"I Don't Believe in Word Senses"

1 author:

Adam Kilgarriff
Lexical Computing Ltd
142 PUBLICATIONS **5,190** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Kelly - KEywords for Language Learning for Young and adults alike View project

"I don't believe in word senses"*

Adam Kilgarriff
ITRI
University of Brighton
Lewes Road
Brighton BN2 4GJ

May 18, 2003

Abstract

Word sense disambiguation assumes word senses. Within the lexicography and linguistics literature, they are known to be very slippery entities. The paper looks at problems with existing accounts of 'word sense' and describes the various kinds of ways in which a word's meaning can deviate from its core meaning. An analysis is presented in which word senses are abstractions from clusters of corpus citations, in accordance with current lexicographic practice. The corpus citations, not the word senses, are the basic objects in the ontology. The corpus citations will be clustered into senses according to the purposes of whoever or whatever does the clustering. In the absence of such purposes, word senses do not exist.

Word sense disambiguation also needs a set of word senses to disambiguate between. In most recent work, the set has been taken from a general-purpose lexical resource, with the assumption that the lexical resource describes the word senses of English/French/..., between which NLP applications will need to disambiguate. The implication of the paper is, by contrast, that word senses exist only relative to a task.

1   Introduction

There is now a substantial literature on the problem of word sense disambiguation (WSD). The goal of WSD research is generally taken to be disambiguation between the senses given in a dictionary, thesaurus or similar. The idea is simple enough and could be stated as follows:

Many words have more than one meaning. When a person understands a sentence with an ambiguous word in it, that understanding is built on the basis of just one

---

*Sue Atkins —Past President, European Association for Lexicography; General Editor, Collins-Robert English/French Dictionary; Lexicographical Adviser, Oxford University Press— responding to a discussion which assumed discrete and disjoint word senses, at 'The Future of the Dictionary' workshop, Uriage-les-Bains, October 1994.

of the meanings. So, as some part of the human language understanding process, the appropriate meaning has been chosen from the range of possibilities.

Stated in this way, it would seem that WSD might be a well-defined task, undertaken by a particular module within the human language processor. This module could then be modelled computationally in a WSD program, and this program, performing, as it did, one of the essential functions of the human language processor, would stand alongside a parser as a crucial component of a broad range of NLP applications. This point of view is clearly represented in Cottrell (1989):

> [Lexical ambiguity] is perhaps the most important problem facing an NLU system. Given that the goal of NLU is understanding, correctly determining the meanings of the words used is fundamental. ... The tack taken here is that it is important to understand how people resolve the ambiguity problem, since whatever their approach, it appears to work rather well. (p 1)

Word meaning is of course a venerable philosophical topic, and questions of the relation between the signifier and the signified will never be far from the theme of the paper. However, philosophical discussions have not addressed the fact of lexicography and the theoretical issues raised by sense distinctions as marked in dictionaries. We often have strong intuitions about words having multiple meanings, and lexicography aims to capture them, systematically and consistently. The philosophy literature does not provide a taxonomy of the processes underpinning the intuition, nor does it analyse the relations between the word sense distinctions a dictionary makes and the primary data of naturally-occurring language. This is a gap that this paper aims to fill.

I show, first, that Cottrell's construal of word senses is at odds with theoretical work on the lexicon (section 2); then, that the various attempts to provide the concept 'word sense' with secure foundations over the last thirty years have all been unsuccessful (section 3). I then consider the lexicographers' understanding of what they are doing when they make decisions about a word's senses, and develop an alternative conception of the word sense, in which it

2

corresponds to a cluster of citations for a word (section 4). Citations are clustered together where they exhibit similar patterning and meaning. The various possible relations between a word's meaning potential and its dictionary senses are catalogued and illustrated with corpus evidence.

The implication for WSD is that there is no reason to expect a single set of word senses to be appropriate for different NLP applications. Different corpora, and different purposes, will lead to different senses. In particular, the sets of word senses presented in different dictionaries and thesauri have been prepared, for various purposes, for various human users: there is no reason to believe those sets are appropriate for any NLP application.

2 Thesis and antithesis: practical WSD and theoretical lexicology

2.1 Thesis

NLP has stumbled into word sense ambiguity.

