Hierarchical MT with Discontinuous Phrases

Miriam Kaeshammer

kaeshammer@phil.uni-duesseldorf.de

University of Düsseldorf

Tenth Workshop on Statistical Machine Translation
EMNLP 2015, Lisbon, Portugal

Thanks to Laura Kallmeyer and Wolfgang Maier

September 2015

Idea

Use Synchronous LCFRS instead of SCFG for translation modeling

LCFRS: Linear Context-Free Rewriting Systems

(Vijay-Shanker et al., 1987; Weir, 1988)

- mildly context-sensitive formalism
- suitable for the direct modeling of discontinuous constituents
- Probabilistic data-driven parsing with LCFRS is feasible. (Maier, 2010; van Cranenburgh and Bod, 2013; Kallmeyer and Maier, 2013)

Idea

Use Synchronous LCFRS instead of SCFG for translation modeling

LCFRS: Linear Context-Free Rewriting Systems

(Vijay-Shanker et al., 1987; Weir, 1988)

- mildly context-sensitive formalism
- suitable for the direct modeling of discontinuous constituents
- Probabilistic data-driven parsing with LCFRS is feasible. (Maier, 2010; van Cranenburgh and Bod, 2013; Kallmeyer and Maier, 2013)

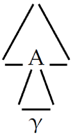
Discontinuous phrase-based SMT (Galley and Manning, 2010)

→ improvement in BLEU score for Chinese-English

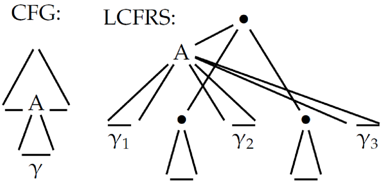
→ this work: hierarchical, tree-based counterpart

Inspiration from Parsing

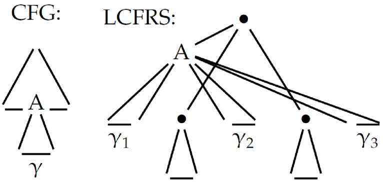
CFG:

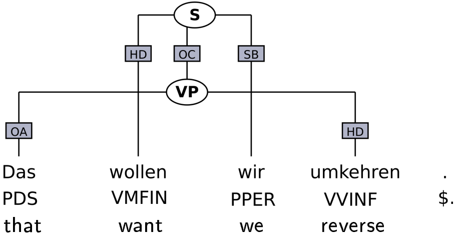


Inspiration from Parsing



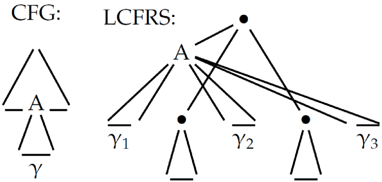
Inspiration from Parsing

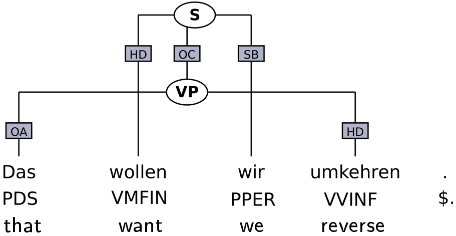




We want to reverse that.

Inspiration from Parsing

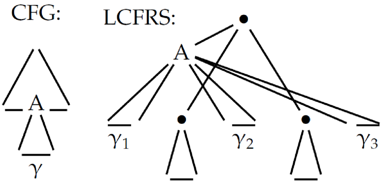


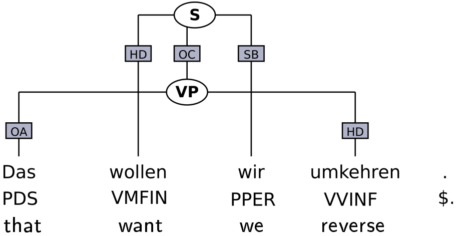


We want to reverse that.

$$VP(X_1, X_2) \rightarrow PDS(X_1)VVINF(X_2)$$

Inspiration from Parsing

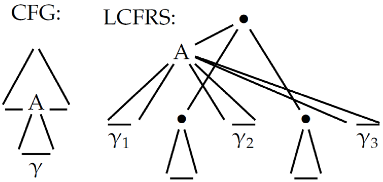


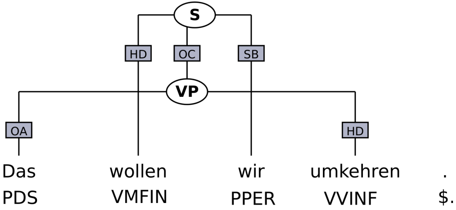


We want to reverse that.

$$VP(X_1, X_2) \rightarrow PDS(X_1)VVINF(X_2)$$
$$S(X_1X_2X_3X_4) \rightarrow VMFIN(X_2)PPER(X_3)VP(X_1, X_4)$$

Inspiration from Parsing

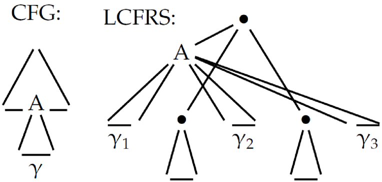


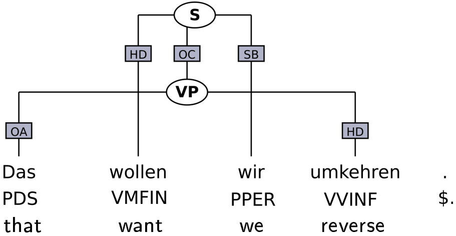


We want to reverse that.

