Cross-lingual Semantic Parsing

Part III: Cross-lingual Learning of an Open-domain Semantic Parser

Kilian Evang

University of Düsseldorf

ESSLLI 2018
30th European Summer School in Logic, Language and Information
August 6-17, 2018, Sofia, Bulgaria

Semantic Parsing: What?

From Words to (Logical) Meaning

She likes to read books →

x1 p1 e1
female(x1)
p1: [x2 e2 | book.n.01(x2), read.v.01(e2), Agent(e2, x1), Theme(e2, x2)]
like.v.02(e1)
Experiencer(e1, x1)
Stimulus(e1, p1)

DRT Kamp (1984)

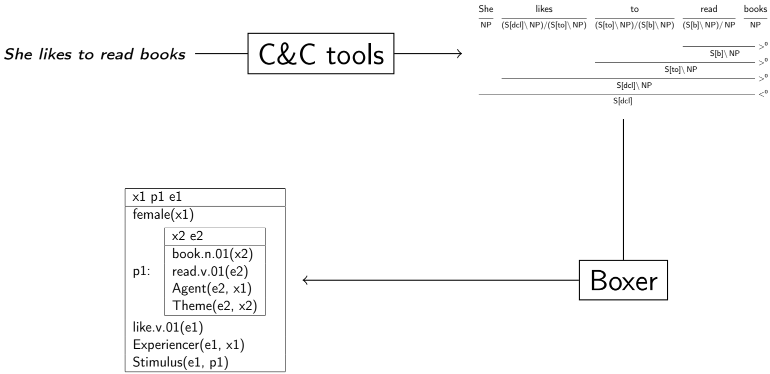
Semantic Parsing: Why?

Translate to something a computer can "understand"

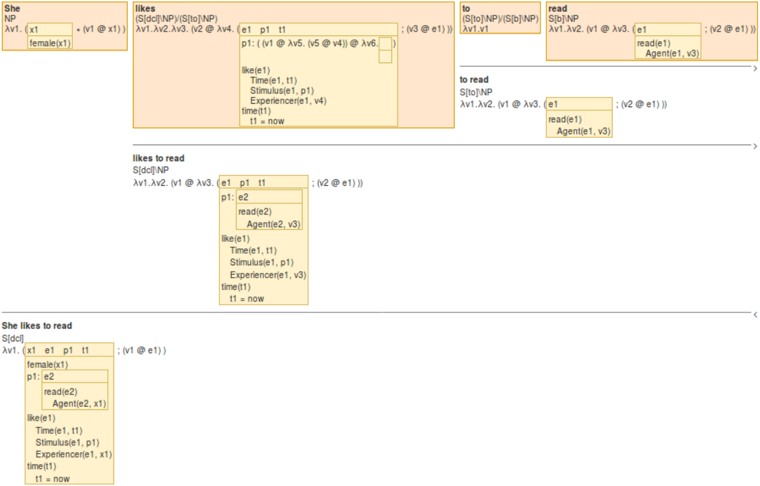
- commands for robots, e.g., Dukes (2014)
- queries for databases, e.g., Reddy et al. (2014)
- formulas for (probabilistic) reasoners, e.g., Beltagy et al. (2015)

Semantic Parsing: How?

System for English (Curran et al., 2007; Bos, 2015)



How Boxer Interprets CCG Derivations



Abbreviated Semantics

She	likes	to	read	books
NP	(S[dcl]\ NP)/(S[to]\ NP)	(S[to]\ NP)/(S[b]\ NP)	(S[b]\ NP)/ NP	NP
she′	*like′*	*to′*	*read′*	*book′*

$>^0$: S[b]\ NP — $read'@book'$

$>^0$: S[to]\ NP — $to'@(read'@book')$

$>^0$: S[dcl]\ NP — $like'@(to'@(read'@book'))$

$<^0$: S[dcl] — $(like'@(to'@(read'@book')))@she'$

System for Other Languages?

