The Creation of a Corpus of English Metalanguage

Shomir Wilson*
Carnegie Mellon University
Pittsburgh, PA 15213, USA
shomir@cs.cmu.edu

Abstract

Metalanguage is an essential linguistic mechanism which allows us to communicate explicit information about language itself. However, it has been underexamined in research in language technologies, to the detriment of the performance of systems that could exploit it. This paper describes the creation of the first tagged and delineated corpus of English metalanguage, accompanied by an explicit definition and a rubric for identifying the phenomenon in text. This resource will provide a basis for further studies of metalanguage and enable its utilization in language technologies.

1 Introduction

In order to understand the language that we speak, we sometimes must refer to the language itself. Language users do this through an understanding of the *use-mention distinction*, as exhibited by the mechanism of *metalanguage*: that is, language that describes language. The use-mention distinction is illustrated simply in Sentences (1) and (2) below:

(1) I watch football on weekends.
(2) Football may refer to one of several sports.

A reader understands that *football* in Sentence (1) refers to a sporting activity, while the same word in Sentence (2) refers to the term *football* itself. Evidence suggests that human communication frequently employs metalanguage (Anderson et al. 2002), and the phenomenon is essential for many activities, including the introduction of new

---

vocabulary, attribution of statements, explanation of meaning, and assignment of names (Saka 2003). Sentences (3) through (8) below further illustrate the phenomenon, highlighted in bold.

(3) This is sometimes called **tough love**.
(4) I wrote "**meet outside**" on the chalkboard.
(5) **Has** is a conjugation of the verb **have**.
(6) The button labeled **go** was illuminated.
(7) That bus, was its name **61C**?
(8) **Mississippi** is fun to spell.

Recognizing a wide variety of metalinguistic constructions is a skill that humans take for granted in fellow interlocutors (Perlis, Purang & Andersen 1998), and it is a core language skill that children demonstrate at an early age (Clark & Schaefer 1989). Regardless of context, topic, or mode of communication (spoken or written), we are able to refer directly to language, and we expect others to recognize and understand when we do so.

The study of the syntax and semantics of metalanguage is well developed for formal languages. However, the study of the phenomenon in natural language is relatively nascent, and its incorporation into language technologies is almost non-existent. Parsing the distinction is difficult, as shown in Figure 1 below: *go* does not function as a verb in Sentence (6), but it is tagged as such. *Delineating* an instance of metalanguage with quotation marks is a common convention, but this often fails to ameliorate the parsing problem. Quotation marks, italic text, and bold text—three common conventions used to highlight metalanguage—are inconsistently applied and are already "overloaded" with several distinct uses.

Moreover, applications of natural language processing generally lack the ability to recognize and interpret metalanguage (Anderson et al. 2002).

Systems using sentiment analysis are affected, as sentiment-suggestive terms appearing in metalanguage (especially in quotation, a form of the phenomenon (Maier 2007)) are not necessarily reflective of the writer or speaker. Applications of natural language understanding cannot process metalanguage without detecting it, especially when upstream components (such as parsers) mangle its structure. Interactive systems that could leverage users' expectations of metalanguage competency currently fail to do so. Figure 2 below shows a fragment of conversation with the *Let's Go!* (Raux et al. 2005) spoken dialog system, designed to help users plan trips on Pittsburgh's bus system.

```
(ROOT
  (S
    (NP
      (NP (DT The) (NN button))
      (VP (VBN labeled)
        (S
          (VP (VB go)))))
    (VP (VBD was)
      (VP (VBN illuminated)))
    (. .)))
```

Figure 1. Output of the Stanford Parser (Klein & Manning 2003) for Sentence (6). Adding quotation marks around *go* alters the parser output slightly (not shown), but *go* remains labeled VB.

```
Let's Go!: Where do you wish to depart
from?
User: Arlington.
Let's Go!: Departing from Allegheny
West. Is this right?
User: No, I said "Arlington".
Let's Go!: Please say where you are
leaving from.
```

Figure 2: A conversation with *Let's Go!* in which the user responds to a speech recognition error.

The exchange shown in Figure 2 is representative of the reactions of nearly all dialog systems: in spite of the domain generality of metalanguage and the user's expectation of its availability, the system does not recognize it and instead "talks past" the user. In effect, language technologies that ignore metalanguage are discarding the most direct source of linguistic information that text or utterances can provide.