$VP(X_1,X_2) \rightarrow PDS(X_1)VVINF(X_2)$
$S(X_1X_2X_3X_4) \rightarrow$
$\quad VMFIN(X_2)PPER(X_3)VP(X_1,X_4)$
$PDS(\text{Das}) \rightarrow \varepsilon$
$VMFIN(\text{wollen}) \rightarrow \varepsilon$
$PPER(\text{wir}) \rightarrow \varepsilon$
$VVINF(\text{umkehren}) \rightarrow \varepsilon$

Inspiration from Parsing



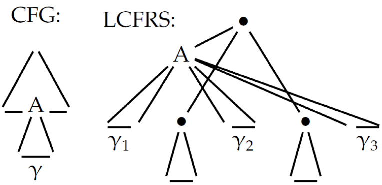


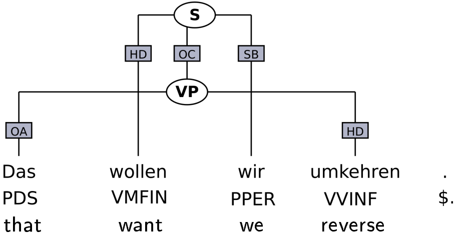
We want to reverse that.

- (u,v)-LCFRS: grammar G with rank u and fan-out v

$$VP(X_1,X_2) \rightarrow PDS(X_1)VVINF(X_2)$$
$$S(X_1X_2X_3X_4) \rightarrow VMFIN(X_2)PPER(X_3)VP(X_1,X_4)$$
$$PDS(\text{Das}) \rightarrow \varepsilon$$
$$VMFIN(\text{wollen}) \rightarrow \varepsilon$$
$$PPER(\text{wir}) \rightarrow \varepsilon$$
$$VVINF(\text{umkehren}) \rightarrow \varepsilon$$

Inspiration from Parsing





We want to reverse that.

- (u,v)-LCFRS: grammar G with rank u and fan-out v
- G with fan-out 1: equivalent to CFG

$VP(X_1,X_2) \rightarrow PDS(X_1)VVINF(X_2)$
$S(X_1X_2X_3X_4) \rightarrow$
$\quad VMFIN(X_2)PPER(X_3)VP(X_1,X_4)$
$PDS(\text{Das}) \rightarrow \varepsilon$
$VMFIN(\text{wollen}) \rightarrow \varepsilon$
$PPER(\text{wir}) \rightarrow \varepsilon$
$VVINF(\text{umkehren}) \rightarrow \varepsilon$

Synchronous LCFRS (SLCFRS)

A tuple $G = (N_s, N_t, T_s, T_t, V_s, V_t, P, S_s, S_t)$ where

- N_s, T_s, V_s, S_s, resp. N_t, T_t, V_t, S_t are defined as for LCFRS
  $\rightarrow$ alphabets for the *source* and *target side* respectively.
- P is a finite set of synchronous rewriting rules $\langle r_s, r_t, \sim \rangle$ where
  - r_s and r_t are LCFRS rewriting rules based on N_s, T_s, V_s and N_t, T_t, V_t respectively, and
  - $\sim$ is a bijective mapping of the non-terminals in the RHS of r_s to the non-terminals in the RHS of r_t.
    $\rightarrow$ co-indexation

Synchronous LCFRS (SLCFRS)

A tuple $G = (N_s, N_t, T_s, T_t, V_s, V_t, P, S_s, S_t)$ where

- N_s, T_s, V_s, S_s, resp. N_t, T_t, V_t, S_t are defined as for LCFRS $\rightarrow$ alphabets for the *source* and *target side* respectively.
- P is a finite set of synchronous rewriting rules $\langle r_s, r_t, \sim \rangle$ where
  - r_s and r_t are LCFRS rewriting rules based on N_s, T_s, V_s and N_t, T_t, V_t respectively, and
  - $\sim$ is a bijective mapping of the non-terminals in the RHS of r_s to the non-terminals in the RHS of r_t.
    $\rightarrow$ co-indexation
- During a derivation, the yields of two co-indexed non-terminals have to be explained from one synchronous rule. $\langle S_s, S_t \rangle$ is the start pair.
- Fan-out v of G: $v_{G_s} + v_{G_t}$ (Notation: $v_{v_{G_s}|v_{G_t}}$)

SLCFRS Example

SCFG:

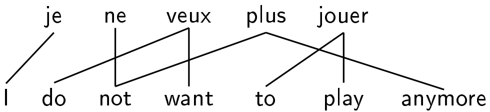
$\langle X \rightarrow \text{ne veux plus } X_{\boxed{1}} \quad , \quad X \rightarrow \text{do not want to } X_{\boxed{1}} \text{ anymore}\rangle$

$\langle X \rightarrow \text{jouer} \quad , \quad X \rightarrow \text{to play}$

SLCFRS:

$\langle X(\text{ne veux plus } Y_1) \rightarrow X_{\boxed{1}}(Y_1) \quad , \quad X(\text{do not want to } Z_1 \text{ anymore}) \rightarrow X_{\boxed{1}}(Z_1)\rangle$

$\langle X(\text{jouer}) \rightarrow \varepsilon \quad , \quad X(\text{to play}) \rightarrow \varepsilon\rangle$



SLCFRS Example

SCFG:

$\langle X \rightarrow \text{ne veux plus } X_{\boxed{1}} \quad , \quad X \rightarrow \text{do not want to } X_{\boxed{1}} \text{ anymore} \rangle$

$\langle X \rightarrow \text{jouer} \quad , \quad X \rightarrow \text{to play} \rangle$

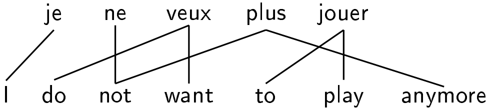
SLCFRS:

$\langle X(\text{ne veux plus } Y_1) \rightarrow X_{\boxed{1}}(Y_1) \quad , \quad X(\text{do not want to } Z_1 \text{ anymore}) \rightarrow X_{\boxed{1}}(Z_1) \rangle$

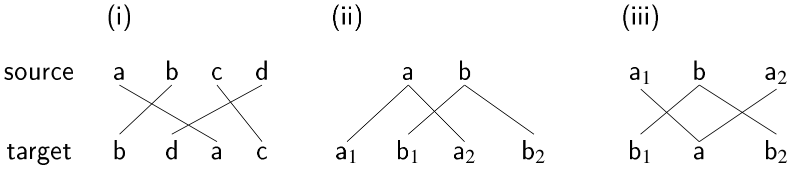
$\langle X(\text{jouer}) \rightarrow \varepsilon \quad , \quad X(\text{to play}) \rightarrow \varepsilon \rangle$

