Light Textual Inference for Semantic Parsing

Kyle Richardson *Jonas Kuhn*
Institute for Natural Language Processing,
University of Stuttgart
{kyle,jonas}@ims.uni-stuttgart.de

ABSTRACT
There has been a lot of recent interest in Semantic Parsing, centering on using data-driven techniques for mapping natural language to full semantic representations (Mooney, 2007). One particular focus has been on learning with ambiguous supervision (Chen and Mooney, 2008; Kim and Mooney, 2012), where the goal is to model language learning within broader perceptual contexts (Mooney, 2008). We look at learning light inference patterns for Semantic Parsing within this paradigm, focusing on detecting speaker commitments about events under discussion (Nairn et al., 2006; Karttunen, 2012). We adapt PCFG induction techniques (Börschinger et al., 2011; Johnson et al., 2012) for learning inference using event polarity and context as supervision, and demonstrate the effectiveness of our approach on a modified portion of the Grounded World corpus (Bordes et al., 2010).

KEYWORDS: Semantic Parsing, Computational Semantics, Detecting Textual Entailment, Grammar Induction.

Proceedings of COLING 2012: Posters, pages 1007–1018,
COLING 2012, Mumbai, December 2012.

1 Overview and Motivation

Semantic Parsing is a subfield in NLP that looks at using data-driven techniques for mapping language expressions to complete semantic representations (Mooney, 2007). A variety of corpora and learning techniques have been developed for these purposes, both for doing supervised learning (Kate et al., 2005; Kwiatkowski et al., 2010) and learning in more complex (ambiguous) settings (Chen and Mooney, 2008, 2011). In many studies, the learning is done by finding alignments between (latent) syntactic patterns in language and parts of the target semantic representations, often using techniques from Statistical Machine Translation (Wong and Mooney, 2006; Jones et al., 2012). Despite achieving impressive results in different domains, learning semantic inference patterns is often not addressed, making it unclear how to apply these methods to tasks like Detecting Textual Entailment. In this work, we show how to learn light (syntactic) inference patterns for textual entailment using loosely-supervised Semantic Parsing methods.

Detecting Textual Entailment is a topic that has received considerable attention in NLP, largely because of its connection to applications such as question answering, summarization, paraphrase generation, and many others. The goal, loosely speaking, is to detect entailment inference relationships between pairs of sentences (Dagan et al., 2005). More recent work on Hedge and Event Detection (Farkas et al., 2010) has focused on similar issues related to determining event certainty, especially in the biomedical domain (Example 3 (Thompson et al., 2011)). Four inferences are shown in Examples 1-4, and relate to implied speaker commitments (Karttunen, 2012; Nairn et al., 2006) about events under discussion.

1. John forgot to help Mary organize the meeting
    (a) $\models$ John didn't help Mary organize the meeting
2. John remembered (to not neglect) to turn off the lights before leaving work
    (a) $\models$ John turned off some lights
3. NF-kappa B p50 alone *fails to* (=*doesn't*) stimulate kappa B-directed transcription
4. The camera {didn't manage, managed} to impress me (=*negative/positive* opinion)

In Example 1, the speaker of the sentence is committed to the belief that the main event (i.e. *helping Mary organize the meeting*) did not occur, whereas the opposite is true in Example 2. This is triggered by the implicative phrases *Forget to X* and *Remember to X*, which affect the polarity of the modified event X. These inferences relate to the semantics of English complement constructions, a topic well studied in Linguistics (Karttunen, 1971; Kiparsky and Kiparsky, 1970). They are also part of a wider range of inference patterns that are syntactic in nature, or visible from language surface form (Dowty, 1994). They have been of interest to studies in proof-theoretic semantics and Natural Logic, which look at doing inference on natural language directly (MacCartney and Manning, 2007; Moss, 2010; Valencia, 1991).

We aim to learn these implicative patterns, building on existing computational work. (Nairn et al., 2006; Karttunen, 2012) provide a classification of implicative verbs according to the effect they have on their surrounding context. They observe that implicative constructions differ in terms of the polarity contexts they occur in, and the effect they have in these contexts.