This paper describes the first substantial study to characterize and gather instances of English metalanguage. Section 2 presents a definition and a rubric for metalanguage in the form of *mentioned language*. Section 3 describes the procedure used to create the corpus and some notable properties of its contents, and Section 4 discusses insights gained into the phenomenon. The remaining sections discuss the context of these results and future directions for this research.

2 Metalanguage and the Use-Mention Distinction[1]

Although the reader is likely to be familiar with the terms *use-mention distinction* and *metalanguage*, the topic merits further explanation to precisely establish the phenomenon being studied. Intuitively, the vast majority of utterances are produced for use rather than mention, as the roles of language-mention are auxiliary (albeit indispensible) to language use. This paper will adopt the term *mentioned language* to describe the literal, delineable phenomenon illustrated in examples thus far. Other forms of metalanguage occur through deictic references to linguistic entities that do not appear in the relevant statement. (For example, consider "That word was misspelled" where the referred-to word resides outside of the sentence.) For technical tractability, this study focuses on mentioned language.

2.1 Definition

Although the use-mention distinction has enjoyed a long history of theoretical discussion, attempts to explicitly define one or both of the distinction's disjuncts are difficult (or impossible) to find. Below is the definition of mentioned language adopted by this study, followed by clarifications.

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.

Here, a *token* is the specific, situated (i.e., as appearing in the sentence) instantiation of a linguistic entity: a letter, symbol, sound, word, phrase, or another related entity. A *property* might

---

[1] The definition and rubric in this section were originally introduced by Wilson (2011a). For brevity, their full justifications and the argument for equivalence between the two are not reproduced here.

be a token's spelling, pronunciation, meaning (for a variety of interpretations of *meaning*), structure, connotation, original source (in cases of quotation), or another aspect for which language is shown or demonstrated. The *type* of T is relevant in most instances of mentioned language, but the *token* itself is relevant in others, as in the sentence below:

(9) "The" appears between quote marks here.

Constructions like (9) are unusual and are of limited practical value, but the definition accommodates them for completeness.

The adoption of this definition was motivated by a desire to study mentioned language with precise, repeatable results. However, it was too abstract to consistently apply to large quantities of candidate phrases in sentences, a necessity for corpus creation. A brief attempt to train annotators using the definition was unsuccessful, and instead a rubric was created for this purpose.

2.2 Annotation Rubric

A human reader with some knowledge of the use-mention distinction can often intuit the presence of mentioned language in a sentence. However, to operationalize the concept and move toward corpus construction, it was necessary to create a rubric for labeling it. The rubric is based on substitution, and it may be applied, with caveats described below, to determine whether a linguistic entity is mentioned by the sentence in which it occurs.

Rubric: Suppose X is a linguistic entity in a sentence S. Construct sentence S' as follows: replace X in S with a phrase X' of the form "that [item]", where [item] is the appropriate term for X in the context of S (e.g., "letter", "symbol", "word", "name", "phrase", "sentence", etc.). X is an instance of mentioned language if, when assuming that X' refers to X, the meaning of S' is equivalent to the meaning of S.

To further operationalize the rubric, Figure 3 shows it rewritten in pseudocode form. To verify the rubric, the reader can follow a positive example and a negative example in Figure 4.

To maintain coherency, minor adjustments in sentence wording will be necessary for some candidate phrases. For instance, Sentence (10) below must be rewritten as (11):

(10) The word *cat* is spelled with three letters.
(11) *Cat* is spelled with three letters.

This is because S' for (10) and (11) are respectively (12) and (13):

(12) The word that word is spelled with three letters.
(13) That word is spelled with three letters.

```
Given S a sentence and X a copy of a
linguistic entity in S:
(1) Create X': the phrase "that [item]",
    where [item] is the appropriate term
    for linguistic entity X in the
    context of S.
(2) Create S': copy S and replace the
    occurrence of X with X'.
(3) Create W: the set of truth
    conditions of S.
(4) Create W': the set of truth
    conditions of S', assuming that X'
    in S' is understood to refer
    deictically to X.
(5) Compare W and W'. If they are equal,
    X is mentioned language in S. Else,
    X is not mentioned language in S.
```

Figure 3: Pseudocode equivalent of the rubric.

```
Positive Example
S: Spain is the name of a European
country.
X: Spain.
(1) X': that name
(2) S': That name is the name of a
    European country.
(3) W: Stated briefly, Spain is the name
    of a European country.
(4) W': Stated briefly, Spain is the
    name of a European country.
(5) W and W' are equal. Spain is
    mentioned language in S.

Negative Example
S: Spain is a European country.
X: Spain.
(1) X': that name
(2) S': That name is a European country.
(3) W: Stated briefly, Spain is a
    European country.
(4) W': Stated briefly, the name Spain
    is a European country.
(5) W and W' are not equal. Spain is not
    mentioned language in S.
```

Figure 4: Examples of rubric application using the pseudocode in Figure 3.