Within the overall shape of a natural language understanding system –morphological analysis, parsing, semantic and pragmatic interpretation– word sense ambiguity first features as an irritation. It does not appear as a matter of particular linguistic interest, and can be avoided altogether simply by treating all words as having just one meaning. Rather, it is a snag: if you have both river *bank* and money *bank* in your lexicon, when you see the word *bank* in an input text you are at risk of selecting the wrong one. There is a practical problem to be solved, and since Margaret Masterman's group started examining it in the 1950s (see, e.g., Sparck Jones (1986)), people have been writing programs to solve it.

NLP has not found it easy to give a very principled answer to the question, "what goes in the lexicon". Before the mid-1980s, many systems made no claims to wide coverage and contained only as many words in the lexicon as were needed for the 'toy' texts that were going to be analysed. A word was only made ambiguous — that is, given multiple lexical entries — if it was one that the researchers had chosen as a subject for the disambiguation study. This was clearly not an approach that was sustainable for wide coverage systems, and

interest developed in dictionaries, as relatively principled, wide-coverage sources of lexical information.

As machine-readable versions of dictionaries started to become available, so it became possible to write experimental WSD programs on the basis of the dictionary's verdict as to what a word's senses were (Lesk, 1986; Jensen and Binot, 1987; Slator, 1988; Veronis and Ide, 1990; Guthrie et al., 1990; Guthrie et al., 1991; Dolan, 1994). Looked at the other way round, WSD was one of the interesting things you might be able to do with these exciting new resources.

Since then, with the advent of language corpora and the rapid growth of statistical work in NLP, the number of possibilities for how you might go about WSD has mushroomed, as has the quantity of work on the subject (Brown et al., 1991; Hearst, 1991; McRoy, 1992; Gale, Church, and Yarowsky, 1992; Yarowsky, 1992; Gale, Church, and Yarowsky, 1993). Clear (1994), Schütze and Pederson (1995) and Yarowsky (1995) are of particular interest because of their approach to the issue of the set of word senses to be disambiguated between. Schütze and Pederson devised high-dimensionality vectors to describe the context of each occurrence of their target word, and then clustered these vectors. They claim that the better-defined of these clusters correspond to word senses, so a new occurrence of the word can be disambiguated by representing its context as a vector and identifying which cluster centroid the vector is closest to. This system has the characteristic that a context may be close to more than one cluster centroid, so at times it may be appropriate to classify it as more than one sense.

Both Clear (1994) and Yarowsky (1995) provide a mechanism for the user to input the senses between which they would like the system to disambiguate. They ask the user to classify a small number of statistically-selected 'seed' collocates, so the user determines the senses to be disambiguated between when deciding on the senses he or she will assign seed collocates to.[1] Clear then finds all the words which tend to co-occur with the node word in a large corpus, and quantifies, for a very large number of words, the evidence that it occurs with each of the seeds, and thus indirectly, with each sense of the nodeword. Disambiguation

4

then proceeds by summing the evidence for each sense provided by each context word.

Yarowsky's method is iterative: first, those corpus lines for the nodeword which contain one of the seed collocates are classified. Then the set of corpus lines so classified is examined for further indicators of one or other of the senses of the word. These indicators are sorted, according to the strength of evidence they provide for a sense. It will now be possible to classify a larger set of corpus lines, so producing more indicators for each sense, and the process can be continued until all, or an above-threshold proportion, of the corpus lines for the word are classified. The ordered list of sense-indicators will then serve as a disambiguator for new corpus lines.

In the Semantic Concordance project at Princeton a lexicographic team has been assigning a WordNet (Miller, 1990) sense to each noun, verb, adjective and adverb in a number of texts, thus providing a 'gold standard' disambiguated corpus which can be used for training and evaluating WSD programs (Landes, Leacock, and Tengi, 1996).

In 1994-95, there was an extended discussion of whether WSD should be one of the tasks in the MUC program.[2] This would have provided for competitive evaluation of different NLP groups' success at the WSD task, as measured against a 'benchmark' corpus, in which each word had been manually tagged with the appropriate WordNet sense number (as in the Semantic Concordance). Some trials took place, but the decision was not to proceed with the WSD task as part of the 1996 MUC6 evaluation, as there was insufficient time to debate and define detailed policies. The theme has recently been taken up by the Lexicons Special Interest Group of the Association for Computational Linguistics, and a pilot evaluation exercise is taking place in 1998: a milestone on the road from research to technology.