$\langle X(\text{veux}) \rightarrow \varepsilon \quad , \quad X(\text{do , want}) \rightarrow \varepsilon \rangle$

$\langle X(\text{ne } Y_1 \text{ plus } Y_2) \rightarrow X_{\boxed{1}}(Y_1) X_{\boxed{2}}(Y_2) \quad , \quad X(Z_1 \text{ not } Z_2 Z_3 \text{ anymore}) \rightarrow X_{\boxed{1}}(Z_1, Z_2) X_{\boxed{2}}(Z_3) \rangle$



Alignment Configurations Beyond SCFG



Inside-out alignment
Wu (1997)

Cross-serial disc. translation unit
Søgaard & Kuhn (2009)

Bonbon alignment
Simard et al. (2005)

⇒ Beyond the alignment capacity of ITG/SCFG of rank 2

- 5% of Chinese-English sentences have IO alignments (Wellington et al., 2006)
- 9% of Spanish-French sentences and 5.5% of English-German sentences are beyond 2-SCFG (Kaeshammer, 2013)

Training

- Rule extraction from a word-aligned parallel corpus as for *hierarchical phrase-based MT* (SCFG)
  1. Extraction of initial phrase pairs
  2. Creation of hierarchical rules by replacing phrase pairs which are contained within other phrase pairs with non-terminals/variables

Training

- Rule extraction from a word-aligned parallel corpus as for *hierarchical phrase-based MT* (SCFG)
  1. Extraction of initial phrase pairs
  2. Creation of hierarchical rules by replacing phrase pairs which are contained within other phrase pairs with non-terminals/variables
- Crucial difference: a phrase is a set of word indices

Training

- Rule extraction from a word-aligned parallel corpus as for *hierarchical phrase-based MT* (SCFG)
  1. Extraction of initial phrase pairs
  2. Creation of hierarchical rules by replacing phrase pairs which are contained within other phrase pairs with non-terminals/variables
- Crucial difference: a phrase is a set of word indices
- Restrictions in addition to the usual ones (Chiang, 2007):
  1. number of words in a gap (10)
  2. no unaligned blocks
  3. number of continuous blocks in a phrase (2), cf. (Kaeshammer, 2013)

Training

- Rule extraction from a word-aligned parallel corpus as for *hierarchical phrase-based MT* (SCFG)
  1. Extraction of initial phrase pairs
  2. Creation of hierarchical rules by replacing phrase pairs which are contained within other phrase pairs with non-terminals/variables
- Crucial difference: a phrase is a set of word indices
- Restrictions in addition to the usual ones (Chiang, 2007):
  1. number of words in a gap (10)
  2. no unaligned blocks
  3. number of continuous blocks in a phrase (2), cf. (Kaeshammer, 2013)
- Additional features: source gap degree and target gap degree (number of gaps during a derivation)

Decoder (1)

- Same methodology as for SCFG-based decoding
- Bottom-up CYK parser using the source side of the translation grammar
  - → monolingual weighted LCFRS parsing
  - → parse items $[A, \boldsymbol{\rho}, v_t]$
  - → Specific $(2,2)$-LCFRS parser because of the specific form of the grammar (rank 2, fan-out $4_{2|2}$) : $\boldsymbol{\rho} = (\langle i_1, j_1 \rangle, \langle i_2, j_2 \rangle)$

Decoder (2)

- Intersection of the parse hypergraph with an n-gram LM: Cube pruning (Huang and Chiang, 2007)
  - → target string of a hypothesis is a tuple of continuous blocks of target words, e.g. ⟨do not want, anymore⟩
  - → score each block separately
  - → store a LM state for each block
- Extraction of k-best translations on the hypergraph after cube pruning
- Implementation in C++, including code from KenLM for language modeling

Experimental Setup

- German-to-English translation
- Data from the WMT 2014 translation task (max. 30 words)
- Standard preprocessing and word alignment
- Filter the translation grammar w.r.t. input data set by extracting per-sentence-grammars
- 3-gram LM, KenLM
- For decoding: cube pruning buffer size 400, no limits on the number of words a non-terminal can span
- Tuning the feature weights with MERT, maximizing BLEU-4, using 200-best translations (ZMERT, mert-moses.pl)
- multi-bleu.perl for calculating BLEU scores (lc), repeating each experiment four times, reporting the average

Results

system	feat	devtest BLEU	test BLEU
sys(1,1)	-	24.13	23.23

Results

system	feat	devtest BLEU	test BLEU
sys(1,1)	-	24.13	23.23
sys(2,2)	-	23.90	22.90

Results

system	feat	devtest BLEU	test BLEU
sys(1,1)	-	24.13	23.23
sys(2,2)	-	23.90	22.90
sys(2,2)	s	24.06	23.17
sys(2,2)	t	**24.20**	23.35
sys(2,2)	s+t	24.18	23.32

Results

system	feat	devtest BLEU	test BLEU
sys(1,1)	-	24.13	23.23
sys(2,2)	-	23.90	22.90
sys(2,2)	s	24.06	23.17
sys(2,2)	t	**24.20**	23.35
sys(2,2)	s+t	24.18	23.32
sys(1,2)	-	23.39	23.24
sys(2,1)	-	24.17	**23.41**

Manual Evaluation

- sys(1,1) vs. sys(2,1) system comparison using Appraise
- 95 sentences where sys(2,1) uses at least one SLCFRS rule
- two native speakers of English with basic knowledge of German

