

Generating Disambiguating Paraphrases for Use in Crowd-Sourced Judgments of Meaning

Ethan Hill, Manjuan Duan, Michael White

{hill.1303, duan.59, white.1240}@osu.edu

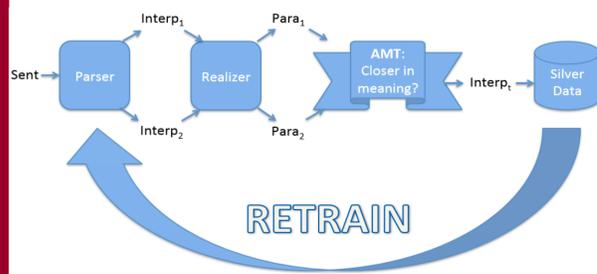
Introduction

Syntactic annotations are expensive and time consuming to obtain. A study by Jha et al. (2010) shows that annotators from Amazon's Mechanical Turk (AMT) can achieve high accuracy in prepositional phrase attachment disambiguation after initial linguistic training. In this project, we have developed a system for automatically generating sentence disambiguation tasks for a greater variety of structural ambiguities where Turkers need only select the paraphrase with the correct meaning. Preliminary results show that the judgments collected from AMT for these disambiguation tasks are significantly more accurate than chance, indicating the potential of the method for creating training corpora to adapt a parser to a new domain. We are currently evaluating the effectiveness of the method on a larger data set and plan to test how well it works for parser adaptation.

Paraphrase Generation Methods

- Collect sentences from the domain of interest
- Parse the sentences using OpenCCG, yielding an n-best list of parses; for each sentence, the top two parses that are amenable to disambiguation via paraphrasing are selected (excluding minor POS and dependency label differences as well as attachment ambiguities involving auxiliaries)
- Apply structure rewrites if possible, and realize the parses using OpenCCG's surface realizer
- Choose realizations which break up the ambiguity and contain the dependencies which contribute to the ambiguity

Fig 1: Big Picture Overview

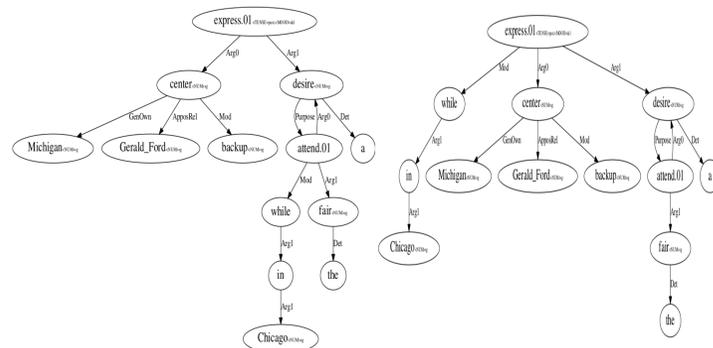


Reverse Realizations (Reversals)

- Parse a sentence to decide whether this sentence is structurally ambiguous
- Realize back into sentences those parses which represent different interpretations of the sentence
- Verify that the realizations still contain the dependencies that contribute the original ambiguity
- Collect AMT judgments on which realization means the same as the original sentence

Reversals (cont.)

Fig 2: Competing parses for Michigan's backup center, Gerald Ford, expressed a desire to attend the fair while in Chicago.



Realizations of the above competing parses.

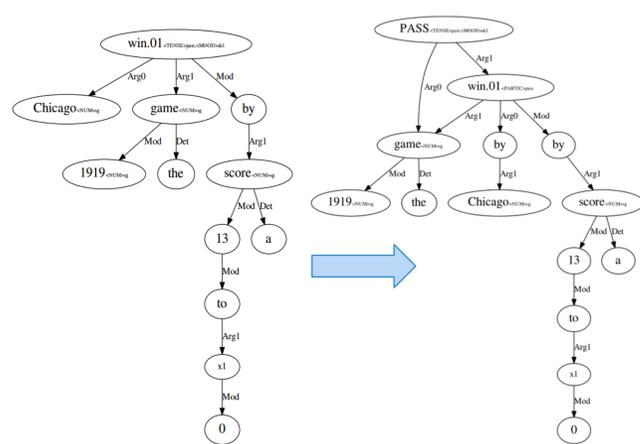
Left: "Michigan's backup center, Gerald Ford, expressed a desire to attend while in Chicago the fair."

Right: "Michigan's backup center, Gerald Ford, expressed while in Chicago a desire to attend the fair."

Passivization and Clefting Rewrites

- Force passive voice and/or a cleft structure onto sentences with attachment ambiguities between the verb and object
- Effective in disambiguating PP attachment ambiguities

Fig 3: Parse and passive rewrite for Chicago won the 1919 game by a 13 to 0 score.

