Title
Inferring the Meaning of Verbs from Context

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Authors
Wiemer-Hastings, Peter
Graesser, Arthur C.
Wiemer-Hastings, Katja

Publication Date
1998

Peer reviewed

Inferring the Meaning of Verbs from Context

Peter Wiemer-Hastings (PWMRHSTN@MEMPHIS.EDU)
Arthur C. Graesser (A-GRAESSER@MEMPHIS.EDU)
Katja Wiemer-Hastings (KWIEMER@CC.MEMPHIS.EDU)
The University of Memphis, Psychology, Campus Box 526400, Memphis TN 38152-6400

Abstract

This paper describes a cross-disciplinary extension of previous work on inferring the meanings of unknown verbs from context. In earlier work, a computational model was developed to incrementally infer meanings while processing texts in an information extraction task setting. In order to explore the space of possible predictors that the system could use to infer verb meanings, we performed a statistical analysis of the corpus that had been used to test the computational system. There were various syntactic and semantic features of the verbs that were significantly diagnostic in determining verb meaning. We also evaluated human performance at inferring the verb in the same set of sentences. The overall number of correct predictions for humans was quite similar to that of the computational system, but humans had higher precision scores. The paper concludes with a discussion of the implications of these statistical and experimental findings for future computational work.

Introduction

Verbs play a critical role in human languages. They constrain and interrelate the entities mentioned in sentences. It is therefore important to understand the processes by which we acquire verb meanings. This paper examines verb acquisition from three directions: (1) a computational system which acquires verb meaning from the linguistic context of real-world texts, (2) a statistical analysis of the predictiveness of various features of the context to the verb, and (3) two experiments on adults to determine their ability to infer missing verbs from context.

Granger (1977) conducted some of the earliest compuational work on verb acquisition from context, and Salveter (1979, 1980) followed close thereafter. However, neither of these systems was tested on real-world domains. Zernik's thesis work (1987) concentrated on verb learning, but mainly on verb-particle combinations and not on real-world texts. Likewise, Siskind's work (1994, 1996) examined verb learning with synthetic corpora that included sentences and representations of their meanings.

Cardie's MayTag system (1993) used a case-based approach and statistical methods for determining relevance of cases, but did not learn verb meanings. Riloff (1993) has recently introduced a computational mechanism that simulates a coarse-grained lexical acquisition from context. The Autoslog system operates in an information extraction task by comparing completed template forms with the sentences that the text was taken from. Autoslog selects just the key pieces of text, performs a simple analysis on the surrounding structure, and then proposes template-like definitions which are later filtered by a human, for example:

> If a SUBJECT is followed by a passive form of "kidnap", put the SUBJECT into the VICTIM slot of a KIDNAPPING template.

One of the important trends has been statistical analyses of corpora. Almost all of it, however, has focused on word-sense discrimination or lexical category training (Brill, 1993, for example). Resnik (1993) explored statistical methods of defining relationships between words, but only briefly mentioned its implications for verb acquisition.

Developmental psycholinguists have known for a long time that there are substantial differences between the learning of verbs and nouns (Gentner, 1978). Researchers (Shatz, 1987; Landau & Gleitman, 1985; Gleitman, 1990) have examined how children, as they develop increasing knowledge of the syntactic structure of a language, use that information to constrain their ideas of what unknown verbs can mean. The recent work by Gleitman (Gleitman & Gillette, 1994; Gleitman & Gleitman, 1997) is quite similar to the human tests reported in this study. Her empirical work examines the contribution of the observational context in addition to the lexical contribution (as reported here).

Camille: A Model of Verb Acquisition

Camille (the Contextual Acquisition Mechanism for Incremental Lexeme LEarning) (Hastings, 1994, 1995, 1996), was developed as an extension of Lytinen's natural language understanding system called Link (Lytinen & Roberts, 1989). When the parser comes across a word that it does not know, Camille infers whatever it can about the meaning of the unknown word. Unknown nouns are handled quite naturally by a unification-based parser because the semantic constraints on slot-fillers provided by the verbs give relatively useful limitations on what those nouns could mean. For example,[1] if the system processed the sentence, "Mary hijacked the limousine," and it didn't know what "limousine" meant, it could

[1] All our examples will come from the domain of ARPA's MUC4 evaluation (Sundheim, 1992) which consists of newspaper articles describing terrorist activity.

conclude that it is some type of vehicle because of the constraints on what can be hijacked. But precisely because the constraints are associated with the verbs, learning unknown verbs is much more difficult. Verb acquisition has thus been the focus of the research on Camille.