Manual Evaluation

- sys(1,1) vs. sys(2,1) system comparison using Appraise
- 95 sentences where sys(2,1) uses at least one SLCFRS rule
- two native speakers of English with basic knowledge of German

	sys(1,1)	sys(2,1)	=
e1	43	49	3
e2	46	47	2

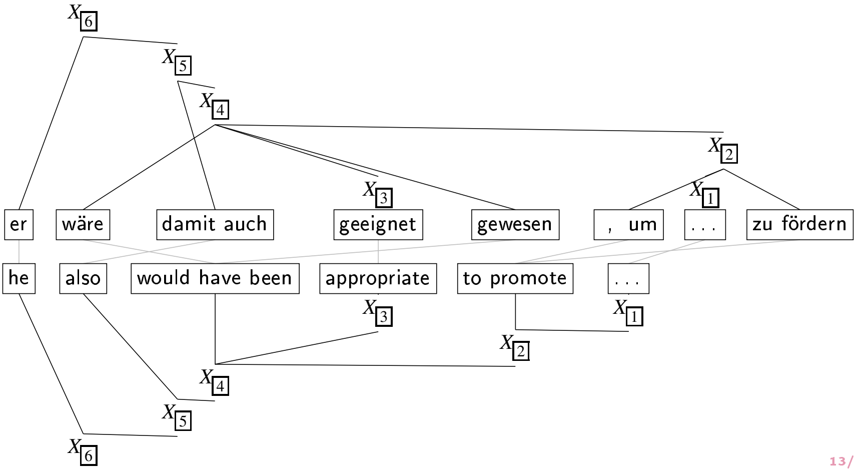
Table: Result of the manual system comparison

Translation Example

Source	er wäre damit auch geeignet gewesen , um die . . . zu fördern
Reference	it would thus be suitable to assist . . .
sys(1,1)	it would also have to be , in order to promote the . . .
sys(2,1)	he also would have been appropriate to promote the . . .

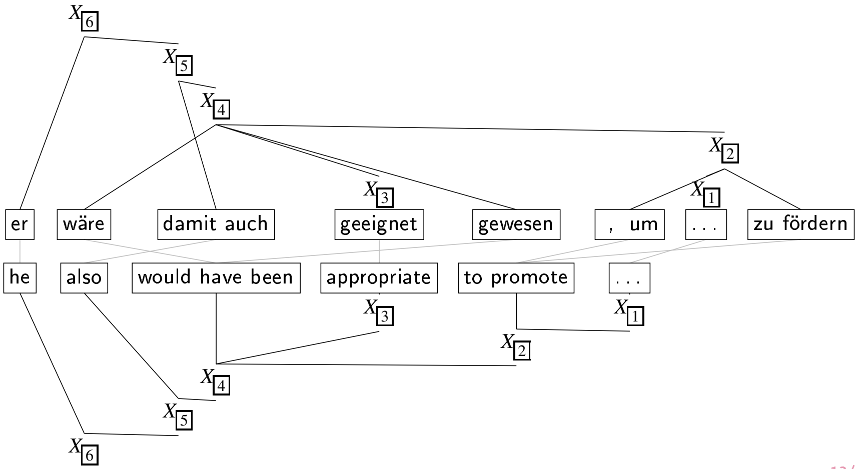
Translation Example

Source	er wäre damit auch geeignet gewesen , um die . . . zu fördern
Reference	it would thus be suitable to assist . . .
sys(1,1)	it would also have to be , in order to promote the . . .
sys(2,1)	he also would have been appropriate to promote the . . .



Translation Example

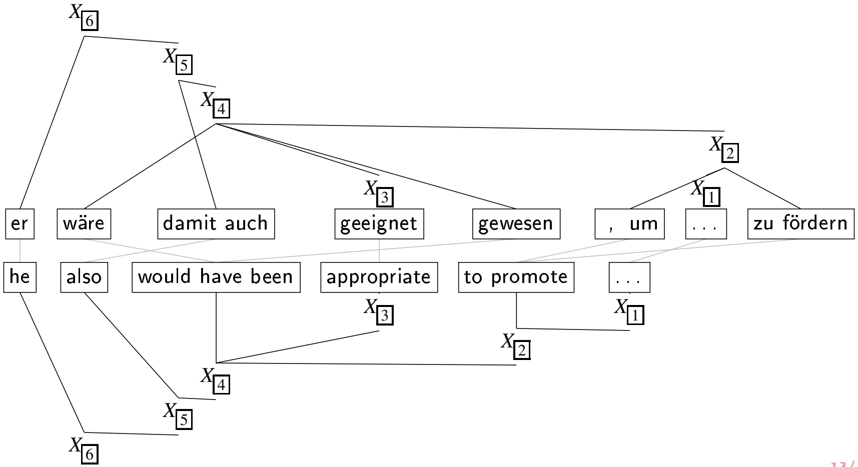
$\langle X(\text{wäre}\ ,Y_1\ \text{gewesen}\ Y_2) \rightarrow X_{\boxed{1}}(Y_1)X_{\boxed{2}}(Y_2)\ ,X(\text{would have been}\ Y_1Y_2) \rightarrow X_{\boxed{1}}(Y_1)X_{\boxed{2}}(Y_2)\rangle$



Translation Example

$\langle X(\text{wäre }, Y_1 \text{ gewesen } Y_2) \rightarrow X_{\boxed{1}}(Y_1)X_{\boxed{2}}(Y_2)\ , X(\text{would have been } Y_1 Y_2) \rightarrow X_{\boxed{1}}(Y_1)X_{\boxed{2}}(Y_2)\rangle$

$\langle X(Y_1 \text{ damit auch } Y_2) \rightarrow X_{\boxed{1}}(Y_1, Y_2)\ , X(\text{also } Y_1) \rightarrow X_{\boxed{1}}(Y_1)\rangle$



Conclusions & Future Work

- Extension of the hierarchical phrase-based MT approach to discontinuous phrases
- SLCFRS as the translation grammar formalism
- Previous work on SCFG-based MT can be directly extended
- Modest improvement in BLEU score over the SCFG baseline
- Slight preference by the human evaluators for the translations produced by the SLCFRS system

Conclusions & Future Work

- Extension of the hierarchical phrase-based MT approach to discontinuous phrases
- SLCFRS as the translation grammar formalism
- Previous work on SCFG-based MT can be directly extended
- Modest improvement in BLEU score over the SCFG baseline
- Slight preference by the human evaluators for the translations produced by the SLCFRS system

- More detailed evaluation
- Experiments with other language pairs

References I

Bertsch, E. and Nederhof, M.-J. (2001). On the complexity of some extensions of RCG parsing. In *Proceedings of IWPT*.

