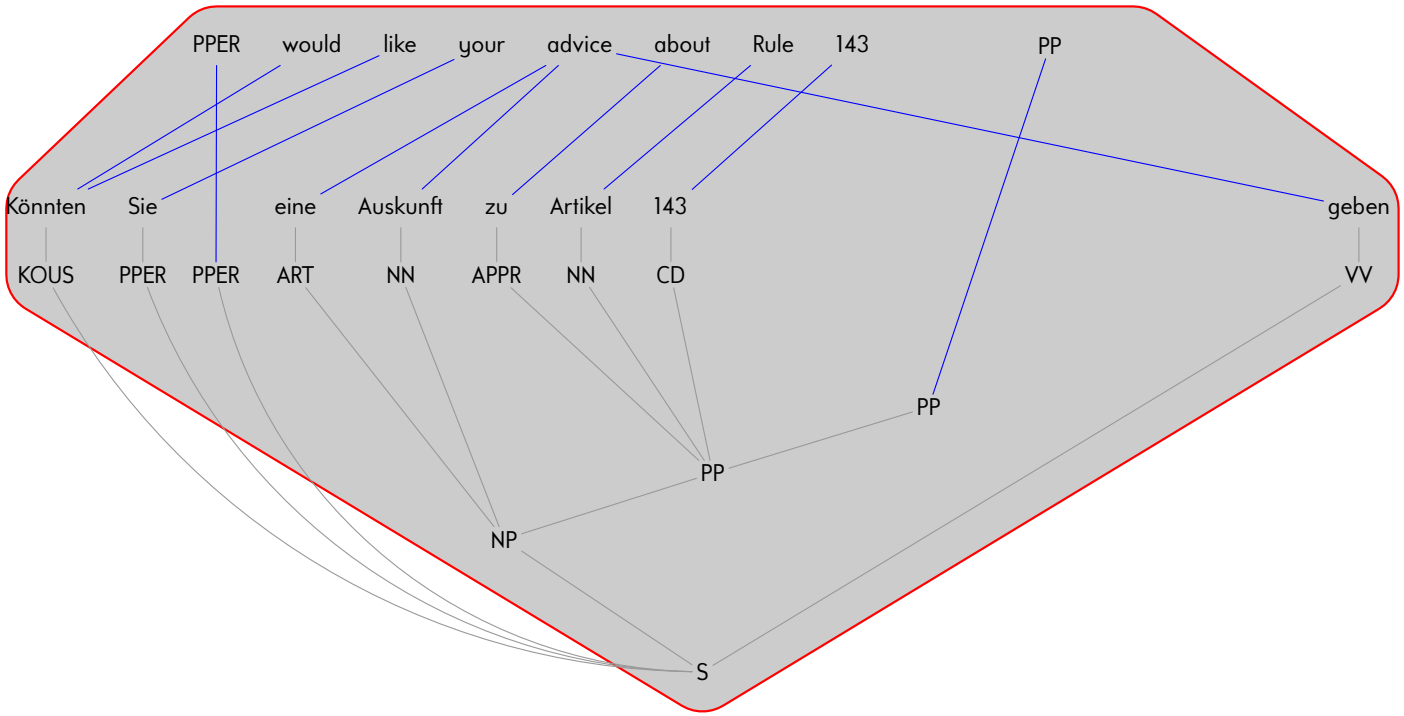
Type	**System**	**BLEU** [vanilla]	**BLEU** [competition]
string-to-string	phrase-based	16.7	20.3
	hierarchical	17.0	—
string-to-tree	STSG	15.2	—
	STSG [GHKM]	17.1	19.4
tree-to-tree	STSG	14.5	—

from [Seemann, Braune, M.: A systematic evaluation of MBOT in statistical machine translation. *Proc. MT-Summit,* 2015] and [Bojar et al.: Findings of the 2013 workshop on statistical machine translation. *Proc. WMT,* 2013]

Overview

1. Background
2. Extending the Expressive Power
3. Investigating their Expressive Power

Production Extraction



- very specific production
- every production for ‘advice’ contains sentence structure

(syntax “in the way”)

Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow \left(r, \langle r_1, \ldots, r_n \rangle\right)$

variant of [M.: Why synchronous tree substitution grammars?. *Proc. NAACL*, 2010]

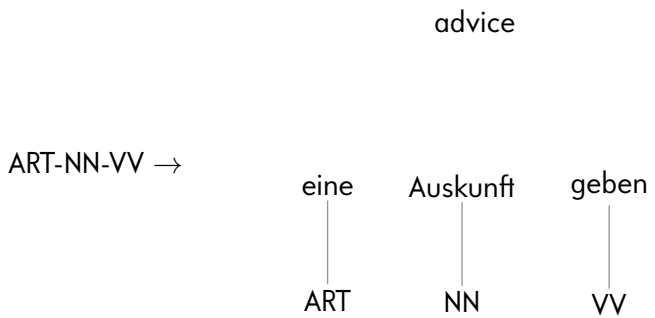
- nonterminal N
- right-hand side r of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production

Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow \left(r, \langle r_1, \ldots, r_n \rangle\right)$

variant of [M.: Why synchronous tree substitution grammars?. *Proc. NAACL*, 2010]

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production

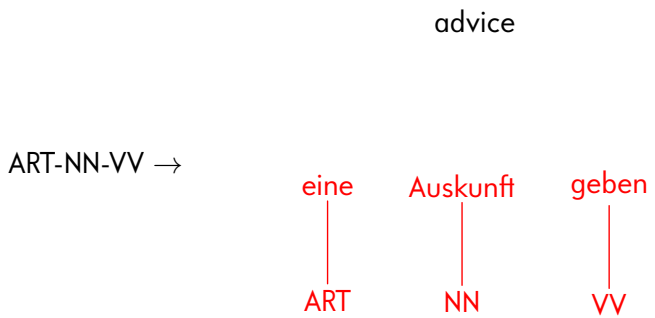


Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow \left(r, \langle r_1, \ldots, r_n \rangle\right)$

variant of [M.: Why synchronous tree substitution grammars?. *Proc. NAACL*, 2010]

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production

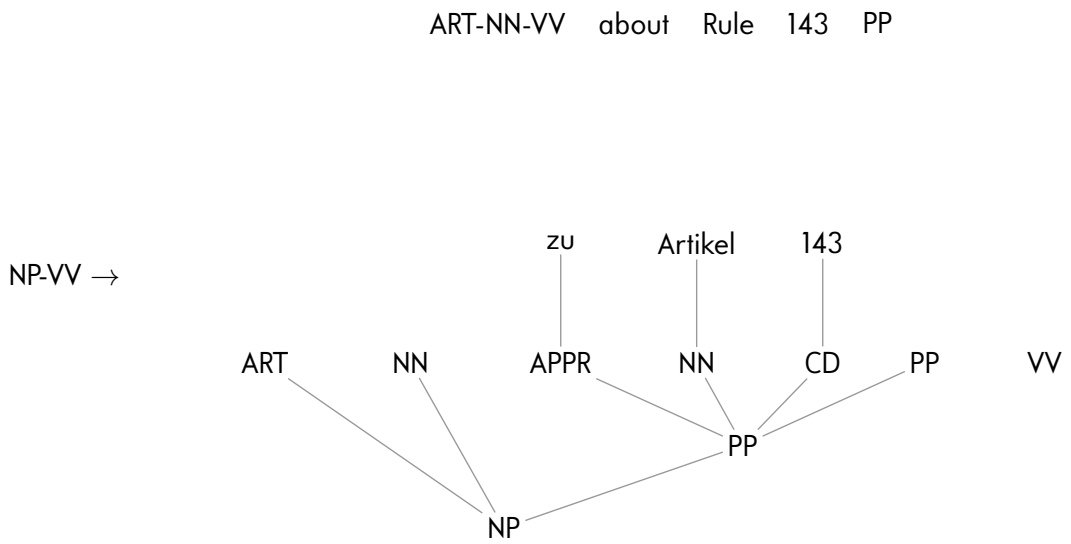


Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow \left(r, \langle r_1, \ldots, r_n \rangle\right)$

variant of [M.: Why synchronous tree substitution grammars?. *Proc. NAACL*, 2010]

