Cross-lingual transfer of a semantic parser via parallel data

Kilian Evang
k.evang@rug.nl

University of Groningen

CLIN26. VU Amsterdam
2015-12-18

Semantic parsing: what?

From Words to (Logical) Meaning

She likes to read books $\rightarrow$

x1 p1 e1
female(x1)
p1: [x2 e2 | book(e2), read(e2), Actor(e2, x1), Theme(e2, x2)]
like(e1)
Actor(e1, x1)
Topic(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]

She likes to read books → tokenizer → C&C tools → Boxer →

x1 p1 e1
female(x1)
p1: [x2 e2 | book(e2), read(e2), Actor(e2, x1), Theme(e2, x2)]
like(e1)
Actor(e1, x1)
Topic(e1, p1)

System for other languages?

Goal

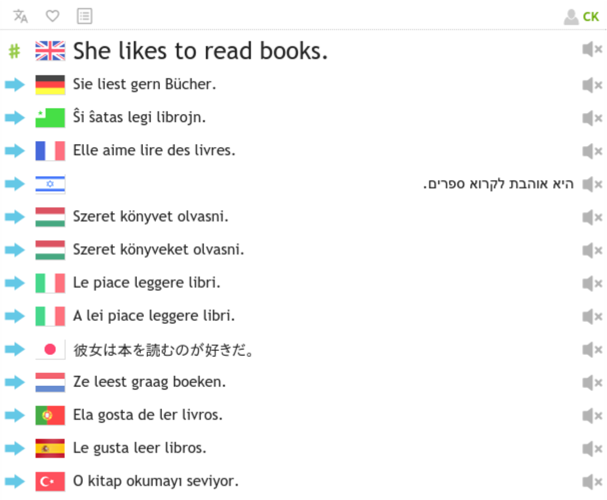
Learn (rudimentary) semantic parser from nothing but

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

Method

1. meaning annotation by proxy
2. inducing lexical items using word alignments
3. shift-reduce parsing

Parallel corpus: Tatoeba.org



Automatic annotation of English sentences

$$
\begin{array}{ccccc}
\text{She} & \text{likes} & \text{to} & \text{read} & \text{books} \\
\hline
NP & (S[\text{dcl}]\backslash NP)/(S[\text{to}]\backslash NP) & (S[\text{to}]\backslash NP)/(S[\text{b}]\backslash NP) & (S[\text{b}]\backslash NP)/NP & NP \\
she' & like' & to' & read' & book'
\end{array}
$$

$$
\frac{(S[\text{b}]\backslash NP)/NP : read' \quad NP : book'}{S[\text{b}]\backslash NP : read'@book'} >_0
$$

$$
\frac{(S[\text{to}]\backslash NP)/(S[\text{b}]\backslash NP) : to' \quad S[\text{b}]\backslash NP : read'@book'}{S[\text{to}]\backslash NP : to'@(read'@book')} >_0
$$

$$
\frac{(S[\text{dcl}]\backslash NP)/(S[\text{to}]\backslash NP) : like' \quad S[\text{to}]\backslash NP : to'@(read'@book')}{S[\text{dcl}]\backslash NP : like'@(to'@(read'@book'))} >_0
$$

$$
\frac{NP : she' \quad S[\text{dcl}]\backslash NP : like'@(to'@(read'@book'))}{S[\text{dcl}] : (like'@(to'@(read'@book')))@she'} <_0
$$

Meaning annotation by proxy

$(like'@(to'@(read'@book')))@she'$
Ze leest graag boeken

Generating candidate lexical items

- [Zettlemoyer and Collins, 2007, Kwiatkowski et al., 2013]: hand-written lexical templates for English
- [Kwiatkowski et al., 2011]: recursively splitting gold-standard meaning representations, heuristics to constrain search space
- this work: from the English parse tree
    - use the same lexical semantics as in English
    - assign them to Dutch words, possibly one to two or two to one
    - drop category subdistinctions (dcl, b, to...)
    - use undirected slashes

Example alignment (ideal)

NP	$(S[\text{dcl}]\backslash NP)/(S[\text{to}]\backslash NP)$	$(S[\text{to}]\backslash NP)/(S[\text{b}]\backslash NP)$	$(S[\text{b}]\backslash NP)/NP$	NP
she'	$like'$	to'	$read'$	$book'$
She	likes	to	read	books

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

Inducing Dutch lexical items

- extract one lexical item per translation unit, keep most frequent ones
- IBM model 4, all translation units from 5-best alignments in both directions
- for each word+POS, cutoff frequency is 0.1 of max

Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results

Shift-reduce parsing

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

Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes

Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes

Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes

Dutch training parse (example)

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

$$
\begin{array}{c}
\underline{\hspace{6em}(S|NP)|NP\ \ \lambda x.(likes'@(to'@x)) \quad <_{\times 1}} \\
(S|NP)|NP \\
\lambda x.(likes'@(to'@(read'@x))) \\
\hline
\quad >_0 \\
S|NP \\
likes'@(to'@(read'@book')) \\
\hline
\quad <_0 \\
S \\
(likes'@(to'@(read'@book')))@book'
\end{array}
$$

Parser training

- Training data: Dutch derivations obtained with forced decoding
- Averaged perceptron with beam search ($b = 16$)
- Early update [Collins and Roark, 2004]
- Features: [Zhang and Clark, 2011]