Goal

Learn (rudimentary) semantic parser from nothing but

- existing source language system (C&C+Boxer)
- parallel data
- unsupervised word aligner
- (POS tagger for target language)

Q: Can it be done?

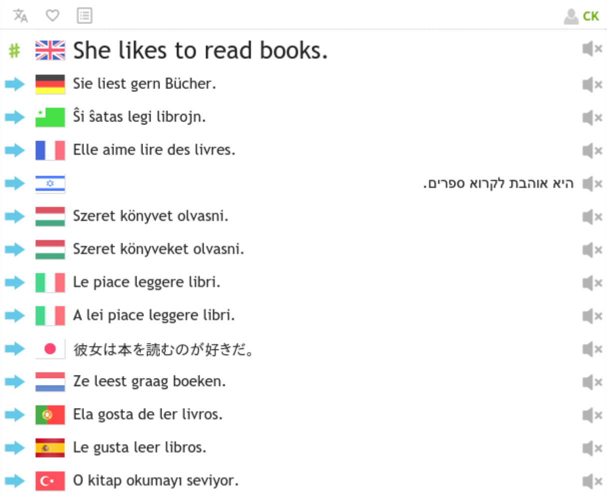
Method

1. CCG category projection → target-language lexicon
2. CCG derivation projection → target-language training data
3. parser training → target-language semantic parser

Ways to Generate Candidate Lexical Items

1. Zettlemoyer and Collins (2007) and much subsequent work: hand-written lexical templates
2. Kwiatkowksi et al. (2010): recursive splitting of gold-standard meaning representations
3. this work: projection from English parse trees to target-language words

Parallel Corpus: Tatoeba.org



Automatic Annotation of English Sentences

She	likes	to	read	books
NP	(S[dcl]\ NP)/(S[to]\ NP)	(S[to]\ NP)/(S[b]\ NP)	(S[b]\ NP)/ NP	NP
she′	*like′*	*to′*	*read′*	*book′*

$>^0$: S[b]\ NP — $read'@book'$

$>^0$: S[to]\ NP — $to'@(read'@book')$

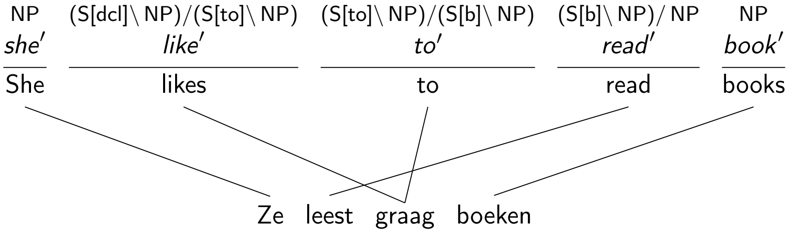
$>^0$: S[dcl]\ NP — $like'@(to'@(read'@book'))$

$<^0$: S[dcl] — $(like'@(to'@(read'@book')))@she'$

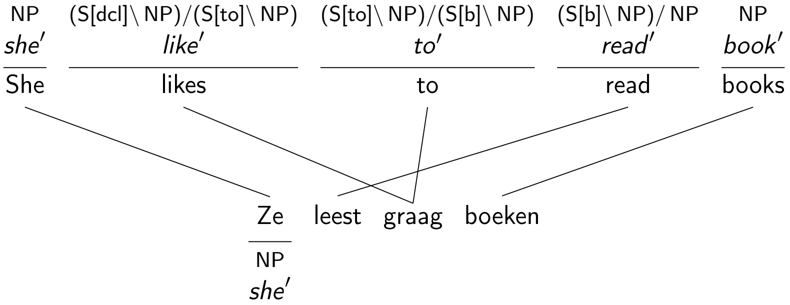
Category Projection: How?