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production

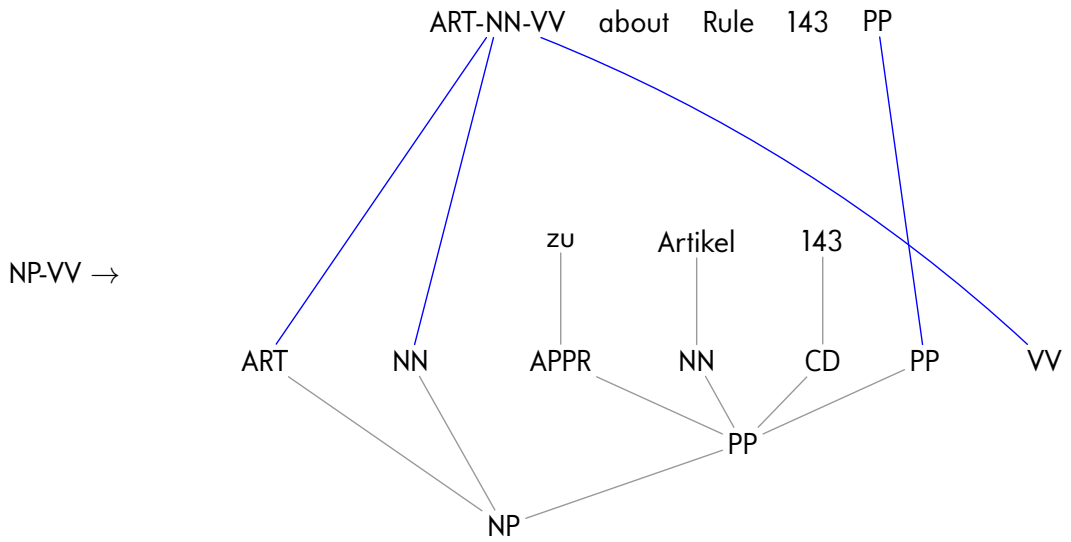


Synchronous Grammars

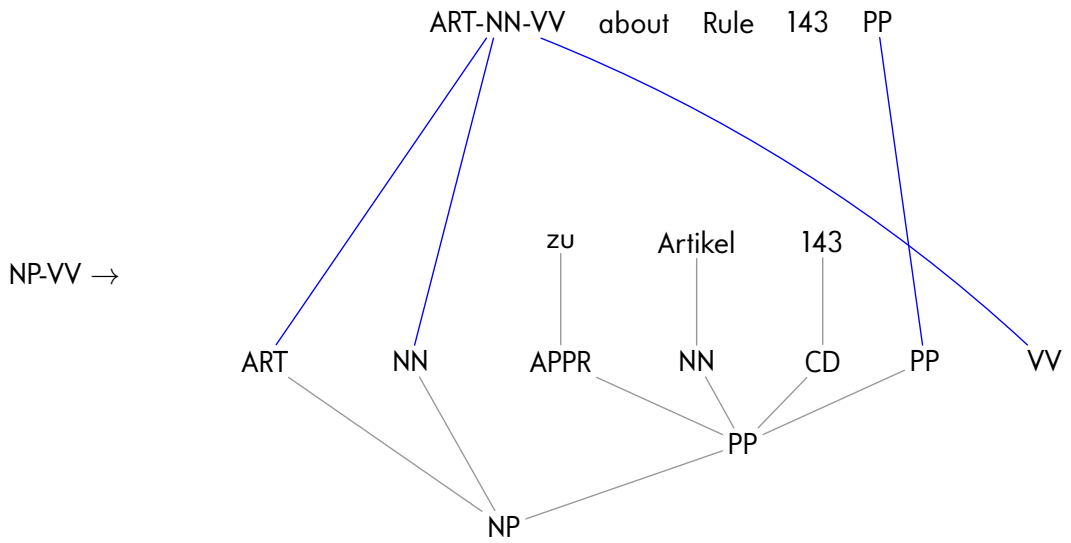
Synchronous multi tree substitution grammar: $N \rightarrow \left(r, \langle r_1, \ldots, r_n \rangle\right)$

variant of [M.: Why synchronous tree substitution grammars?. *Proc. NAACL*, 2010]

- nonterminal N
- right-hand side r of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production
- synchronization via map NT $r_1, \ldots, r_n$ to NT r