When a sentence containing an unknown verb is encountered, Camille puts a default "definition" for the verb into the parse. During the parsing process, as the other elements of the sentence are attached as slot-fillers of the verb, Camille compares each filler with the semantic constraints [for example, (OBJECT = VEHICLE) for "hijacking"] on the various action concepts in the domain's concept representation.[2] For any example, there will be a set of action concepts which are logically consistent with the slot fillers. But Camille eliminates from consideration all but the most specific concepts, that is, those that most closely match the actual slot fillers This extreme inductive approach is necessary because of the no-negative-evidence problem: Camille doesn't get examples of how a word is *not* used, just how it *is* used. Camille therefore must make the extreme hypotheses because they are the most falsifiable. It relies on later (positive) examples of the word and on its incremental learning mechanism to correct erroneous hypotheses.

Camille was tested on several real-world domains within information extraction tasks. We used the scoring methods from MUC4 (Chinchor, 1992) which calculate the recall and precision of the system. Camille often guessed multiple concepts per word because the semantic constraints are not sufficient to distinguish them. For the lexical acquisition task, recall is defined as the percentage of correct hypotheses. A hypothesis was counted as correct if one of the concepts in the hypothesis matched the target concept. Precision is defined as the number of correct hypotheses divided by the total number of concepts generated in all the hypotheses. Thus, there tends to be a recall/precision tradeoff: By guessing many concepts per hypothesis, Camille can increase its recall because it has a better chance of "hitting" the right concept. However, as the total number of concepts gets larger, precision is reduced. Camille has achieved a recall of 42% and precision of 19% on a set of 50 randomly-selected sentences containing 17 different verbs. This was respectable performance, but we thought that the system should be able to do better.

Statistical Corpus Analysis

One deficit of Camille's lexical acquisition mechanism is its failure to make much use of the syntactic context of example sentences. It does use syntax, but only indirectly. Specifically, the parser performs a syntactic and semantic parse of the sentence constituents and then passes the semantic role-fillers on to Camille. But Camille has no information about the syntactic structure of the sentence. Some of the verbs in the terrorism corpus, "accused" for example, occur within a very particular syntactic frame, as in, "John accused Mary of bombing the building."[3] Camille should be able to use such information in its learning process.

We looked up the verbs' frames in Wordnet (Miller, 1990) in order to evaluate the respective contributions of syntactic and semantic features of linguistic context. The frames for the verb "deny" are shown below:

```
Somebody denies something
Somebody denies that CLAUSE
Somebody denies somebody something
Somebody denies something to somebody
```

Next we deleted the distinction between "somebody" and "something" in order to distinguish purely syntactic information from semantic information. We eliminated those frames that didn't occur in our corpus of 259 example sentences. Finally, we broke down the templates into separate syntactic features to allow the different parts of the templates to be evaluated independently. The resulting 11 boolean features (including some surface features that were not specified in the templates) are shown below.

- There is a syntactic subject (occurred in every case)
- There is a syntactic object
- There is an indirect object
- Sentence is passive
- There is a clausal object (sentential complement)
- The clausal object precedes the subject
- There is a "to" + INFINITIVE clause
- There is a gerund object
- There is a "with" prepositional phrase modifying the verb
- There is an "of" + VERBing prepositional phrase modifying the verb
- There is a reflexive construction

To evaluate the semantic features of the corpus, we analyzed the sentences based on the semantic categories of the slot fillers. In order to maintain consistency with the Camille test, we used the categories that occur in Link's semantic constraints. These categories, separated by semantic role, are listed in Table 1. Many of the categories subsume other categories, but we left it to the statistical analyses to sort these out. In fact, some interesting results can occur when an ancestor node is a strong diagnostic factor in favor of a particular verb, but one of its descendants is diagnostic against a particular verb. This constitutes both a least upper bound and a greatest lower bound for the possible meaning space, which could not be achieved with the simple Camille algorithm because of the no-negative evidence problem.

