Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based SMT

Matthias Huck, Hieu Hoang, Philipp Koehn

School of informatics
University of Edinburgh

25 October 2014

Introduction

Feature-based integration of syntactic information
into GHKM string-to-tree translation

- Preference grammars:
  soft target-side syntax
  - Target syntax as a feature
    rather than via labeled non-terminals in the SCFG

- Soft syntactic constraints:
  non-restrictive source-syntactic enhancement
  - No hard source-syntactic constraints
    (as in standard tree-to-tree translation)
    imposed in extraction or decoding

Our empirical evaluation: English→German WMT task

Related Work

- GHKM string-to-tree translation: *Galley et al. (2004)*
- Open-source Moses implementation for GHKM translation
  - GHKM rule extraction: *Williams and Koehn (2012)*
  - Decoding with CYK+ parsing and cube pruning: *Hoang et al. (2009)*
  - Competitive results for European language pairs: *Nadejde et al. (2013); Williams et al. (2014)*
- Preference grammars: beneficial as a syntactic extension of hierarchical systems *(Venugopal et al., 2009; Stein et al., 2010)*
- Soft syntactic constraints: related source-syntactic techniques improved hierarchical *(Marton and Resnik, 2008; Vilar et al., 2008; Hoang and Koehn, 2010)* and other syntax-based systems *(Zhang et al., 2011; Huang et al., 2013)* on Chinese→English and Arabic→English tasks

Preference Grammars

- Target-side non-terminals not decorated with syntactic labels, but with a single generic non-terminal symbol

baseline	**preference grammar system**
$X, ADJD \rightarrow \langle \text{present}, \text{anwesend} \rangle$ $X, ADV \rightarrow \langle \text{present}, \text{anwesend} \rangle$ $X, AP\text{-}PD \rightarrow \langle \text{present}, \text{anwesend} \rangle$ $\ldots$	$X, X \rightarrow \langle \text{present}, \text{anwesend} \rangle$

- Distribution of implicit target label vectors stored as additional information with each translation rule

$$X, X \rightarrow \langle \text{present}, \text{anwesend} \rangle \;\#\; (ADJD)\ 0.98\ (ADV)\ 0.001\ (AP\text{-}PD)\ 0.01\ \ldots$$

- Computation of a tree-wellformedness feature during decoding

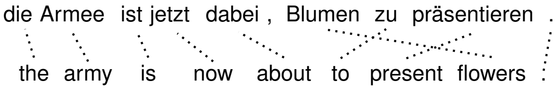
Soft Source Syntactic Constraints

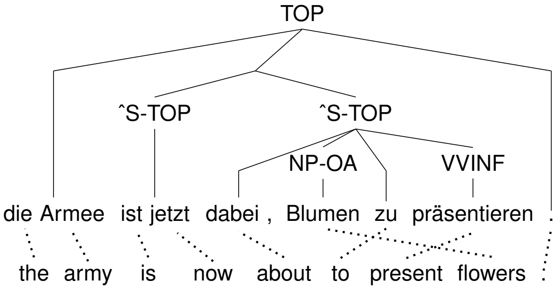
- Parse source-side data as well
- GHKM extractor stores an additional rule property: source syntactic label vectors

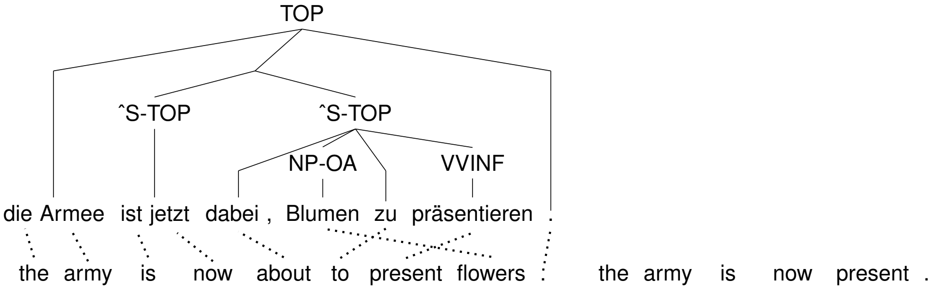
- Provide parsed input data to the decoder
- During decoding, score matches and mismatches of source label vectors and input labels
- Soft features, no hard constraints

