# Toward Automatic Processing of English Metalanguage

**Shomir Wilson***

School of Informatics
University of Edinburgh
10 Crichton Street
Edinburgh EH8 9AB, UK

School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA

shomir@cs.cmu.edu

## Abstract

The metalinguistic facilities of natural language are crucial to our ability to communicate, but the patterns behind the appearance of metalanguage—and thus the clues for how we may instruct computers to detect it—have remained relatively unknown. This paper describes the first results on the feasibility of automatically identifying metalanguage in English text. A core metalinguistic vocabulary has been identified, supporting intuitions about the phenomenon and aiding in its detection and delineation. These results open the door to applications that can extract the direct, salient information that metalanguage encodes.

## 1 Introduction

In linguistic communication it is sometimes necessary to refer to features of language, such as orthography, vocabulary, structure, pragmatics, or meaning. *Metalanguage* enables a speaker to select a linguistically-relevant referent over (or in addition to) other typical referents (Audi, 1995). Metalanguage is illustrated in sentences such as

(1) ***Graupel*** refers to a kind of precipitation.
(2) The name is actually ***Rolla***.
(3) ***Keep tabs on*** is a colloquial phrase.
(4) He wrote "**All gone**" and nothing more.

The roles of the bold substrings in the above sentences contrast with those in (5)-(8) below:

(5) **Graupel** fell on the weary hikers.
(6) **Rolla** is a small town.
(7) **Keep tabs on** him, will you?
(8) They were **all gone** when I returned.

Conventional stylistic cues, such as italics in (1), (2), and (3) and quotation marks in (4), sometimes help the audience to recognize metalinguistic statements in written language. In spoken language or in written contexts where stylistic cues are not used, the audience is expected to identify metalinguistic statements using paralinguistic cues (such as intonation, when speaking) or context and meaning.

Metalanguage is both pervasive and, paradoxically, the subject of limited attention in research on language technologies. The ability to produce and understand metalanguage is a core linguistic competence that allows humans to converse flexibly, unrestricted by domain (Anderson et al., 2002). Humans use it to establish grounding, verify audience understanding, and maintain communication channels (Anderson et al., 2004). Metalanguage encodes direct and salient information about language, but many typical examples thwart parsers with novel word usage or arrangement (Wilson, 2011a). Metalanguage is difficult to classify through the interpretive lens of word senses, given that conventional word senses have little relevance when a word appears chiefly "as a word". The roles of metalanguage in L2 language acquisition (Hu, 2010), expression of sentiment toward others' utterances (Jaworski et al., 2004), and irony (Sperber and Wilson, 1981) have also been noted.

This paper describes the results of the first effort to automatically identify instances of metalanguage in English text. *Mentioned language*, a common variety of metalanguage, is focused upon for its explicit, direct nature, which makes its structure and meaning easily accessible once an instance is identified. Section 2 reviews a prior project by Wilson (2012) to create a corpus of instances of metalanguage, a necessary resource for the present effort. Section 3 describes an approach to distinguishing sentences that contain

* This research was performed during a prior affiliation with the University of Maryland at College Park.

metalanguage from those that do not, a task referred to as *detection* for brevity. Results show that the performance of this approach roughly matches an implied performance ceiling of inter-annotator agreement. Section 4 describes an approach to *delineate* sequences of words that are directly mentioned by a metalinguistic statement; although the results are preliminary, its accuracy shows promise for future development. Together, these results on detection and delineation show the feasibility of enabling language technologies to extract the salient information about language that metalanguage contains.

## 2 Background

The reader is likely to be familiar with the concept of metalanguage, but a discussion is appropriate to ground the concept and connect to previous work. Section 2.1 summarizes a prior study (Wilson, 2012) to collect instances of metalanguage, and 2.2 reviews some related efforts.

### 2.1 Prior Work

A diverse variety of phenomena in natural language satisfy the intuitive criteria that we associate with metalanguage. The prior study focused on identifying sentences that contained *mentioned language*, a phenomenon defined below:

*Definition: For T a token or a set of tokens in a sentence, if T is produced to draw attention to a property of the token T or the type of T, then T is an instance of mentioned language.*[1]

Here, a *token* is an instantiation of a linguistic entity (e.g., a letter, symbol, sound, word, phrase, or other related entity), and a *property* is an ostension of language (García-Carpintero, 2004; Saka, 2006), such as spelling, pronunciation, meaning (for a variety of interpretations of *meaning*), structure, connotation, or quotative source. Generally attention is drawn to the *type* of T (for example, in Sentences (1)-(4)), but it can be drawn to the *token* of T for self-reference, as in Sentence (9):

(9) "The" appears between quote marks.