Also, quotation marks around or inside of a candidate phrase require special attention, since their inclusion or exclusion in *X* can alter the meaning of *S'*. For this discussion, quotation marks

and other stylistic cues are considered informal cues which aid a reader in detecting mentioned language. Style conventions may call for them, and in some cases they might be strictly necessary, but a competent language user possesses sufficient skill to properly discard or retain them as each instance requires (Saka 1998).

3   The Mentioned Language Corpus

"Laboratory examples" of mentioned language (such as the examples thus far in this paper) only begin to illustrate the variation in the phenomenon. To conduct an empirical examination of mentioned language and to study the feasibility of automatic identification, it was necessary to gather a large, diverse set of samples. This section describes the process of building a series of three progressively more sophisticated corpora of mentioned language. The first two were previously constructed by Wilson (2010; 2011b) and will be described only briefly. The third was built with insights from the first two, and it will be described in greater detail. This third corpus is the first to *delineate* mentioned language: that is, it identifies precise subsequences of words in a sentence that are subject to the phenomenon. Doing so will enable analysis of the syntax and semantics of English metalanguage.

3.1   Approach

The article set of *English Wikipedia[2]* was chosen as a source for text, from which instances were mined using a combination of automated and manual efforts. Four factors led to its selection:

1) *Wikipedia is collaboratively written.* Since any registered user can contribute to articles, Wikipedia reflects the language habits of a large sample of English writers (Adler et al. 2008).
2) *Stylistic cues that sometimes delimit mentioned language are present in article text.* Contributors tend to use quote marks, italic text, or bold text to delimit mentioned language[3], thus following conventions respected across many domains of writing (Strunk & White 1979; Chicago Editorial Staff 2010; American Psychological Association. 2001). Discussion

boards and other sources of informal language were considered, but the lack of consistent (or any) stylistic cues would have made candidate phrase collection untenably time-consuming.
3) *Articles are written to introduce a wide variety of concepts to the reader.* Articles are written informatively and they generally assume the reader is unfamiliar with their topics, leading to frequent instances of mentioned language.
4) *Wikipedia is freely available.* Various language learning materials were also considered, but legal and technical obstacles made them unsuitable for creating a freely available corpus.

To construct each of the three corpora, a general procedure was followed. First, a set of current article revisions was downloaded from Wikipedia. Then, the main bodies of article text (excluding discussion pages, image captions, and other peripheral text) were scanned for sentences that contained instances of highlighted text (i.e., text inside of the previously mentioned stylistic cues). Since stylistic cues are also used for other language tasks, candidate instances were heuristically filtered and then annotated by human readers.

3.2   Previous Efforts

In previous work, a pilot corpus was constructed to verify the fertility of Wikipedia as a source for mentioned language. From 1,000 articles, 1,339 sentences that contained stylistic cues were examined by a human reader, and 171 were found to contain at least one instance of mentioned language. Although this effort verified Wikipedia's viability for the project, it also revealed that the hand-labeling procedure was time-consuming, and prior heuristic filtering would be necessary.

Next, the "Combined Cues" corpus was constructed to test the combination of stylistic filtering and a new lexical filter for selecting candidate instances. A set of 23 "mention-significant" words was gathered informally from the pilot corpus, consisting of nouns and verbs:

Nouns: letter, meaning, name, phrase, pronunciation, sentence, sound, symbol, term, title, word

Verbs: ask, call, hear, mean, name, pronounce, refer, say, tell, title, translate, write

Instances of highlighted text were only promoted to the hand annotation stage if they contained at least one of these words within the three-word phrase directly preceding the

---

[2] Described in detail at
http://en.wikipedia.org/wiki/English_Wikipedia.
[3] These conventions are stated in Wikipedia's style manual, though it is unclear whether most contributors read the manual or follow the conventions out of habit.

highlighted text. From 3,831 articles, a set of 898 sentences were found to contain 1,164 candidate instances that passed the combination of stylistic and lexical filters. Hand annotation of those candidates yielded 1,082 instances of mentioned language. Although the goal of the filters was only to ease hand annotation, it could be stated that the filters had almost 93% precision in detecting the phenomenon. It did not seem plausible that the set of mention-significant words was complete enough to justify that high percentage, and concerns were raised that the lexical filter was rejecting many instances of mentioned language.