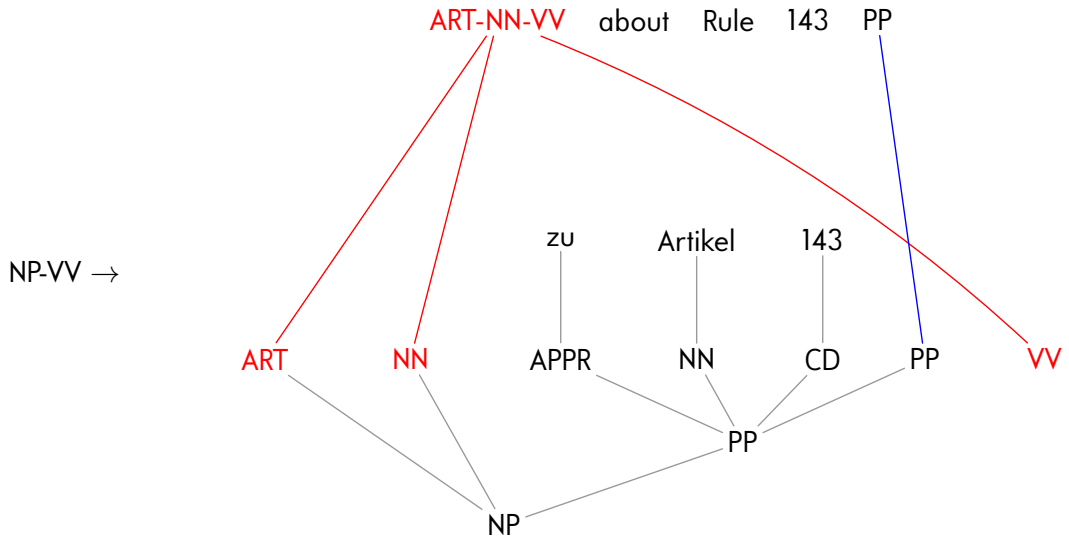
Synchronous Grammars



Production application

1. synchronous nonterminals

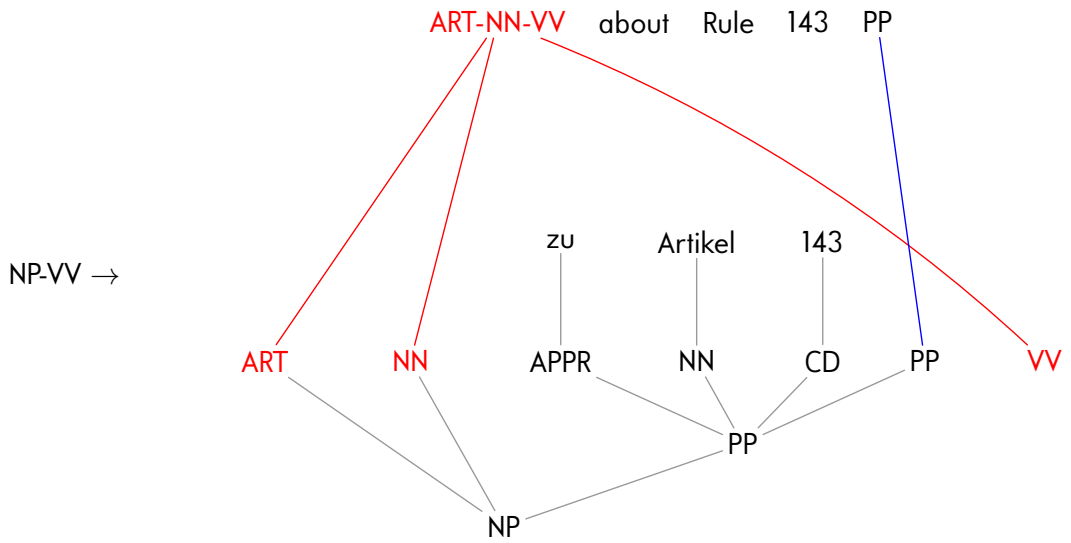
Synchronous Grammars



Production application

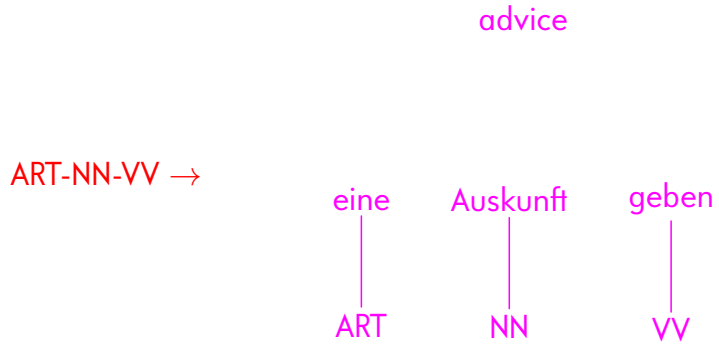
1. synchronous nonterminals

Synchronous Grammars

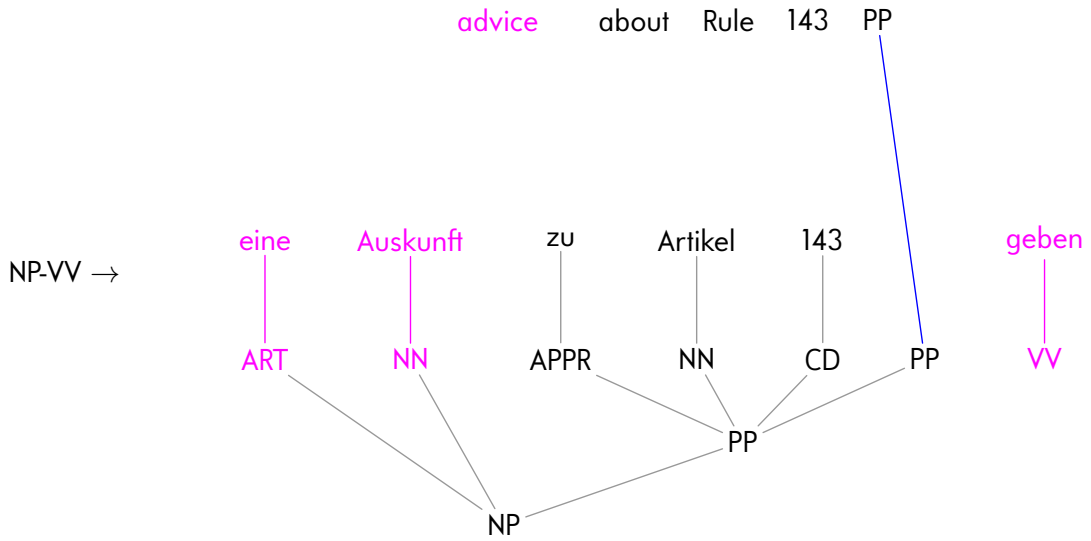


Production application

1. synchronous nonterminals
2. suitable production

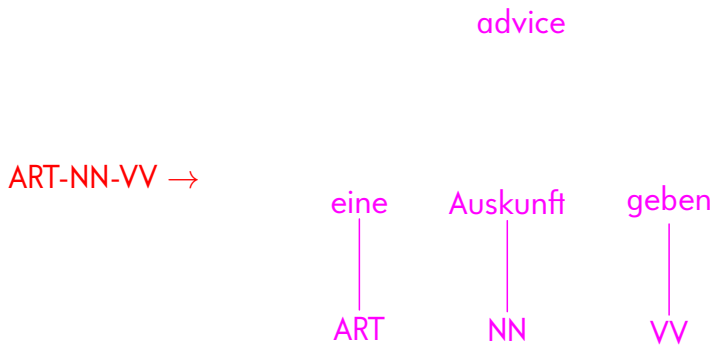


Synchronous Grammars

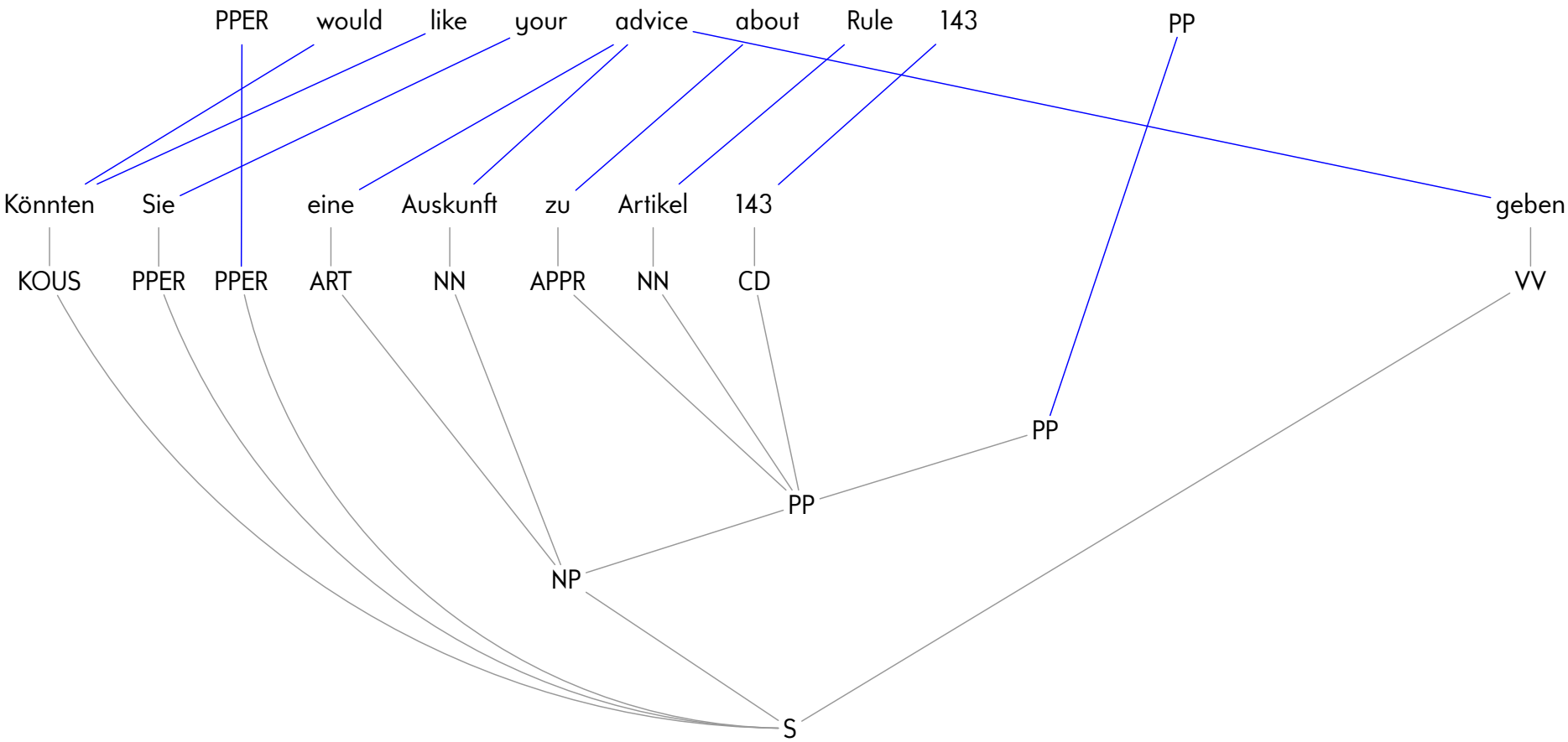


Production application

1. synchronous nonterminals
2. suitable production
3. replacement

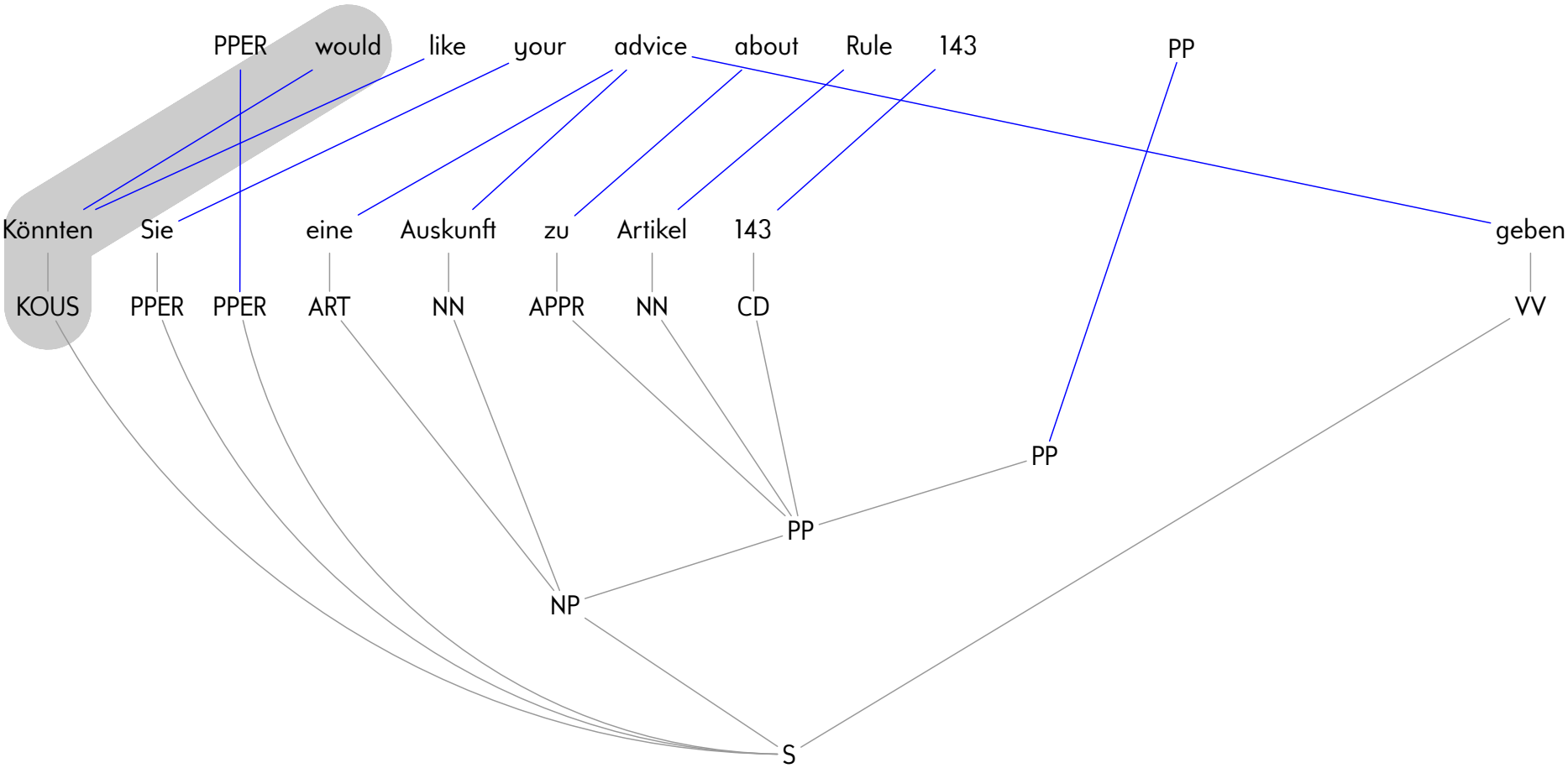


Production Extraction



variant of [M.: How to train your multi bottom-up tree transducer. *Proc. ACL*, 2011]