Although constructions like (9) are unusual and carry less practical value, the definition accommodates them for completeness.

Mentioned language is a common form of metalanguage, used to perform the full variety of language tasks discussed in the introduction. However, other metalinguistic constructions draw attention to tokens *outside of* the referring sentence. Some examples of this are (10)-(12) below. Supporting contexts are not shown for these sentences, though such contexts are easily imagined:

(10) Disregard the last thing I said.
(11) That spelling, is it correct?
(12) People don't use those words lightly.

In each of the above three sentences, a linguistic entity (an utterance, a sequence of letters, and a sequence of words, respectively) is referred to, but the referent is contained in a separate sentence. The referent may have been produced by a different utterer or appeared in a different medium (e.g., speaking aloud while referring to written text). These "extra-sentential" forms of metalanguage have clear value to understanding discourse and coreference. The focus on mentioned language is a limitation to the present work, to utilize an existing corpus and to apply tractable boundaries to the identification tasks.

The mentioned language corpus of the prior study[2] was constructed by filtering a large volume of sentences with a heuristic, followed by annotation by a human reader. A randomly selected subset of articles from English Wikipedia was chosen as a source for text because of its representation of a large sample of English writers (Adler et al., 2008), the rich frequency of mentioned language in its text, and the frequent use of stylistic cues in its text that delimit mentioned language (i.e., bold text, italic text, and quotation marks). Sentences were sought that contained at least one of these stylistic cues and a *mention-significant* word in close proximity. Mention-significant words were a set of 8,735 words and collocations with potential metalinguistic significance (e.g., *word, symbol, call*), extracted from the WordNet lexical ontology (Fellbaum, 1998). Phrases highlighted by the stylistic cues were considered *candidate instances*, and these were labeled by a human reader, who determined that 629 sentences were *mention sentences* (i.e., containing instances of mentioned language) and the remaining 1,764 were not. Mention sentences were categorized based on functional properties that emerged during categorization. Table 1 shows some examples of collected mention sentences in each category.

[1] This definition was introduced by Wilson (2011a) along with a practical rubric for evaluating candidate sentences. For brevity, its full justification is not reproduced here.

[2] The corpus is available at http://www.cs.cmu.edu/~shomir/um_corpus.html.

| Category | Examples |
|---|---|
| Words as Words | The IP Multimedia Subsystem architecture uses the term **transport plane** to describe a function roughly equivalent to the routing control plane. |
| | The material was a heavy canvas known as **duck**, and the brothers began making work pants and shirts out of the strong material. |
| Names as Names | **Digeri** is the name of a Thracian tribe mentioned by Pliny the Elder, in The Natural History. |
| | Hazrat Syed Jalaluddin Bukhari's descendants are also called **Naqvi al-Bukhari**. |
| Spelling or Pronunciation | The French changed the spelling to **bataillon**, whereupon it directly entered into German. |
| | Welles insisted on pronouncing the word apostles with a hard **t**. |
| Other Mentioned Language | He kneels over Fil, and seeing that his eyes are open whispers: **brother**. |
| | During Christmas 1941, she typed **The end** on the last page of Laura. |

Table 1: Examples of mentioned language from the corpus. Instances of the phenomenon appear in bold, with the original stylistic cues removed.

To verify the reliability of the corpus and the definition of mentioned language, three additional expert annotators independently labeled a shuffled set of 100 sentences, consisting of 54 randomly selected mention sentences and 46 randomly selected non-mention sentences. All three agreed with the primary annotator on 46 mention sentences and 30 non-mention sentences, with an average pairwise Kappa of 0.74. Kappa between the primary annotator and a hypothetical "majority voter" of the additional annotators was 0.90. These results were seen as a moderate indication of reliability and a potential performance ceiling for automatic identification.

### 2.2 Related Work

The present effort is believed to be the first to automatically identify a natural variety of metalanguage in English text. Aside from the corpus described above, the only other significant corpus of metalanguage was created by Anderson et al. (2004), who collected metalinguistic utterances in conversational English. A lack of phrase-level annotations in their corpus as well as substantial noise made it suboptimal for the present effort. However, it is possible (if not likely) that indicators of metalanguage differ between written and spoken English, lending importance to the Anderson corpus as a resource.