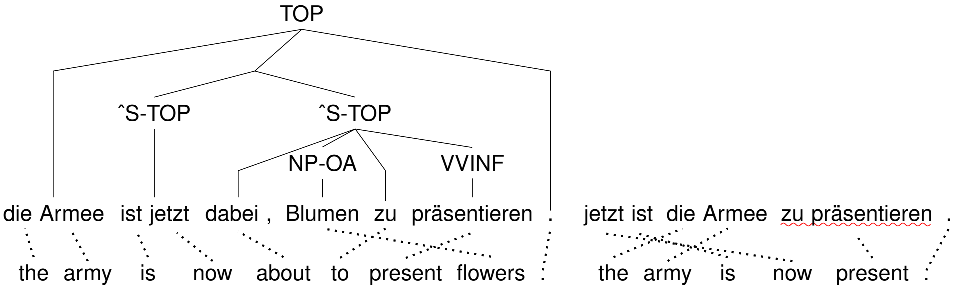
rule	**source label vectors**
$X, \mathit{VP\text{-}OC} \rightarrow \langle$present, zu präsentieren$\rangle$	$(\mathit{VB}), (\mathit{NN})$
$X, \mathit{ADJD} \rightarrow \langle$present, anwesend$\rangle$	$(\mathit{ADJP}), (\mathit{ADVP}), \ldots$
$X, \mathit{TOP} \rightarrow \langle X^{\sim 0}$ is now $X^{\sim 1}$. , jetzt ist $\mathit{NP\text{-}SB}^{\sim 0}$ $\mathit{VP\text{-}OC}^{\sim 1}$.$\rangle$	$(\mathit{TOP}, \mathit{NP}, \mathit{ADJP}), \ldots$
$X, \hat{~}\mathit{S\text{-}TOP} \rightarrow \langle$is $X^{\sim 0}$ $X^{\sim 1}$, ist $\mathit{ADV}^{\sim 0}$ $\mathit{ADJD}^{\sim 1}\rangle$	$(\mathit{VP}, \mathit{ADVP}, \mathit{ADJP}), \ldots$

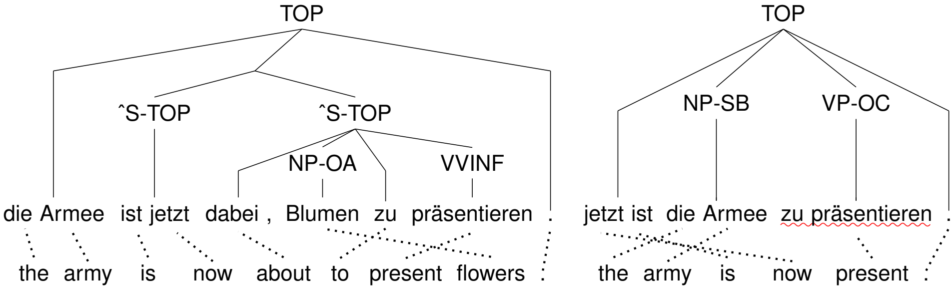
the army is now about to present flowers .


die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .

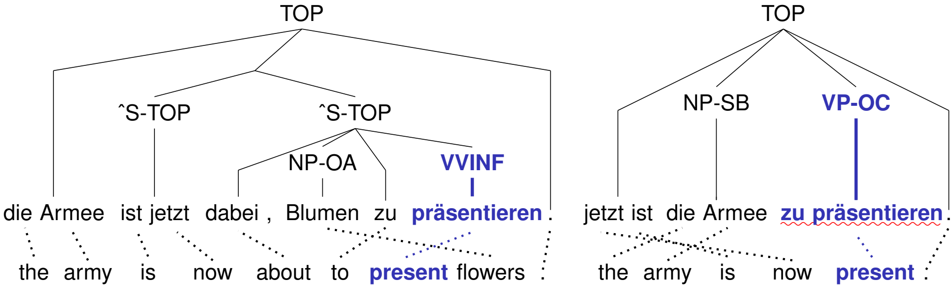

TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .

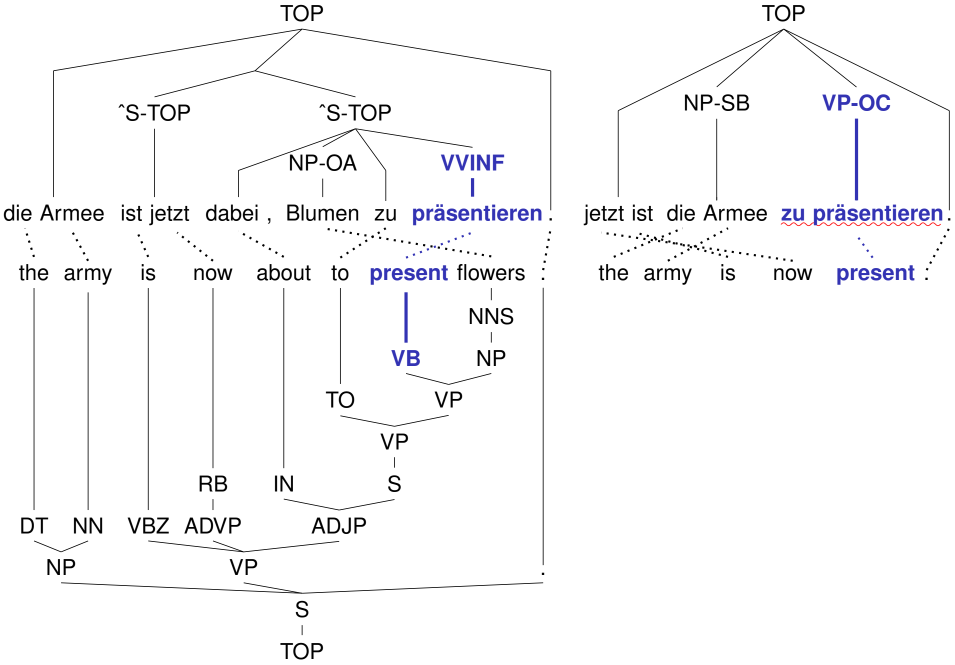

TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers :
the army is now present .

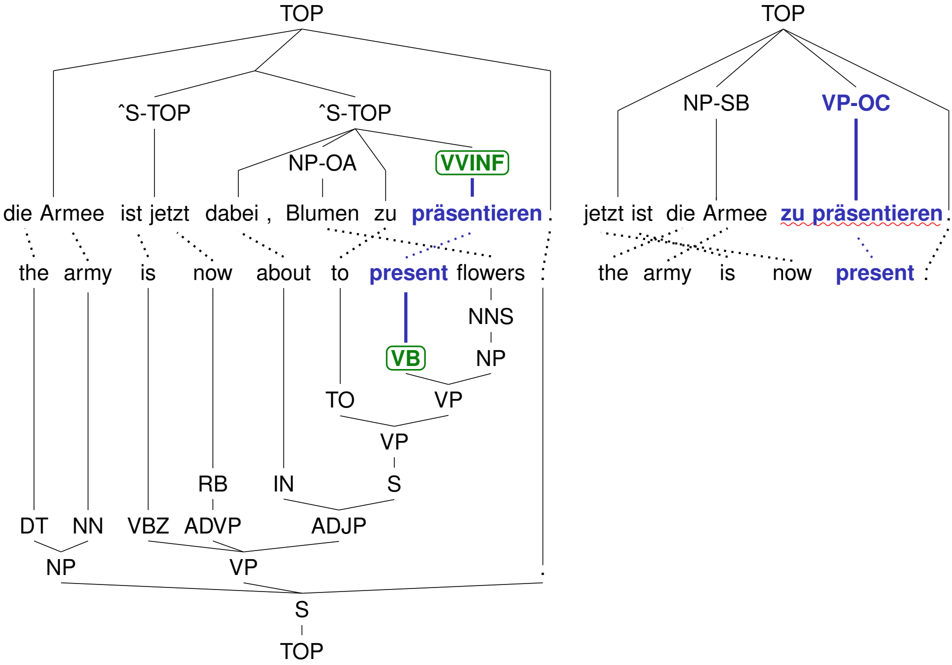

TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
jetzt ist die Armee zu präsentieren .
the army is now present .