2.2   Antithesis

Since the publication of *Metaphors We Live By* (Lakoff and Johnson, 1980) and *Women, Fire and Dangerous Things* (Lakoff, 1987), there has been one approach to linguistics – cognitive linguistics – for which metaphor has been a central phenomenon. Metaphor is, amongst other things, a process whereby words spawn additional meanings, and cognitive linguists

are correspondingly interested in polysemy. Lakoff's analysis of the polysemy of *mother* is hugely cited. Word sense ambiguity can often be seen as a trace of the fundamental processes underlying language understanding (Sweetser, 1990). The structures underlying the distinct meanings of words are at the heart of the cognitive linguistics enterprise (Geeraerts, 1990; Taylor, 1989).

Working in this framework, Cruse (1995) gives a detailed typology of polysemy. He distinguishes polysemy, defined according to distinctness of meaning, from polylexy, which is where, in addition to distinctness of meaning, distinct lexical entries are required. A word is polysemous but not polylexic where its non-base meanings are predictable, so they can be generated as required and need not be stored. He also addresses where readings are antagonistic and where they are not, and the characteristics of the different semantic properties, or 'facets', of a sense. He uses ambiguity tests to tease out a number of issues, and a full Cruse lexical entry would contain: a specification of polysemous senses; their lexical relations including their relations to each other; whether they were antagnostic or not; the facets, shared or otherwise, of each, and the extent to which distinct facets of meaning could operate autonomously, so approach the status of senses on their own. He considers several varieties of 'semi-distinct' readings.

Lexical ambiguity has also moved centre-stage within theoretical and computational linguistics. Both AAAI and ACL have recently devoted workshops to the topic.[3] When Pustejovsky and others discuss the generative lexicon (Pustejovsky, 1991; Briscoe, Copestake, and Boguraev, 1990), the generative processes they have in mind are, again, ones whereby words spawn additional meanings (or, at least, additional uses). Regular polysemy (Apresjan, 1974) has recently been discussed, and computational mechanisms for addressing it proposed, by Ostler and Atkins (1991), Lyons (1995) and Copestake and Briscoe (1995), *inter alia.* Levin and colleagues have also been finding systematicity in lexical ambiguity, in relation to verb classes, their patterns of subcategorisation, and their patterns of alternation (Levin and Rappoport Hovav, 1991; Levin, 1993; Levin, Song, and Atkins, 1997).

This combination of circumstances leads to an odd situation. Much WSD work proceeds

on the basis of there being a computationally relevant, or useful, or interesting, set of word senses in the language, approximating to those stated in a dictionary. To the WSD community, word senses are, more or less, as the dictionary says.[4] (This is not, of course, to say that WSD authors have not noted the theoretical problems associated with dictionary's word senses.) WSD research has gone a long way on this basis: it is now common for papers to present quantitative comparisons between the performance of different systems. Meanwhile, the theoreticians provide various kinds of reason to believe there is no such set of senses. To get beyond this impasse, we need to look more closely at the question, "what is a word sense?".

3 What is a word sense?

> No entity without identity. (Quine, 1969)

Or, to know what something is, is to know when something is it. To know what a word sense s_1 is, is to know which uses of the word are part of s_1 and which are not, probably because they are part of s_i where $i \neq 1$. If we are to know what word senses are, we need operational criteria for distinguishing them.

3.1 Selection and modulation

A good starting point is Cruse's textbook on Lexical Semantics (Cruse, 1986). 'Lexical units' are the object of his enquiry, and he devotes two substantial chapters to specifying what they are. He states the heart of the problem thus:

> One of the basic problems of lexical semantics is the apparent multiplicity of semantic uses of a single word form (without grammatical difference).

He addresses in some detail the difference between those cases where the context **selects** a distinct unit of sense, from those where it **modulates** the meaning. In the pair

> Have you put the money in the bank?
> The rabbit climbed up the bank.

the two sentences select different meanings of *bank*, whereas in

> He doesn't often oil his bike.
>
> Madeleine dried off her bike.
>
> Boris's bike goes like the wind.

different aspects of the bicycle —its mechanical parts; its frame, saddle and other large surfaces; its (and its rider's) motion— are highlighted in each case. The meaning of *bike* is modulated differently by each context.[5]

3.2 Ambiguity tests

The selection/modulation distinction is closely related to the distinction between ambiguity and generality, also referred to as 'vagueness', 'indeterminacy' and 'lack of specification'.[6] Where a word is ambiguous, a sense is selected. Where a word-meaning is general between two readings, any particular context may or may not modulate the word-meaning to specify one or other of the readings. Thus, *hand* is unspecified between right hands and left hands; some sentences modulate the meaning to specify a right or left hand, as in "When saluting, the hand should just touch the forehead", while others do not.[7]

Clearly, *bank* is ambiguous between the readings demonstrated above; *bike* is not. But for many reading-pairs, the answer is not clear:[8]

- I planted out three rows of beans yesterday.
  Cook the beans in salted water.