3.3   The "Enhanced Cues" Corpus

The construction of the present corpus (referred to as the "Enhanced Cues" Corpus) was similar to previous efforts but used a much-enlarged set of mention-significant nouns and verbs gathered from the *WordNet* (Fellbaum 1998) lexical ontology. For each of the 23 original mention-significant words, a human reader started with its containing synset and followed hypernym links until a synset was reached that did not refer to a linguistic entity. Then, backtracking one synset, all lemmas of all descendants of the most general linguistically-relevant synset were gathered. Figure 5 illustrates this procedure with an example.
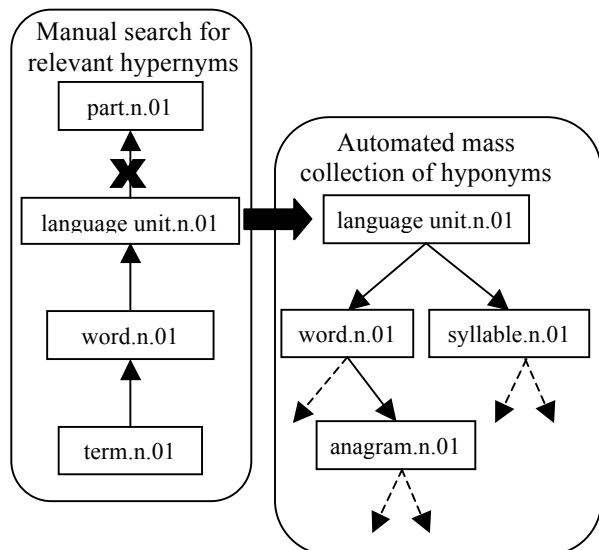

Figure 5: Gathering mention-significant words from WordNet using the seed noun "term". Here, "Language unit", "word", "syllable", "anagram", and all their descendants are gathered.

Using the combination of stylistic and lexical cues, 2,393 candidate instances were collected, and the researcher used the rubric and definition from Section 2 to identify 629 instances of mentioned language [4]. The researcher also identified four categories of mentioned language based on the nature of the substitution phrase X' specified by the rubric. These categories will be discussed in the following subsection. Figure 6 summarizes this procedure and the numeric outcomes.
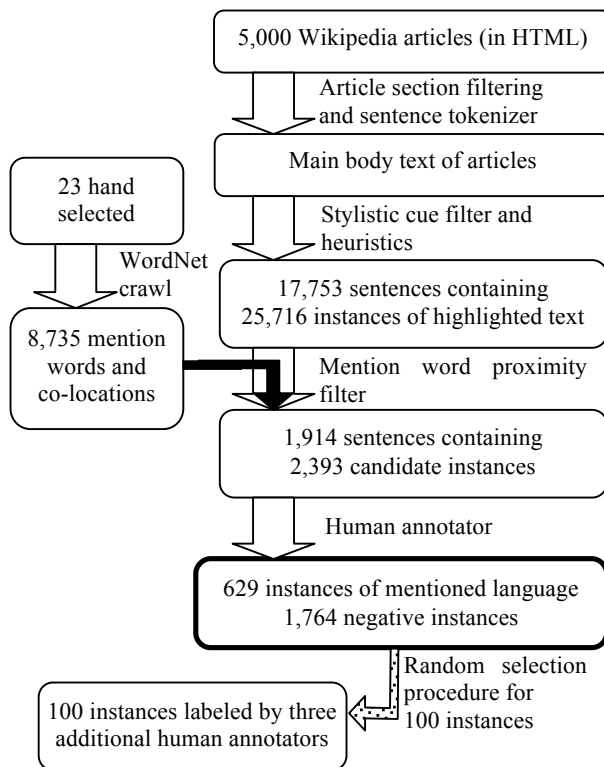

Figure 6: The procedure used to create the Enhanced Cues Corpus.