As illustrated in Table 1, one-way implicatives occur in a single polarity, whereas two-way implicatives occur in both. For example, *Forget to X* in Example 1 switches polarity in a positive context to negative, and has the opposite effect in a negative context, giving it the implicative signature (+)(-), (-)(+) (i.e. start context, result).

Implicatives can be productively stacked together as shown in Example 2. Determining the resulting inference for an arbitrary nesting of implicatives requires computing the relative polarity of each smaller phrase, which is the idea behind the polarity propagation algorithm (Nairn et al., 2006). This can be done directly from syntax by traversing a tree annotated with polarity information and calculating the polarity interactions incrementally. This general strategy for doing inference, which relies on syntactic and lexical features alone, avoids a full semantic analysis and translation into logic (Bos and Markert, 2005), and has been successfully applied to more general textual entailment tasks (MacCartney and Manning, 2007, 2008).

One problem with the approach of (Nairn et al., 2006), however, is that the implicative signatures of verbs must be manually compiled, as there are no standard datasets available for doing learning. To our knowledge, there has been little work on learning these specific patterns (some related studies (Danescu-Niculescu-Mizil et al., 2009; Cheung and Penn, 2012)), which would be useful for applying these methods to languages and domains where resources are not available. Further, their algorithm encodes the lexical properties as hard facts, making it hard to model potential uncertainty and ambiguity associated with these inferences (e.g. if *John was* able to *do X*, how certain are we that he actually *did X*?)

The semantics of implicative expressions can often be inferred from non-linguistic context. Knowing that *managed to X* implies X is something we can learn from hearing this utterance in contexts where X holds. Recent studies on learning from ambiguous supervision for Semantic Parsing (Chen and Mooney, 2008, 2011) has looked at incorporating perceptual context (Mooney, 2008) of this sort into the learning process (see also (Johnson et al., 2012)). Work on the Sportscaster Corpus (Chen and Mooney, 2008) considers interpreting soccer commentary in ambiguous contexts where several closely occurring events are taking place. Their data is taken from a set of simulated soccer games extended with human commentary. Each comment is paired with a set of grounded events occurring in the game around the time of the comment. Using these ambiguous contexts as supervision, they learn how to map novel sentences to the correct grounded semantic representations.

We look at learning implicative inference in a similar grounded learning scenario, using ambiguous contexts and the polarity of events as supervision. We use a modified portion of the Grounded World corpus (Bordes et al., 2010), which was extended to have phrasal implicatives and ambiguous contexts. Three training examples are displayed in Figure 1, and an illustration of the analysis we aim to learn. Each example is situated in a virtual house environment and a context, and describes events taking place in the house. Details of the corpus and learning procedure are described in the next section.

2 Experiments

2.1 Materials

The original Grounded World corpus (Bordes et al., 2010) is a set of English descriptions situated within a virtual house, and was designed for doing named entity recognition and situated pronoun resolution. Inside the house is a fixed set of domain objects, including a set of actors (e.g. *father, brother*), a set of furniture pieces (e.g. *couch, table*), a set of rooms (e.g.

Type	Examples	Effect on Polarity
Two-way implicatives	*manage to*	(+)(+) | (–)(–)
	forget to	(+)(–) | (–)(+)
One-way +implicatives	*force to*	(+)(+)
	refuse to	(+)(–)
One-way -implicatives	*attempt to*	(–) (–)
	hesitate to	(–)(+)

Table 1: Types of Implicative Verbs from (Nairn et al., 2006; Karttunen, 2012)

# Sentences	# Token Gold Relations	Aver. Context Size
7,010 Total *(6,065 (85%) unique)*	**2,444** *(63 unique concepts)*	**2.17** *(90% > 1)*
1,863 Implicative Sentences *(26%)*		

Frequent Verb Tokens: refuse to, manage to, decline to, admit to, remember to, dare to
Complex Constructions: fail to neglect to, didn't refrain from, refuse to remember to
Examples:
Their grandmother [admitted$_{++}$ to]$_{+}$ drinking a little wine.
The brother [didn't$_{+-}$ dare$_{++}$ to]$_{-}$ move into the bedroom.
Their mom [remembered$_{++}$ to not$_{-+}$ forget$_{+-}$ to]$_{+}$ grab their toy from the closet

Table 2: Details of the extended Grounded World Corpus. The average context size is the average number of events in the ambiguous training contexts. On the bottom are some corpus examples with implicative constructions.

living room, bathroom), and a set of small objects (e.g. *doll, chocolate*), plus a set of 15 event types (e.g. *eating, sleeping, and drinking*).