Metalanguage has a long history of theoretical treatments, which chiefly explained the mechanics of selected examples of the phenomenon. Many addressed it through the related topic of *quotation* (Cappelen and Lepore, 1997; Davidson, 1979; Maier, 2007; Quine, 1940; Tarski, 1933), and others previously cited in this paper discussed it directly as *metalanguage* or the *use-mention distinction*. The definition of mentioned language in Section 2.1 was a synthesis of the most empirically-compatible theoretical treatments, and the present effort to automatically identify metalanguage builds on that synthesis.

## 3 Detection of Mentioned Language

The corpus-building effort used a heuristic to accelerate the collection of mentioned language, but its low precision is impractical for automatic identification. Moreover, the stylistic cues that the heuristic relied upon are often inconsistently applied (or entirely absent in informal contexts), and they are sometimes unavailable for the writer to use or for the audience to extract. This section presents an approach to the *detection* task, to discriminate between mention and non-mention sentences. Early examination of the corpus suggested that mention sentences tend not to have distinct structural differences from non-mention sentences, so a lexical approach was first taken, although combinations of lexical and structural approaches are later explored indirectly through the delineation task. In this section a sentence is assumed to be a sequence of words without stylistic cues for mentioned language.

### 3.1 Approach

To establish performance baselines, a matrix of feature sets and classifiers was run on the corpus with ten-fold cross validation. The feature sets were bags of the following: stemmed words (SW), unstemmed words (UW), stemmed words plus stemmed bigrams (SWSB), and unstemmed words plus unstemmed bigrams (UWUB). Classifiers were chosen to reflect a variety of approaches to supervised learning; as implemented in Weka (Hall et al., 2009), these were Naive Bayes (John and Langley, 1995), SMO (Keerthi

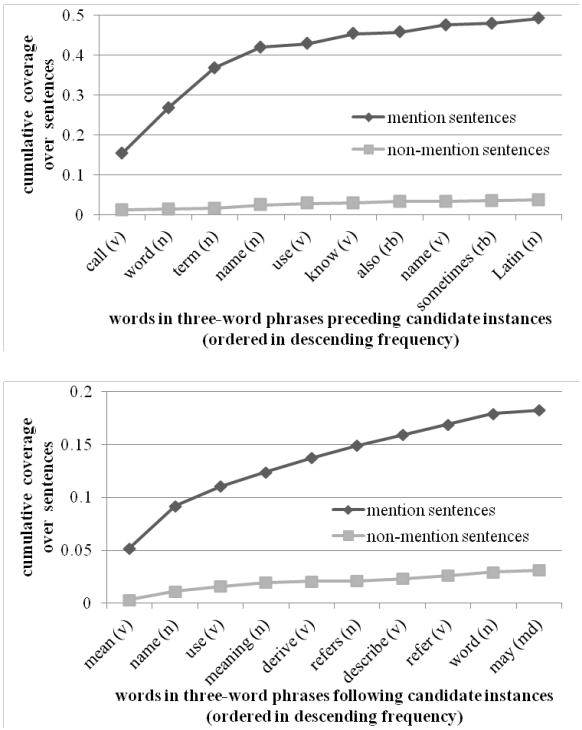


Figure 1. Cumulative coverage over sentences by the most common words before (top) and after (bottom) instances of mentioned language.

et al., 2001), J48 (Quinlan, 1993), IBk (Aha and Kibler, 1991), and Decision Table (Kohavi, 1995).

Prior observations suggested that a small set of approximately ten words significant to metalanguage ("metawords", informally) appear near most instances of mentioned language (Wilson, 2011b). The metalanguage corpus described in Section 2 provided an opportunity to explore this observation. The sentences in the corpus were part-of-speech tagged and stemmed (using NLTK (Bird, 2006)). Sets were collected of all unique (stemmed) words in the three-word phrases directly preceding and following candidate instances, and (respective of position) these were ranked by frequency. The appearance or non-appearance of these words was then determined over all mention and non-mention sentences. Figure 1 shows the cumulative coverage (i.e., appearance at least once) over sentences for the top ten words appearing before and after candidate instances. For example, *call*, *word*, or *term* were the three most common words before candidate instances; they appear at least once in 36% of mention sentences, but they appear in only 1.6% of non-mention sentences.

The high frequencies of intuitive metawords, combined with their difference in coverage over mention and non-mention sentences, informed the approach taken to the detection task. To attempt to improve over the baseline, the SW feature set was ranked by information gain, and all features except the top ten were discarded to create the metawords feature set (MW). Feature selection was done using the training set for each cross-validation fold, and the testing data for each fold was pruned correspondingly. The five selected classifiers were then applied to the data.

