

Cross-lingual CCG Induction

Kilian Evang
@texttheater

University of Düsseldorf

2019-06-04
NAACL-HLT

Outline

Introduction

Derivation Projection

Experiments

Conclusions

Outline

Introduction

Derivation Projection

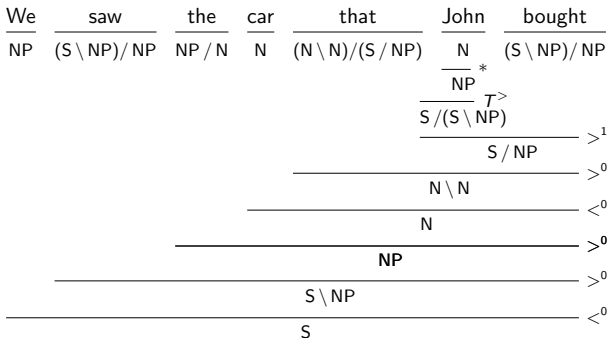
Experiments

Conclusions

Combinatory Categorical Grammar

$$\frac{\frac{\text{We}}{\text{NP}} \quad \frac{\text{sang}}{\text{S} \setminus \text{NP}}}{\text{S}} \leftarrow^0$$

Combinatory Categorical Grammar



Appeal

- coordination
- universal rules
- syntax-semantics interface

Most CCG Parsers

- trained on large treebanks, or
- hand-crafted



David Blackwell, CC-BY-NC

What about low-resource languages?

Unsupervised CCG Induction?

$$\begin{array}{c} \text{target-language text} \\ + \\ \text{magic} \\ = \\ \text{target-language CCG parser} \end{array}$$

(Bisk and Hockenmaier, 2013; Bisk et al., 2015)

Cross-lingual CCG Induction?

English CCG parser
+
parallel corpus
+
magic
=
target-language CCG parser

Cross-lingual CCG Induction via Derivation Projection

$$\begin{aligned} & \text{parallel corpus} \\ & + \\ & \text{English CCG derivations} \\ & + \\ & \text{word alignments} \\ & + \\ & \text{derivation projection} \\ & = \\ & \text{target-language CCG derivations} \\ & = \\ & \text{target-language training data} \end{aligned}$$

Outline

Introduction

Derivation Projection

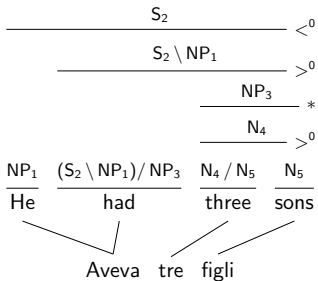
Experiments

Conclusions

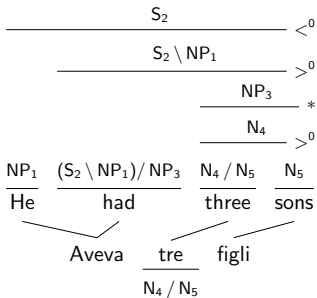
Derivation Projection

- project lexical categories along word alignments
- n:1 alignment → merge
- word order difference → flip slash