- automatically word-align English and target-language sentence
- assign categories, interpretations to target-language words (1:1, 1:2, 2:1)
- drop category subdistinctions (dcl, b, to...)
- use undirected slashes

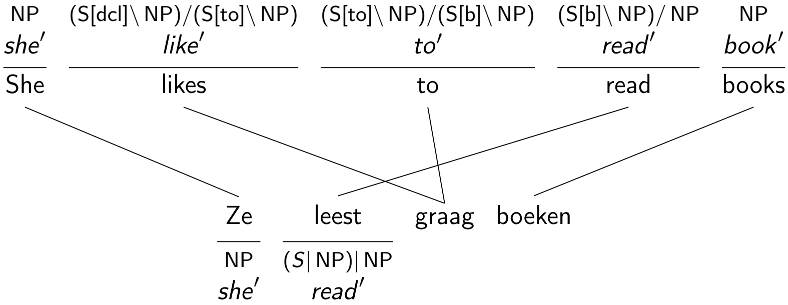
Example


NP
she′
She
(S[dcl]\ NP)/(S[to]\ NP)
like′
likes
(S[to]\ NP)/(S[b]\ NP)
to′
to
(S[b]\ NP)/ NP
read′
read
NP
book′
books
Ze leest graag boeken

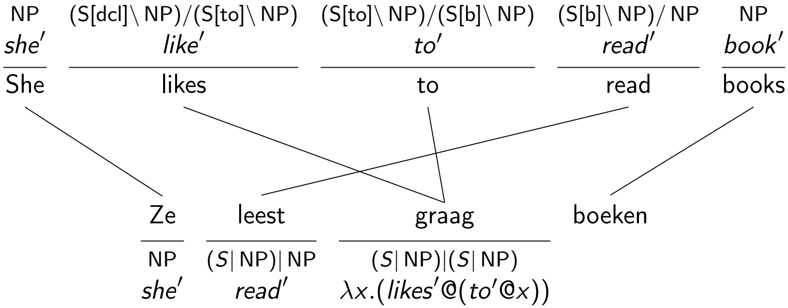
Example


NP
she′
She
(S[dcl]\ NP)/(S[to]\ NP)
like′
likes
(S[to]\ NP)/(S[b]\ NP)
to′
to
(S[b]\ NP)/ NP
read′
read
NP
book′
books
Ze
leest
graag
boeken
NP
she′

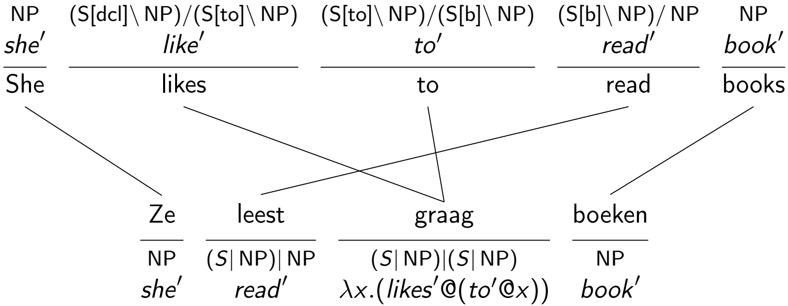
Example


NP
she′
She
(S[dcl]\ NP)/(S[to]\ NP)
like′
likes
(S[to]\ NP)/(S[b]\ NP)
to′
to
(S[b]\ NP)/ NP
read′
read
NP
book′
books
Ze
NP
she′
leest
(*S*| NP)| NP
read′
graag
boeken

Example


NP
she′
She
(S[dcl]\ NP)/(S[to]\ NP)
like′
likes
(S[to]\ NP)/(S[b]\ NP)
to′
to
(S[b]\ NP)/ NP
read′
read
NP
book′
books
Ze
NP
she′
leest
(S| NP)| NP
read′
graag
(S| NP)|(S| NP)
λx.(likes′@(to′@x))
boeken

Example


NP
she′
She
(S[dcl]\ NP)/(S[to]\ NP)
like′
likes
(S[to]\ NP)/(S[b]\ NP)
to′
to
(S[b]\ NP)/ NP
read′
read
NP
book′
books
Ze
NP
she′
leest
(S| NP)| NP
read′
graag
(S| NP)|(S| NP)
λx.(likes′@(to′@x))
boeken
NP
book′

