Augmenting FrameNet Via PPDB

Pushpendre Rastogi1 And Benjamin Van Durme1,2

1Center for Language and Speech Processing
2Human Language Technology Center of Excellence

Introduction

Frame semantic parsers analyze a sentence and annotate its frame evoking words and the corresponding frame elements/arguments. The task of frame semantic parsing was introduced by Gildea and Jurafsky (2002) shortly after the release of FrameNet (Fillmore and Baker, 2001). CMU’s SEMAFOR system (Das et al., 2013) is the current state-of-the-art. Like other semantic resources, the coverage of FrameNet is incomplete. State-of-the-art frame semantic parsers thus employ heuristics to identify the frame evoked by out-of-vocabulary items (OOVs).

For example the word jettison does not appear in FrameNet. A parser may look up its synonym from WordNet, e.g., Abandon and its associated frame in FrameNet and assign to jettison that frame.

PPDB

Ganitkevitch et al. (2013) released a large collection of lexical, phrasal and syntactic paraphrases collectively called PPDB.*

Lexical : Two words with the same meaning.

Phrasal : Two strings of words with the same meaning.

Syntactic : Two strings consisting of words and non-terminal categories that have the same meaning, e.g., [S/PP] ||| NP/Noun explosion VP/PP ||| NP/Noun blast VP/PP

A lexical rule in PPDB looks like the following: [VB] ||| help ||| assist ||| p(e)=2.832, p(f|e)=1.551, ... This rule conveys that the log-probability that help would be paraphrased by the word assist is -2.832 but the log probability of assist being paraphrased as help is -1.551.

References


Conclusion and Error Analysis

We found that over a sample of 25 target words the top three paraphrases produced by PPDB XL evoked the correct frame and were grammatical 65% of the time.*

We found two major reasons for ungrammaticality of lexical paraphrases.

1. Within FrameNet some sentences will have a single token annotated as trigger, when in fact it is part of a multi-word expression. For example, it was grammatically infelicitous to replace for a new state health lab , millions of doses of antiviral drugs and a fund to help meet basic needs after a disaster, a legislative panel recommended Thursday.

2. The other major source of error was inaccuracies in PPDB itself. We found that for a large number of cases when PPDB XL did not have a high number of paraphrases, the paraphrases were wrong (e.g., PPDB XL had only 2 paraphrases for the words lab and millions.)

We have released the generated corpus as well as the manual annotations at http://cs.jhu.edu/~prastog3/res/ftp/ppdb.html

Acknowledgments

This material is based on research sponsored by the NSF under grant IIS-1249516 and DARPA under agreement number FA8750-13-2-0017.

Contact

pushpendre@jhu.edu, vandurne@cs.jhu.edu

References


Conclusion and Error Analysis

We found that over a sample of 25 target words the top three paraphrases produced by PPDB XL evoked the correct frame and were grammatical 65% of the time.*

We found two major reasons for ungrammaticality of lexical paraphrases.

1. Within FrameNet some sentences will have a single token annotated as trigger, when in fact it is part of a multi-word expression. For example, it was grammatically infelicitous to replace for a new state health lab , millions of doses of antiviral drugs and a fund to help meet basic needs after a disaster, a legislative panel recommended Thursday.

2. The other major source of error was inaccuracies in PPDB itself. We found that for a large number of cases when PPDB XL did not have a high number of paraphrases, the paraphrases were wrong (e.g., PPDB XL had only 2 paraphrases for the words lab and millions.)

We have released the generated corpus as well as the manual annotations at http://cs.jhu.edu/~prastog3/res/ftp/ppdb.html

Acknowledgments

This material is based on research sponsored by the NSF under grant IIS-1249516 and DARPA under agreement number FA8750-13-2-0017.

Contact

pushpendre@jhu.edu, vandurne@cs.jhu.edu

References


Conclusion and Error Analysis

We found that over a sample of 25 target words the top three paraphrases produced by PPDB XL evoked the correct frame and were grammatical 65% of the time.*

We found two major reasons for ungrammaticality of lexical paraphrases.

1. Within FrameNet some sentences will have a single token annotated as trigger, when in fact it is part of a multi-word expression. For example, it was grammatically infelicitous to replace for a new state health lab , millions of doses of antiviral drugs and a fund to help meet basic needs after a disaster, a legislative panel recommended Thursday.

2. The other major source of error was inaccuracies in PPDB itself. We found that for a large number of cases when PPDB XL did not have a high number of paraphrases, the paraphrases were wrong (e.g., PPDB XL had only 2 paraphrases for the words lab and millions.)

We have released the generated corpus as well as the manual annotations at http://cs.jhu.edu/~prastog3/res/ftp/ppdb.html

Acknowledgments

This material is based on research sponsored by the NSF under grant IIS-1249516 and DARPA under agreement number FA8750-13-2-0017.

Contact

pushpendre@jhu.edu, vandurne@cs.jhu.edu

References