The verbs did not occur equally frequently in our corpus of 259 randomly selected sentences. Table 2 shows the number of occurrences of each target verb in the corpus.

[2]These constraints are there in order to limit the complexity of the parsing process, not for the purpose of verb acquisition.

[3]The line between the syntactic features of a verb (e.g. "accuse" occurs in a frame like "Somebody accuses someone of something") and the meaning of the verb becomes a bit hazy here. We recognize that not all of these features are related to meaning, but some are, and one goal of this research project is to find out which features are related to meaning.

Table 1: Semantic categories by role

actor	none, human, human-or-official, human-or-organization, terrorist, physical-object, explosive
object	none, action, human-or-official, death, human-or-organization, responsibility, effect-or-action, human-or-place, bomb, physical-target, vehicle, building, tangible
instrument	none, explosive, gun
location	place
of-object	action
time	time

Table 2: Number of sentences per verb in corpus

accused	8	killed	50
attacked	24	machinegunned	4
claimed	8	murdered	11
denied	8	reported	61
destroyed	11	riddled	1
died	4	stated	14
dynamited	2	threatened	3
exploded	13	wounded	30
kidnapped	8		

The results of a series of multiple regression analyses, one analysis for each specific verb, are shown in table 3. The presence (1) or absence (0) of a particular verb in the sentence was the criterion variable in each regression analysis. The predictor variables were categorial variables for all of the syntactic and semantic features of the corpus sentences. Table 3 shows the beta weights for those features that were significant at the .05 level. All of the verbs were in the simple past tense form, although some of them (for example, mac-gun = "machinegunned") are abbreviated here.

Table 3 shows that diagnostic syntactic and semantic cues can be identified for each verb. For example, a terrorist actor and human semantic object are both diagnostic of the word "kidnapped" For "claimed", the two diagnostic features are a semantic object of RESPONSIBILITY, and a syntactic "to" + INFINITIVE clause.

One interesting case involved the verb "attack" A TANGIBLE object is diagnostic for that particular verb, whereas the descendant HUMAN-OR-OFFICIAL, PHYSICAL-TARGET, and VEHICLE categories are diagnostic against it. This is rather counterintuitive. We might suspect that "attack" would be used with human objects, but in this terrorist activity corpus, that is apparently not so. Thus, this type of statistical analysis, on a domain-specific basis, should help fine tune lexical acquisition, knowledge representation, and parsing mechanisms.

We performed a manual cluster analysis on these data, grouping the verbs with their related beta weights into the syntactic/semantic groups of communication verbs, intransitive verbs, and terrorist actions. Some interesting clusters of diagnosticity emerged. For example, the two intransitive verbs, "died" and "exploded", show a very similar pattern. Not surprisingly, the feature indicating the lack of a semantic object is diagnostic for both of these verbs. A semantic actor of type EXPLOSIVE is strongly indicative of "exploded" and against "died" A semantic actor of PHYSICAL-OBJECT is weakly indicative of "exploded" and against "died". However, none of the syntactic features are diagnostic of these verbs.

The syntactic features have little diagnosticity for the terrorist-act verbs as well. They play a much more pronounced role for the communication verbs, especially for "accused", for which an "of VERBing" complement is highly diagnostic. The semantic actor features have a somewhat smaller but still significant diagnosticity for the communication verbs, and less for the terrorist acts. A HUMAN-OR-OFFICIAL semantic object is only diagnostic for the terrorist-act verbs.[4] There is also a cluster of diagnosticity among the semantic object features for the subset of terrorist actions with physical targets. These statistial analyses suggest that there is both syntactic and semantic information available to help language learners discriminate meanings. We explored human ability to make use of this information in the studies described below.

Testing human performance

In order to test how well humans infer verb meanings, we performed an experiment using the same sentences that were analyzed statistically. We used the Cloze procedure in which the target word is replaced by a blank, as shown below:

> A mercenary group _________ an interior ministry worker.

We are interpreting the participants' ability to identify the correct verb in this situation as an indicator of the extent to which contextual information constrains the possible verbs, and we use that to estimate how this information facilitates verb acquisition.[5]

The participants (N=14) were told that the sentences were taken from a set of newspaper articles that describe terrorist activity in order to compensate for the domain-specific nature of Camille. Camille's concept representation space contains primarily those concepts that are required in that domain

[4] As previously stated, a semantic object of type HUMAN-OR-OFFICIAL is diagnostic against "attacked".