Chiang, D. (2007). Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Galley, M. and Manning, C. D. (2010). Accurate non-hierarchical phrase-based translation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 966–974.

Huang, L. and Chiang, D. (2007). Forest rescoring: Faster decoding with integrated language models. In *Annual Meeting-Association For Computational Linguistics*, volume 45, page 144.

Kaeshammer, M. (2013). Synchronous linear context-free rewriting systems for machine translation. In *Proceedings of the Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 68–77. Association for Computational Linguistics.

Kallmeyer, L. and Maier, W. (2013). Data-driven parsing using Probabilistic Linear Context-Free Rewriting Systems. *Computational Linguistics*, 39(1).

Maier, W. (2010). Direct parsing of discontinuous constituents in German. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*.

Melamed, I. D., Satta, G., and Wellington, B. (2004). Generalized multitext grammars. In *Proceedings of the 42nd Annual Conference of the Association for Computational Linguistics (ACL)*.

Simard, M., Cancedda, N., Cavestro, B., Dymetman, M., Gaussier, E., Goutte, C., Yamada, K., Langlais, P., and Mauser, A. (2005). Translating with non-contiguous phrases. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 755–762.

Søgaard, A. (2010). Can inversion transduction grammars generate hand alignments? In *Procceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT)*.

References II

Søgaard, A. and Kuhn, J. (2009). Empirical lower bounds on alignment error rates in syntax-based machine translation. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation (SSST '09)*. Association for Computational Linguistics.

Søgaard, A. and Wu, D. (2009). Empirical lower bounds on translation unit error rate for the full class of inversion transduction grammars. In *Proceedings of the 11th International Conference on Parsing Technologies*, pages 33–36.

van Cranenburgh, A. and Bod, R. (2013). Discontinuous parsing with an efficient and accurate dop model. In *Proceedings of the International Conference on Parsing Technologies (IWPT 2013)*.

Vijay-Shanker, K., Weir, D., and Joshi, A. K. (1987). Characterizing structural descriptions used by various formalisms. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*.

Weir, D. (1988). *Characterizing Mildly Context-Sensitive Grammar Formalisms*. PhD thesis, University of Pennsylviania, Philadelphia, PA.

Wellington, B., Waxmonsky, S., and Melamed, I. D. (2006). Empirical lower bounds on the complexity of translational equivalence. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 977–984.

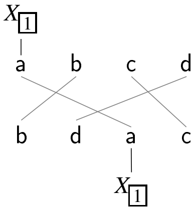
Wu, D. (1997). Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational linguistics*, 23(3):377–403.

Results

system	feat	devtest BLEU	test BLEU
sys(1,1)	-	24.13	23.23
sys(2,2)	-	23.90	22.90
sys(2,2)	s	24.06	23.17
sys(2,2)	t	**24.20**	23.35
sys(2,2)	s+t	24.18	23.32
sys(1,2)	-	23.39	23.24
sys(2,1)	-	24.17	**23.41**
moses		24.33	23.34

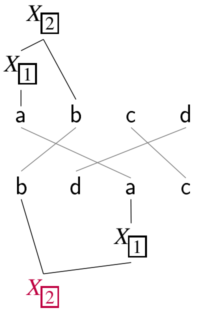
Derivation

$$\langle X(\mathsf{a}) \rightarrow \varepsilon \qquad , X(\mathsf{a}) \rightarrow \varepsilon \rangle$$

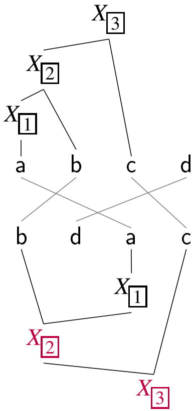


Derivation

$$\langle X(\mathsf{a}) \rightarrow \varepsilon \quad , X(\mathsf{a}) \rightarrow \varepsilon \rangle$$
$$\langle X(Y\mathsf{b}) \rightarrow X_{\boxed{1}}(Y) , X(\mathsf{b}, Z) \rightarrow X_{\boxed{1}}(Z) \rangle$$

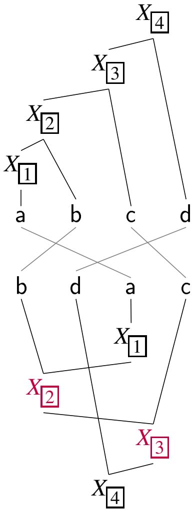


Derivation



$$\langle X(\mathsf{a}) \rightarrow \varepsilon \quad , X(\mathsf{a}) \rightarrow \varepsilon \rangle$$
$$\langle X(Y\mathsf{b}) \rightarrow X_{\boxed{1}}(Y) \ , X(\mathsf{b}, Z) \rightarrow X_{\boxed{1}}(Z) \rangle$$
$$\langle X(Y\mathsf{c}) \rightarrow X_{\boxed{1}}(Y) \ , X(Z_1, Z_2\mathsf{c}) \rightarrow X_{\boxed{1}}(Z_1, Z_2) \rangle$$