- The cottage was charming.
  Our hosts were charming.

- Bother! I was about to talk to John, but now he's disappeared! (NOT-HERE)
  I can't find it anywhere, it seems to have disappeared. (CAN'T-FIND)

A number of tests have been proposed for determining whether a word is ambiguous or general between two meanings. They are catalogued in Zwicky and Sadock (1975), Cruse

(1986), and Geeraerts (1993). Here, I shall describe only one of the more successful tests, the 'crossed readings' one.

Mary arrived with a pike and so did Agnes.

could mean that each arrived with a carnivorous fish, or that each arrived bearing a long-handled medieval weapon, but not that the one arrived with the fish and the other with the weapon. On the other hand, in

Tom raised his hand and so did Dick.

each might have raised a right hand, each might have raised a left, or one might have raised his right, and the other, his left. The question now is, in

Ellen bought some beans, and so did Harry.

is it possible that Ellen bought plants and Harry, food? If so, then the conclusion to be drawn from the test is that *bean* is ambiguous between the readings, and if not, then it is not.[9]

3.2.1 Criticisms of the tests

The tests are generally presented with the aid of an unproblematical example of ambiguity and an unproblematical example of vagueness. This is done in order to demonstrate what the test is and what the two contrasting outcomes are. However, this is not to use the tests in anger. What we want of a test is that it is consistent with our intuitions, where our intuitions are clear, and that it resolves the question, where our intuitions are unclear. The cross-reading test fares tolerably well in meeting the consistency condition (though see (Geeraerts, 1993) for a contrary view). But do the tests help where intuitions are unclear? There is little if any evidence that they do. Here I discuss three classes of problems.

Firstly, it must be possible to construct a plausible test sentence. The word in its two uses must be able to occur with the same syntax and the same lexico-grammatical environment. Consider the transitive and intransitive uses of *eat*, as in "John ate the apple" and "John ate". Is this a case of ambiguity or vagueness?

9

*Mary ate, and John, the apple.

is unacceptable, but the reason is that elided constituents must have the same syntax and subcategorisation in both their expressed and elided occurrences. It might be desirable to treat all words with alternative subcategorisation possibilities as ambiguous. But whether or not that is done, the test still fails to elucidate on the topic of a word's meaning, where the word has different syntax in different uses. The test can only be posed where the two uses are syntactically similar.

The *disappear* example displays a different variant of this problem. The CAN'T-FIND and NOT-HERE readings have different aspectual characteristics: CAN'T-FIND is stative while NOT-HERE is a punctual 'achievement' verb.

Martha disappeared and so did Maud.

does not permit a crossed reading, but that is because we cannot construct a viable aspectual interpretation for the conjoined sentence, compare

? I evicted and knew her.[10]

It is not evident whether there is a conclusion to be drawn regarding polysemy.

In general, one can apply more or less effort into trying to find a test sentence (and associated context) in which the crossed reading is plausible. A test is clearly flawed, if, the more ingenuity the investigator displays, the more of one particular outcome he or she will get. (The crossed reading test is the test which suffers least from this flaw, but it is nonetheless in evidence.)

The second point is more general and theoretical. A certain amount of interpretation of an utterance must have been undertaken before an acceptability judgement can be made. Three parts of the interpretation process are lexical access, parsing, and 'pragmatic interpretation', the final stage of incorporating the new information into the discourse model. The premise behind acceptability judgements is that a subject can report on the outcome of the first two stages, irrespective of what goes on in the third. For a wide range of syntactic questions, the methodology is widely used and has proved its worth.

Nunberg's (1978) arguments illustrate the hazards of the premise for questions in lexical semantics. Consider

The newspaper costs 25p and sacked all its staff.