For our study, we used a subset of 7,010 examples from the original training set, and modified the sentences to have syntactic alternations and paraphrases not seen in the initial corpus. 1,863 of these sentences were modified to have implicative constructions (using 70 unique constructions from 20 verb types, see examples in Table 2)[1] that relate to the original content of the sentence, in some cases creating negated forms of the original sentences. We expanded the original named-entity annotations to normalized semantic representations, and produced a set of distractor events (or observable contexts) for each example to make the data ambiguous.

Three training examples are shown in Figure 1. In the first example, the sentence is situated in three observable events (*sleeping*, *getting* and *bringing*). These can be viewed as events in the current context or the speaker's belief state. Additional information about the world state (i.e. location of objects) is provided from the original corpus for pronoun resolution, which we ignore. The last two examples have implicative constructions, the first one leading to a negative inference (*the sister is not sleeping in the bedroom/guestroom*). The last example leads to a positive inference (*the sister got a toy from the closet/storage*). We show the annotations from the original corpus for comparison.

Expanding the relations from the overall corpus and situating them within ambiguous contexts

[1] we used the phrasal implicative lexicon available at http://www.stanford.edu/group/csli_lnr/Lexical_Resources /phrasal-implicatives/, compiled by the authors of (Nairn et al., 2006; Karttunen, 2012)

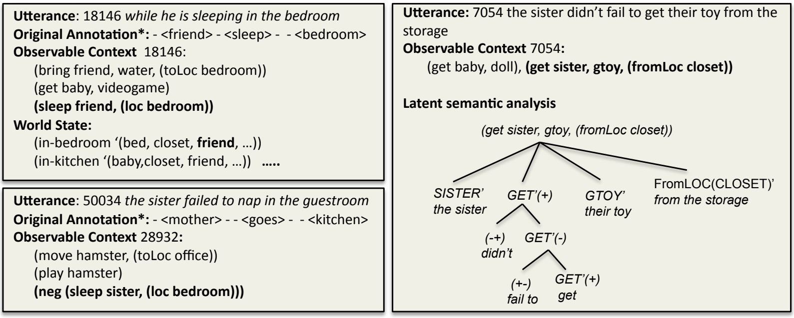

Figure 1: Ambiguous training examples from the extended corpus. The *latent semantic analysis* on the right is the representation we aim to learn from the observable context.

makes the learning task much harder. The overall aim is to use the ambiguous contexts and event polarity to construct a latent semantic analysis (see Figure 1), that derives the appropriate relation and inference (for a similar idea, see (Angeli et al., 2012)). In other words, we want to learn, merely from ambiguous supervision, how to map novel sentences to their correct semantic representations (the typical goal in Semantic Parsing), while also making the correct inferences. Notice that the target analysis is a kind of syntactic analysis, keeping to the idea that such inferences are visible from the surface.

2.2 Method

Many approaches to Semantic Parsing start by assigning rich structure to the target semantic representations, which can be used for finding alignments with latent structures in the language. Well known work by (Wong and Mooney, 2006) uses Statistical Machine Translation methods for finding alignments between semantic representations structured as trees and syntactic patterns in language. These alignments constitute the domain lexicon, and can be modeled using synchronous grammars. A number of such alignment-based learning methods have been proposed, using a variety of tools (Kate and Mooney, 2006; Jones et al., 2012; Wong and Mooney, 2006; Liang et al., 2011; Kwiatkowski et al., 2010).

