In *Proceedings of the First Student Conference in Computational Linguistics in Montreal,* **1996, pages 80-87.**

Evoking Meaning by Choosing the Right Words

Résumé

Choisir le mot juste est difficile. Une raison est que l'effet du contexte sur le sens exprimé par un mot est très complexe. En particulier, quand un mot est utilisé dans un contexte qui n'est pas normal, le mot peut évoquer un sens spécial. Dans cet article, je présente un processus de 'lexical choice' qui choisis le mot meilleur parmis un groupe de synonymes qui produit les effets désirés dans le contexte donné. Le processus dépend d'un modèle regroupé de savoir lexical qui unit un modèle statitistique (pour déterminer quand l'usage d'un mot sera marquer), avec un modèle sur le savoir (pour déterminer les effets qui se produiront).

Abstract

Choosing the right word is difficult. One reason is that the context affects the meaning expressed by a word in complex ways. In particular, when a word is used in a context that is not normal for the word, it may evoke a special meaning. This paper presents a lexical choice process that chooses the word from a set of near-synonyms that best produces the desired effects in the given context. It relies on a clustered representation of lexical knowledge that unites both a statistical model of word co-occurrence (for determining when a word use will be marked), and a knowledge-based model (for determining what specific effects will occur).

In *Proceedings of the First Student Conference in Computational Linguistics in Montreal,* **1996, pages 80-87.**

Evoking Meaning by Choosing the Right Words

Philip Edmonds

University of Toronto

1. Context and lexical choice

Choosing the right word is difficult for computers, let alone for people. Effective communication requires that one seek to express more than the explicit meanings that words carry. One can evoke powerful effects by conveying a myriad of implicit meanings, connotations, overtones, and feelings simply by choosing the right word in the right context. Indeed, context is such an important factor in word choice that it can sometimes preordain the choice of a word, but often it will just set up a preference for one word over another. Moreover, context can affect the meaning expressed by a word in complicated ways.[1]

When choosing words, the usual primary concern is to express the desired denotation, but often many subtly different nearly-synonymous words (or phrases) are available. Now, when the typical or normal word for a particular context is chosen, it will convey just its inherent meaning (which is relative to the meaning of its synonyms); but, when a less typical word is chosen, its use will clash with the context to a certain degree, i.e., its use will be *marked*, and this will cause the evocation of a special effect that is a function of the context and the difference between the typical and chosen words. That is, the combination of context and word meaning actually creates a (temporary) meaning that is not intrinsic to the word itself. This characterization of one type of connotation is one consequence of the view that context creates meaning from words (see footnote 1). For example, while a formal word in a formal context is expected, the same word in an informal context may sound overly serious or pedantic, which may be desired, but it may also point to social ineptitude in the speaker, which may be unwanted.

Effects like these are not arbitrary. People use them in order to achieve specific communicative goals that include conveying the right information, evoking and expressing the right feelings, and avoiding unwanted implications. I will discuss how to account for such evoked meanings in a sophisticated lex-

[1] There is a growing acceptance that words have very little meaning in isolation from text, except in an abstract sense. Instead, the view is that words attain their meaning by use in context (e.g., (Ruhl, 1989; Pustejovsky, 1991; Saint-Dizier and Viegas, 1995)).

ical choice process of a natural language generation (NLG) system.

2. Evoked meaning

Cruse (1986) writes that some “restrictions” on co-occurrence can be flouted deliberately for producing certain effects such as *evoked meaning*. In particular, when the register of a word clashes with the register of the text, there is potential for an evoked meaning. For instance, in a moderately formal (in style) text, ‘cease to live’ would be lexicalized most naturally as *die*; *pass away*, on the other hand, would stress feelings of solemnity (or the like); whereas *kick the bucket* would stress irreverence.