It is anomalous. We cannot place the origin of the anomaly in the lexicon unless we grant the word two lexical entries, one for a copy of the newspaper and one for the owner or corporate entity. Then the size of our lexicon will start to expand, as we list more and more of the possible kinds of referent for the word, and still it will never be complete. So the origin of the anomaly must be the interpretation process. But the anomaly seems similar to the anomaly that occurs with *bank*. In a case lying between *newspaper* and *bank*, how would we know whether the source of the anomaly was the lexicon or the interpretation process? In the general case the point at which the lexical process becomes a general-purpose interpretative one cannot be identified. There is no accessible intermediate representation in which lexical ambiguities are resolved (for acceptable sentences) but in which the contents of the sentence has not been incorporated into the hearer's interpretation of the discourse. Geeraerts (1993) presents an extensive critique of the tests along these lines, presenting evidence that the different tests give contradictory results, and that even if we constrain ourselves to looking at just one of the tests, they can all be made to give contradictory results by manipulating the context in which the item under scrutiny is set.

The third problem is simply the lack of evidence that the tests give stable results. It will sometimes happen that, for the same reading-pair, an informant will deem crossed readings possible for some test sentences and not for others. Or different informants will have conflicting opinions. There are, remarkably, no careful discussions of these issues in the literature. The merit of the method of acceptability judgements for syntax rests on the relative stability of their outcomes: they work (to the extent they do) because linguists agree where the stars belong. Preliminary investigations into the stability of outcomes in lexical semantics suggest that it is severely lacking.

11

3.3 Psycholinguistics and 'semantic priming'

There is a set of findings in psycholinguistics which might allow us to base an account of 'word sense' directly on the mental lexicon. The experimental paradigm is called 'semantic priming'. It is well-established that, if I have just heard the word *doctor* (the 'prime'), and then a sequence of letters (the 'target') is flashed up on a screen and I am asked to identify whether it is a word or not, I respond faster if it is a word and it is *nurse* than if it is a word but unrelated to *doctor*.[11]

If an ambiguous prime such as *bank* is given, it turns out that both *river* and *money* are primed for. If *bank* is presented in isolation, priming for both *river* and *money* is found for another second or two. In a context which serves to make only one of these appropriate, after something between 50 and 200 ms a choice is made and after that only the appropriate target is primed for.

So, for ambiguous words, priming behaviour has a distinct 'signature'. Perhaps it is possible to identify whether a word is vague or ambiguous by seeing whether it exhibits this signature.

The hypothesis is explored by Williams (1992). He looked at adjectives, for example *firm*, for which the two readings were represented by *solid* and *strict*. After confirming that the prime, *firm*, in isolation, primed equally for *solid* and *strict*, he tested to see if *solid* was primed for when *firm* occurred in a STRICT context, and *vice versa*, after delays of 250, 500 and 850 ms.

His results were asymmetrical. He identified central meanings (SOLID) and non-central ones (STRICT). Where the context favoured the central reading, the non-central-sense targets were not primed for. But when the context favoured the non-central reading, central targets were. The experiments provide evidence that the various meanings of polysemous words are not functionally independent in language comprehension, and that not all senses are equal, in their representation in the mental lexicon. Williams discusses the asymmetrical results in terms of hierarchical meaning structures.

Priming experiments do show potential for providing a theoretical grounding for distinguishing ambiguity and generality, but more work needs to be done, and the outcome would not be a simple, two-way, ambiguous/general distinction. Also, the method would never be practical for determining the numbers of senses for a substantial number of words. The results of the experiments are just not sufficiently stable: as Williams says, the priming task "suffers from a large degree of item and subject variability" (p 202).

4   Lexicographers, dictionaries, and authority

> What set of procedures do lexicographers have available to them to pin down those
> protean entities, 'meanings'? Faced with the almost unimaginable diversity of the
> language they are trying to describe, with the knowledge that what for the sake of
> convenience we are pleased to call a language is in many ways a synthesis of shifting
> patterns that change from year to year, from locality to locality, from idiolect to
> idiolect, how do they arrive at those masterpieces of consensus, dictionaries? How
> do they decide what, for the purposes of a dictionary, constitutes the meaning of
> a word, and where, in the case of polysemous words, one meaning ends and the
> next begins? (Ayto, 1983, p 89)

In the middle of this debate stand the lexicographers. The word senses that most WSD researchers aim to discriminate are the product of their intellectual labours. But this is far from the purpose for which the dictionary was written.

Firstly, any working lexicographer is well aware that, every day, they are making decisions on whether to 'lump' or 'split' senses that are inevitably subjective:[12] frequently, the alternative decision would have been equally valid. In fact, most dictionaries encode a variety of relations in the grey area between "same sense" and "different sense": see Kilgarriff (1993) for a description of the seven methods used in (LDOCE, 1987).