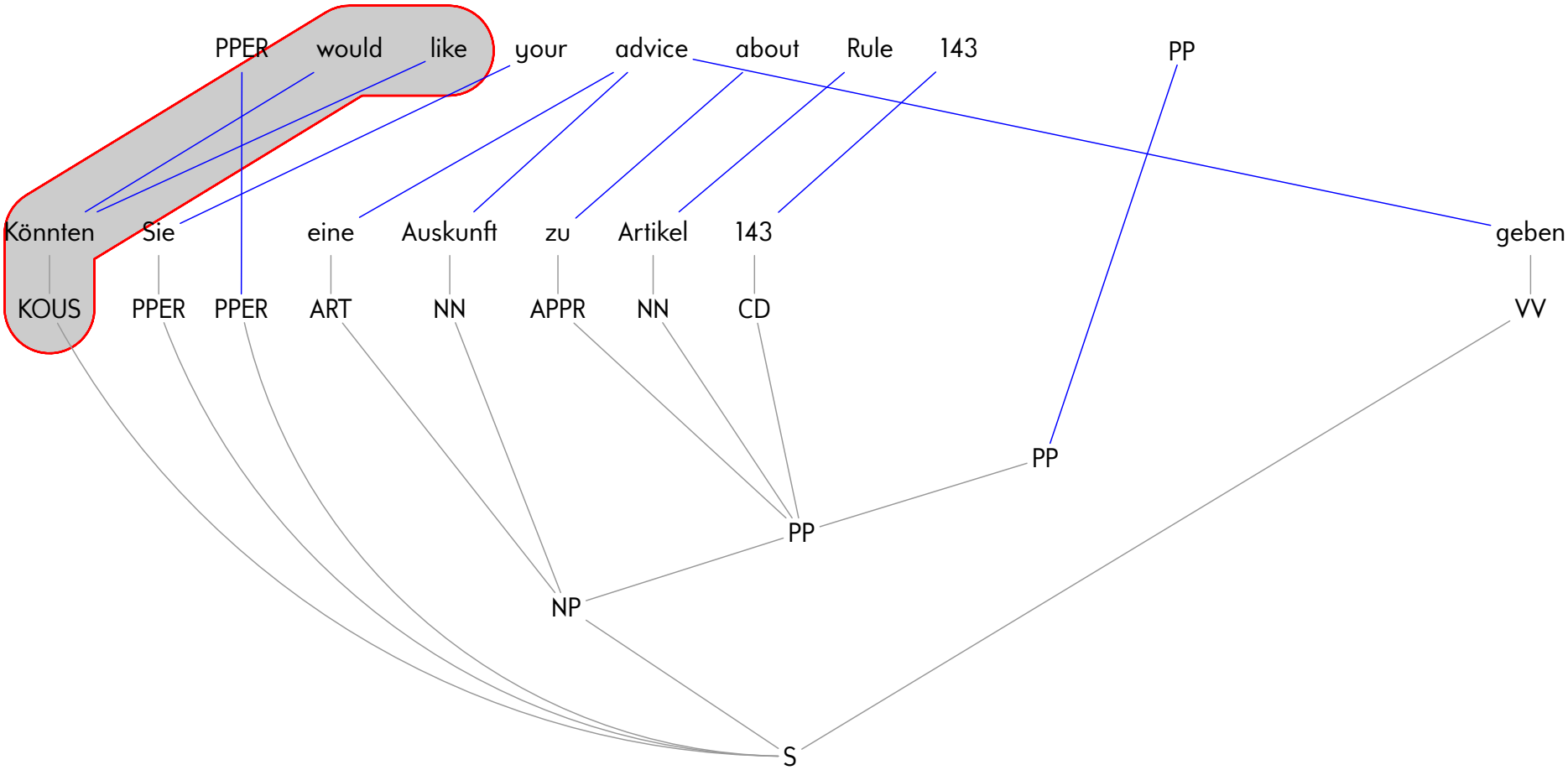
Production Extraction



variant of [M.: How to train your multi bottom-up tree transducer. *Proc. ACL*, 2011]

Production Extraction

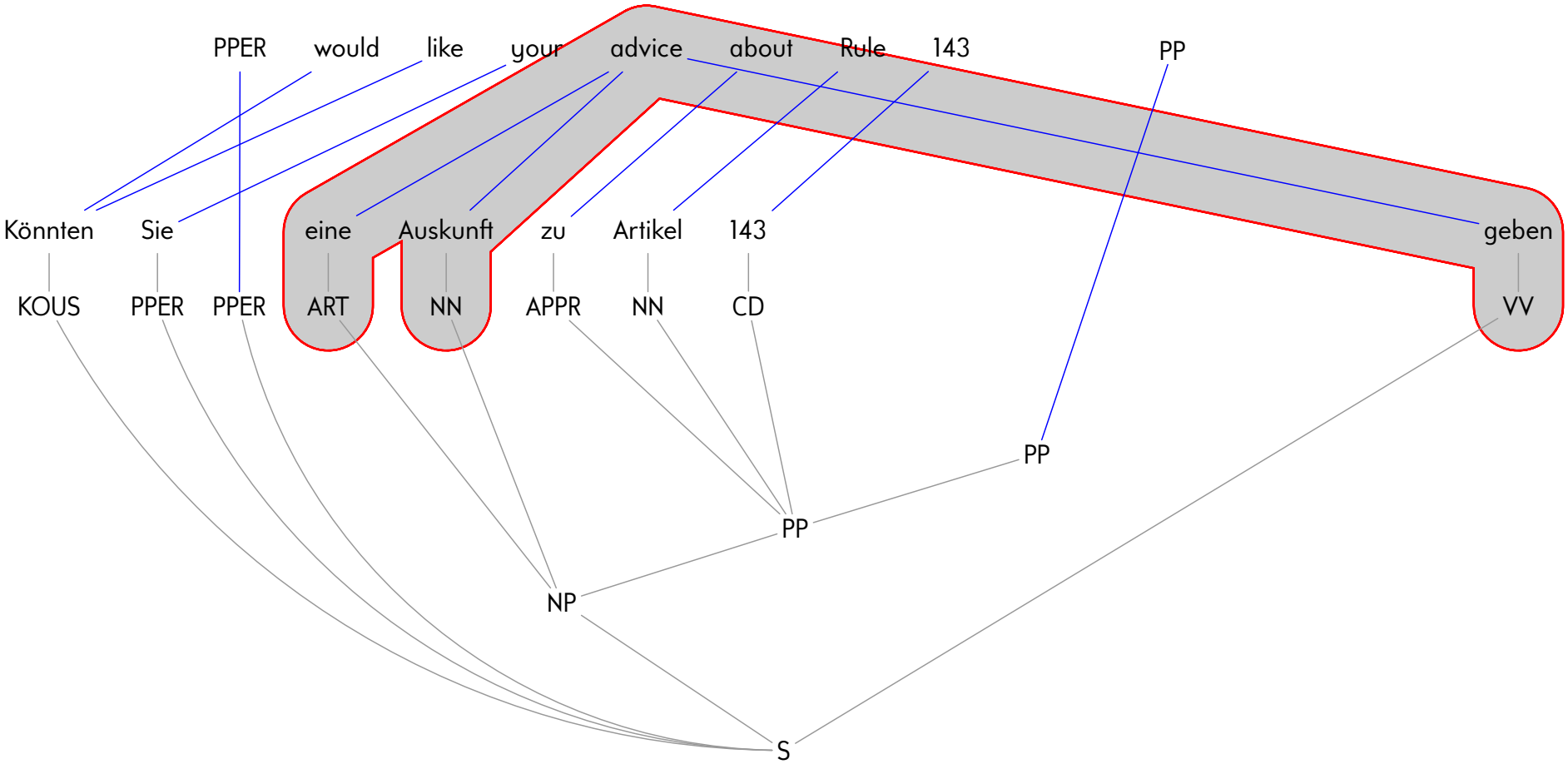
(extractable productions marked in red)



variant of [M.: How to train your multi bottom-up tree transducer. *Proc. ACL*, 2011]

Production Extraction

(extractable productions marked in red)



variant of [M.: How to train your multi bottom-up tree transducer. *Proc. ACL*, 2011]

Synchronous Multi Tree Substitution Grammars

Advantages

- complicated discontinuities
- implemented in framework ‘Moses’
  [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]
- binarizable, composable