3.4   Corpus Composition

As stated previously, categories for mentioned language were identified based on intuitive relationships among the substitution phrases created for the rubric (e.g., "that word", "that title", "that symbol"). The categories are:

1) *Words as Words (WW)*: Within the context of the sentence, the candidate phrase is used to refer to the word or phrase itself and not what it usually refers to.

---

2) *Names as Names (NN)*: The sentence directly refers to the candidate phrase as a proper name, nickname, or title.
3) *Spelling or Pronunciation (SP)*: The candidate text appears only to illustrate spelling, pronunciation, or a character symbol.
4) *Other Mention/Interesting (OM)*: The candidate phrase is an instance of mentioned language that does not fit the above three categories.
5) *Not Mention (XX)*: The candidate phrase is not mentioned language.

Table 1 presents the frequencies of each category in the Enhanced Cues corpus, and Table 2 provides examples for each from the corpus. WW was by far the most common label to appear, which is perhaps an artifact of the use of Wikipedia as the text source. Although Wikipedia articles contain many names, NN was not as common, and informal observations suggested that names and titles are not as frequently introduced via metalanguage. Instead, their referents are introduced directly by the first appearance of the referring text. Spelling and pronunciation were particularly sparse; again, a different source might have yielded more examples for this category. The OM category was occupied mostly by instances of speech or language production by an agent, as illustrated by the two OM examples in Table 2.

Category	Code	Frequency
Words as Words	WW	438
Names as Names	NN	117
Spelling or Pronunciation	SP	48
Other Mention/Interesting	OM	26
Not Mention	XX	1,764

Table 1: The by-category composition of candidate instances in the Enhanced Cues corpus.

In the interest of revealing both lexical and syntactic cues for mentioned language, part-of-speech tags were computed (using NLTK (Loper & Bird 2002)) for words in all of the sentences containing candidate instances. Tables 3 and 4 list the ten most common words (as POS-tagged) in the three-word phrases before and after (respectively) candidate instances. Although the heuristics for collecting candidate instances were not intended to function as a classifier, figures for precision are shown for each word: these represent

the percentage of occurrences of the word which were associated with candidates identified as mentioned language. For example, 80% of appearances of the verb *call* preceded a candidate instance that was labeled as mentioned language.

Code	Example
WW	The IP Multimedia Subsystem architecture uses the term <u>transport plane</u> to describe a function roughly equivalent to the routing control plane.
	The material was a heavy canvas known as <u>duck</u>, and the brothers began making work pants and shirts out of the strong material.
NN	<u>Digeri</u> 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 <u>Naqvi al-Bukhari</u>.
SP	The French changed the spelling to <u>bataillon</u>, whereupon it directly entered into German.
	Welles insisted on pronouncing the word apostles with a hard <u>t</u>.
OM	He kneels over Fil, and seeing that his eyes are open whispers: <u>brother</u>.
	During Christmas 1941, she typed <u>The end</u> on the last page of Laura.
XX	<u>NCR</u> was the first U.S. publication to write about the clergy sex abuse scandal.
	Many Croats reacted by <u>expelling</u> all words in the Croatian language that had, in their minds, even distant Serbian origin.

Table 2: Two examples from the corpus for each category. Candidate phrases appear underlined, with the original stylistic cues removed.

Many of these words appeared as mention words for the Combined Cues corpus, indicating that prior intuitions about framing metalanguage were correct. In particular, *call (v)*, *word(n)*, and *term (n)* were exceptionally frequent and effective at associating with mentioned language. In contrast, the distribution of frequencies for the words *following* candidate instances exhibited a "long tail", indicating greater variation in vocabulary.

Rank	Word	Freq.	Precision (%)
1	call (v)	92	80
2	word (n)	68	95.8
3	term (n)	60	95.2
4	name (n)	31	67.4
5	use (v)	17	70.8
6	know (v)	15	88.2
7	also (rb)	13	59.1
8	name (v)	11	100
9	sometimes (rb)	9	81.9
10	Latin (n)	9	69.2

Table 3: The top ten words appearing in the three-word sequences *before* candidate instances, with precisions of association with mentioned language.