Denotative meanings can also be stressed. For example, normally a doctor would *order*, a colonel would *command*, while a robber might *demand*, each word emphasizing the secondary concepts of ‘ordering’, including ‘authority’ and ‘urgency’, to a greater or lesser degree (see below). But if a doctor *commands*, he puts special emphasis on (i.e., he stresses) urgency, and if he *demands*, he stresses urgency even more, and also adds an air of force, all the while stressing the seriousness of the order, perhaps in order to pressure for compliance.[2]

The restrictions that are broken clearly relate to the syntagmatic relations between words, which we can encode using a model of collocation. ‘Statistical’ approaches to collocation are more appropriate here compared to ‘linguistic’ approaches, such as Benson et al.’s (1986) dictionary, or Mel’čuk’s (1987) lexical functions, for several reasons. For one, they treat collocations more like *preferences* for what is typical, rather than as relatively fixed combinations that *constrain* word choice, which is too strong. For example, some evoked meanings don’t necessarily rely on well-delineated distinctions between ‘domains’, as between the doctor and colonel domains. For instance, a robber who *orders* his victim, or a child who *orders* his parents, is able to stress the robber’s or child’s (attempted) use of authority, because the verb is used atypically; and yet, no constraint is broken.

Second, the linguistic approaches, and the statistical approaches, for that matter, do not explain what happens when a restriction is broken, which is certainly possible, and even allowed. In particular, the semantics carried by a lexical function are two vague to represent subtle variations in word usage. For instance, regarding the **Oper** (‘performance’) lexical function, what is the difference between *asking* a

[2] The effect is reminiscent of metaphor. For example, Kittay (1987) argues that a metaphor necessarily requires an incongruity in the literal or conventional interpretation of an utterance. She further argues that the incongruity results from the rules that are broken—compare Cruse’s understanding of evoked meaning—when a word from one semantic field, the vehicle of the metaphor, is used in the context of a different field, the topic field.

question, and *putting* a question?

Thus, we use a statistical model of word co-occurrence to model how a word might clash with its context, and we unite it with a traditional knowledge-based model of lexical meaning in order to represent the differences between synonyms. From the above examples, we see that meaning can be evoked (or stressed) along explicit dimensions of meaning including denotative, stylistic, and affective dimensions, which are inherently part of a word's meaning, and expressive (or pragmatic) dimensions that arise in context, such as seriousness. Our model *clusters* synonyms by their similarity and represents their differences along the dimensions explicitly, which follows directly from Saussure's paradigmatic view of lexis that "each of a set of synonyms … has its particular value only because they stand in contrast with one another" (Saussure, 1983, p. 114).

2.1. Word families

Firth (1957) regarded the syntagmatic pattern of collocation as one of the two central aspects of lexical meaning—the other being the context of situation. In essence, he said that lexical items are in a relation of mutual expectancy; thus, the use of a word is actually predicted by the others in its local context. Church et al. (1991) have found that the best co-occurrence metric is the mutual information statistic that measures the ratio of the probability of finding two words within a given window over the probabilities of finding each of the words alone, i.e., by chance (which is very similar to what Sinclair (1966) proposed in his early attempt to validate Firth's claims).

Now, I will associate a *word family* with each word in a cluster of synonyms. In its simplest form, a word family is defined as the set of words that collocate significantly with the head word as determined by significant mutual information scores. For example, the following table shows the top five words in the families of three words in the *order* cluster. (Mutual information was computed in the *Brown Corpus* for words within a window a 6 words (on both sides) of the source word and considered significant only if the joint frequency was greater than 2.)

command:	strategic, centers, post, offices, statement, …
demand:	backlog, industry, skills, regional, increases, …
order:	disorder, obeyed, chronological, restraining, congruence, …

Given a local context, i.e., the set of words already chosen, we can rank the candidate synonyms by how close each of their families matches the words in the context. The closer the match, the more typical the candidate is, which allows us to predict the most typical word, and to identify if a word might clash with