(Börschinger et al., 2011) recast the problem in terms of an unsupervised PCFG induction problem, an idea also explored in (Johnson et al., 2012; Angeli et al., 2012; Kim and Mooney, 2012). They develop a method for automatically generating PCFGs from semantic relations, by decomposing parts of the relations into rewrite rules. Formally, a semantic PCFG G is $\left\langle V_{Non}, V_{Term}, Con, S_R, R, P \right\rangle$, where $S_R \in V_{Non}$ is the set of start symbols corresponding to the full semantic representations in a corpus, $Con \in V_{Non}$ is the set of contexts, R is the set of productions $X \rightarrow \beta$ for $X \in V_{Non}, \beta \in V^*$, and P is a probability function over R. A schema of the rules in R is shown at the top of Figure 2. Words in the training data (in V_{Term}) are assigned to all pre-terminals (i.e. semantic concepts) with equal probability, and the parameters are learned using EM and the ambiguous contexts as supervision.

We build a large PCFG from the semantic relations in our data using the method above. Rules

$$
\begin{array}{lcll}
\textit{S-Rel}(arg_1,\ldots,arg_n) & \longrightarrow & Contexts\ \{Phrase_{Rel},\ Phrase_{arg_1},\ \ldots,\ Phrase_{arg_n}\} & \\
Phrase_O & \longrightarrow & Word_O & Rel\,(arg_1\ldots,arg_n) \in Corpus \\
Phrase_O & \longrightarrow & PhX_O\ \ Word_O & O \in \{Rels, args\} \\
PhX_O & \longrightarrow & Word_O & \\
PhX_O & \longrightarrow & PhX_O\ \ Word_O & \\
PhX_O & \longrightarrow & PhX_O\ \ Word_{null} & \\
PhX_O & \longrightarrow & Word_{null} & \\
Word_O & \longrightarrow & w & w \in \{words\,in\,corpus\} \\
Word_{null} & \longrightarrow & w & \\
 & & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\
Phrase_{Rel} & \longrightarrow & Phrase_{pos-pos}\ \ Phrase_{Rel} & \\
Phrase_{Rel} & \longrightarrow & Phrase_{neg-pos}\ \ Phrase_{negRel} & \\
Phrase_{negRel} & \longrightarrow & Phrase_{pos-neg}\ \ Phrase_{Rel} & \\
Phrase_{negRel} & \longrightarrow & Phrase_{neg-neg}\ \ Phrase_{negRel} & \\
Phrase_Z & \longrightarrow & Phrase_{MON} & Z \in \{pos-pos, neg-neg\} \\
Phrase_W & \longrightarrow & Phrase_{NMON} & W \in \{pos-neg, neg-pos\} \\
Phrase_P & \longrightarrow & Word_P & P \in \{NMON, MON\} \\
Phrase_P & \longrightarrow & PhX_P\ \ Word_P & \\
PhX_P & \longrightarrow & PhX_P\ \ Word_P & \\
\ldots\ldots\ldots & & & \\
\ldots\ldots\ldots & & & \\
Word_P & \longrightarrow & w &
\end{array}
$$

Figure 2: PCFG schema (Börschinger et al., 2011) extended with rules for implicative phrases shown under the dotted lines. Note that word order is not modeled. The top most rule encodes all combinations of rules on the right in the brackets.

for detecting implicative patterns are specified at the bottom of Figure 2. Like the rules in the top part of the figure, every word in the corpus has an equal chance of being in an implicative phrase. We distinguish between two types of implicative phrases, ones that reverse polarity in the opposite direction (NMONPhrase), and ones that keep the polarity the same (MONPhrase). The rules $Phrase_Z$ and $Phrase_W$ specify that both types can have different effects (e.g. MONPhrase can be (+)(+), (-)(-)), which gets settled once the neighboring polarity is determined. The top rules specify that each event or relation is subject to modification by an implicative phrase, which allows for an arbitrary nesting of implicative phrases.

