Semantic Representations of Near-Synonyms

Philip Edmonds
Department of Computer Science, University of Toronto
Toronto, Canada, M5S 3G4
pedmonds@cs.toronto.edu

February 26, 1998

Abstract

We present a new model of lexical semantics that clusters near-synonyms and so facilitates sophisticated lexical choice. The semantics of a word arise from a combination of its core denotation (shared with its near-synonyms) and a set of explicit differences from its near-synonyms.

Submission Type: student session 3 pages

Topic Areas: lexical semantics, lexical choice.

Author of Record: Philip Edmonds

Under consideration for other conferences (specify)? NO

Semantic Representations of Near-Synonyms

Abstract

We present a new model of lexical semantics that clusters near-synonyms and so facilitates sophisticated lexical choice. The semantics of a word arise from a combination of its core denotation (shared with its near-synonyms) and a set of explicit differences from its near-synonyms.

1 The problem

If we are ever to have accurate machine translation systems and articulate natural language generation systems we will need a very sophisticated and principled model of lexical choice. The difficulty is that a word can express much more than its literal dictionary meaning; it can also indirectly convey a myriad of implications, connotations, and attitudes depending on the situation. And for a particular concept, there are often many *near-synonyms* that differ precisely in these kinds of minor ideas. For example, consider the synonyms of *error* shown in table 1. If one chooses poorly, then one risks at best being ineffective and at worst giving a wrong or even antagonistic impression—for instance, if one chooses *blunder* when one doesn't intend to label a person as "stupid".

Error implies a straying from a proper course and suggests guilt as may lie in failure to take proper advantage of a guide …

Mistake implies misconception, misunderstanding, a wrong but not always a blameworthy judgment; it expresses less severe criticism that *error*.

Blunder is harsher than *mistake* or *error*; it commonly implies ignorance or stupidity, sometimes blameworthiness.

Slip carries a stronger implication of inadvertence or accident than *mistake* …

Table 1: An abridged entry from *Webster's New Dictionary of Synonyms* (Gove, 1973)

Our ultimate goal is to develop a sophisticated computational lexical choice process that can choose from a group of near-synonyms the one that best achieves the desired effects in the current context. The process is to be broadly applicable, but we are implementing it within a current NLG system, known for its ability to paraphrase, developed by Stede (1996).

In order to achieve this goal, we must first formalize lexical variation. This paper presents a new theory of lexical semantics that was inspired by the structure and information contained in synonym usage notes (e.g., in *Webster's New Dictionary of Synonyms* (Gove, 1973).)[1] The semantics of a cluster of near-synonyms is represented in two levels: a single core denotation that is shared, and a set of peripheral concepts that the words differ in denoting, implying, and/or emphasizing.

2 Synonymy in the lexicon

The conventional model of lexical knowledge, used in many computational systems, is not suitable for representing the fine-grained and fuzzy distinctions between near-synonyms (Hirst, 1995). In this model, knowledge of the world is represented by interrelated concepts that are often organized into an inheritance hierarchy. The denotation of a lexical item is represented as a concept, or a configuration of concepts, and amounts to a direct link between word and concept. So except for polysemy and (absolute) synonymy, there is no logical difference between a lexical item and a concept. Therefore, words that are nearly synonymous have to be linked each to their own slightly different concepts. The problem comes in trying to represent these slightly different concepts and the relationships between them. We end up with an awkward proliferation of concepts—not unlike the proliferation of word senses seen in the problem of logical polysemy (Cruse, 1986; Pustejovsky, 1991)—that neglects many of the important relationships, because of the enforced hierarchy. For instance, should *mistake* be a kind of *error*, or vice versa? (See also Hirst's (1995) example: *falsehood, lie, fib, misrepresentation, untruth*).

An alternative representation is suggested by the principle behind the synonym usage notes of *Webster's New Dictionary of Synonyms*. Words are grouped into an entry if they have the same essential meaning, i.e., that they "can be defined in the same terms up to a certain point" (Gove, 1973, p. 25a) and differ only in terms of minor ideas involved in their meanings. We combine this principle with Saussure's paradigmatic view that "each of a set of synonyms … has its particular value only because they stand in contrast with one another" (Saussure, 1983, p. 114) and envision a representation in which the meaning of a word arises from a combination of its essential denotation (shared with other words) and a set of explicit differences to its near-synonyms.