Synchronous Multi Tree Substitution Grammars

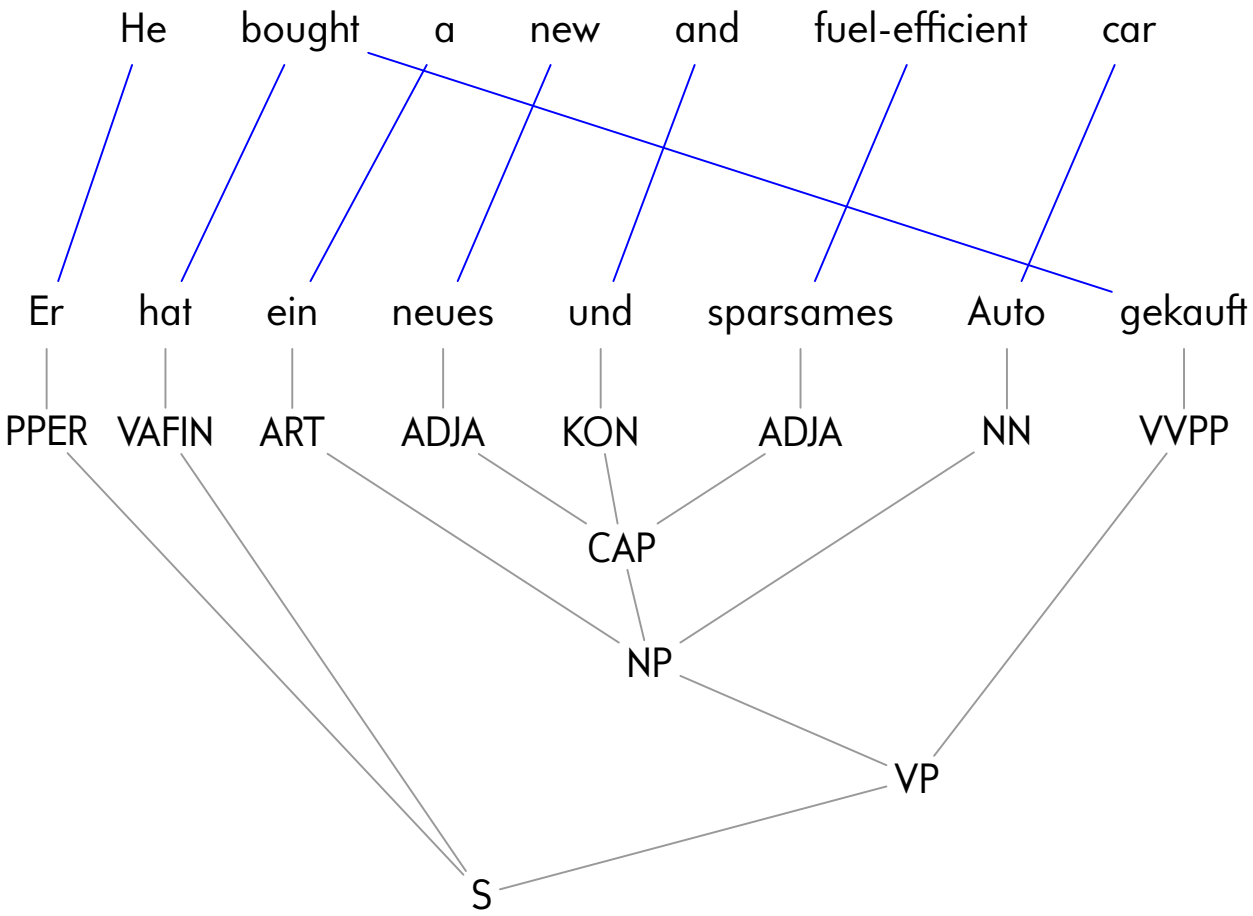
Advantages

- complicated discontinuities
- implemented in framework ‘Moses’
  [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]
- binarizable, composable

Disadvantages

- output non-regular (trees) or non-context-free (strings)
- not symmetric (input context-free; output not)

Discontinuity



Evaluation

Task	System	BLEU
English → German	STSG	15.2
	SMTSG	*15.9
English → Arabic	STSG	48.3
	SMTSG	*49.1
English → Chinese	STSG	17.7
	SMTSG	*18.4
English → Polish	STSG	21.3
	SMTSG	*23.4
English → Russian	STSG	24.7
	SMTSG	*26.1

from [Seemann, Braune, M.: A systematic evaluation of MBOT in statistical machine translation. *Proc. MT-Summit*, 2015] and [Seemann, M.: Discontinuous statistical machine translation with target-side dependency syntax. *Proc. WMT*, 2015]

Overview

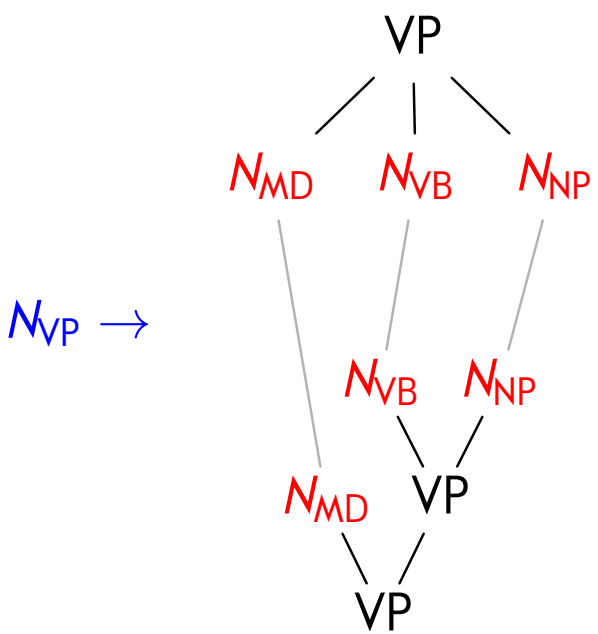
1. Background
2. Extending the Expressive Power
3. Investigating their Expressive Power

Synchronous Grammars

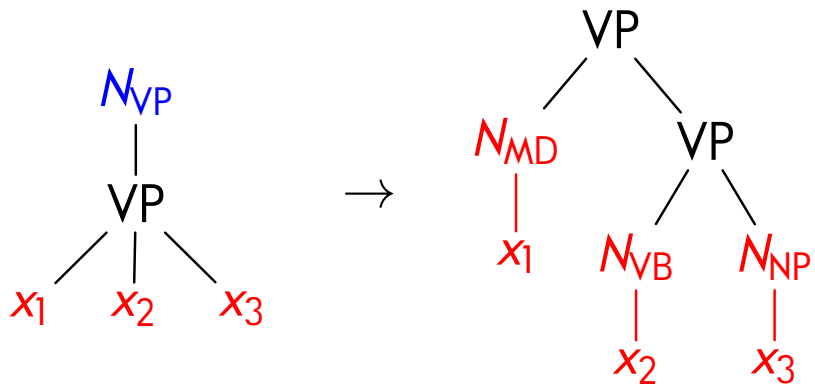
Notes

- tree-to-tree models easier for theoretical investigation
- strongly related to tree transducers
- we disallow trivial input sides of just a nonterminal (ε-free)

Synchronous grammar:



Tree transducer:



Synchronous Grammars

Major linear tree transducers:

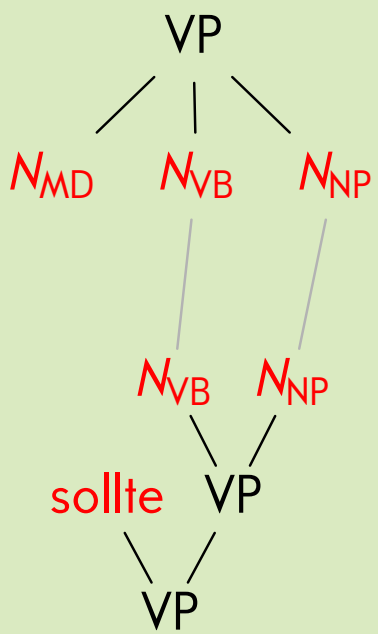
synchronization / input sides	**bijective**	**injective** (output → input)
shallow	nondeleting top-down ...	top-down ...
general	nondeleting extended ...	extended ...

Further distinction

- allow productions on disconnected input nonterminals
  → regular look-ahead
- allow arbitrary trees for disconnected input nonterminals
  → no look-ahead

Synchronous Grammars

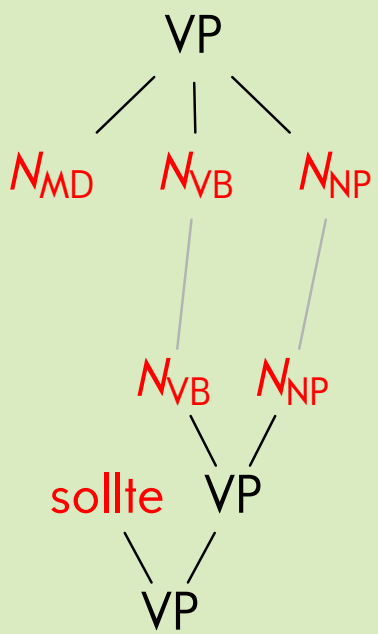
Illustration



- no look-ahead: can plug any (terminal) tree for N_{MD}
  [e.g., NP(DT(the), NN(tower))]

Synchronous Grammars

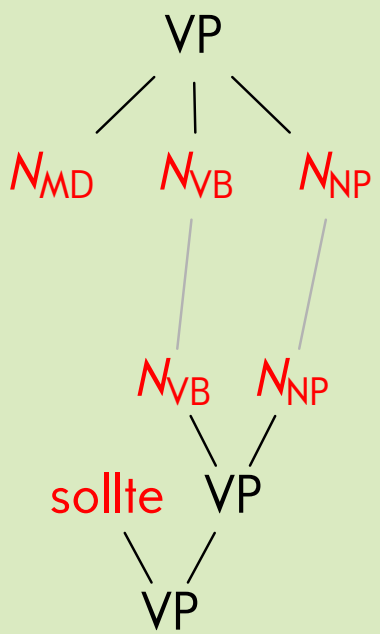
Illustration



- no look-ahead: can plug any (terminal) tree for N_{MD}
  [e.g., NP(DT(the), NN(tower))]
- regular look-ahead: use special "no-output"-productions $N \to (r)$
  [e.g., $N_{MD} \to$ (MD(should))]

Synchronous Grammars

Illustration

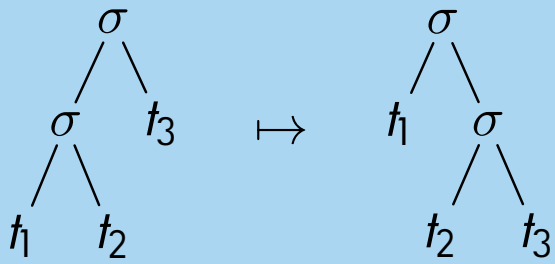


- no look-ahead: can plug any (terminal) tree for N_{MD}
  [e.g., NP(DT(the), NN(tower))]
- regular look-ahead: use special "no-output"-productions $N \to (r)$
  [e.g., $N_{MD} \to$ (MD(should))]
- SMTSG always have regular look-ahead
  (any number of components includes 0)

Synchronous Grammars

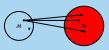
Evaluation Criteria


rotations implementable? (for arbitrary t_1, t_2, t_3)




symmetric?


domain regular?


range regular?