For example, in the fragment *didn't bother to remember to eat*, *didn't* reverses polarity (NMON-Phrase), whereas *bother* and *remember* preserve polarity (MONPhrase). Equation (1) shows how the polarity of the verb is propagated through a derivation in our grammar. Because the verb gets transformed back into its original phrase when it encounters a MONPhrase with the signature *pp*, it is again subject to modification. This is consistent with how inferences are computed in the polarity propagation algorithm, and stays within the syntactic analysis.

$$
notEat \leftarrow \left[Eat_n \leftarrow (didn't_{pn}\left(Eat_p \leftarrow \left(bother_{pp}\left(Eat_p \leftarrow \left(remember_{pp}\left(Eat_p \right)\right)\right)\right)\right)\right] \tag{1}
$$

For training, we perform cross validation by making four different splits in our 7,010 sentence set (5,010 for training, and 2,000 for testing). As in (Börschinger et al., 2011), we train the

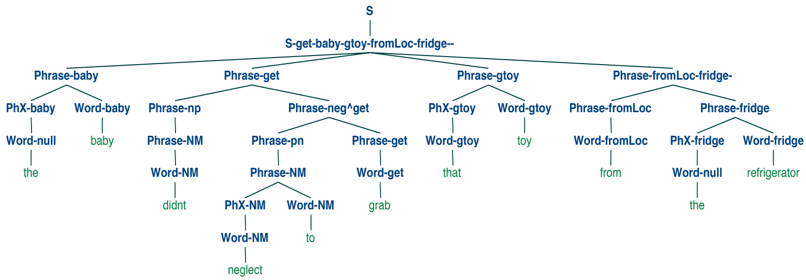

Figure 3: Example output of an analysis after training

Set	Pronoun Precision	Implicative Precision	Overall Precision	Recall	F-Score
1	**0.3859** (203/526)	**0.8277** (471/569)	**0.788** (1576/2000)	1.0	0.8814
2	**0.38878** (208/535)	**0.7489** (373/498)	**0.769** (1538/2000)	1.0	0.8694
3	**0.39405** (199/505)	**0.83116** (448/539)	**0.8005** (1601/2000)	1.0	0.8891
4	**0.333** (177/531)	**0.730** (376/515)	**0.75** (1500/2000)	1.0	0.8571
Av.	**0.3755**	**0.7845**	**0.7768**		**0.874**

Table 3: Results on extended Grounded World data

grammars on each split using the Inside-Out Algorithm (Lari and Young, 1990) [1], a variant of the EM algorithm often used for PCFG induction. The main thrust of the learning algorithm is that sentences are parsed with their contexts, which provides a top-down constraint on the possible analyses. Implicatives that lead to negative inference, for example, will consistently be observed in negative contexts, which forces the learner to construct latent parses that lead to such inferences. Over time, the probability that the associated words receive the correct analysis increases.

Once the grammars are trained on the different splits, information about the original contexts is removed, and the remaining unseen sentences are parsed. Like in (Börschinger et al., 2011), the derived relation (or S-node) for each parse is evaluated against a gold standard relation and marked correct if it matches this relation exactly. An example parse after training is shown in Figure 3, where the resulting relation is *(get baby, gtoy, (fromLoc fridge))*. All words related to the inference, in addition to words corresponding to other semantic concepts, were learned to have the correct analysis (e.g. *didn't*$_{-+}$, *neglect to*$_{+-}$, *grab*$_{get'}$) , which allows us to recursively compute the inference in the manner described above.

2.3 Results and Discussion

The results are provided in Table 3 and are broken down into each training-testing split. Sentences are counted as correct when the main relation in the parse matches exactly a gold-standard annotation. In terms of evaluating inference, getting the right relation means that

[1] we used Mark Johnson's CKY and Inside-Out implementation available at http://web.science.mq.edu.au/ mjohnson/Software.htm.

the correct inference is achieved. As mentioned above, a large portion of the original corpus contains sentences with pronouns, and we isolate sentences with pronouns, as well as with implicative phrases, and measure the overall precision for each set.