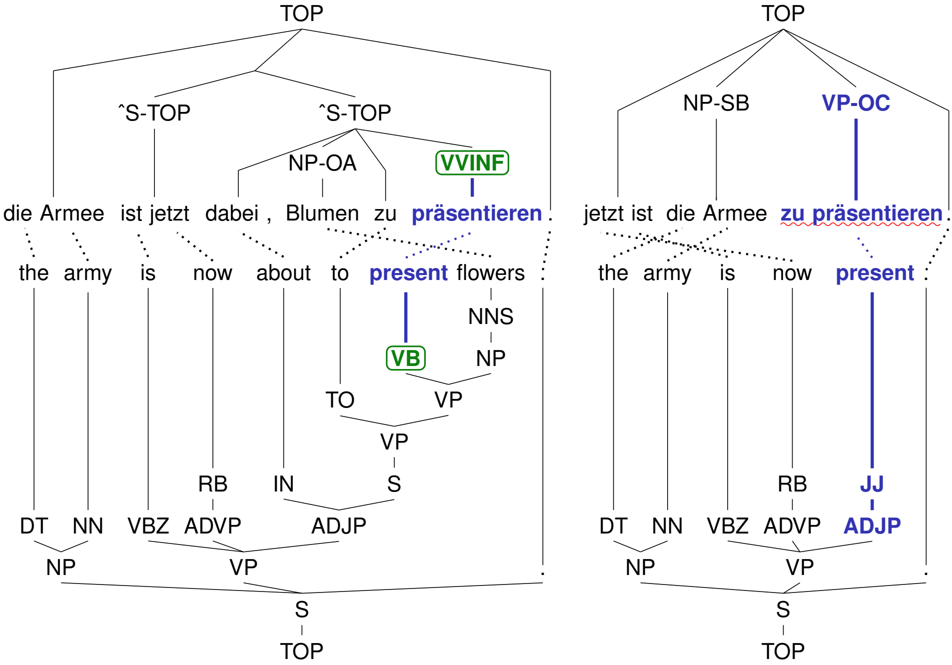
TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .


TOP
^S-TOP
^S-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .

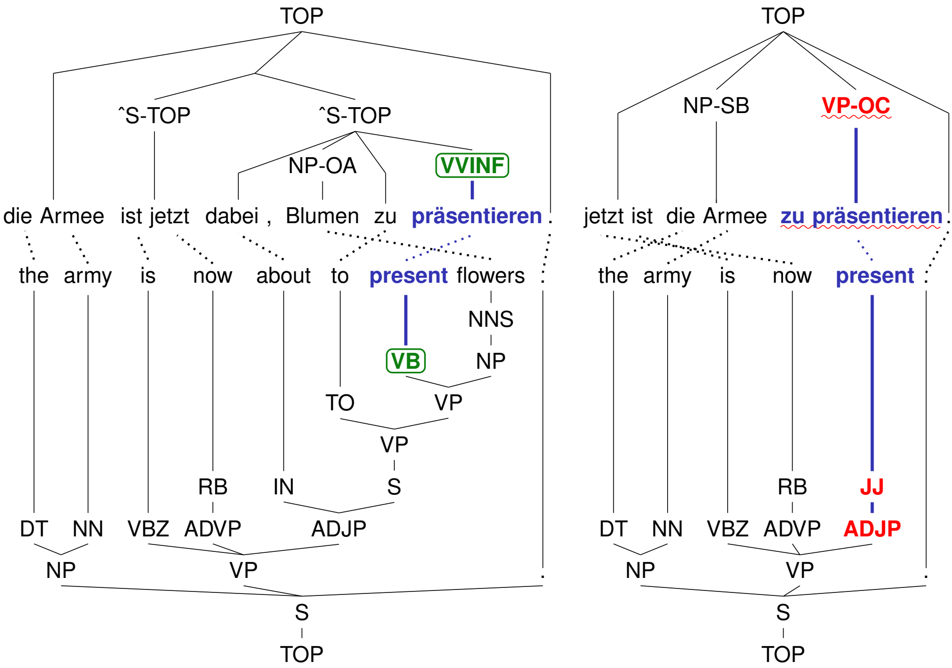

TOP
^S-TOP
^S-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .

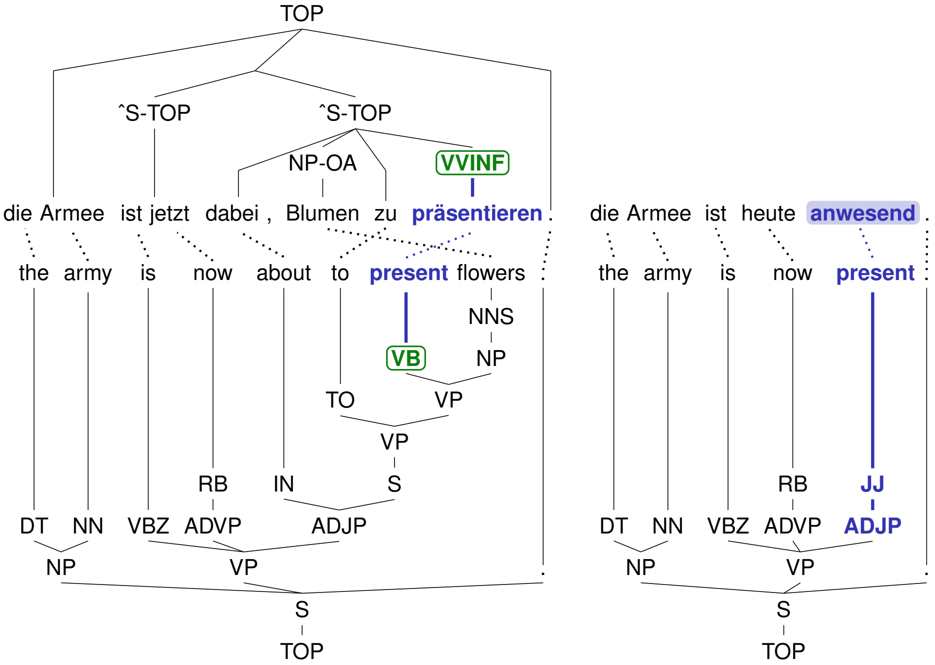


TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .

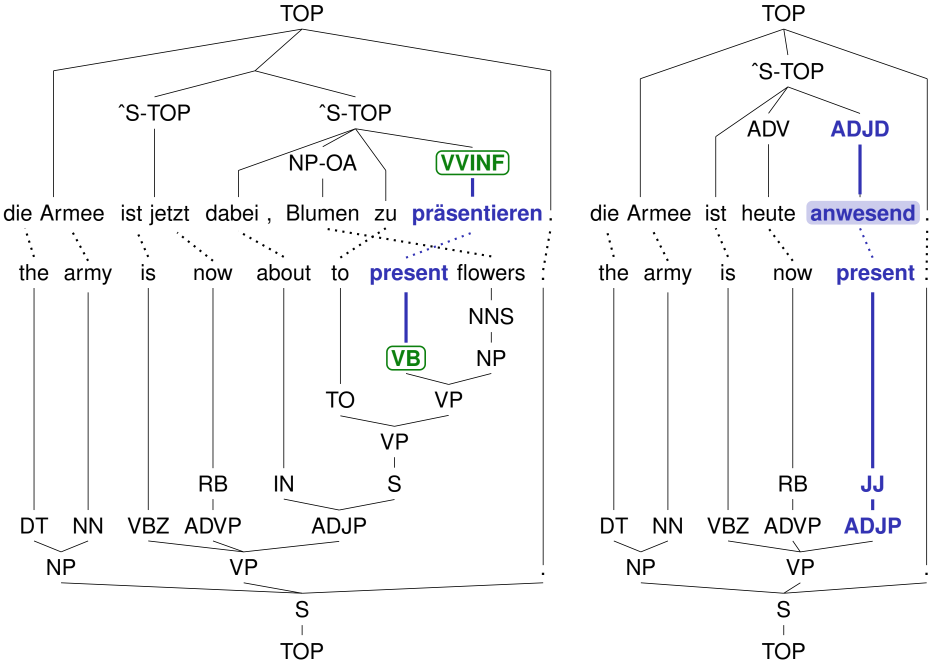


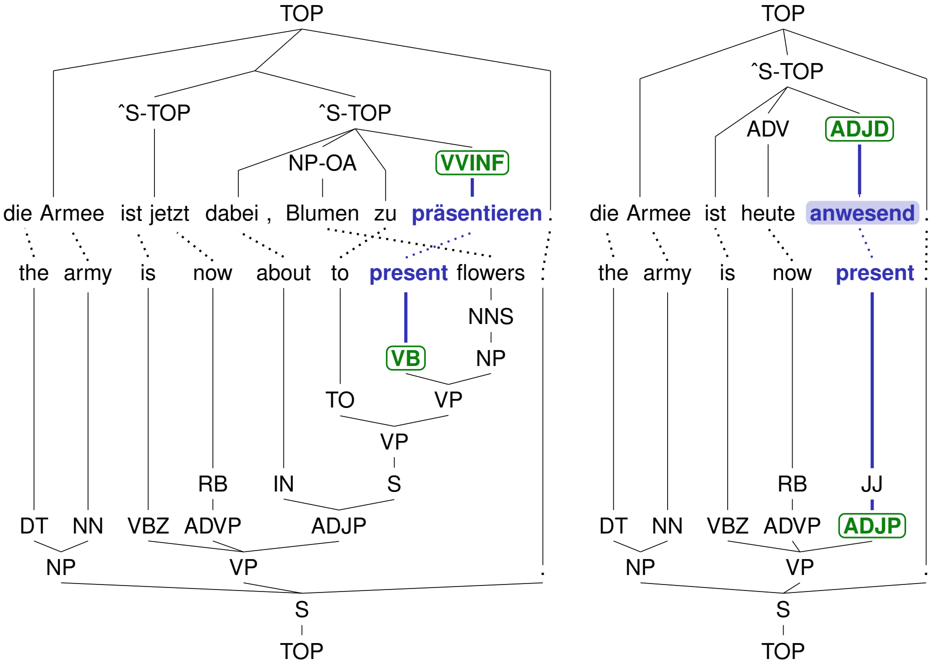
TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .
RB
JJ
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP


TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
NP-SB
VP-OC
jetzt ist die Armee zu präsentieren .
the army is now present .
RB
JJ
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP



TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
die Armee ist heute anwesend .
the army is now present .
RB
JJ
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP


TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
S
RB
IN
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
ˆS-TOP
ADV
ADJD
die Armee ist heute anwesend .
the army is now present .
RB
JJ
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP



TOP
ˆS-TOP
ˆS-TOP
NP-OA
VVINF
die Armee ist jetzt dabei , Blumen zu präsentieren .
the army is now about to present flowers .
NNS
VB
NP
TO
VP
VP
RB
IN
S
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP
TOP
ˆS-TOP
ADV
ADJD
die Armee ist heute anwesend .
the army is now present .
RB
JJ
DT
NN
VBZ
ADVP
ADJP
NP
VP
.
S
TOP

Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based SMT

School of informatics

Matthias Huck
mhuck@inf.ed.ac.uk

Hieu Hoang
hhoang@inf.ed.ac.uk

Philipp Koehn
pkoehn@inf.ed.ac.uk

The University of Edinburgh

Motivation

Feature-based integration of syntactic information into GHKM string-to-tree statistical machine translation

- The hard target-side syntactic constraints that are imposed by the target non-terminal labels might be too restrictive. Should we soften them?

> **Preference grammars** promote syntactic well-formedness on the target language side while also allowing for derivations that are not linguistically motivated (as in hierarchical translation)

- Tree-to-tree translation often underperforms. How can we effectively enhance a strong string-to-tree baseline with source-side syntactic information?