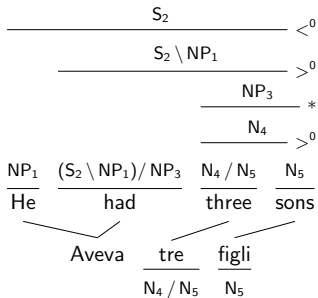
Example 1/3



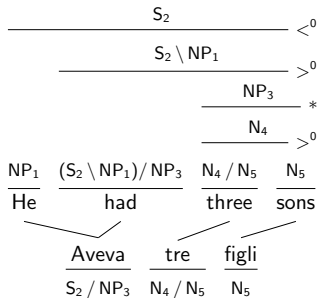
Example 1/3



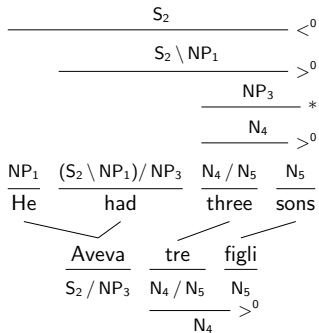
Example 1/3



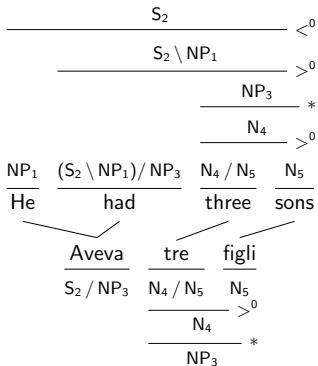
Example 1/3



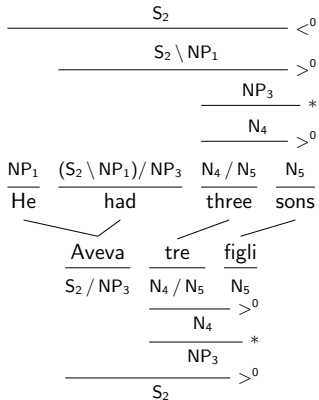
Example 1/3



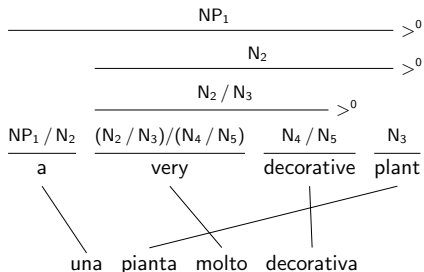
Example 1/3



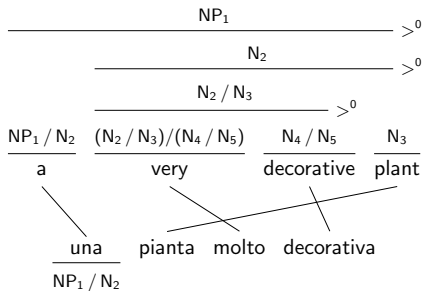
Example 1/3



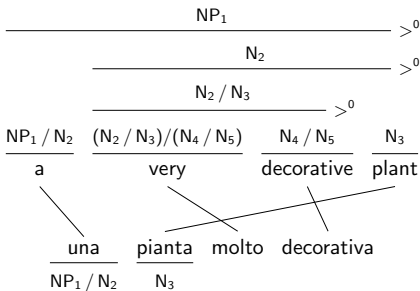
Example 2/3



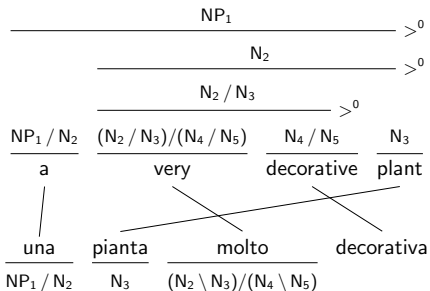
Example 2/3



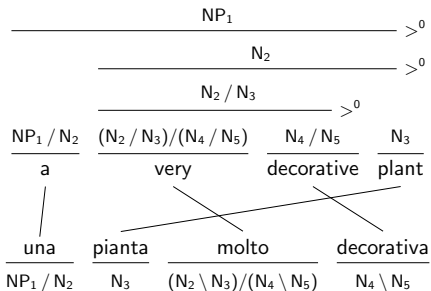
Example 2/3



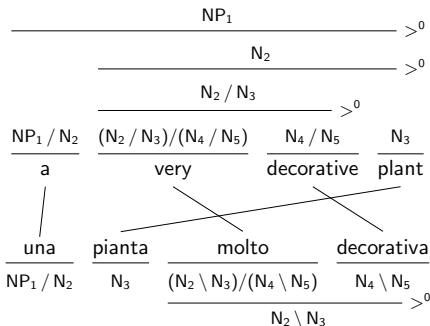
Example 2/3



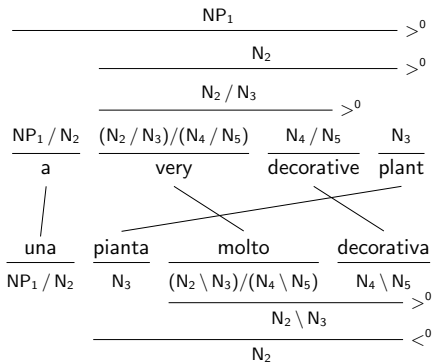
Example 2/3



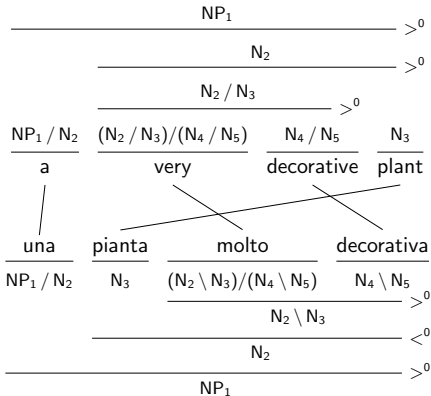
Example 2/3



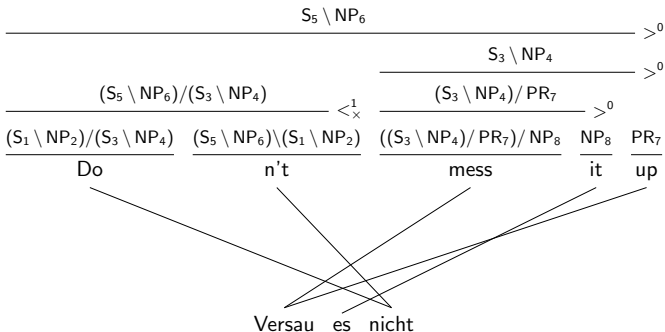
Example 2/3



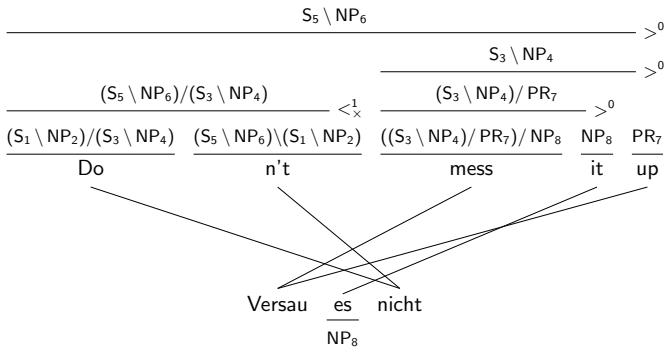
Example 2/3



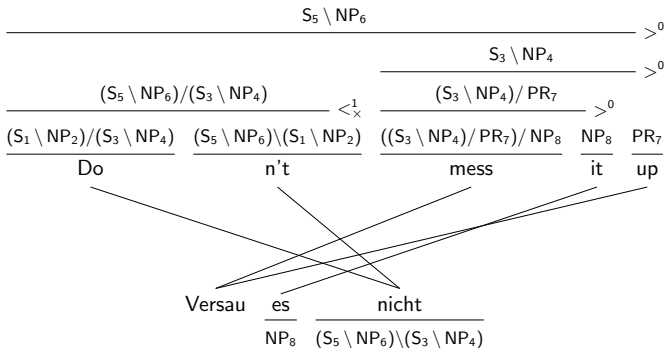
Example 3/3



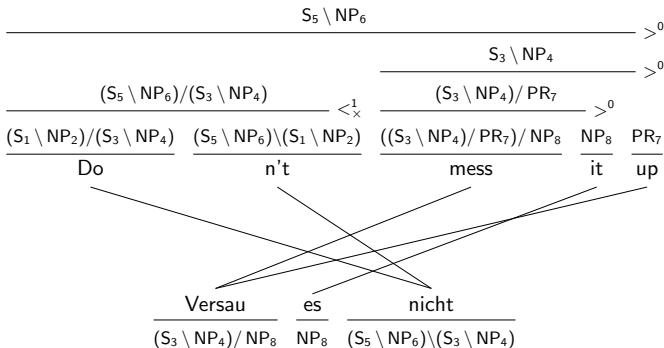
Example 3/3



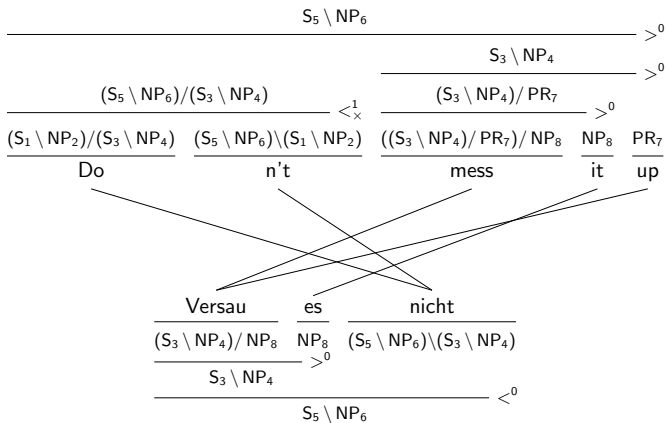
Example 3/3



Example 3/3



Example 3/3



Outline

Introduction

Derivation Projection

Experiments

Conclusions

Training

- **Parallel corpus:** tatoeba.org