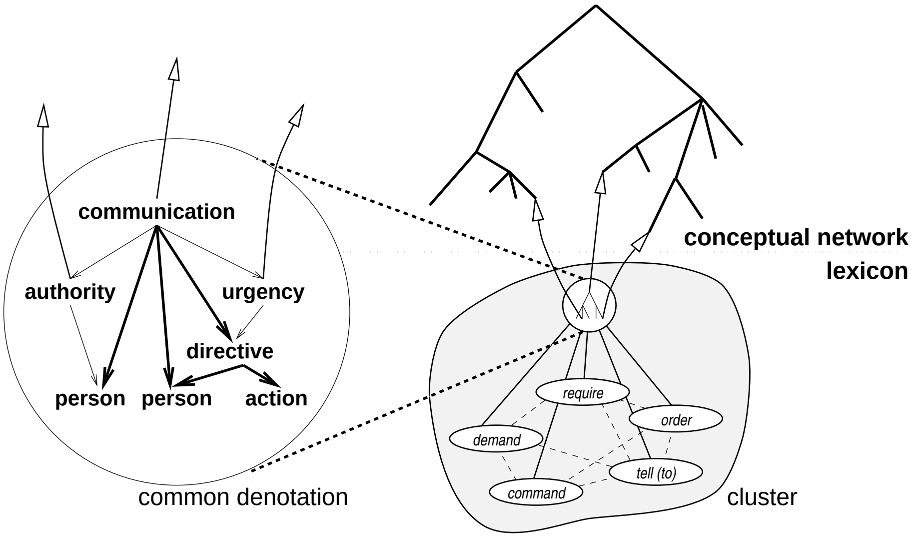

Figure 1: The clustered representation of lexical knowledge: the *order* cluster is shown with its common denotation expanded on the left, and several (unspecified) relations of difference shown as dashed lines.

the context.[3] However, it is likely that such simplistic word families will be inadequate in encoding the typical context a word occurs in; so, more elaborate methods must be experimented with. Nevertheless, what is important is the idea that synonyms should be ranked according to typicalness in context.

2.2. A clustered representation of lexical knowledge

Now, it is difficult to account for fine-grained and fuzzy differences between near-synonyms in a traditional taxonomic model of the lexicon; thus, we have developed a clustered representation of lexical knowledge, depicted in figure 1, that is strongly inspired by DiMarco, Hirst and Stede's (1993) and Hirst's (1995) work. The lexical entries of near-synonyms are *clustered* together underneath a *common denotation* that, crucially, is a configuration of concepts and roles of a conceptual network.[4]

Following Hirst, DiMarco, and Stede, differences in denotation will be represented explicitly, but not as differences on simple feature-like dimensions as their work might suggest one do. Instead, we no-

[3] Church et al. (1994) show that comparing only the mutual information scores of two words can be misleading, since the significance of the difference is not taken into account. They advocate using a t-test, but unfortunately, computing t-scores is inefficient in time and space.

[4] While this model of lexical knowledge is controversial, it has been argued to be both computationally practical and feasible (Hirst, 1995; Stede, 1996). The model does have its problems though. Because sets of synonyms are represented statically, it cannot account for context-dependent cases of synonymy.

tice that differences can predicate over different aspects and roles of the denotation and can be reasoned about, e.g., ‘authority’ (see figure 1) predicates over the ‘person’ who orders. Thus, we consider a lexical item as denoting not only a core concept with its roles (shown in bold arrows in figure 1), but also several secondary or peripheral concepts (shown in thin arrows) that represent the dimensions of possible variation. Near-synonyms clearly denote the same core concept, but following Saussure, they also denote the same secondary concepts, although emphasized to different degrees relative to each other.

Dimensions of variation can be adduced from many of the reference books created for this very purpose (e.g., (Hayakawa, 1987; Gove, 1984)), but, really, it is a (very daunting) lexicographic task. Some dimensions will be very general, applicable to many clusters (such as ‘intentional/accidental’ of *stare at*/*glimpse* (DiMarco, Hirst, and Stede, 1993)), while others will be specific to only a few clusters. For example, we found that the reference books discriminate words of the *order* cluster in terms of at least authority and urgency. In our model, the differences themselves are represented explicitly in the cluster as partial orderings of words on each of the dimensions, like so: (these orderings were identified from the references books)

Authority:	command, order, require $\succ$ direct $\succ$ instruct $\succ$ tell (to)
Urgency:	demand, require $\succ$ command $\succ$ order

The differences are represented in purely relative terms, with no notion of an absolute, because (1) it is difficult to assign emphasis in terms of fixed points on an absolute scale—the dictionaries themselves are often quite vague—and (2) the emphatic relations among words in one cluster are independent of those in other clusters. With this representation we can easily find a word of, say, high authority, relative to its synonyms.