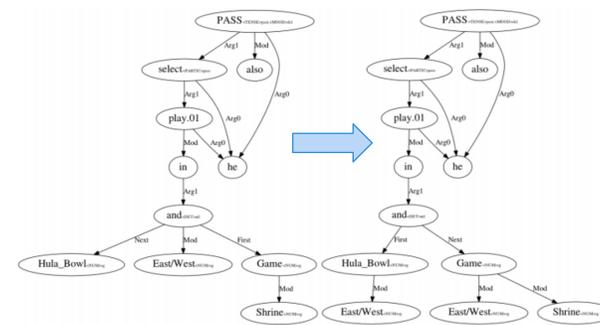


Realization of correct parse after passive rewrite - "The 1919 game was won by Chicago by a 13 to 0 score."

Coordinating Conjunction Rewrites

- Apply modifiers, determiners dependent on conjunction to all arguments of the conjunction
- Reverse the order of the conjunction arguments
- Forces verbosity for effective disambiguation

Fig 4: Parse and coordination rewrite for He was also selected to play in East/West Shrine game and Hula bowl.



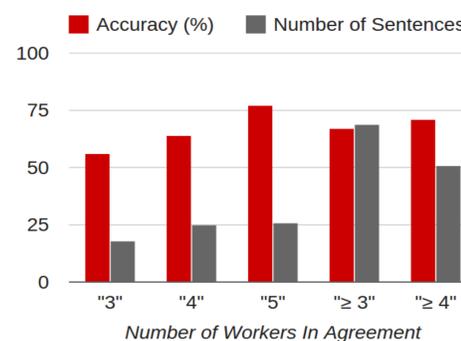
Realization of incorrect parse after rewrite - "He also was selected to play in East/West Hula bowl and East/West Shrine game."

Annotation Experiment

- For the preliminary results below, 92 sentences from Wikipedia of length 5-20 words were randomly chosen from two domains (*B1G football* and *prehistoric reptiles*)
- 5 judgments were collected from AMT for each sentence
- The correct parses were annotated by the researchers with "Top", "Next" and "Neither" (neither parse correct)
- The inter-annotator agreement was 82.5% for three-way agreement and 90.8% for agreement excluding the cases involving "neither"
- Annotation accuracy was evaluated on the sentences annotated as Top/Next

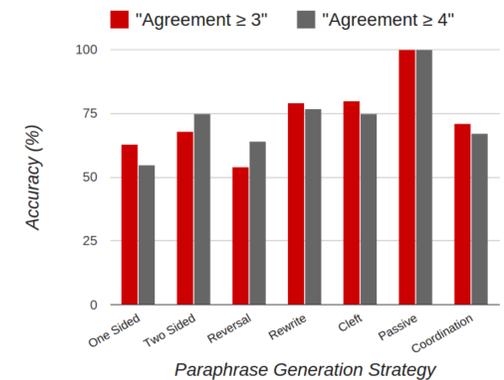
Preliminary Results

Chart 1: Accuracy in terms of AMT worker agreement



Preliminary Results (cont.)

Chart 2: Accuracy in terms of paraphrase strategy



Conclusions

- Annotation accuracy is significantly better than chance ($p < 0.01$, binomial test)
- Two-sided paraphrases work better than one-sided ones, and rewrites work better than reversals
- Accuracy for PP-attachment rewrites is in the same ballpark as Jha et al. (2010)
- About 29% of correctly annotated parses are non-top parses, representing a substantial correction of the current parser on these domains

Ongoing work

- Investigate cases where reversals do not work well
- Collect 10x more judgments from the two domains
- Retrain parser with annotations from the new judgments and evaluate parser domain adaptation
- Investigate other rewrite methods which might increase overall disambiguation coverage of ambiguities

Acknowledgments

We would like to thank the National Science Foundation (IIS-1319318) for providing funding for this project.

References

Mukund Jha, Jacob Andreas, Kapil Thadani, Sara Rosenthal, and Kathleen McKeown, (2010). Corpus creation for new genres: A crowdsourced approach to pp attachment, *Workshop on Creating Speech and Text Language Data with Amazon's Mechanical Turk at NAACL-HLT*, Los Angeles, California

Michael White, (2006). Efficient Realization of Coordinate Structures in Combinatory Categorical Grammar, *Research on Language and Computation*, 4:1, 39-75

Michael White and Rajakrishnan Rajkumar, (2009). Perceptron Reranking for CCG Realization, *Proc. of EMNLP 2009*, Singapore.410-419.

Manjuan Duan and Michael White, (2014). That's Not What I Meant! Using Parsers to Avoid Structural Ambiguities in Generated Text, *Proc. of ACL 2014*. Baltimore, Maryland. 413-423