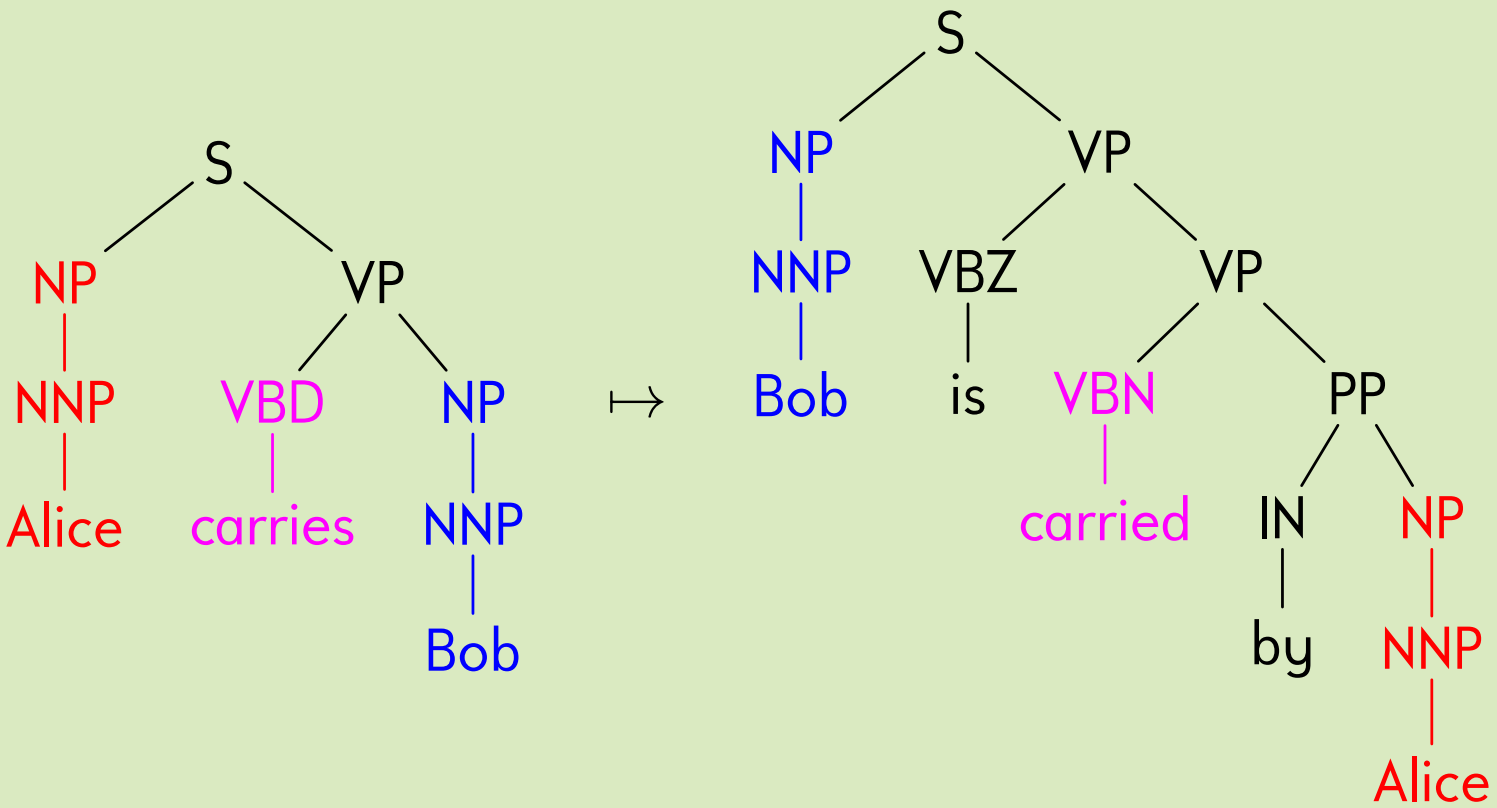

closed under composition?

following [Knight: Capturing practical natural language transformations. *Machine Translation* 21(2), 2007]
and [May, Knight, Vogler: Efficient inference through cascades of weighted tree transducers. *Proc. ACL*, 2010]

Icons by interactivemania (`http://www.interactivemania.com/`) and UN Office for the Coordination of Humanitarian Affairs

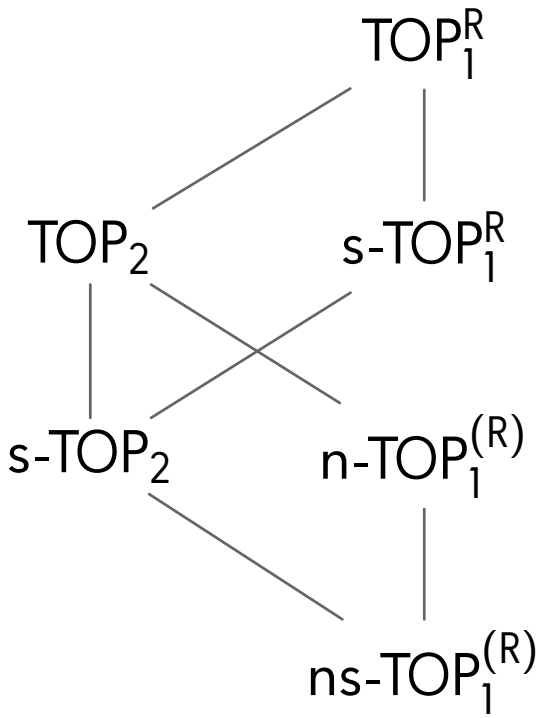
Synchronous Grammars

Illustration of rotations

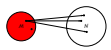
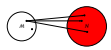


Top-down Tree Transducer

Hasse diagram with composition closure indicated in subscript:

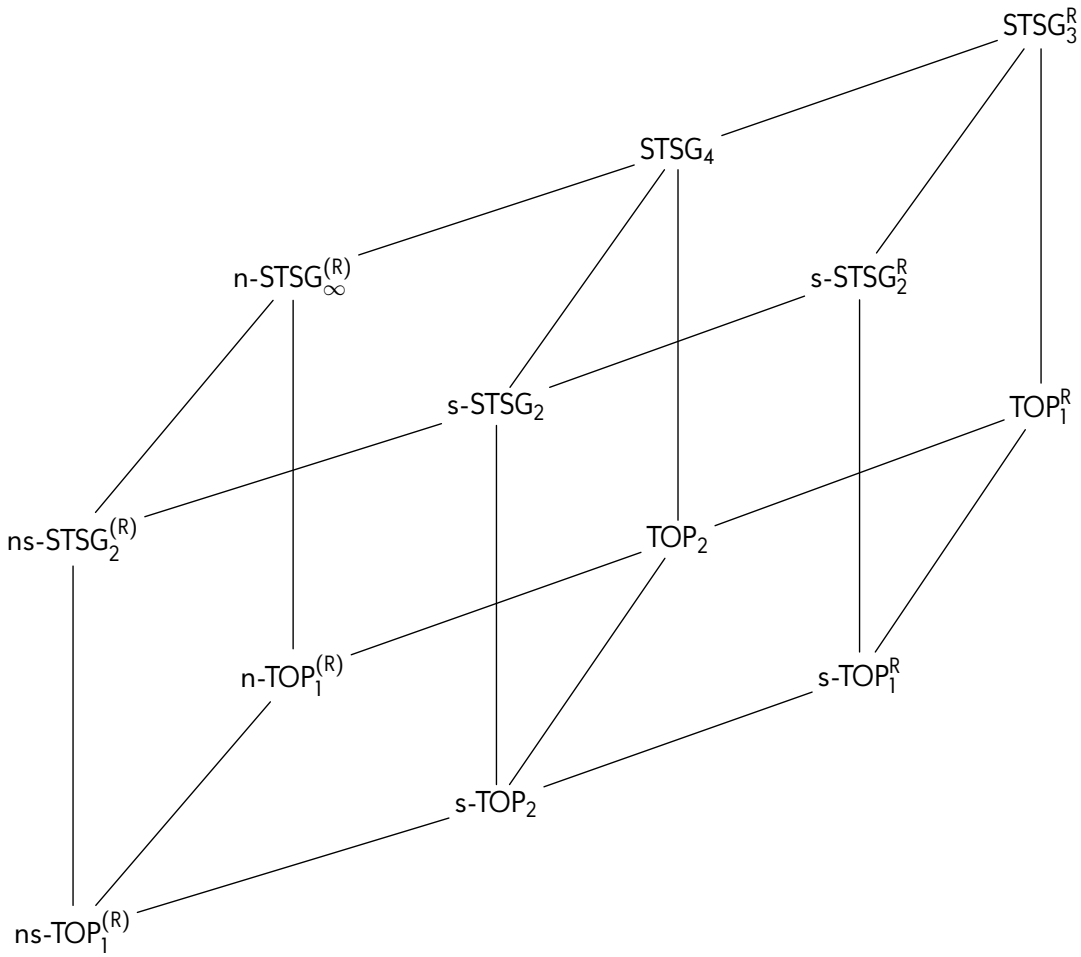


Top-down Tree Transducer

Model \ Criterion					
ns-TOP	✗	✗	✓	✓	✓
n-TOP	✗	✗	✓	✓	✓
s-TOP	✗	✗	✓	✓	$\times_2$
s-TOP^R	✗	✗	✓	✓	✓
TOP	✗	✗	✓	✓	$\times_2$
TOP^R	✗	✗	✓	✓	✓