Experiment

- 13K English-Dutch sentence pairs
- C&C/Boxer to make derivations for English sentences
- IBM Model 4 to align sentences, 3-best alignments in both directions
- for each word, cutoff is 0.1
- result: 24K Dutch candidate lexical items

Derivation Projection

Need target-language training derivations

- parse target-language sentences
- use target-language candidate lexical items from category projection
- constraint: parse must have same semantics as the English one

Example

Ze	leest	graag	boeken
NP	$(S\|\mathrm{NP})\|\mathrm{NP}$	$(S\|\mathrm{NP})\|(S\|\mathrm{NP})$	NP
she'	$read'$	$\lambda x.(likes'@(to'@x))$	$book'$

Example

$$
\begin{array}{cccc}
\text{Ze} & \text{leest} & \text{graag} & \text{boeken} \\
\hline
\mathrm{NP} & (S|\mathrm{NP})|\mathrm{NP} & (S|\mathrm{NP})|(S|\mathrm{NP}) & \mathrm{NP} \\
\mathit{she}' & \mathit{read}' & \lambda x.(\mathit{likes}'@(\mathit{to}'@x)) & \mathit{book}' \\
\hline
 & \multicolumn{2}{c}{<^{1}_{\times}} & \\
 & \multicolumn{2}{c}{(S|\mathrm{NP})|\mathrm{NP}} & \\
 & \multicolumn{2}{c}{\lambda x.(\mathit{likes}'@(\mathit{to}'@(\mathit{read}'@x)))} &
\end{array}
$$

Example

$$
\begin{array}{cccc}
\text{Ze} & \text{leest} & \text{graag} & \text{boeken} \\
\hline
\mathrm{NP} & (S|\mathrm{NP})|\mathrm{NP} & (S|\mathrm{NP})|(S|\mathrm{NP}) & \mathrm{NP} \\
she' & read' & \lambda x.(likes'@(to'@x)) & book'
\end{array}
$$

$$
\dfrac{(S|\mathrm{NP})|\mathrm{NP} \quad read' \qquad (S|\mathrm{NP})|(S|\mathrm{NP}) \quad \lambda x.(likes'@(to'@x))}{(S|\mathrm{NP})|\mathrm{NP} \quad \lambda x.(likes'@(to'@(read'@x)))} <^{1}_{\times}
$$

$$
\dfrac{(S|\mathrm{NP})|\mathrm{NP} \quad \lambda x.(likes'@(to'@(read'@x))) \qquad \mathrm{NP} \quad book'}{S|\mathrm{NP} \quad likes'@(to'@(read'@book'))} >^{0}
$$

Example

$$
\begin{array}{cccc}
\text{Ze} & \text{leest} & \text{graag} & \text{boeken} \\
\hline
\text{NP} & (S|\text{NP})|\text{NP} & (S|\text{NP})|(S|\text{NP}) & \text{NP} \\
\textit{she}' & \textit{read}' & \lambda x.(\textit{likes}'@(\textit{to}'@x)) & \textit{book}'
\end{array}
$$

$$
\frac{(S|\text{NP})|\text{NP} \quad \textit{read}' \qquad (S|\text{NP})|(S|\text{NP}) \quad \lambda x.(\textit{likes}'@(\textit{to}'@x))}{(S|\text{NP})|\text{NP} \quad \lambda x.(\textit{likes}'@(\textit{to}'@(\textit{read}'@x)))} <^{1}_{\times}
$$

$$
\frac{(S|\text{NP})|\text{NP} \quad \lambda x.(\textit{likes}'@(\textit{to}'@(\textit{read}'@x))) \qquad \text{NP} \quad \textit{book}'}{S|\text{NP} \quad \textit{likes}'@(\textit{to}'@(\textit{read}'@\textit{book}'))} >^{0}
$$

$$
\frac{\text{NP} \quad \textit{she}' \qquad S|\text{NP} \quad \textit{likes}'@(\textit{to}'@(\textit{read}'@\textit{book}'))}{S \quad (\textit{likes}'@(\textit{to}'@(\textit{read}'@\textit{book}')))@\textit{she}'} <^{0}
$$