[5] Of course, the participants also know a great deal about the verbs that could be put in the blanks, including both their meaning and their subcategorization constraints. The participants might simply match the syntactic templates of a set of verbs with the syntactic structure of the sentence. Because many of the verbs in the corpus have quite similar syntactic frames, however, we suspect that this would not provide sufficient discriminating evidence to account for the performance levels that the participants achieve. This issue will be addressed by the planned experiment described in the final section of this paper.

Table 3: Significant beta weights in multiple regression ($p < .05$)

accused	1.00	syn "of" VERBing clause	exploded	.72	sem actor is EXPLOSIVE
attacked	.11	sem actor is TERRORIST		.27	no sem object
	.23	instrument is GUN		.11	sem actor is PHYSICAL-OBJECT
	-.14	no location		.10	sem object is BUILDING
	-1.01	sem object is HUMAN-OR-OFFICIAL		-.08	no time
	1.09	sem object is TANGIBLE	kidnap	.23	sem object is HUMAN-OR-OFFICIAL
	-.31	sem object is PHYSICAL-TARGET		.18	sem actor is TERRORIST
	-.16	sem object is VEHICLE	killed	.58	sem object is HUMAN-OR-OFFICIAL
	-.12	syn form is passive		-.15	"of" ACTION clause
	-.10	syn "of" VERBing clause		.12	syn reflexive form
claimed	.68	sem object is RESPONSIBILITY	mac-gun	.37	sem object is VEHICLE
	.39	syn "to" INFINITIVE clause	murdered	.25	sem object is HUMAN-OR-OFFICIAL
denied	.40	syn indirect object		-.12	no time
	-.61	sem object is TANGIBLE	reported	.65	sem object is EFFECT-OR-ACTION
	.52	sem object is ACTION		.24	sem object is ACTION
	-.35	no sem object		.12	syn indirect object
destroy	.55	sem object is PHYSICAL-TARGET		-.09	sem actor is PHYSICAL-OBJECT
	.18	sem object is VEHICLE		-.09	syn "to" INFINITIVE clause
	-.12	sem actor is TERRORIST	riddled	.45	sem instrument is GUN
	.11	no location		-.15	sem object is PHYSICAL-TARGET
died	.76	no sem object	stated	.32	syn object is CLAUSE
	-.51	sem actor is EXPLOSIVE		.20	sem actor is PHYSICAL-OBJECT
	.11	sem actor is HUMAN-OR-OFFICIAL	threaten	.58	syn "with" clause
dynamite	.52	sem object is PHYSICAL-TARGET		.22	no sem object
	-.29	sem object is VEHICLE		.15	sem actor is EXPLOSIVE
	-.24	sem object is BUILDING		.10	no sem actor
			wounded	.44	sem object is HUMAN-OR-OFFICIAL
				-.12	syn "of" VERBing clause

to expedite parsing. Thus, the system tends to infer primarily meanings for the verbs that are related to that domain. The participants were also told that they could fill in multiple guesses if they could not arrive at a single best answer. In practice, however, they rarely entered more than one word per blank.

The subjects were each given 17 sentences, one for each verb in the corpus, but they were not told that there was one sentence per verb. The sentences were randomly chosen from the corpus by verb and by syntactic frame, and they were presented to the subjects in random order.

We scored the answers using the same basic method used for evaluating Camille, using the recall and precision measures. Because the subjects were guessing *words* and not *concepts*, all words which would normally be defined as mapping to the target concepts were accepted.

The average recall score was quite close to Camille's: 42% (standard deviation 11%) which compares favorably to 42% for Camille. However, the human subjects had a strong tendency against entering multiple words per sentence, with only three additional guesses for the entire set of 14*17 = 238 sentences. As a consequence, their precision score was virtually identical to their recall at 41%, as opposed to Camille's 19% precision score. By generating almost 2 concepts per hypothesis, Camille increases its chance of identifying the correct concept, but ends up with less usable hypotheses.