Synchronous Tree Substitution Grammars

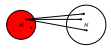
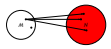
Hasse diagram with the composition closure indicated in subscript:



composition closures by

[Engelfriet, Fülöp, M.: Composition closure of linear extended top-down tree transducers. *Theory of Computing Systems*, to appear 2015]

Synchronous Tree Substitution Grammars

Model \ Criterion					
n-TOP	✗	✗	✓	✓	✓
TOP	✗	✗	✓	✓	$✗_{2}$
TOP^{R}	✗	✗	✓	✓	✓
ns-STSG	✓	✓	✓	✓	$✗_{2}$
n-STSG	✓	✗	✓	✓	$✗_{\infty}$
$s\text{-STSG}^{(R)}$	✓	✗	✓	✓	$✗_{2}$
STSG	✓	✗	✓	✓	$✗_{4}$
$STSG^{R}$	✓	✗	✓	✓	$✗_{3}$

Synchronous Multi Tree Substitution Grammars

Advantages of SMTSG

- always have regular look-ahead
- can always be made nondeleting & shallow
- closed under composition

[Engelfriet, Lilin, M.: Extended multi bottom-up tree transducers — composition and decomposition. *Acta Informatica* 46(8), 2009]

Synchronous Multi Tree Substitution Grammars

Advantages of SMTSG

- always have regular look-ahead
- can always be made nondeleting & shallow
- closed under composition

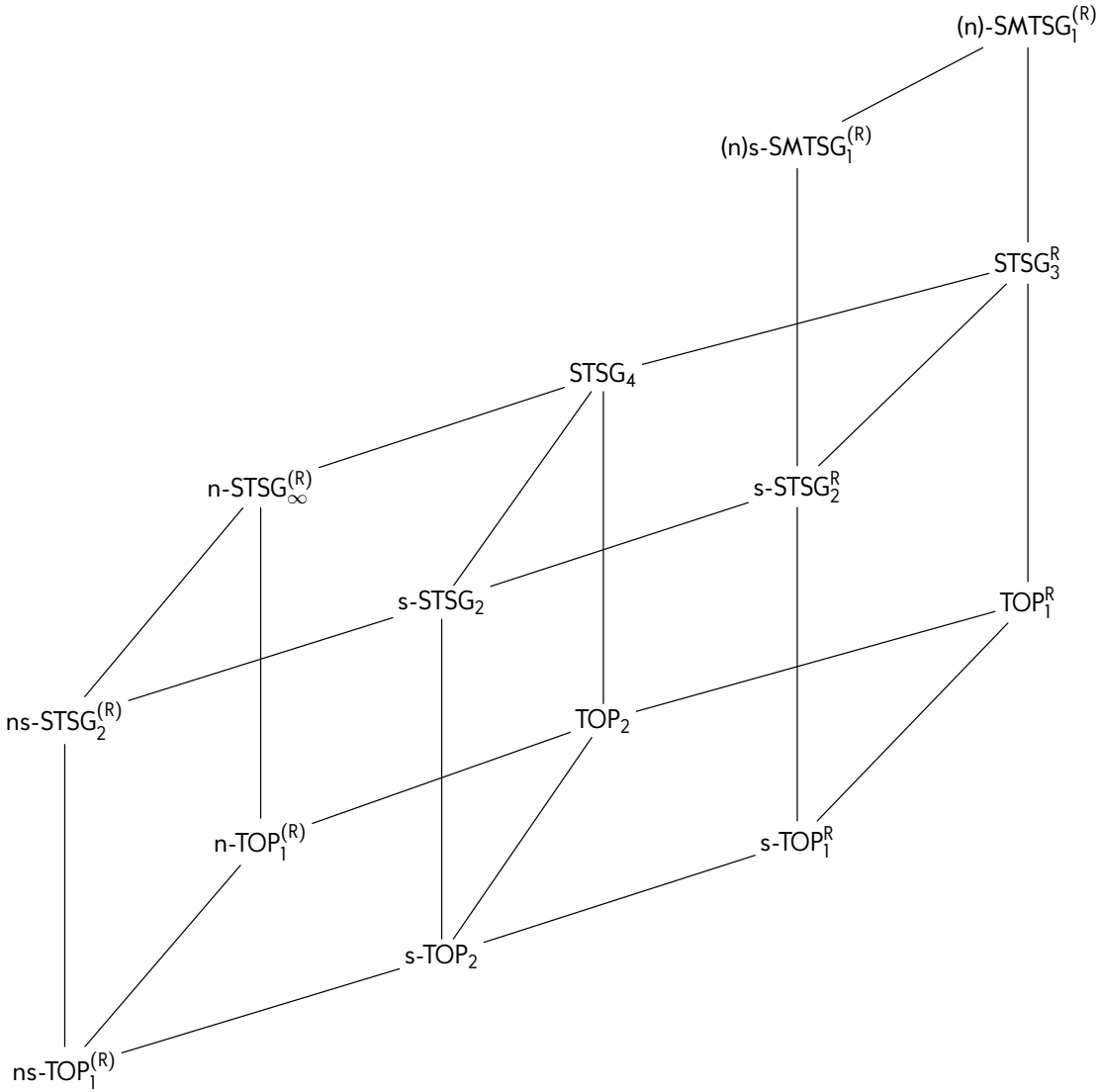
Disadvantages of SMTSG

- non-regular range

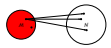
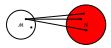
[Engelfriet, Lilin, M.: Extended multi bottom-up tree transducers — composition and decomposition. *Acta Informatica* 46(8), 2009]

Synchronous Multi Tree Substitution Grammars

Hasse diagram with the composition closure indicated in subscript:



Synchronous Multi Tree Substitution Grammars

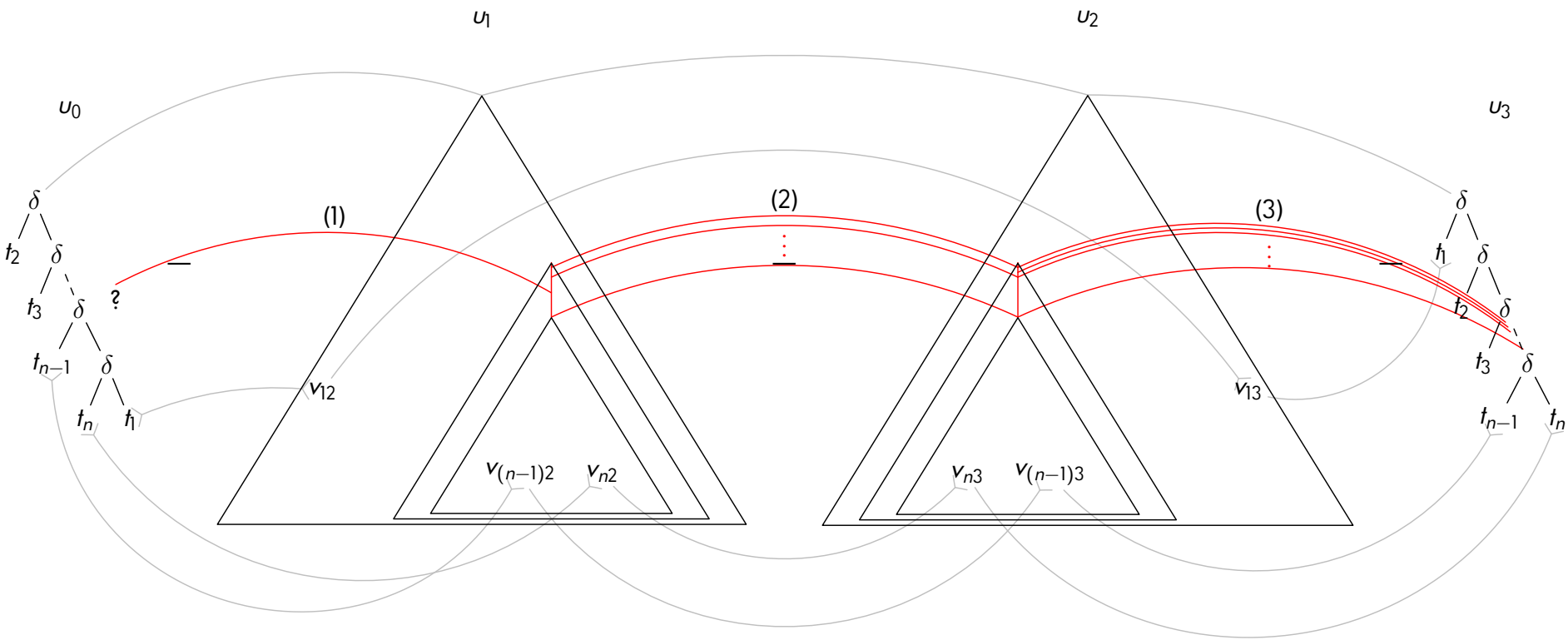
Model \ Criterion					
n-TOP	✗	✗	✓	✓	✓
TOP	✗	✗	✓	✓	$\times_2$
TOP^R	✗	✗	✓	✓	✓
ns-STSG	✓	✓	✓	✓	$\times_2$
n-STSG	✓	✗	✓	✓	$\times_\infty$
s-STSG$^{(R)}$	✓	✗	✓	✓	$\times_2$
STSG	✓	✗	✓	✓	$\times_4$
STSG^R	✓	✗	✓	✓	$\times_3$
(n)s-SMTSG$^{(R)}$	✓	✗	✓	✗	✓
(n)-SMTSG$^{(R)}$	✓	✗	✓	✗	✓
reg.-range SMTSG	✓	✗	✓	✓	✓
symmetric SMTSG	✓	✓	✓	✓	✓