### 3.2 Results and Discussion

The combination of five feature sets and five classifiers produced 25 sets of annotations, for which precision, recall and F1 were calculated for detecting mentioned language. For brevity, we present the highlights and contrast the metawords approach with baseline performances.

Table 2 compares classifier performances using MW with SW, its closest relation. MW produced improvement for all classifiers except Naive Bayes[3]. The J48-MW combination had the highest F1 and recall of any feature set-classifier combination, though some combinations exceeded its precision. For all feature set-classifier pairs, precision was higher than recall, by as little as 0.024 (IBK-MW) and as much as 0.22 (Decision Table-UW). For the baseline feature sets, the best classifier was consistently SMO, with F1 scores of 0.70, 0.70, 0.73, and 0.71 for SW, UW, SWSB, and UWUB, respectively. J48 was consistently the second best, with F1 scores within 0.01 of SMO for each feature set.

Table 3 lists differences between F1 scores using the MW feature set and each baseline feature set. MW resulted in improvements over the baseline feature sets for nearly all classifiers, and statistically significant improvements (using one-tailed T-tests across the populations of validation folds, $p<0.05$) were observed for eleven of the sixteen combinations. IBk appeared to benefit the most, with significant improvements over all baseline feature sets, and Naive Bayes the least. In general, recall benefited more than precision.

Examining the MW features confirmed that most were intuitive metawords. Nine words appeared in all ten folds of MW: *name, word, call, term, mean, refer, use, derive,* and *Latin*. The last two words are perhaps artifacts of the encyclopedic nature of the source text, but the rest generalize easily. Future research using additional

[3] It seems likely that the method used to create the MW feature set aggravated the Naive Bayes assumption of feature independence.

| Classifier | Precision | Recall | F1 |
|---|---|---|---|
| Naive Bayes | .76 / **.75** | .63 / **.60** | .69 / **.66** |
| SMO | .74 / **.75** | .67 / **.70** | .70 / **.73** |
| IBk | .69 / **.74** | .64 / **.72** | .66 / **.73** |
| Decision Table | .76 / **.74** | .61 / **.68** | .67 / **.71** |
| J48 | .72 / **.75** | .69 / **.73** | .70 / **.74** |

Table 2. The performances of classifiers using the SW and MW (in bold) feature sets.

| Classifier | SW | UW | SWSB | UWUB |
|---|---|---|---|---|
| Naive Bayes | -.024 | -.018 | .005 | .007 |
| SMO | **.023*** | **.026*** | .000 | .015 |
| IBk | **.067*** | **.088*** | **.07*** | **.108*** |
| Decision Table | **.038*** | **.047*** | .027 | **.052*** |
| J48 | **.037*** | **.046*** | .025 | .034 |

Table 3. Differences between F1 scores from using the MW feature set and baseline feature sets. Statistically significant improvements are starred.

text sources will be necessary to fully verify whether the MW approach and these specific metalinguistic terms are widely applicable.

It also appears that 20% to 30% of instances of mentioned language resist identification using word and bigram-based features alone. Many of the false negatives from this experiment appeared to lack the common metawords that the detection approach relied upon. The sentences below (taken from the corpus) illustrate this lack:

(13) Other common insulting modifiers include “dog”, “filthy”, etc.
(14) To note, in the original version the lyrics read “Jim crack corn”.

While *modifier* in (13) and *read* in (14) have intuitive metalinguistic value, they also have common non-metalinguistic senses. This suggests that an approach incorporating word senses may further improve upon the MW performances, and such an approach is preliminarily explored in the following section.

Finally, it is notable that the best MW performances approach the Kappa score observed between the additional annotators. Although this is an indication of some success, the higher “majority vote” Kappa score of 0.90 remains a meaningful goal for future research efforts.

## 4 Toward Delineation

After identifying a mention sentence, the task remains to determine the specific sequence of words subject to direct reference (e.g., the bold words in Sentences (1) through (4) and in other examples throughout this paper). This task, in addition to detection, is necessary to ascribe the information encoded in a metalinguistic statement to a specific linguistic entity.

### 4.1 Approach

Manual examination of the corpus showed two frequent relationship patterns between metawords and mentioned language. The first was *noun apposition*, in constructions like (15) and (16), where the metaword-noun appears in italics and the mentioned word in bold:

(15) The *term* **auntie** was used depreciatively.
(16) It comes from the root *word* **conficere**.