Rank	Word	Freq.	Precision (%)
1	mean (v)	31	83.4
2	name (n)	24	63.2
3	use (v)	11	55
4	meaning (n)	8	57.1
5	derive (v)	8	80
6	refers (n)	7	87.5
7	describe (v)	6	60
8	refer (v)	6	54.5
9	word (n)	6	50
10	may (md)	5	62.5

Table 4: The top ten words appearing in the three-word sequences *after* candidate instances, with precisions of association with mentioned language.

3.5 Reliability and Consistency of Annotation

To provide some indication of the reliability and consistency of the Enhanced Cues Corpus, three additional expert annotators were recruited to label a subset of the candidate instances. These additional annotators received guidelines for annotation that included the five categories, and they worked separately (from each other and from the primary annotator) to label 100 instances selected randomly with quotas for each category.

Calculations first were performed to determine the level of agreement on the mere presence of mentioned language, by mapping labels WW, NN, SP, and OM to *true* and XX to *false.* All four annotators agreed upon a *true* label for 46 instances and a *false* label for 30 instances, with an average pairwise Kappa (computed via NTLK) of 0.74. Kappa between the primary annotator and a

hypothetical "majority voter" of the three additional annotators was 0.90. These results were taken as moderate indication of the reliability of "simple" use-mention labeling.

However, the per-category results showed reduced levels of agreement. Kappa was calculated to be 0.61 for the original coding. Table 5 shows the Kappa statistic for binary re-mapping for each of the categories. This was done similarly to the "XX versus all others" procedure described above.

Code	Frequency	K
WW	17	0.38
NN	17	0.72
SP	16	0.66
OM	4	0.09
XX	46	0.74

Table 5: Frequencies of each category in the subset labeled by additional annotators and the values of the Kappa statistic for binary relabelings.

The low value for remapped OM was expected, since the category was small and intentionally not well-defined. The relatively low value for WW was not expected, though it seems possible that the redaction of specific stylistic cues made annotators less certain when to apply this category. Overall, these numbers suggest that, although annotators tend to agree whether a candidate instance is mentioned language or not, there is less of a consensus on how to qualify positive instances.

4 Discussion

The Enhanced Cues corpus confirms some of the hypothesized properties of metalanguage and yields some unexpected insights. Stylistic cues appear to be strongly associated with mentioned language; although the examination of candidate phrases was limited to "highlighted" text, informal perusal of the remainder of article text confirmed this association. Further evidence can be seen in examples from other texts, shown below with their original stylistic cues intact:

• Like so many words, the meaning of "addiction" has varied wildly over time, but the trajectory might surprise you.[5]

---

[5] News article from CNN.com:
http://www.cnn.com/2011/LIVING/03/23/addicted.to.addiction/index.html

- Sending a signal in this way is called a **speech act**.[6]
- M1 and M2 are Slashdot shorthand for "moderation" and "metamoderation," respectively.[7]
- He could explain foreordination thoroughly, and he used the terms "baptize" and "Athanasian."[8]
- They use *Kabuki* precisely because they and everyone else have only a hazy idea of the word's true meaning, and they can use it purely on the level of insinuation.[9]

However, the connection between mentioned language and stylistic cues is only valuable when stylistic cues are available. Still, even in their absence there appears to be an association between mentioned language and a core set of nouns and verbs. Recurring patterns were observed in how mention-significant words related to mentioned language. Two were particularly common:

- *Noun apposition between a mention-significant noun and mentioned language*. An example of this appears in Sentence (5), consisting of the noun *verb* and the mentioned word *have*.
- *Mentioned language appearing in appropriate semantic roles for mention-significant verbs*. Sentence (3) illustrates this, with the verb *call* assigning the label *tough love* as an attribute of the sentence subject.

With further study, it should be possible to exploit these relationships to automatically detect mentioned language in text.

5 Related Work

The use-mention distinction has enjoyed a long history of chiefly theoretical discussion. Beyond those authors already cited, many others have addressed it as the formal topic of *quotation* (Davidson 1979; Cappelen & Lepore 1997; García-Carpintero 2004; Partee 1973; Quine 1940; Tarski 1933). Nearly all of these studies have eschewed empirical treatments, instead hand-picking illustrations of the phenomenon.

One notable exception was a study by Anderson et al. (2004), who created a corpus of metalanguage from a subset of the British National Corpus, finding that approximately 11% of spoken utterances contained some form (whether explicit or implicit) of metalanguage. However, limitations in the Anderson corpus' structure (particularly lack of word- or phrase-level annotations) and content (the authors admit it is noisy) served as compelling reasons to start afresh and create a richer resource.