The average overall precision is 0.7768, with 0.7845 average precision on implicative phrases and 0.3755 on sentences with pronouns. The latter precision is the lowest, and is to be expected since we simply assign pronouns the most probable referent based on training (no further resolution is done). Recall in all cases is 1.0, since we build the semantic relations from the total corpus following (Börschinger et al., 2011). This avoids having out-of-grammar issues when parsing test sentences, but limits the parsers to only the relations seen in the corpus. This is one downside to a grammar approach, which is discussed and improved upon in (Kim and Mooney, 2012) and will be a main focus of future work.

We emphasize that the evaluation, following (Börschinger et al., 2011; Kim and Mooney, 2012) and others, is done by looking at the resulting semantic relations (S-Node), and ignores the rest of the syntactic analysis. The parser therefore might make wrong decisions while arriving at the correct inference and relations. For example, the analysis in Figure 3 might have *didn't neglect to* as a single implicative phrase marked as *pp* (as opposed to two), which leads to the same inference. Future work will look at evaluating this and employing unsupervised learning methods for ensuring that the domain lexicon is properly inferred.

Despite these issues, the results are encouraging and show that learning light inference can be done using standard Semantic Parsing techniques with loose ambiguous supervision. This result is not altogether surprising, given that the inference patterns we consider are types of syntactic patterns, and are therefore similar to the other patterns we induce. Future work will look at scaling this up to more complex types of inference in an open-domain. One particular direction might be looking at more complex forms of negation, as studied in, for example, (Blanco and Moldovan, 2012). Another direction is using these techniques, which require very little supervision, to help learn inference patterns for unresourced languages and domains.

3 Conclusions

This work complements recent work on Semantic Parsing, specifically within the ambiguous learning paradigm, and shows how to integrate light syntactic inference into the learning using event polarity and context as loose supervision. The main focus has been on learning implicative verb constructions, which have well-understood semantic properties relating to speaker commitment. The strategy we adopted follows that of (Nairn et al., 2006), and keeps inference computation within the syntax. We adapted current PCFG-based grammar induction techniques for Semantic Parsing, and demonstrated the effectiveness of our inference learning method on a modified portion of the Grounded World corpus. Future work will concentrate on extending these results to open-domain textual entailment problems, and on inference learning for unresourced languages and domains.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) on the project SFB 732, "Incremental Specification in Context". We thank Sina Zarriess for useful suggestions and discussions, and Annie Zaenen and Cleo Condoravdi for earlier discussions about the overall idea and method.

References

Angeli, G., Manning, C. D., and Jurafsky, D. (2012). Parsing time: Learning to interpret time expressions. In *Proceedings of NAACL-HLT-12*, pages 446–455, Montreal, Canada.

Blanco, E. and Moldovan, D. (2012). Fine-grained focus for pinpointing positive implicit meaning from negated statements. In *Proceedings of NAACL-HLT-12*, pages 456–465, Montreal, Canada.

Bordes, A., Usunier, N., Collobert, R., and Weston, J. (2010). Towards understanding situated natural language. In *Proceedings of AISTATS-10*, pages 628–635, Sardinia, Italy.

Börschinger, B., Jones, B. K., and Johnson, M. (2011). Reducing grounded learning tasks to grammatical inference. In *Proceedings of EMNLP-11*, pages 1416–1425, Edinburgh, United Kingdom.

Bos, J. and Markert, K. (2005). Recognising textual entailment with logical inference. In *Proceedings of HLT-EMNLP-05*, pages 628–635, Vancouver, Canada.

Chen, D. L. and Mooney, R. J. (2008). Learning to sportscast: A test of grounded language acquisition. In *Proceedings of the ICML-08*, pages 128–135, Helsinki, Finland.

Chen, D. L. and Mooney, R. J. (2011). Learning to interpret natural language navigation instructions from observations. In *Proceedings of the AAAI-11*, pages 859–865, San Francisco, CA.

Cheung, J. C. K. and Penn, G. (2012). Unsupervised detection of downward-entailing operators by maximizing classification certainty. In *Proceedings of EACL-12*, pages 696–705, Avignon, France.

Dagan, I., Glickman, O., and Magnini, B. (2005). The pascal recognizing textual entailment challenge. In *Proceedings of PASCAL Challenges Workshop on Recognizing Textual Entailment*, pages 177–190.