Example

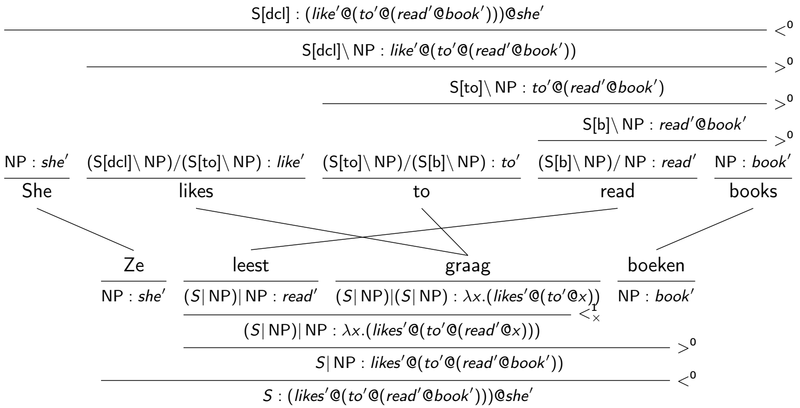
Ze leest graag boeken →

x1 p1 e1	
female(x1)	
p1:	x2 e2 book.n.01(x2) read.v.01(e2) Agent(e2, x1) Theme(e2, x2)
like.v.02(e1) Experiencer(e1, x1) Stimulus(e1, p1)	

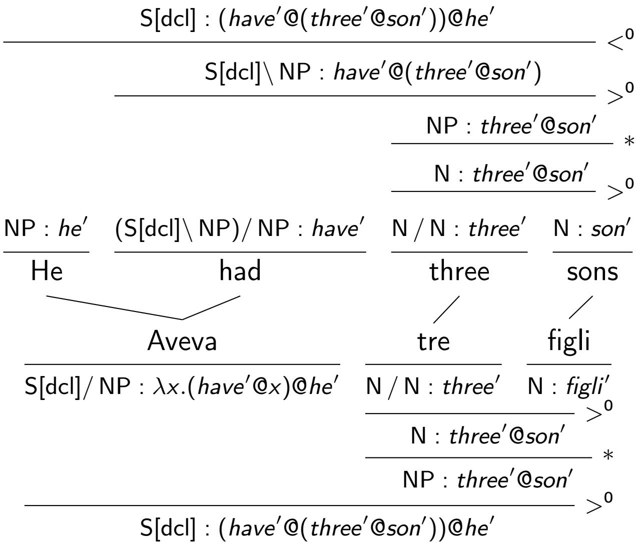
Experiment

- trying to project 13K training derivations
- sometimes, no parse found
- sometimes, search space explodes despite pruning
  - abort if agenda > 256
- successful for 8K derivations