[1]The idea of formalizing usage notes was first put forth by DiMarco and Hirst (1993).

```
(defcluster error_C
  :core (ROOT Action (ACTOR (V1 Person))
               (ATTRIBUTE (V2 Deviation)))
  :peripheral (
    (P1 Stupidity (ATTRIBUTE-OF V1))
    (P2 Blameworthiness (ATTRIBUTE-OF V1))
    (P3 Criticism (ACTEE V1))
    (P4 Misconception (CAUSE-OF V2)
                      (ACTOR V1))
    (P5 Accident (CAUSE-OF V2)
                 (ACTOR V1))))
```

Figure 1: Representation of the "error" cluster

Implication	249
Suggestion	123
Stress	54
Emphasis	12
Connotation	53
Presupposition	7
Denotation	84
Application	25
Total	601

Table 2: Methods of distinguishing synonyms in 99 usage notes of *Webster's New Dictionary of Synonyms.*

Thus, we propose a clustered model of lexical knowledge that has two levels of representation: a core denotation and peripheral concepts. The *core denotation* is the same as a denotation in the conventional model—a configuration of concepts that functions as a necessary applicability condition (for choice)—but it is shared among the near-synonyms in the cluster. All of the *peripheral concepts* that the words may differ in denoting, implying, or emphasizing are also represented as configurations of concepts, but crucially, they are not necessary conditions. The differences themselves are represented explicitly in the lexical entries, as described in the next section.

Figure 1 shows, in the representation language of our system, the core denotation and a selection of peripheral concepts for the cluster of "error" synonyms.

We claim that the clustered model is the only practical solution to modelling lexical knowledge to this level of granularity. Even though the conceptual hierarchy is more coarse-grained, it still retains its inferential power (from inheritance) to a level consistent with the current systems. It also solves the 'proliferation of concepts' problem, while enabling a principled account of lexical variation.

3 Lexical Variation

In order to formalize lexical variation, we analyzed the first 99 synonym usage notes in *Webster's New Dictionary of Synonyms*, and found 601 distinctions, which we classified into 8 categories according to the terminology used in the dictionary; see table 2. We did not include distinctions of style (e.g., of formality).

The meaning of these terms is far from clear. For instance, *suggest* and *imply* are themselves differentiated in a usage note: "very often the difference between *suggest* and *imply* is not clear, though *suggest* often connotes the necessity of delicate perception while *imply* connotes the need of inference" (Gove, 1973). And fine-grained denotations *are* implications, in at least one sense of implication. So, we could possibly use a single term for all three methods, but we are reluctant to throw away the information that the lexicographers have carefully and deliberately recorded.

However, at least one major distinction can be made. There is a continuum of *directness* of expression, from denotation at one end, to suggestion at the other. This continuum could equally be called one of necessity (see Cruse's (1986) definitions of *criterial*, *expected*, and *possible* traits), or of strength of implication. In fact, the strength of an implication (or denotation, or suggestion) is sometimes explicitly stated in the usage note (see *slip* in table 1). So, while fine-grained denotational distinctions are logically entailed by a word, its implications and suggestions are mere possibilities that a reader might or might not understand in an expression. Which method is used by a writer will of course have pragmatic ramifications on the discourse. For instance, the more indirect an expression is, the less likely the reader is to infer it and pay attention to it, and the more easily it can be cancelled by an explicit statement.

Nearly half of the denotations, implications, and suggestions (192 of 456) in table 2 are qualified by a frequency term such as "not always", "commonly", "usually", "seldom", etc. (compare table 1). As the usage notes are descriptive, these terms represent the frequency with which a synonym expresses the given concept in the corpora that the lexicographers used. Whether or not a word actually expresses a peripheral concept in a given context is a very difficult issue that the lexicographers resolved by using their intuition. We are not yet in a position to formalize this kind of contextual effect, but we can make use of the frequency, at least, by treating it as *potential*. We formalize potential on a five-point scale ranging from "never" to "always", and assign every distinction a value on this scale. Potential is used during lexical choice to decide how suitable a particular word is for satisfying a preference to express a peripheral concept.