The second pattern was the appearance of mentioned language in the *semantic role of a metaword-verb*, as in (17) below:

(17) We sometimes *call* it **the alpha profile**.

Notably these patterns do not guarantee the correct delineation of mentioned language, but their applicability made them suitable for the task.

To assess the applicability of phrase structures and semantic roles to the automatic delineation of mentioned language, case studies were performed on the sets of sentences in the corpus containing the nouns *term* and *word* and the verb *call*. *All* sentences containing these three metawords (appearing as their respective targeted parts of speech) were examined, including those that did not contain mentioned language, since it was believed that methods of delineation could indirectly perform detection as well. Because of the limited data available, formal experiments were not possible, although the results still have illustrative value.

The noun apposition pattern described above was formalized for *term* and *word* using TRegex search strings (Levy and Andrew, 2006). The 91 sentences in the corpus containing *term* and the 107 containing *word* were parsed using the Stanford Parser (Marneffe et al., 2006), and the TRegex strings were applied to each sentence; when a match occurred, the result was a prediction that a specific sequence of words was mentioned language (delineation), as well as a prediction that the sentence contained mentioned language (detection). The semantic role pattern for *call* was explored similarly using the Illinois Semantic Role Labeler (SRL) (Punyakanok et al., 2008). Each of the 158 sentences in the corpus containing *call* as a verb was processed by

SRL, and when the output contained the appropriate semantic role (i.e., SRL's "attribute of arg1") with respect to the metaword, the phrase fulfilling that role was considered a predicted delineation of mentioned language. By proxy, such matching also implied a prediction that the phenomenon was present in the sentence.

### 4.2 Results and Discussion

Delineation was evaluated with respect to the correctness of *label scope*: that is, for a sentence that contained an instance of mentioned language, whether the predicted word sequence exactly matched the sequence labeled in the corpus, overlabeled it (i.e., included the instance of mentioned language plus additional words), or underlabeled it (i.e., did not include the entire instance). To avoid confounding detection and delineation, the statistics on label scope do not include instances when the appropriate pattern failed to annotate *any* phrase in a sentence that contained mentioned language, or annotated a phrase when no mentioned language was present. Such instances are instead represented through *pattern applicability* statistics: when one of the sought relationships between a chosen metaword and a phrase appeared in a sentence, it was considered a positive prediction of the presence of mentioned language. Table 4 shows performance metrics from each of the three case studies.

Noun apposition with either *term* or *word* appeared to be adept at predicting scope, with perfect labels for 97% and 89% of instances, respectively. The instances of overlabeling and underlabeling for these two were mostly due to parsing errors, which occurred prior to applying the TRegex pattern. Overlabeling was a greater problem for *call*, for which 80% of labels were perfect and nearly the rest were overlabeled. Manual examination revealed that the prediction often would "spill" far past the actual end of mentioned language, due to the boundaries of the semantic role in SRL's output. For example, the entire phrase in bold in (18) below was erroneously predicted to be mentioned language, instead of simply *snow-eaters*:

(18) Winds of this type are called ***snow-eaters*** *for their ability to make snow melt or sublimate rapidly*.

Re-examining the detection task through pattern applicability, noun appositions with *term* and *word* exhibited perfect precision. The false negatives that lowered the recall were again mostly due to parse errors. Precision and recall for *call* suffered from two sources of errors: incorrect applications of the semantic role and applications of it that, while valid, did not involve mentioned language.

| Metaword | **Label Scope** | | |
|---|---|---|---|
| | Overlabeled | Underlabeled | Exact |
| *term* (n) | 0 | 2 | 57 |
| *word* (n) | 3 | 4 | 57 |
| *call* (v) | 16 | 1 | 68 |

| Metaword | **Pattern Applicability** | | |
|---|---|---|---|
| | Precision | Recall | F1 |
| *term* (n) | 1.0 | 0.89 | 0.90 |
| *word* (n) | 1.0 | 0.94 | 0.97 |
| *call* (v) | 0.87 | 0.76 | 0.81 |

Table 4: Performance statistics for delineation (in the form of label scope) and detection (pattern applicability) for the case studies.