(string-level) range characterization by
[Gildea: On the string translations produced by multi bottom-up tree transducers. *Computational Linguistics* 38(3), 2012]

Synchronous Multi Tree Substitution Grammars

Theorem

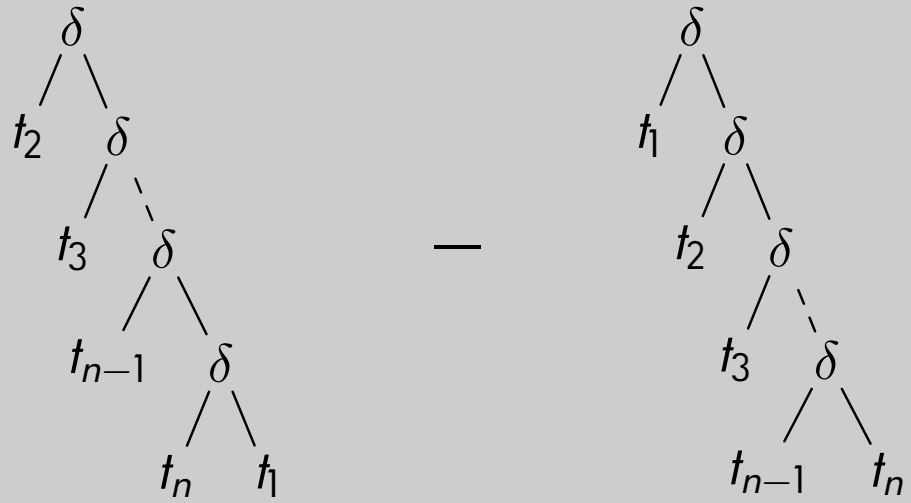
$$\left(\mathrm{STSG}^{\mathrm{R}}\right)^3 \subsetneq \text{reg.-range SMTSG}$$



[M.: The power of weighted regularity-preserving multi bottom-up tree transducers. *Int. J. Found. Comput. Sci.*, to appear 2015]

Synchronous Multi Tree Substitution Grammars

Counterexample relation



- abstracts a well-known linguistic transformation called topicalization
- implementable by SMTSG, but not by any composition of STSG

Synchronous Multi Tree Substitution Grammars

Illustration of topicalization

- It rained yesterday night.

  Topicalized: Yesterday night, it rained.

Synchronous Multi Tree Substitution Grammars

Illustration of topicalization

- It rained yesterday night.

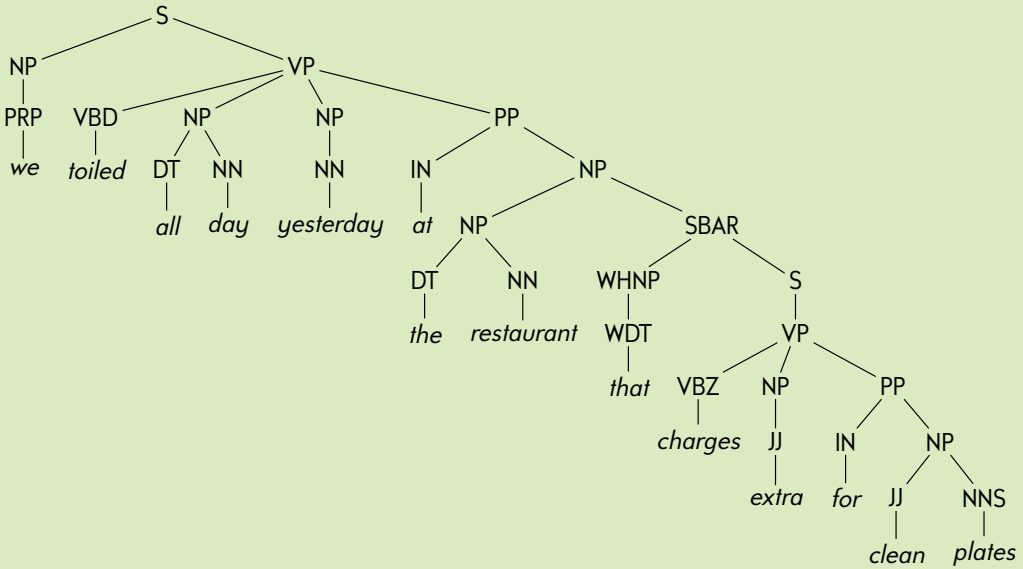
  Topicalized: Yesterday night, it rained.

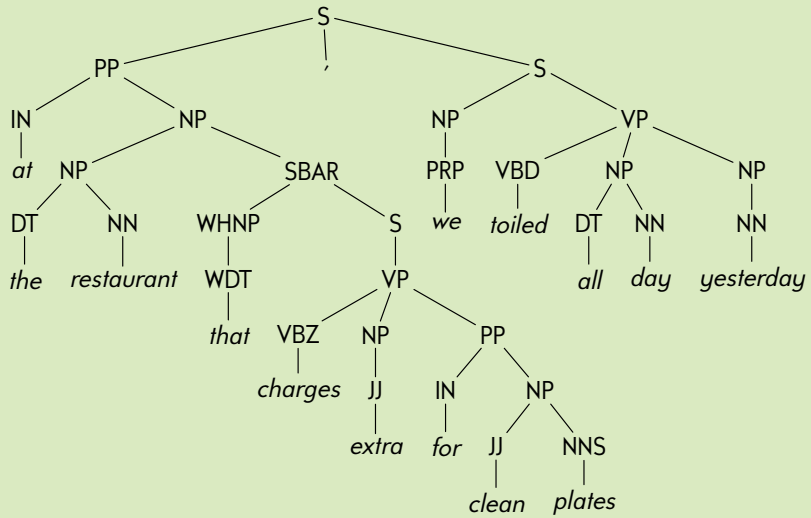
- We toiled all day yesterday
  at the restaurant that charges extra for clean plates.

  Topicalized: At the restaurant that charges extra for clean plates,
  we toiled all day yesterday.

Synchronous Multi Tree Substitution Grammars

On the tree level





Summary

Contributions

- SMTSG implementation and evaluation
  [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]
  [Seemann, Braune, M.: String-to-tree multi bottom-up tree transducers. *Proc. ACL*, 2015]
  [Seemann, Braune, M.: A systematic evaluation of MBOT in statistical machine translation. *Proc. MT-Summit*, 2015]
- characterization of expressive power of STSG and SMTSG
  [Engelfriet, Lilin, M.: Extended multi bottom-up tree transducers — composition and decomposition. *Acta Informatica* 46(8), 2009]
  [Engelfriet, Fülöp, M.: Composition closure of linear extended top-down tree transducers. *Theory of Computing Systems*, 2015]
  [M.: The power of weighted regularity-preserving multi bottom-up tree transducers. *Int. J. Found. Comput. Sci.*, 2015]
- new proof technique (based on synchronization links)
  [Fülöp, M.: Linking theorems for tree transducers. Manuscript, 2014]
  similar ideas used in
  [Bojanczyk: Transducers with origin information. *Proc. ICALP*, 2014]
  [Filiot, Maneth, Reynier, Talbot: Decision problems of tree transducers with origin. *Proc. ICALP*, 2015]

Summary

Contributions

- SMTSG implementation and evaluation
  [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]
  [Seemann, Braune, M.: String-to-tree multi bottom-up tree transducers. *Proc. ACL*, 2015]
  [Seemann, Braune, M.: A systematic evaluation of MBOT in statistical machine translation. *Proc. MT-Summit*, 2015]
- characterization of expressive power of STSG and SMTSG
  [Engelfriet, Lilin, M.: Extended multi bottom-up tree transducers — composition and decomposition. *Acta Informatica* 46(8), 2009]
  [Engelfriet, Fülöp, M.: Composition closure of linear extended top-down tree transducers. *Theory of Computing Systems*, 2015]
  [M.: The power of weighted regularity-preserving multi bottom-up tree transducers. *Int. J. Found. Comput. Sci.*, 2015]
- new proof technique (based on synchronization links)
  [Fülöp, M.: Linking theorems for tree transducers. Manuscript, 2014]
  similar ideas used in
  [Bojanczyk: Transducers with origin information. *Proc. ICALP*, 2014]
  [Filiot, Maneth, Reynier, Talbot: Decision problems of tree transducers with origin. *Proc. ICALP*, 2015]

Thank you for the attention.