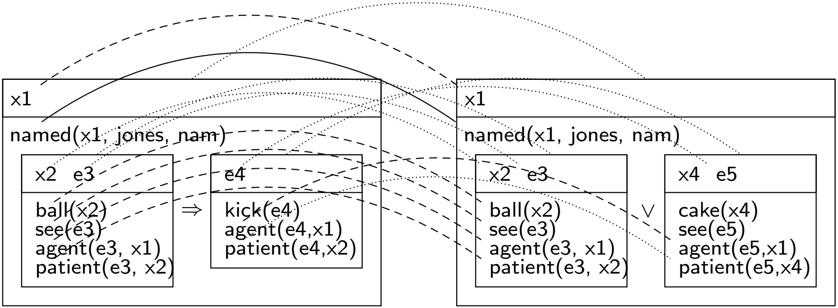
Dealing with unknown words at test time

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

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

Evaluation: graph match measure

[Allen et al., 2008, 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__

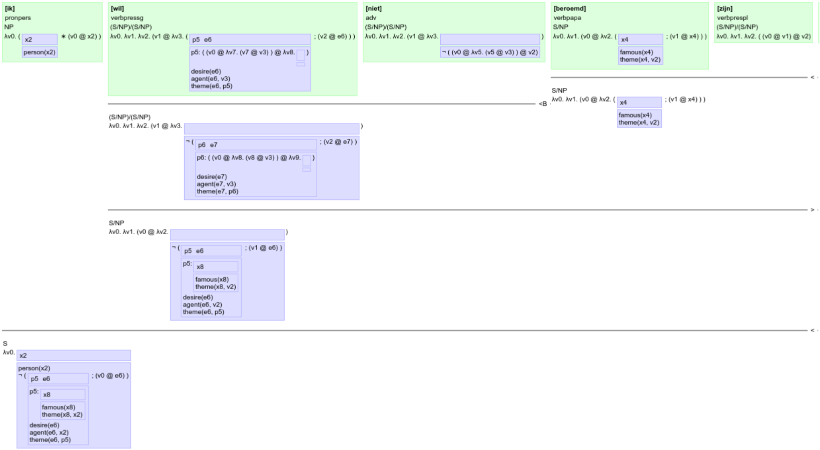
Evaluation: baseline and upper bound

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

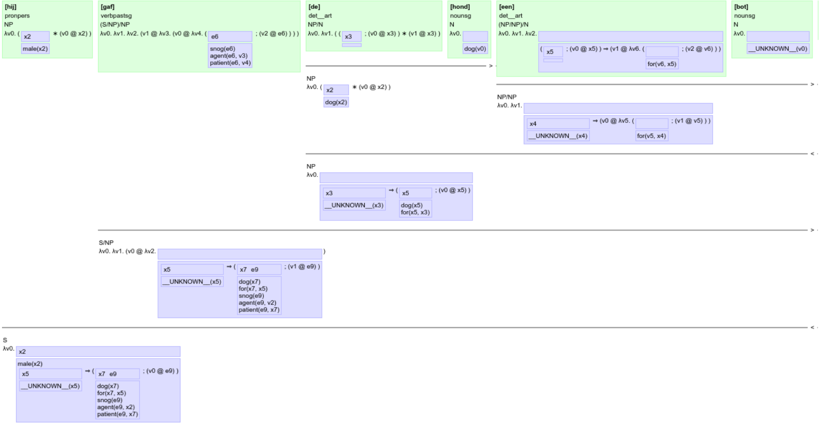
Results on development test set (1,641 sentences)

		rec	prec	f1
baseline		.338	.344	.341
training iterations:	0	.362	.384	.372
	1	.507	.503	.505
	2	.504	.510	.507
	3	.508	.514	.511
	4	.510	.516	.513
	5	.507	.512	.510
			...	
upper bound		.962	.896	0.928

Where it goes well



Where it doesn't



Conclusions

- CCG suitable formalism for cross-lingual semantic parser induction
- Reasonable grammar learned Dutch
- Important areas for future work
  - Lexicon induction: tweak to get more training data
  - Treat English parses, word alignments as latent
  - Morphology
  - Lexical semantics

Interested in collaborating? Let me know!

Conclusions

- CCG suitable formalism for cross-lingual semantic parser induction
- Reasonable grammar learned Dutch
- Important areas for future work
  - Lexicon induction: tweak to get more training data
  - Treat English parses, word alignments as latent
  - Morphology
  - Lexical semantics

Interested in collaborating? Let me know!

References I

Allen, J. F., Swift, M., and de Beaumont, W. (2008).
Deep Semantic Analysis of Text.
In Bos, J. and Delmonte, R., editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, volume 1 of *Research in Computational Semantics*, pages 343–354. College Publications.

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.

Collins, M. and Roark, B. (2004).
Incremental parsing with the perceptron algorithm.
In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*.

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, Prague, Czech Republic.

Dukes, K. (2014).
Semeval-2014 task 6: Supervised semantic parsing of robotic spatial commands.
SemEval 2014, page 45.

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.

References II

Kwiatkowski, T., Choi, E., Artzi, Y., and Zettlemoyer, L. (2013).
Scaling semantic parsers with on-the-fly ontology matching.
In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1545–1556. Association for Computational Linguistics.

Kwiatkowski, T., Zettlemoyer, L., Goldwater, S., and Steedman, M. (2011).
Lexical generalization in CCG grammar induction for semantic parsing.
In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1512–1523. Association for Computational Linguistics.

Le, P. and Zuidema, W. (2012).
Learning compositional semantics for open domain semantic parsing.
In *Proceedings of COLING 2012*, volume 1 of *COLING 2012*, page 33–50. Association for Computational Linguistics.

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.

Zettlemoyer, L. S. 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-2007)*, pages 678–687.

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. Association for Computational Linguistics.

References III

Zhao, K. and Huang, L. (2015).
Type-driven incremental semantic parsing with polymorphism.
In *Proceedings of NAACL HLT*.