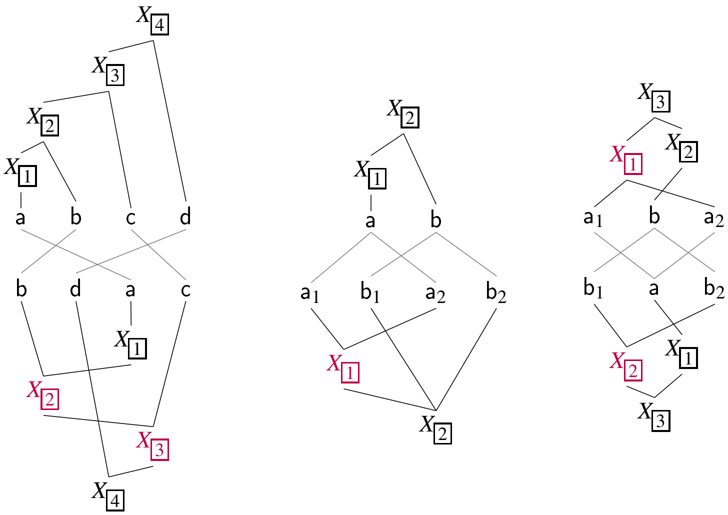
Derivation



$$\langle X(\mathsf{a}) \rightarrow \varepsilon \quad , X(\mathsf{a}) \rightarrow \varepsilon \rangle$$
$$\langle X(Y\mathsf{b}) \rightarrow X_{\boxed{1}}(Y) \ , X(\mathsf{b}, Z) \rightarrow X_{\boxed{1}}(Z) \rangle$$
$$\langle X(Y\mathsf{c}) \rightarrow X_{\boxed{1}}(Y) \ , X(Z_1, Z_2\mathsf{c}) \rightarrow X_{\boxed{1}}(Z_1, Z_2) \rangle$$
$$\langle X(Y\mathsf{d}) \rightarrow X_{\boxed{1}}(Y) \ , X(Z_1\mathsf{d}Z_2) \rightarrow X_{\boxed{1}}(Z_1, Z_2) \rangle$$

or ...

Derivation


X 4
X 3
X 2
X 1
a b c d
b d a c
X 1
X 2
X 3
X 4
X 2
X 1
a b
a1 b1 a2 b2
X 1
X 2
X 3
X 1 X 2
a1 b a2
b1 a b2
X 2 X 1
X 3

Related Work

- Studies addressing the alignment coverage of a formalism w.r.t. gold alignments (Søgaard and Wu, 2009; Søgaard and Kuhn, 2009; Wellington et al., 2006; Søgaard, 2010; Kaeshammer, 2013)
  - → 5% of Chinese-English sentences have IO alignments
  - → 9% of Spanish-French sentences and 5.5% of English-German sentences are beyond 2-SCFG
- SLCFRS are equivalent to Simple Range Concatenation Transducers (Bertsch and Nederhof, 2001) and Generalized Multitext Grammars (Melamed et al., 2004)
- Discontinuous phrase-based SMT (Galley and Manning, 2010)
  - → improvement in BLEU score for Chinese-English
  - → this work: hierarchical, tree-based counterpart

Features

- Standard features: direct and inverse translation probabilities, lexical translation probabilities, number of rules etc.
- MLE on the distribution of the extracted rules to obtain the translation probabilities
- Additional: number of gaps during a derivation (source gap degree and target gap degree of a rule)

Experimental Setup

- German-to-English translation
- Data from the WMT 2014 translation task (max. 30 words)
- Punctuation normalization, tokenization, truecasing, compound splitting for German with the Moses scripts
- Multi-threaded GIZA++ and `grow-diag-final-and` heuristics for word-aligning the training data
- Filter the translation grammar w.r.t. input data set by extracting per-sentence-grammars
- 3-gram LM, KenLM
- For decoding: cube pruning buffer size 400, no limits on the number of words a non-terminal can span
- Tuning the feature weights with MERT, maximizing BLEU-4, using 200-best translations (ZMERT, mert-moses.pl)
- multi-bleu.perl for calculating BLEU scores (lc), repeating each experiment four times, reporting the average

Manual Evaluation

- sys(1,1) vs. sys(2,1) system comparison using Appraise
- 95 sentences where sys(2,1) uses at least one SLCFRS rule
- two native speakers of English with basic knowledge of German

	sys(1,1)	sys(2,1)	=
e1	43	49	3
e2	46	47	2

Table: Result of the manual system comparison

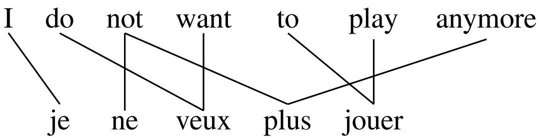
		e2		
		sys(1,1)	sys(2,1)	=
e1	sys(1,1)	**29**	13	1
	sys(2,1)	15	**33**	1
	=	2	1	0

Table: Confusion matrix of the decisions of the manual evaluation, Cohen's $\kappa = 0.338$

Notion of Alignment Capacity

Same as in previous related work

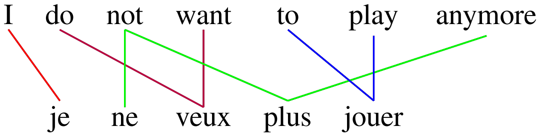
- A **translation unit** (TU) is a maximally connected subgraph of a given alignment structure.
- Alignment structure is divided into disjoint TUs.



Notion of Alignment Capacity

Same as in previous related work

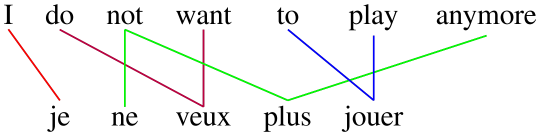
- A **translation unit** (TU) is a maximally connected subgraph of a given alignment structure.
- Alignment structure is divided into disjoint TUs.