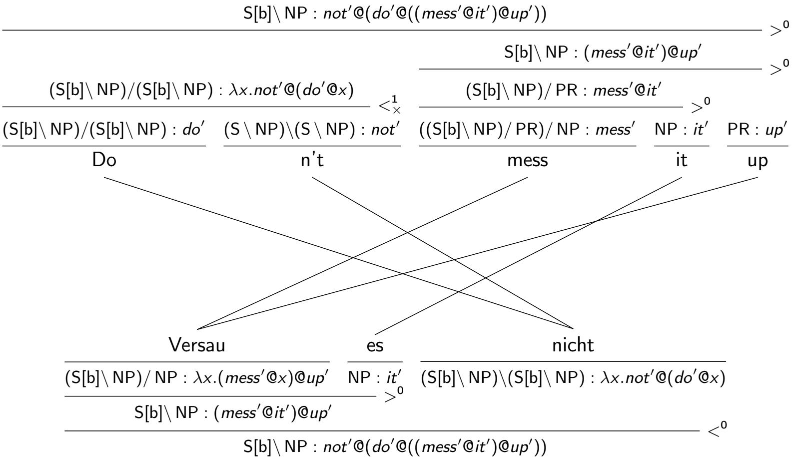
Verb + complementizer vs. adverb



Pro-drop


S[dcl] : (have′@(three′@son′))@he′
<⁰
S[dcl]\ NP : have′@(three′@son′)
>⁰
NP : three′@son′
*
N : three′@son′
>⁰
NP : he′
(S[dcl]\ NP)/ NP : have′
N / N : three′
N : son′
He
had
three
sons
Aveva
tre
figli
S[dcl]/ NP : λx.(have′@x)@he′
N / N : three′
N : figli′
>⁰
N : three′@son′
*
NP : three′@son′
>⁰
S[dcl] : (have′@(three′@son′))@he′

Particle vs. Simplex Verb



Shift-reduce Parsing

- Based on English CCG parser of Zhang and Clark (2011)
- Action types: shift, combine, unary, skip, finish
- Allows fragmentary parses

Parser Training

- Training data: 8K Dutch derivations obtained in derivation projection
- Features: Zhang and Clark (2011)
- Averaged perceptron with early update and beam search ($b = 16,\ T = 10$)

Dealing with OOV Words at Test Time

Pick schematic lexical entries for POS extracted from lexicon, e.g.,

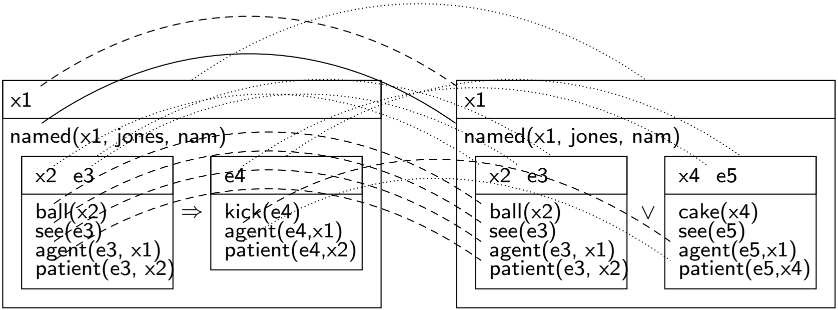
$$\frac{\text{getuige/nounsg}}{N : f} \quad \text{where } f = \lambda x. \begin{array}{|c|}\hline \\ \hline \_\_\text{UNKNOWN}\_\_(x) \\ \hline\end{array}$$

Test Data

- 150 English-Dutch sentence pairs
- Discourse Representation Structures automatically produced by C&C/Boxer
- manually corrected by 2 annotators, then adjudicated

Evaluation: Graph Match Measure

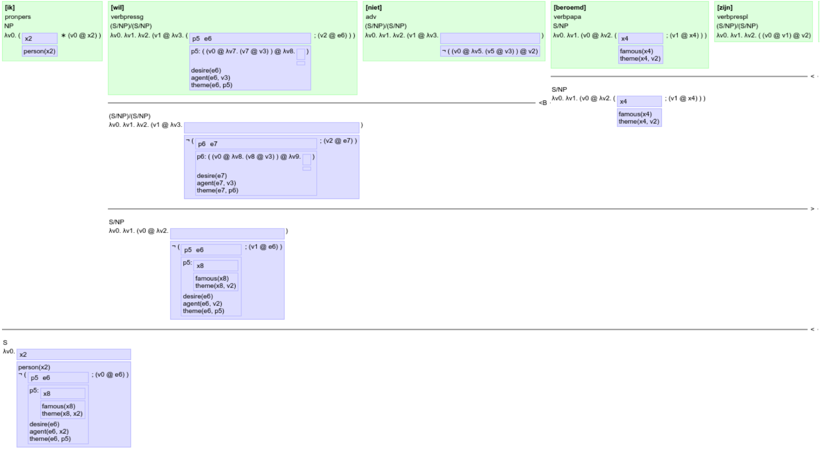
Le and Zuidema (2012)