- **English parser:** EasyCCG trained on CCGrebank
- **Word alignments:** GIZA++
- **Target-language parser:** EasyCCG trained on projected derivations

Evaluation

- PASCAL challenge on unsupervised grammar induction:
Arabic, Czech, Danish, Basque, Dutch, Portuguese, Slovenian,
Swedish
- unlabeled dependency f-score








Training (cont.)

	ara	ces	dan	eus	nld	por	slv	swe
sentence pairs	20K	11K	21K	2K	44K	161K	835	24K
projected	7K	4K	11K	590	18K	50K	364	12K

Baselines

- **BH13:** CCG induction from raw text + POS tags (Bisk and Hockenmaier, 2013)
- **BCH15:** CCG induction from raw text (Bisk et al., 2015)

Results

Language	ara	ces	dan	eus	nld	por	slv	swe
<i>Monolingual training on PASCAL</i>								
Train tokens	5K	436K	25K	81K	79K	159K	54K	62K
BH13	.651	.507	.585	.450	.544	.629	.464	.669
BCH15	.437	.324	.377	.352	.438	.516	.236	.529
<i>Cross-lingual training on Tatoeba</i>								
Train tokens	20K	11K	21K	2K	44K	161K	835	24K
this work	.468	.449	.630	.290	.614	.678	.350	.637
								

Induced Lexicons

		eng	deu	ita	nld
SOV	$(S \setminus NP) \setminus NP$	-	+	-	+
right adj	$N \setminus N, (N \setminus N) / (N \setminus N)$	-	-	+	-
pro-drop	$S, S / NP$	-	-	+	-

Outline

Introduction

Derivation Projection

Experiments

Conclusions

Take-home Message

- CCG derivations can be automatically projected along word alignments
- Cross-lingual supervision helps CCG induction
- Induces linguistically plausible lexicons
- Used for bootstrapping the Parallel Meaning Bank:
<https://pmb.let.rug.nl>

Bibliography I

- Bisk, Y., Christodoulopoulos, C., and Hockenmaier, J. (2015). Labeled grammar induction with minimal supervision. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 870–876. Association for Computational Linguistics.
- Bisk, Y. and Hockenmaier, J. (2013). An HDP model for inducing combinatorial categorial grammars. *Transactions of the Association for Computational Linguistics*, 1:75–88.

Bibliography II

- Honnibal, M., Curran, J. R., and Bos, J. (2010). Rebanking CCGbank for improved NP interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 207–215. Association for Computational Linguistics.
- Lewis, M. and Steedman, M. (2014). A* CCG parsing with a supertag-factored model. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 990–1000, Doha, Qatar.
- Steedman, M. (2001). *The Syntactic Process*. The MIT Press.