In a followup experiment, we attempted to determine the effect of the amount of context on verb identification. The participants (N=27) were again given a set of 17 sentences from the same corpus. This time, there were two different sentence conditions. In one condition, we used the sentences as they appeared in the original corpus. These sentences were long and complex, with an average length of 28.6 words. In the other condition, the sentences were pared down until only one proposition remained. These sentences had an average length of 8.8 words. Examples of the original and shortened versions[6] of a sentence from the corpus, with the verbs removed, are shown below:

Original: "National Army spokesmen said that one noncommissioned officer and six soldiers were ________, four were wounded, and 15 others were injured during the attacks against military bases and police installations in Segovia by the 4th Front of the FARC."

Shortened: "Six soldiers were ________ in Segovia by the 4th Front of the FARC."

[6]In some cases the original sentence expressed a single proposition. In this case the original and shortened versions were the same.

The participants were given 8 sentences from one condition, and 9 from the other.

We calculated overall recall and precision for the participants' answers. This time, both scores were slightly lower but still comparable to those in the first experiment: recall = 37% and precision = 36%. We also calculated the scores separately for the original and shortened sentences, to examine whether the full context facilitates the identification of the missing verb. In fact, however, the scores were slightly lower for the longer versions. 35% (longer), compared with 39% (shorter) for recall, and 34% and 37% respectively for precision. Thus, the larger amount of context did not facilitate verb identification.

Discussion and future work

The goal of this research was to find out how a (computational or human) agent can learn the meanings of verbs from context. A symbolic incremental learning system can perform fairly well on this task by comparing input sentences to the selectional constraints on action concepts in its domain knowledge. In fact, we find that humans in a similar setting perform at relatively similar recall levels. However, the human results were significantly different in their precision level. Humans almost without exception produced a single verb for a sentence, whereas Camille produced an average of just over two. Camille could not distinguish these concepts based purely on semantic constraints.

We are left with the question of what type of information humans use in order to distinguish between candidate verbs. One possibility is the syntactic content of the sentence, which Camille did not use, but which a statistical corpus analysis showed could contribute significantly to verb inference. This possibility would argue against the claims of some researchers in corpus-based semantic representations like LSA (Landauer & Dumais, 1997; Landauer, Laham, Rehder, & Schreiner, 1997) who suggest that word ordering plays an insignificant role in determining the meaning of a text. It would also raise questions about Siskind's recent work (1996), in which he showed that it was theoretically possible for a learning system to acquire word meanings in particular situations without any knowledge of syntax. This may be true, but we also know that human language learners, even very young ones, do know some elements of syntax (Hirsh-Pasek & Golinkoff, 1993; Naigles, 1990). It remains to be seen exactly how these syntactic features are brought into play.

Another factor that could help humans distinguish between verb choices is word frequency. As previously stated, there is a wide range of variation in the frequencies just within the terrorism domain. Presumably this variation would be even more pronounced within the experience of the human participants. If humans do take familiarity of a word into account, they may be more likely to infer that the verb is one with high familiarity, and might not even think of some of the more obscure words ("machine-gunned" for example). Furthermore, the sentences in this corpus were all written by journalists, and none of the participants (as far as we know) is a journalist. Perhaps some of the target words are unlikely ever to be uttered by a non-journalist.

One implication of this research for natural language processing is its focus on domain-specific inference methods. Camille requires a fairly complete domain concept representation, but beyond that, its inference methods can be considered "weak methods", i.e. they don't rely on any domain-specific heuristics — just a general search procedure. The statistical analyses performed here are also clearly domain specific. We wouldn't want to extend our inferences about the occurrences of "killed" and "denied", for example, to another domain. But this very domain specificity can be an asset. General purpose parsers have been an unachievable grail for NLP. Perhaps by performing fairly simple statistical analyses of corpora in specific domains, we can augment parsers to make them more effective in specific areas, as in the recent MUC evaluations. Furthermore, by narrowing the processing focus to a sublanguage, perhaps we can also facilitate the acquisition of unknown words.