3. Choosing the right word

The lexical choice process is to be integrated into an NLG system that has a sentence planning stage (for instance (Elhadad, 1992; Stede, 1996)). It takes as input the local context, which is the set of words that have already been chosen (or are potentials for choice), a configuration of concepts and roles that is to be verbalized, and a set of parameters that specify communicative intentions such as ‘stress that the order is serious’. (Of course, the parameters available will depend on the denotation and the situation.) Its output is the word (or phrase) that best produces the desired effects in the context.

The choice procedure, very briefly, is as follows:

(1) Find the synonym cluster that conveys the input denotation.

(2) Eliminate all candidate words constrained by the syntactic structure, or otherwise.

(3) Rank the candidate words by how typical they are in the context by matching their families to the context.

(4) If no parameters are specified, output the most typical word for the context.

(5) Otherwise, find the word that differs on the dimension(s) specified in the parameters *relative to* the most typical word, and output it.

4. Conclusion

The representation of lexical knowledge is the heart of the lexical choice process. By uniting two seemingly disparate models of lexical meaning we are able to account for evoked meaning: when a word is placed into an atypical context (as determined by its representation in the 'Firthian' model), we look to its denotation (in the traditional model) to see what effects occur. Much work remains to be done including a deeper investigation of word families, and the development of a model of expressive dimensions, just barely touched on here.

Acknowledgements. Thanks are due to Graeme Hirst for comments on earlier drafts, and to François Pitt for translation advice. This research was supported by the University of Toronto, and by a grant from the Ministry of Education and Training of Ontario.

References

Benson, Morton, Evelyn Benson, and Robert Ilson. 1986. *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. John Benjamins, Amsterdam.

Church, Kenneth Ward, William Gale, Patrick Hanks, and Donald Hindle. 1991. Using statistics in lexical analysis. In Uri Zernik, editor, *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*. Lawrence Erlbaum Associates, pages 115–164.

Church, Kenneth Ward, William Gale, Patrick Hanks, Donald Hindle, and Rosamund Moon. 1994. Lexical substitutability. In B.T.S. Atkins and A. Zampolli, editors, *Computational Approaches to the Lexicon*. Oxford University Press, pages 153–177.

Cruse, D. Alan. 1986. *Lexical Semantics*. Cambridge University Press.

DiMarco, Chrysanne, Graeme Hirst, and Manfred Stede. 1993. The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI Spring Symposium on Building Lexicons for Machine Translation*, pages 114–121, Stanford, CA, March.

Elhadad, Michael. 1992. *Using Argumentation to Control Lexical Choice: A Functional Unification Implementation*. Ph.D. thesis, Columbia University.

Firth, J. R. 1957. A synopsis of linguistic theory, 1930–55. In Philological Society, London, editor, *Studies in Linguistic Analysis*. Oxford Blackwell, pages 1–31. Reprinted in Palmer, F. R. (ed.), *Selected Papers of J. R. Firth 1952–59*. pp. 168–205. Longmans, Green, and Co. Ltd., London, 1968.

Gove, Philip, editor. 1984. *Webster's New Dictionary of Synonyms*. Merriam-Webster, Springfield, Mass.

Hayakawa, S. I., editor. 1987. *Choose the Right Word: A Modern Guide to Synonyms*. Harper & Row, New York, perennial library edition.

Hirst, Graeme. 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.

Kittay, Eva Feder. 1987. *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford University Press.

Mel'čuk, Igor and Alain Polguère. 1987. A formal lexicon in meaning-text theory (or how to do lexica with words). *Computational Linguistics*, 13(3–4):261–275.

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

Ruhl, Charles. 1989. *On Monosemy: A Study in Linguistic Semantics*. State University of New York Press, Albany.

Saint-Dizier, Patrick and Evelyn Viegas, editors. 1995. *Computational Lexical Semantics*. Cambridge University Press, Cambridge, England.

Saussure, Ferdinand de. 1983. *Course in General Linguistics*. G. Duckworth, London. Translation by Roy Harris of *Cours de linguistique générale*.

Sinclair, John McH. 1966. Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins, editors, *In Memory of J. R. Firth*. Longmans, Green, and Co. Ltd., London, pages 410–430.

Stede, Manfred. 1996. *Lexical Semantics and Knowledge Representation in MultiLingual Sentence Generation*. Ph.D. thesis, University of Toronto.