6 Future Work

As explained in the introduction, the long-term goal of this research program is to apply an understanding of metalanguage to enhance language technologies. However, the more immediate goal for creating this corpus was to enable (and to begin) progress in research on metalanguage. Between these long-term and immediate goals lies an intermediate step: methods must be developed to detect and delineate metalanguage automatically.

Using the Enhanced Cues Corpus, a two-stage approach to automatic identification of mentioned language is being developed. The first stage is *detection*, the determination of whether a sentence contains an instance of mentioned language. Preliminary results indicate that approximately 70% of instances can be detected using simple machine learning methods (e.g., bag of words input to a decision tree). The remaining instances will require more advanced methods to detect, such as word sense disambiguation to validate occurrences of mention-significant words. The second stage is *delineation*, the determination of the subsequence of words in a sentence that functions as mentioned language. Early efforts have focused on the associations discussed in Section 5 between mentioned language and mention-significant words. The total number of such associations appears to be small, making their collection a tractable activity.

Acknowledgements

---

[6] Page 684 of Russell and Norvig's 1995 edition of *Artificial Intelligence*, a textbook.
[7] Frequently Asked Questions (FAQ) list on Slashdot.org: http://slashdot.org/faq/metamod.shtml
[8] Novel *Elmer Gantry* by Sinclair Lewis.
[9] Opinion column on Slate.com: http://www.slate.com/id/2250081/

References

Adler, B. Thomas, Luca de Alfaro, Ian Pye & Vishwanath Raman. 2008. Measuring author contributions to the Wikipedia. In *Proc. of WikiSym '08*. New York, NY, USA: ACM.

American Psychological Association. 2001. *Publication Manual of the American Psychological Association*. 5th ed. Washington, DC: American Psychological Association.

Anderson, Michael L, Yoshi A Okamoto, Darsana Josyula & Donald Perlis. 2002. The use-mention distinction and its importance to HCI. In *Proc. of EDILOG 2002*. 21–28.

Anderson, Michael L., Andrew Fister, Bryant Lee & Danny Wang. 2004. On the frequency and types of meta-language in conversation: A preliminary report. In *Proc. of the 14th Annual Conference of the Society for Text & Discourse*.

Cappelen, H & E Lepore. 1997. Varieties of quotation. *Mind* 106(423). 429 –450.

Chicago Editorial Staff. 2010. *The Chicago Manual of Style*. 16th ed. University of Chicago Press.

Clark, Herbert H. & Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13(2). 259–294.

Davidson, Donald. 1979. Quotation. *Theory and Decision* 11(1). 27–40.

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.

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

Klein, Dan & Christopher D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. *Advances in Neural Information Processing Systems* 15.

Loper, Edward & Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics* 1. 63–70. Association for Computational Linguistics.

Maier, Emar. 2007. Mixed quotation: Between use and mention. In *Proc. of LENLS 2007*.

Partee, Barbara. 1973. The syntax and semantics of quotation. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart, Winston.

Perlis, Donald, Khemdut Purang & Carl Andersen. 1998. Conversational adequacy: Mistakes are the essence. *International Journal of Human-Computer Studies* 48(5). 553–575.

Quine, W. V. O. 1940. *Mathematical Logic*. Cambridge, MA: Harvard University Press.

Raux, Antoine, Brian Langner, Dan Bohus, Alan W Black & Maxine Eskenazi. 2005. Let's Go public! Taking a spoken dialog system to the real world. In *Proc. of Interspeech 2005*.

Saka, Paul. 1998. Quotation and the use-mention distinction. *Mind* 107(425). 113 –135.

Saka, Paul. 2003. Quotational constructions. *Belgian Journal of Linguistics* 17(1).

Strunk, Jr. & E. B. White. 1979. *The Elements of Style, Third Edition*. Macmillan.

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

Wilson, Shomir. 2010. Distinguishing use and mention in natural language. In *Proc. of the NAACL HLT 2010 Student Research Workshop*, 29–33. Association for Computational Linguistics.

Wilson, Shomir. 2011a. *A Computational Theory of the Use-Mention Distinction in Natural Language*. Ph.D. dissertation, University of Maryland at College Park.

Wilson, Shomir. 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.