In future work, we plan to examine a larger corpus to ensure the reliability of our statistical analyses. We also want to test how well humans perform with multiple examples of a word because Camille relies on getting multiple examples in order to fix erroneous initial inferences. We have recently begun a study to try to tease apart the different aspects of the semantic and syntactic context. The materials are similar to those described in the human testing section here. But we will substitute, for the arguments of the verb, phrases which have less information, for example, "the 4th Front of the FARC" will be replaced with "a terrorist organization," "a group," or "an entity". This should give us a better idea of the contribution of different types of contextual information.

Finally, we will perform a qualitative analysis of the human responses versus Camille's guesses to see if any patterns can be found. In the end, we hope to be able to implement a computational lexical acquisition mechanism which incorporates what we've learned about the aspects of linguistic context which are important for verb acquisition.

Acknowledgements

This work was supported in part by grant number SBR9720314 from the National Science Foundation's Learning and Intelligent Systems Unit.

References

Brill, E. (1993). *A corpus-based approach to language learning*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Cardie, C. (1993). A case-based approach to knowledge acquisition for domain-specific sentence analysis. In *Proceedings of the 11^{th} National Conference on Artificial Intelligence*, pp. 798–803.

Chinchor, N. (1992). MUC-4 Evaluation Metrics. In *Proceedings of the Fourth Message Understanding Conference* San Mateo, CA. Morgan Kaufmann Publishers, Inc.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49*, 988–998.

Gleitman, L. (1990). The structural sources of verb meaning. *Language Acquisition, 1*(1), 3–55.

Gleitman, L., & Gillette, J. (1994). The role of syntax in verb learning. In Fletcher, P., & MacWhinney, B. (Eds.), *The Handbook of Child Language*. Blackwell, Oxford UK.

Gleitman, L., & Gleitman, H. (1997). What is a language made out of?. *Lingua, 99*.

Granger, R. (1977). FOUL-UP: A program that figures out meanings of words from context. In *Proceedings of Fifth International Joint Conference on Artificial Intelligence*.

Hastings, P. (1994). *Automatic Acquisition of Word Meaning from Context*. Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Hastings, P. (1995). Use of context in an automatic lexical acquisition mechanism. In Iwanska, L. (Ed.), *Proceedings of the Workshop on Context in Natural Language of the 14th International Joint Conference on Artificial Intelligence*.

Hastings, P. (1996). Implications of an automatic lexical acquisition mechanism. In Wermter, S., Riloff, E., & Scheler, C. (Eds.), *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*. Springer-Verlag, Berlin.

Hirsh-Pasek, K., & Golinkoff, R. (1993). Skeletal supports for grammatical learning: What infants bring to the language learning task. *Advances in Infancy Research, 8*, 299–338.

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press, Cambridge, MA.

Landauer, T., Laham, D., Rehder, B., & Schreiner, M. (1997). How well can passage meaning be derived without using word order?. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, pp. 412–417 Mahwah, NJ. Lawrence Erlbaum Publishers.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Lytinen, S., & Roberts, S. (1989). Unifying Linguistic Knowledge. Unpublished technical report, AI Laboratory, University of Michigan, Ann Arbor, MI 48109.

Miller, G. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography, 3*(4).

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*, 357–374.

Resnik, P. (1993). *Selection and Information: A class-based approach to lexical relationships*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Riloff, E. (1993). Automatically constructing a dictionary for information extraction tasks. In *Proceedings of the 11^{th} National Conference on Artificial Intelligence*, pp. 811–816.

Salveter, S. (1979). Inferring conceptual graphs. *Cognitive Science, 3*, 141–166.

Salveter, S. (1980). Inferring conceptual graphs. In *Proceedings of the 18^{th} Annual Meeting of the Association for Computational Linguistics*, pp. 13–15.

Shatz, M. (1987). Bootstrapping operations in child language. *Children's Language, 6*, 1–22.

Siskind, J. (1994). Lexical Acquisition in the presence of noise and homonymy. In *Proceedings of the 12^{th} National Conference on Artificial Intelligence*, pp. 760–766 Cambridge, MA. MIT Press.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39–91.

Sundheim, B. (1992). Overview of the Fourth Message Understanding Evaluation and Conference. In *Proceedings of the Fourth Message Understanding Conference* San Mateo, CA. Morgan Kaufmann Publishers.

Zernik, U. (1987). Strategies in Language Acquisitions: Learning Phrases from Examples in Context. Tech. rep. UCLA-AI-87-1, UCLA.