For the selected metawords, it appeared patterns in noun apposition and semantic roles were moderately effective at delineating as well as detecting mentioned language. However, the accuracy of these patterns was a reflection of the dependability of the underlying language tools, and the case studies in aggregate covered only 33% of the sentences containing mentioned language in the corpus. To create a comprehensive method for delineation, more relationships must be identified between metawords and mentioned language. A perusal of the corpus suggests that these patterns are small in variety but large in quantity: metawords are diverse, and some have non-metalinguistic senses that must be accounted for, as shown by Sentences (13) and (14) and others that resisted detection.

## 5 Conclusion

The detection and delineation methods presented in this paper demonstrate the feasibility of identifying metalanguage in English text. The next goals of this project will be to assimilate metalanguage from additional text sources and integrate the detection and delineation tasks. This will improve performance and provide a richer structural knowledge of metalanguage, which will enable practical systems to incorporate processing of the phenomenon and exploit the linguistic information that it encodes.

## References


Adler, B. T., de Alfaro, L., Pye, I., & Raman, V. (2008). Measuring author contributions to the Wikipedia. In *Proceedings of the 4th International Symposium on Wikis* (pp. 15:1–15:10). New York, NY, USA: ACM. doi:10.1145/1822258.1822279

Aha, D. W., & Kibler, D. (1991). Instance-based learning algorithms. In *Machine Learning* (pp. 37–66).

Anderson, M. L., Fister, A., Lee, B., & Wang, D. (2004). On the frequency and types of meta-language in conversation: A preliminary report. In *14th Annual Conference of the Society for Text and Discourse*.

Anderson, M. L., Okamoto, Y. A., Josyula, D., & Perlis, D. (2002). The Use-Mention Distinction and Its Importance to HCI. *In Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialog*, 21–28.

Audi, R. (1995). *The Cambridge Dictionary of Philosophy*. Cambridge University Press.

Bird, S. (2006). NLTK: The natural language toolkit. In *Proceedings of the COLING/ACL on Interactive Presentation Sessions* (pp. 69–72). Stroudsburg, PA, USA: Association for Computational Linguistics.

Cappelen, H., & Lepore, E. (1997). Varieties of quotation. *Mind*, *106*(423), 429–450. doi:10.1093/mind/106.423.429

Davidson, D. (1979). Quotation. *Theory and Decision*, *11*(1), 27–40. doi:10.1007/BF00126690

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.

García-Carpintero, M. (2004). The deferred ostension theory of quotation. *Noûs*, *38*(4), 674–692.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*, 10–18. doi:10.1145/1656274.1656278

Hu, G. (2010). A place for metalanguage in the L2 classroom. *ELT Journal*. doi:10.1093/elt/ccq037

Jaworski, A., Coupland, N., & Galasinski, D. (Eds.). (2004). *Metalanguage: Language, Power, and Social Process*. De Gruyter.

John, G., & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338–345). Morgan Kaufmann.

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, *13*, 637–649. doi:10.1162/089976601300014493

Kohavi, R. (1995). The Power of Decision Tables. In *Proceedings of the European Conference on Machine Learning* (pp. 174–189). Springer Verlag.

Levy, R., & Andrew, G. (2006). TRegex and TSurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Maier, E. (2007). Mixed quotation: between use and mention. In *Logic and Engineering of Natural Language Semantics Workshop*. Retrieved from http://ncs.ruhosting.nl/emar/em_lenls_quot.pdf

Marneffe, M. D., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the International Conference on Language Resources and Evaluation,* pp. 449–454.

Punyakanok, V., Roth, D., & Yih, W. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, *34*(2), 257–287. doi:10.1162/coli.2008.34.2.257

Quine, W. V. O. (1940). *Mathematical logic*. Cambridge, MA: Harvard University Press.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. San Mateo, Calif.: Morgan Kaufmann Publishers.

Saka, P. (2006). The Demonstrative and Identity Theories of Quotation. *Journal of Philosophy*, *103*(9), 452–471.

Sperber, D., & Wilson, D. (1981). Irony and the Use-Mention Distinction. In *Radical Pragmatics* (pp. 295–318). New York.

Tarski, A. (1933). The concept of truth in formalized languages. In J. H. Woodger (Ed.), *Logic, Semantics, Mathematics*. Oxford: Oxford University Press.

Wilson, S. (2011a). *A computational theory of the use-mention distinction in natural language*. University of Maryland at College Park. PhD Thesis, College Park, MD, USA.

Wilson, S. (2011b). In search of the use-mention distinction and its impact on language processing tasks. *International Journal of Computational Linguistics and Applications*, *2*(1-2), 139–154.

Wilson, S. (2012). The Creation of a Corpus of English Metalanguage. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, (pp. 638–646).