> **Soft syntactic constraints** augment the system with additional source-side syntax features while not modifying the set of string-to-tree translation rules or the baseline feature scores

Preference Grammars

Training:

- Target-side non-terminals not decorated with syntactic labels, but with a **single generic non-terminal symbol**
- Extracted rules which differ only with respect to their non-terminal labels are collapsed to a single entry in the rule table, and their rule counts are pooled

baseline	preference grammar system
$X, \mathit{ADJD} \rightarrow \langle$present, anwesend$\rangle$ $X, \mathit{ADV} \rightarrow \langle$present, anwesend$\rangle$ $X, \mathit{AP\text{-}PD} \rightarrow \langle$present, anwesend$\rangle$ …	$X, X \rightarrow \langle$present, anwesend$\rangle$

- **Distribution of implicit target label vectors** stored as additional information with each translation rule

$X, X \rightarrow \langle$present, anwesend$\rangle$ # (ADJD) 0.98 (ADV) 0.001 $(\mathit{AP\text{-}PD})$ 0.01 …

Decoding:

- Computation of a **tree-wellformedness feature**

Soft Source Syntactic Constraints

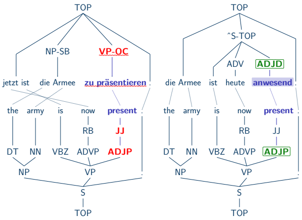
Training:

- Provide **syntactic parses of the source side** of the data
- GHKM extractor collects the **source syntactic labels that cover the source-side span of non-terminals**
- Sets of source syntactic label vectors are memorized with the rules as an **additional property**

Decoding:

- Input data parsed in a preprocessing step
- Computation of three dense **features which score matches and mismatches of input labels** and source label vectors that are associated with translation rules

rule	source label vectors
$X, \mathit{VP\text{-}OC} \rightarrow \langle$present, zu präsentieren$\rangle$	$(\mathit{VB}), (\mathit{NN})$
$X, \mathit{ADJD} \rightarrow \langle$present, anwesend$\rangle$	$(\mathit{ADJP}), (\mathit{ADVP}), \ldots$
$X, \mathit{TOP} \rightarrow \langle X^{\sim 0}$ is now $X^{\sim 1}$. , jetzt ist $\mathit{NP\text{-}SB}^{\sim 0}\ \mathit{VP\text{-}OC}^{\sim 1}$.$\rangle$	$(\mathit{TOP}, \mathit{NP}, \mathit{ADJP}), \ldots$
$X, \hat{}\mathit{S\text{-}TOP} \rightarrow \langle$is $X^{\sim 0}\ X^{\sim 1}$, ist $\mathit{ADV}^{\sim 0}\ \mathit{ADJD}^{\sim 1}\rangle$	$(\mathit{VP}, \mathit{ADVP}, \mathit{ADJP}), \ldots$



Dense features:

- a source syntactic label vector fully matches the input labels
- left-hand side non-terminal label mismatch
- number of right-hand side non-terminals label mismatches

Experimental Setup

- English→German WMT task (4.5 M sentence pairs)
- Syntactic annotation: BitPar for German, Berkeley Parser for English
- Right binarization of target parse trees
- SAMT-style composite labels on source side
- Singleton hierarchical rules are discarded
- No more than 50 most frequent label vectors per rule stored
- Decoding with CYK+ and cube pruning
- Tuning with batch MIRA
- Development set: 2000 selected sentences from newstest2008-2012

Experimental Results (English→German)

system	dev		newstest2013		newstest2014	
	BLEU	TER	BLEU	TER	BLEU	TER
GHKM string-to-tree baseline	34.7	47.3	20.0	63.3	19.4	65.6
+ hard source syntactic constraints	34.6	47.4	19.9	63.4	19.4	65.6
+ soft source syntactic constraints	35.1	47.0	20.3	62.7	19.7	64.9
string-to-string (GHKM syntax-directed extraction)	33.8	48.0	19.3	63.8	18.7	66.2
+ preference grammar	33.9	47.7	19.3	63.7	18.8	66.0
+ soft source syntactic constraints	34.6	47.0	19.8	62.9	19.5	65.2