Secondly, any particular dictionary is written with a particular target audience in mind, and with a particular editorial philosophy in relation to debates such as 'lumping *vs.* splitting',

so the notion of specifying a set of word senses for a language in isolation from any particular user group will be alien to them.

Thirdly, many are aware of the issues raised by Lakoff, Levin, Pustejovsky and others, with several lexicographers bringing valuable experience of the difficulties of sense-division to that literature (see below).

Fourthly, the weight of history: publishers expect to publish, bookshops expect to sell, and buyers expect to buy and use dictionaries which, for each word, provide a (possibly nested) list of possible meanings or uses. Large sums of money are invested in lexicographic projects, on the basis that a dictionary has the potential to sell hundreds of thousands of copies. Investors will not lightly adopt policies which make their product radically different to the one known to sell. However inappropriate the nested list might be as a representation of the facts about a word, for all but the most adventurous lexicographic projects, nothing else is possible.[13]

The division of a word's meaning into senses is forced onto lexicographers by the economic and cultural setting within which they work. Lexicographers are obliged to describe words as if all words had a discrete, non-overlapping set of senses. It does not follow that they do, nor that lexicographers believe that they do.

4.1 Lexicographical literature

Lexicographers write dictionaries rather than writing about writing dictionaries. Little has been written that answers the challenge posed by Ayto in the quotation above. Zgusta's influential *Manual* (1971), while stating that the specification of word meaning is the central task for the lexicographer (p 23) and the division of a word's meanings into senses is a central part of that, gives little guidance beyond admonishments to avoid making too many, or too few, distinctions (pp 66–67).

Ayto's own offering in the 1983 paper is the 'classical' or 'analytic' definition, comprising genus and differentiae. In choosing the genus term, the lexicographer must take care to neither select one that is too general —*entity* would not do as a genus term for *tiger*— nor too specific,

if the specific genus term is likely to be unknown by the dictionary users. Where two meanings of a word have different genus terms, they need treating as different senses. The next task is to identify the differentiae required to separate out senses falling under the same genus term. He discusses *cup*, and argues that there are three senses, one for the 'trophy' sense, one for the varieties standardly made of china or earthenware, and one for the prototypically plastic or paper varieties. But his consideration of the arguments for treating the second and third of these as distinct ends in a welter of open questions.

Stock (1983) is a response to Ayto's piece, and finds it wanting, firstly, in the circularity involved in using different genus terms to identify distinct senses —the lexicographer will only look for distinct genus terms after determining there are distinct senses— and secondly, in that the model cannot be applied to many words. She looks closely at *culture*, noting how different dictionaries have divided the territory that the word covers in quite different ways, and observes,

> It is precisely the lack of clarity in our use of the word *culture* which makes it such a handy word to have at one's disposal. It offers, as it were, semantic extras just because in most uses its possible meanings are not clearly disambiguated. ...What can the dictionary maker do to reflect this state of affairs? ...They do not, cannot by their very structure, show that there is slippage between some of the senses that they give but not between others. (p. 139)

Hanks (1994), looking at *climb*, and Fillmore and Atkins (1992), studying the semantic field centred on *risk*, make similar comments about the inadequacies of dictionary conventions, and appeal to prototype theory and frame semantics for richer frameworks to describe the relationships between the different ways a word (or word-family) is used.

Stock, Hanks and Atkins were all involved in the early stages of the COBUILD project, which, in the early 1980s, broke new ground in lexicography through its use of very large computerised language corpora (Sinclair, 1987). Good lexicographic practice had long used huge citation indexes, but being able to see hundreds of instances of a word in context, ordi-

nary and extraordinary examples thrown together, was a radical development. It has changed how lexicographers think about meaning. Where Ayto's paper offers semantic analysis, Stock presents corpus evidence. The lexicographer's primary source of evidence for how a word behaves switches from subjective to objective; from introspection to looking at contexts.

4.2 A corpus-based model of word senses

This suggests a quite different answer to the question, "what is a word sense?" Corpus lexicography proceeds approximately as follows. For each word, the lexicographer

1. calls up a concordance[14] for the word;

2. divides the concordance lines into clusters, so that, as far as possible, all members of each cluster have much in common with each other, and little in common with members of other clusters;

3. for each cluster, works out what it is that makes its members belong together, re-organising clusters as necessary;

4. takes these conclusions and codes them in the highly constrained language of a dictionary definition.