Danescu-Niculescu-Mizil, C., Lee, L., and Ducott, R. (2009). Without a "doubt"? unsupervised discovery of downward-entailing operators. In *Proceedings of HLT-NAACL-09*, pages 137–145, Boulder, Colorado.

Dowty, D. (1994). The role of negative polarity and concord marking in natural language reasoning. In *Proceedings of Semantics and Linguistics Theory (SALT)*, pages 114–144, Ithaca, New York.

Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010). The conll-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of CoNNL-10: Shared Task*, pages 1–12, Uppsala, Sweden.

Johnson, M., Demuth, K., and Frank, M. (2012). Exploiting social information in grounded language learning via grammatical reduction. In *Proceedings of the ACL-12*, pages 883–891, Jeju Island, Korea.

Jones, B. K., Johnson, M., and Goldwater, S. (2012). Semantic parsing with bayesian tree transducers. In *Proceedings of ACL-12*, pages 488–496, Jeju Island, Korea.

Karttunen, L. (1971). Implicative verbs. *Language*, 47(2):340–358.

Karttunen, L. (2012). Simple and phrasal implicatives. In *Proceedings of *SEM-12*, pages 124–131, Montreal, Canada.

Kate, R. and Mooney, R. (2006). Learning language semantics using string kernels. In *Proceedings of Coling-ACL-06*, pages 913–920, Sydney, Australia.

Kate, R. J., Wong, Y. W., and Mooney, R. J. (2005). Learning to transform natural to formal languages. In *Proceedings of AAAI-05*, pages 1062–1068, Pittsburgh, Pennsylvania.

Kim, J. and Mooney, R. J. (2012). Unsupervised pcfg induction for grounded language learning with highly ambiguous supervision. In *Proceedings of EMNLP-CoNLL-12*, pages 883–891, Jeju Island, Korea.

Kiparsky, P. and Kiparsky, C. (1970). Fact. In Bierwisch, M. and Heidolph, K., editors, *Progress in Linguistics*, pages 143–173. Mouton,Hague.

Kwiatkowski, T., Zettlemoyer, L., Goldwater, S., and Steedman, M. (2010). Inducing probabilistic ccg grammars from logical form with higher-order unification. In *Proceedings of EMNLP-10*, pages 1223–1233, Cambridge, Massachusetts.

Lari, K. and Young, S. (1990). The estimation of stochastic context-free grammars using the inside-outside algorithm. *Computer Speech and Language*, 4(1):35–56.

Liang, P., Jordan, M. I., and Klein, D. (2011). Learning dependency-based compositional semantics. In *Proceedings of ACL-11*, pages 590–599, Portland, Oregon.

MacCartney, B. and Manning, C. (2008). Modeling semantic containment and exclusion in natural language inference. In *Proceedings of Coling-08*, pages 521–528, Manchester, United Kingdom.

MacCartney, B. and Manning, C. D. (2007). Natural logic for textual inference. In *Proceedings of ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 193–200, Prague, Czech Republic.

Mooney, R. (2007). Learning for semantic parsing. In *Proceedings of CICLing-07*, pages 311–324, Mexico City, Mexico.

Mooney, R. (2008). Learning to connect language and perception. In *Proceedings of AAAI-08*, pages 1598–1601, Chicago, Illinois.

Moss, L. (2010). Natural logic and semantics. In *Proceedings of 17th Amsterdam Colloquium*, pages 84–93.

Nairn, R., Condoravdi, C., and Karttunen, L. (2006). Computing relative polarity for textual inference. In *Proceedings of ICoS-5 (Inference in Computational Semantics)*, pages 63–76, Buxton, UK.

Thompson, P., Nawaz, R., McNaught, J., and Ananiadou, S. (2011). Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics*, 12:393.

Valencia, V. S. (1991). *Studies on Natural Logic and Categorial Grammar*. PhD thesis, University of Amsterdam.

Wong, Y. W. and Mooney, R. J. (2006). Learning for semantic parsing with statistical machine translation. In *Proceedings of HLT-NAACL-06*, pages 439–446, New York City, NY.