Sparse Features for Soft Syntactic Constraints

- Large number of binary features which depend on the label identity
- Separate weight tuned for each of them
- Optionally: Restrict the number of sparse features by specifying a core set of labels

core = non-composite – plain constituent labels as given by the syntactic parser (no SAMT-style composite labels)
core = dev-min-occ100 – labels in the input data on the development set with minimum occurrence count threshold 100

system (tuned on newstest2012)	newstest2012		newstest2013		newstest2014	
	BLEU	TER	BLEU	TER	BLEU	TER
GHKM string-to-tree baseline	17.9	65.7	19.9	63.2	19.4	65.3
+ soft source syntactic constraints	18.2	65.3	20.3	62.6	19.7	64.7
+ sparse features	18.6	64.9	20.4	62.5	19.8	64.7
+ sparse features (core = non-composite)	18.4	65.1	20.3	62.7	19.8	64.7
+ sparse features (core = dev-min-occ100)	18.4	64.8	20.6	62.2	19.9	64.4

References I

Galley, M., Hopkins, M., Knight, K., and Marcu, D. (2004). What's in a translation rule? In *Proc. of the Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics (HLT-NAACL)*, pages 273–280, Boston, MA, USA.

Hoang, H. and Koehn, P. (2010). Improved Translation with Source Syntax Labels. In *Proc. of the Workshop on Statistical Machine Translation (WMT)*, pages 409–417, Uppsala, Sweden.

Hoang, H., Koehn, P., and Lopez, A. (2009). A Unified Framework for Phrase-Based, Hierarchical, and Syntax-Based Statistical Machine Translation. In *Proc. of the Int. Workshop on Spoken Language Translation (IWSLT)*, pages 152–159, Tokyo, Japan.

Huang, Z., Devlin, J., and Zbib, R. (2013). Factored Soft Source Syntactic Constraints for Hierarchical Machine Translation. In *Proc. of the Conf. on Empirical Methods for Natural Language Processing (EMNLP)*, pages 556–566, Seattle, WA, USA.

References II

Marton, Y. and Resnik, P. (2008). Soft Syntactic Constraints for Hierarchical Phrased-Based Translation. In *Proc. of the Annual Meeting of the Assoc. for Computational Linguistics (ACL)*, pages 1003–1011, Columbus, OH, USA.

Nadejde, M., Williams, P., and Koehn, P. (2013). Edinburgh's Syntax-Based Machine Translation Systems. In *Proc. of the Workshop on Statistical Machine Translation (WMT)*, pages 170–176, Sofia, Bulgaria.

Stein, D., Peitz, S., Vilar, D., and Ney, H. (2010). A Cocktail of Deep Syntactic Features for Hierarchical Machine Translation. In *Proc. of the Conf. of the Assoc. for Machine Translation in the Americas (AMTA)*, Denver, CO, USA.

Venugopal, A., Zollmann, A., Smith, N. A., and Vogel, S. (2009). Preference Grammars: Softening Syntactic Constraints to Improve Statistical Machine Translation. In *Proc. of the Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics (HLT-NAACL)*, pages 236–244, Boulder, CO, USA.

References III

Vilar, D., Stein, D., and Ney, H. (2008). Analysing Soft Syntax Features and Heuristics for Hierarchical Phrase Based Machine Translation. In *Proc. of the Int. Workshop on Spoken Language Translation (IWSLT)*, pages 190–197, Waikiki, HI, USA.

Williams, P. and Koehn, P. (2012). GHKM Rule Extraction and Scope-3 Parsing in Moses. In *Proc. of the Workshop on Statistical Machine Translation (WMT)*, pages 388–394, Montréal, Canada.

Williams, P., Sennrich, R., Nadejde, M., Huck, M., Hasler, E., and Koehn, P. (2014). Edinburgh's Syntax-Based Systems at WMT 2014. In *Proc. of the Workshop on Statistical Machine Translation (WMT)*, pages 207–214, Baltimore, MD, USA.

Zhang, J., Zhai, F., and Zong, C. (2011). Augmenting String-to-Tree Translation Models with Fuzzy Use of Source-side Syntax. In *Proc. of the Conf. on Empirical Methods for Natural Language Processing (EMNLP)*, pages 204–215, Edinburgh, Scotland, UK.