Notion of Alignment Capacity

Same as in previous related work

- A **translation unit** (TU) is a maximally connected subgraph of a given alignment structure.
- Alignment structure is divided into disjoint TUs.
- Synchronously recognized or generated terminals are aligned $\rightarrow$ TU



LCFRS (1)

(Vijay-Shanker et al., 1987; Weir, 1988)

A tuple $G = (N, T, V, P, S)$ where

- N: a finite set of non-terminals with a function dim: $N \to \mathbb{N}$ determining the **fan-out** of each $A \in N$;
- T and V: disjoint finite sets of terminals and variables;
- $S \in N$: start symbol with $dim(S) = 1$;
- P: a finite set of rewriting rules

LCFRS (1)

(Vijay-Shanker et al., 1987; Weir, 1988)

A tuple $G = (N, T, V, P, S)$ where

- N: a finite set of non-terminals with a function dim: $N \rightarrow \mathbb{N}$ determining the **fan-out** of each $A \in N$;
- T and V: disjoint finite sets of terminals and variables;
- $S \in N$: start symbol with $dim(S) = 1$;
- P: a finite set of rewriting rules

$$A(\alpha_1, \ldots, \alpha_{dim(A)}) \rightarrow A_1(X_1^{(1)}, \ldots, X_{dim(A_1)}^{(1)}) \cdots A_m(X_1^{(m)}, \ldots, X_{dim(A_m)}^{(m)})$$

where

- $A, A_1, \ldots, A_m \in N$, $X_j^{(i)} \in V$ for $1 \leq i \leq m$, $1 \leq j \leq dim(A_i)$, and
- $\alpha_i \in (T \cup V)^*$ for $1 \leq i \leq dim(A)$, for a **rank** $m \geq 0$.

LCFRS (2)

$$A(\alpha_1, \dots, \alpha_{dim(A)}) \rightarrow A_1(X_1^{(1)}, \dots, X_{dim(A_1)}^{(1)}) \cdots A_m(X_1^{(m)}, \dots, X_{dim(A_m)}^{(m)})$$

- Every variable X in $r \in P$ occurs exactly once in the LHS and exactly once in the RHS of r.
- r describes how the yield of the LHS non-terminal is computed from the yields of the RHS non-terminals.
- The yield of S is the language of the grammar.
- Rank u of G: the maximal rank of any of its rules
- Fan-out v of G: the maximal fan-out of any of its non-terminals.

Normal Form

Conditions:

1. $u_G \leq 2$
2. For all $r \in P$ it holds that that the LHS arguments of r_s and r_t contain either terminals or variables but not mixture of both.

NF-ITG $\leftrightarrow$ $(2, 2_{1|1})$-SLCFRS in normal form
ITG of rank 2 $\leftrightarrow$ $(2, 2_{1|1})$-SLCFRS

Different alignment capacity of normal form and full class

Bitext Parsing Complexity

SLCFRS in normal form with fan-out v

$\mathcal{O}(n^{3v})$

(assuming that $n_s \approx n_t$)

Bitext Parsing Complexity

SLCFRS in normal form with fan-out v

$\mathcal{O}(n^{3v})$

(assuming that $n_s \approx n_t$)

Which fan-out v is required to cover the alignment configurations that occur in manually aligned data?

Data & Results

Graça	en-fr
	en-pt
	en-es
	pt-fr
	pt-es
	es-fr
Martin	en-ro
	en-hi
	en-iu
Pado	en-de
Mihal.	en-fr
CDT	da-en
	da-de
	da-es
	da-it
Holmqv.	en-sv
Schoen.	en-de
Lambert	en-es
Macken	en-nl

Data & Results

		NF $v = 2_{1\|1}$ = NF-ITG	$u = 2$
Graça	en-fr	73.00	
	en-pt	76.00	
	en-es	82.00	
	pt-fr	73.00	
	pt-es	90.00	
	es-fr	74.00	
Martin	en-ro	45.07	
	en-hi	82.73	
	en-iu	40.66	
Pado	en-de	73.74	
Mihal.	en-fr	67.56	
CDT	da-en	72.90	
	da-de	64.87	
	da-es	66.61	
	da-it	69.01	
Holmqv.	en-sv	82.83	
Schoen.	en-de	29.15	
Lambert	en-es	47.15	
Macken	en-nl	57.14	

Data & Results

		NF	$u = 2$
		$v = 2_{1\|1}$ = NF-ITG	$v = 2_{1\|1}$ = ITG
Graça	en-fr	73.00	95.00
	en-pt	76.00	98.00
	en-es	82.00	96.00
	pt-fr	73.00	92.00
	pt-es	90.00	99.00
	es-fr	74.00	91.00
Martin	en-ro	45.07	95.07
	en-hi	82.73	96.36
	en-iu	40.66	100.00
Pado	en-de	73.74	94.41
Mihal.	en-fr	67.56	95.30
CDT	da-en	72.90	97.80
	da-de	64.87	94.94
	da-es	66.61	97.50
	da-it	69.01	97.95
Holmqv.	en-sv	82.83	95.60
Schoen.	en-de	29.15	76.11
Lambert	en-es	47.15	94.85
Macken	en-nl	57.14	94.86

Data & Results

		NF	$u=2$	Søgaard (2010)	
		$v=2_{1\|1}$ = NF-ITG	$v=2_{1\|1}$ = ITG	NF-ITG	ITG
Graça	en-fr	73.00	95.00	65.00	68.00
	en-pt	76.00	98.00	65.00	67.00
	en-es	82.00	96.00	73.00	74.00
	pt-fr	73.00	92.00	63.00	63.00
	pt-es	90.00	99.00	80.00	81.00
	es-fr	74.00	91.00	68.00	68.00
Martin	en-ro	45.07	95.07		
	en-hi	82.73	96.36		
	en-iu	40.66	100.00		
Pado	en-de	73.74	94.41		
Mihal.	en-fr	67.56	95.30		
CDT	da-en	72.90	97.80		
	da-de	64.87	94.94		
	da-es	66.61	97.50		
	da-it	69.01	97.95		
Holmqv.	en-sv	82.83	95.60		
Schoen.	en-de	29.15	76.11		
Lambert	en-es	47.15	94.85		
Macken	en-nl	57.14	94.86		