Putting the concordance lines into clusters is data-driven rather than theory-driven. The lexicographer may or may not be explicitly aware of the criteria according to which he or she is clustering.[15] (It is a requirement for corpus lexicography software that it supports manual clustering (Atkins, 1993; CorpusBench, 1993; Schulze and Christ, 1994).) Stage 3 is just a fallible *post hoc* attempt to make the criteria explicit. The senses that eventually appear in the dictionary are the result, at several removes, of the basic clustering process.

Ambiguity tests failed to provide us with an account of what it meant for two uses of a word to belong to the same word sense. Once we operationalise 'word sense' as 'dictionary word sense', we now have a test that meets the challenge. The identity test for a word sense in a particular dictionary is that two usages of the word belong to it if and only if the lexicographer would have put them in the same cluster.[16]

We can now present a different perspective on the ambiguity/generality debate. Where a word's uses fall into two entirely distinct clusters, it is ambiguous, but where the clusters are less well-defined and distinct, 'vague' or 'unspecified' may be a more appropriate description. There is no reason to expect to find any clear distinction between the two types of cases.

5   Use, frequency, predictability, and the word sense

'Clustering' is a metaphor. It regards corpus lines as points in space with measurable distances between them. To give the account substance, more must be said about the ways in which corpus lines may be 'close'. In this section, I classify the types of relationships that hold between a word's patterns of usage, and consider how these considerations relate to lexicography.[17]

There are five knowledge sources which come into play for understanding how a word contributes to the meaning or communicative intent of the utterance or discourse it occurs in. If a word in context is interpretable by a language user, it will be by virtue of these knowledge sources.

Whether a dictionary provides a word sense that matches an instance of use of the word, is dictated by considerations of frequency and predictability: if the instance exemplifies a pattern of use which is sufficiently frequent, and is insufficiently predictable from other meanings or uses of the word, then the pattern qualifies for treatment as a dictionary sense. A use is predictable, to the extent that a person reading or hearing it for the first time can understand it (in all its connotations). Clearly, different dictionaries have different thresholds of frequency and predictability.

To illustrate the various processes whereby new types of usage may be added to the repertoire for a word, let us consider the simple single-sense word, *handbag*

> a small bag, used by women to carry money and personal things (British; American
> English translation: purse)(LDOCE3)

As the 715 examples in the British National Corpus (BNC)[18] make plain, typical uses involve

things being put into, or taken out of, or looked for in handbags, or handbags being lost, found, stolen, manufactured, admired, bought or sold. But a couple of dozen examples stretch the limits of the definition or fall outside it altogether.

First, a proper name, and a reference to a unique object:

> the Drowning Handbag, an up-market eatery in the best part of town
> an inimitable rendering of the handbag speech in The Importance of Being Earnest

Next, metonymy, visual metaphor, simile:

> She moved from handbags through gifts to the flower shop
> "How about you? Did the bouncing handbag find you?"[19]
> a weird, menacing building with bats hanging in the trees like handbags
> Skin generally starting to age like old handbag or bodywork of car

Next, Mrs Thatcher:

> from Edward Heath's hip-pocket to Margaret Thatcher's handbag and on to Mr Major's glass
>   of warm beer
> "Thousands . . . will be disgusted at the way she [Thatcher] is lining her handbag"
> send out Mrs Thatcher with a fully-loaded handbag
> "If you want to define the Thatcher-and-after era in a single phrase", he muses, " 'accountants
>   with plenary powers' says it." Well now– I would have gone for something a little snappier:
>   'A mad cow with a handbag,' comes to mind as a first attempt.
> She [Thatcher] cannot see an institution without hitting it with her handbag.

The last of these is cited in another citation as the launching-point of verbal *handbag*. Of the three verbal citations, all were species of hitting and in two of them, Mrs. Thatcher was the perpetrator.

Next, and closely related to Mrs. Thatcher, 'handbag as weapon':

> Meg swung her handbag.
> determined women armed with heavy handbags
> it was time to race the old ladies back to the village for the tea and scones of Beck Hall. I
>   beat them, but only just– those handbags are lethal.
> old ladies continue to brandish their handbags and umbrellas at the likes of Giant Haystacks
> the blue rinse brigade . . . will be able to turn out in force without having to travel and give
>   poor Louis Gerstner the handbagging of his life.
> Peterborough manager Chris Turner added: "Evidently one of their players caught one of our
>   players and it was handbags at 10 paces and then someone threw a punch."