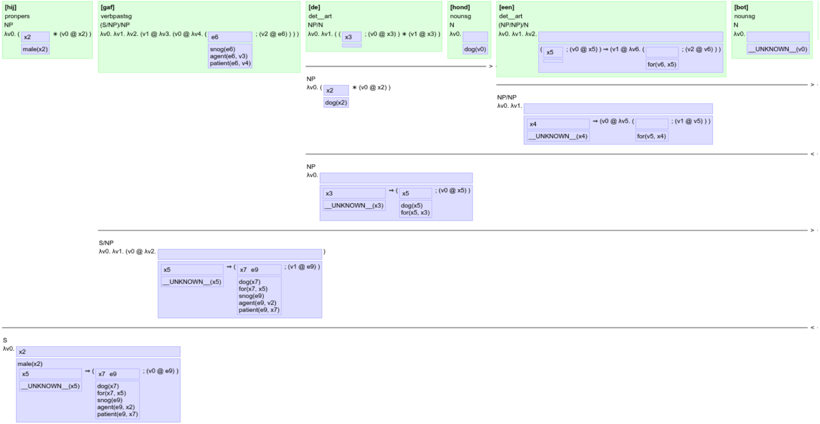
Evaluation: Baseline and Upper Bound

- baseline: most frequent lexical entry/schema for each word, all unconnected
- upper bound: silver standard, unseen symbols replaced with __UNKNOWN__

Where It Goes Well



Where It Doesn't



Results

Language System	English C&C/Boxer	Dutch Our system
Structure	78.88	69.22
Structure+concepts	68.10	48.35
Structure+relations	69.06	60.23
Structure+concepts+relations	58.40	42.99

Conclusions

- Learning an open-domain semantic parser cross-lingually: it can be done, but...
  - Need larger vocabulary
  - Need to study category/derivation projection in more detail
- CCG handles translation divergences well

2018 Update

- new experiments on Dutch, German, Italian
- also learning slash directions
- only syntax, for now
- promising results
- projected parses used to bootstrap the Parallel Meaning Bank (`http://pmb.let.rug.nl`)

References I

Beltagy, I., Roller, S., Cheng, P., Erk, K., and Mooney, R. J. (2015). Representing meaning with a combination of logical form and vectors. *CoRR*, abs/1505.06816.

Bos, J. (2015). Open-domain semantic parsing with boxer. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, pages 301–304.

Curran, J., Clark, S., and Bos, J. (2007). Linguistically Motivated Large-Scale NLP with C
&C and Boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Companion Volume: Proceedings of the Demo and Poster Sessions*, pages 33–36.

Dukes, K. (2014). SemEval-2014 task 6: Supervised semantic parsing of robotic spatial commands. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 45–53.

References II

Kamp, H. (1984). A Theory of Truth and Semantic Representation. In Groenendijk, J., Janssen, T. M., and Stokhof, M., editors, *Truth, Interpretation and Information*, pages 1–41. FORIS, Dordrecht, Holland/Cinnaminson, U.S.A.

Kwiatkowksi, T., Zettlemoyer, L., Goldwater, S., and Steedman, M. (2010). Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1223–1233.

Le, P. and Zuidema, W. (2012). Learning compositional semantics for open domain semantic parsing. In *Proceedings of COLING 2012*, page 33–50.

Reddy, S., Lapata, M., and Steedman, M. (2014). Large-scale semantic parsing without question-answer pairs. *Transactions of the Association for Computational Linguistics*, 2:377–392.

References III

Zettlemoyer, L. and Collins, M. (2007). Online learning of relaxed ccg grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Zhang, Y. and Clark, S. (2011). Shift-reduce CCG parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 683–692.