Sometimes a distinction between words is relative. For example, a word may have a stronger or weaker implication relative to another word, or the implied concept itself might involve relativity (e.g., *mistake* implies less severe criticism than *error*). Explicit differences involving relativity are formalized by having an 'operation' associated

```
(deflex mistake_l
  :cluster error_C
  :language english
  :distinctions (
    (nil always medium implication P4)
    (nil usually medium implication P2)
    ( (compared-to error_l)
      always medium implication
      (P3 (ATTRIBUTE (Severity
                       (VALUE less)))))
    ( (compared-to slip_l)
      always weaker implication P5))
```

Figure 2: Lexical entry for *mistake*.

with every distinction.

In sum, the distinctions for a near-synonym will be represented as a list of structures of the following form:

(OP POTENTIAL STRENGTH METHOD CONCEPT)

in which each variable can have the values shown in table 3. For example, see the lexical entry for *mistake* in figure 2. Collectively, the distinctions for all the near-synonyms in a cluster form an explicit set of differences between pairs of words, which can be reasoned with during lexical choice.

OP:	compared-to, nil
POTENTIAL:	never, seldom, sometimes, usually, always
STRENGTH:	weak, medium, strong, weaker, stronger
METHOD:	denotation, implication, suggestion
CONCEPT:	any peripheral concept

Table 3: Possible values in the template.

There are of course many complexities not dealt with here, including disjunctions in conceptual configurations, usage constraints, emphasis, connotations, and style, but there is no reason that they could not be accommodated in this formalism.

4 Lexical choice

Space allows only a brief example of how lexical choice would work. After finding that the "error" cluster is applicable to an input situation, the choice process analyzes *preferences*, also given as input. It then attempts to choose a word the maximizes the number of preferences satisfied.

For example, if we include the following preference to convey the idea of misconception (a preference is an instance of a concept) then the preference analysis process would determine that only *mistake* would be appropriate:

```
(misconceive1 (ACTOR john
               CAUSE-OF deviation1))
```

On the other hand, the following preference would indicate that *error* is more appropriate:

```
(criticism1 (ACTEE john
             ATTRIBUTE (severity1
                         (VALUE high))))
```

If we specified both preferences, then a decision about which one to satisfy would be made according to the individual importance of the preferences. But, it might also be possible to satisfy both preferences at once. For instance, the preference that is not satisfied by *mistake* could be expressed explicitly by generating a phrase or other syntactic structure, or it could be satisfied indirectly by a peripheral concept of another word in the output. In this way, preferences are like Elhadad et al's (1997) floating constraints, except that they can be satisfied to different degrees (or not at all), and more than one can be specified.

The exact language for representing preferences, and how preferences interact with the POTENTIAL and STRENGTH attributes during lexical choice is currently being developed.

5 Conclusion

We proposed a clustered model of lexical knowledge whose basis is synonym usage notes. In it, a coarse-grained taxonomy of concepts has as leaves lexical clusters. The semantics of the near-synonyms in each cluster are represented in two levels: a core denotation and peripheral concepts. The most common types of lexical variation are explicitly represented, which makes implementing a sophisticated lexical choice process possible.

References

D. Alan Cruse. 1986. *Lexical Semantics*. Cambridge U.P.

Chrysanne DiMarco and Graeme Hirst. 1993. Usage notes as the basis for a representation of near-synonymy for lexical choice. In *9th Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary and Text Research*, pages 33–43, Oxford.

Michael Elhadad, Kathleen McKeown, and Jacques Robin. 1997. Floating constraints in lexical choice. *Computational Linguistics*, 2(23):195–240.

Philip Gove, editor. 1973. *Webster's New Dictionary of Synonyms*. Merriam-Webster, Springfield, Mass.

Graeme Hirst. 1995. Near-synonymy and the structure of lexical knowledge. In *AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*, pages 51–56, Stanford, CA, March.

James Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4):409–441.

Ferdinand de Saussure. 1983. *Course in General Linguistics*. G. Duckworth, London.

Manfred Stede. 1996. Lexical paraphrases in multilingual sentence generation. *Machine Translation*, 11:75–107.