Data & Results

		NF		$u=2$	Søgaard (2010)	
		$v=2_{1\|1}$ = NF-ITG	$v=4_{2\|2}$	$v=2_{1\|1}$ = ITG	NF-ITG	ITG
Graça	en-fr	73.00	**100.00**	95.00	65.00	68.00
	en-pt	76.00	**100.00**	98.00	65.00	67.00
	en-es	82.00	**100.00**	96.00	73.00	74.00
	pt-fr	73.00	**97.00**	92.00	63.00	63.00
	pt-es	90.00	**99.00**	99.00	80.00	81.00
	es-fr	74.00	**100.00**	91.00	68.00	68.00
Martin	en-ro	45.07	**97.85**	95.07		
	en-hi	82.73	**100.00**	96.36		
	en-iu	40.66	**95.60**	100.00		
Pado	en-de	73.74	**100.00**	94.41		
Mihal.	en-fr	67.56	**98.88**	95.30		
CDT	da-en	72.90	**98.93**	97.80		
	da-de	64.87	**98.42**	94.94		
	da-es	66.61	**97.68**	97.50		
	da-it	69.01	**97.65**	97.95		
Holmqv.	en-sv	82.83	**99.78**	95.60		
Schoen.	en-de	29.15	**94.74**	76.11		
Lambert	en-es	47.15	**97.83**	94.85		
Macken	en-nl	57.14	**98.86**	94.86		

Data & Results

		NF		$u = 2$	Søgaard (2010)	
		$v = 2_{1\|1}$ = NF-ITG	$v = 4_{2\|2}$	$v = 2_{1\|1}$ = ITG	NF-ITG	ITG
Graça	en-fr	73.00	**100.00**	95.00	65.00	68.00
	en-pt	76.00	**100.00**	98.00	65.00	67.00
	en-es	82.00	**100.00**	96.00	73.00	74.00
	pt-fr	73.00	**97.00**	92.00	63.00	63.00
	pt-es	90.00	**99.00**	99.00	80.00	81.00
	es-fr	74.00	**100.00**	91.00	68.00	68.00
Martin	en-ro	45.07	**97.85**	95.07		
	en-hi	82.73	**100.00**	96.36		
	en-iu	40.66	**95.60**	100.00		
Pado	en-de	73.74	**100.00**	94.41		
Mihal.	en-fr	67.56	**98.88**	95.30		
CDT	da-en	72.90	**98.93**	97.80		
	da-de	64.87	**98.42**	94.94		
	da-es	66.61	**97.68**	97.50		
	da-it	69.01	**97.65**	97.95		
Holmqv.	en-sv	82.83	**99.78**	95.60		
Schoen.	en-de	29.15	**94.74**	76.11		
Lambert	en-es	47.15	**97.83**	94.85		
Macken	en-nl	57.14	**98.86**	94.86		

Data & Results

		NF		$u = 2$		Søgaard (2010)	
		$v = 2_{1\|1}$ $= \text{NF-ITG}$	$v = 4_{2\|2}$	$v = 2_{1\|1}$ $= \text{ITG}$	$v = 4_{2\|2}$	NF-ITG	ITG
Graça	en-fr	73.00	**100.00**	95.00	**100.00**	65.00	68.00
	en-pt	76.00	**100.00**	98.00	**100.00**	65.00	67.00
	en-es	82.00	**100.00**	96.00	**100.00**	73.00	74.00
	pt-fr	73.00	**97.00**	92.00	**100.00**	63.00	63.00
	pt-es	90.00	**99.00**	99.00	**100.00**	80.00	81.00
	es-fr	74.00	**100.00**	91.00	**100.00**	68.00	68.00
Martin	en-ro	45.07	**97.85**	95.07	**100.00**		
	en-hi	82.73	**100.00**	96.36	**100.00**		
	en-iu	40.66	**95.60**	100.00	**100.00**		
Pado	en-de	73.74	**100.00**	94.41	**100.00**		
Mihal.	en-fr	67.56	**98.88**	95.30	**100.00**		
CDT	da-en	72.90	**98.93**	97.80	**100.00**		
	da-de	64.87	**98.42**	94.94	**100.00**		
	da-es	66.61	**97.68**	97.50	**100.00**		
	da-it	69.01	**97.65**	97.95	**100.00**		
Holmqv.	en-sv	82.83	**99.78**	95.60	**100.00**		
Schoen.	en-de	29.15	**94.74**	76.11	**100.00**		
Lambert	en-es	47.15	**97.83**	94.85	**100.00**		
Macken	en-nl	57.14	**98.86**	94.86	**100.00**		

Data & Results

		NF		$u = 2$		Søgaard (2010)	
		$v = 2_{1\|1}$ = NF-ITG	$v = 4_{2\|2}$	$v = 2_{1\|1}$ = ITG	$v = 4_{2\|2}$	NF-ITG	ITG
Graça	en-fr	73.00	**100.00**	95.00	**100.00**	65.00	68.00
	en-pt	76.00	**100.00**	98.00	**100.00**	65.00	67.00
	en-es	82.00	**100.00**	96.00	**100.00**	73.00	74.00
	pt-fr	73.00	**97.00**	92.00	**100.00**	63.00	63.00
	pt-es	90.00	**99.00**	99.00	**100.00**	80.00	81.00
	es-fr	74.00	**100.00**	91.00	**100.00**	68.00	68.00
Martin	en-ro	45.07	**97.85**	95.07	**100.00**		
	en-hi	82.73	**100.00**	96.36	**100.00**		
	en-iu	40.66	**95.60**	100.00	**100.00**		
Pado	en-de	73.74	**100.00**	94.41	**100.00**		
Mihal.	en-fr	67.56	**98.88**	95.30	**100.00**		
CDT	da-en	72.90	**98.93**	97.80	**100.00**		
	da-de	64.87	**98.42**	94.94	**100.00**		
	da-es	66.61	**97.68**	97.50	**100.00**		
	da-it	69.01	**97.65**	97.95	**100.00**		
Holmqv.	en-sv	82.83	**99.78**	95.60	**100.00**		
Schoen.	en-de	29.15	**94.74**	76.11	**100.00**		
Lambert	en-es	47.15	**97.83**	94.85	**100.00**		
Macken	en-nl	57.14	**98.86**	94.86	**100.00**		