The final, quite distinct group relates to discos, and the lexical unit *dance round your handbag*, a pejorative phrase for the behaviour of certain exclusively female groups at discotheques and dances where —prototypically— they dance in a circle with their handbags on the floor

in the middle. The conversational speech subcorpus of the BNC provides two instances of the full form while in the written corpus, the two related corpus lines, both from music journalism, make only fleeting references to the collocation, and strikingly indicate a process of lexicalisation:

> The shoot was supposed to be a secret, but word got out and Hitman regulars travelled down to Manchester. Two thousand couldn't get into the club, and tension mounted between trendy regulars (locked out of their own club) and the Hitman's handbag brigade (shut out of their programme).
> New Yawk drawling rap over Kraftwerk's 'The Model' just does not work, no way, no how. Handbag DJs will love it.

All these uses can be traced back to the standard sense: the potential for using the word in the nonstandard way, is (in varying degrees) **predictable** from

- its standard meaning and use

- general linguistic knowledge (eg. of processes of metonymy, regular polysemy, and ellipsis, etc., and, in this case, the relation between words for goods and words for shops or departments of shops where those goods are sold),

- general world knowledge (eg. regarding Mrs. Thatcher, or juvenile female behaviour at discotheques) and

- knowledge of related collocations (eg. "lining their pockets", "WEAPON at NUMBER paces")

- taxonomic knowledge

These five knowledge sources define the conceptual space within which lexical creativity and productivity, and the idea of a 'word sense', are located.[20]

Needless to say, they frequently interact in complex ways. In "handbags at ten paces", the speaker[21] assumes the addressee's awareness of handbag-as-weapon. Note that "*briefcases at ten paces" and "*shoulder-bags at ten paces" do not carry the same meaning. Although briefcases and shoulder-bags are just as viable weapons as handbags, the words *briefcase* and *shoulder-bag* do not carry the 'weapon' connotations which make the citation immediately

understandable. Handbag-as-weapon is a feature of the word, over and above the extent to which it is a feature of the denotation.

In the citation's context, there is no overt reason for a reference to *handbag*; the people involved are men, not women, so not prototypical handbag-users, and there is no other reference to femininity. It would appear that the speaker is aiming to both distance himself from and minimise the significance of the incident by treating it as a joke. The 'duel' metaphor is itself a joke, and the oddity of *handbag* in the context of either football or duel, along with its associations with femininity and Mrs. Thatcher, contributes to the effect. Moreover, there is a sexist implication that the men were behaving like women and thereby the matter is laughable.

Interpreting "handbags at ten paces" requires lexical knowledge of "handbag-as-weapon", collocational knowledge of both form and meaning of "WEAPON at NUMBER paces", and (arguably) knowledge of the association between handbags and models of masculinity and femininity.

The 'music journalism' use displays some further features. *Handbag* was lexicalised in the clubbing world in ca. 1990 as a music genre: the genre that, in the 1970s and 1980s, certain classes of young women would have danced round their handbags to.[22] The coinage emanates from the gay and transvestite club scene and is redolent with implications, from the appropriation of the handbag as a symbol of gay pride, to changes in the social situation of women over the last twenty years (and its expression in fashion accessories), to transvestite fantasies of being naive seventeen-year-old girls in a more innocent age.

To restrain ourselves to more narrowly linguistic matters: the license for the coinage is via the "dance round your handbag" collocation, not directly from handbags. As shown by the spoken corpus evidence, the regular, non-ironic use of the collocation co-exists with the music-genre use. It is of much wider currency: all but two of a range of informants knew the collocation, whereas only two had any recollection of the music-genre use. Also, 'handbag' music (or at least the use of that label) was a 1990-91 fashion, and the term is no longer current: 1996 uses of it will probably refer back to 1990-91 (as well as back to the 1970s and

1980s).

Syntactically, the most information-rich word of the collocation has been used as a nominal premodifier for other nouns: in the music-genre sense, it is used as other music-genre words, as an uncountable singular noun, usually premodifying but potentially occurring on its own: "Do you like jazz/house/handbag?"

5.1 Frequency

These arguments make clear that there is a *prima facie* case for including handbag-as-weapon and handbag-as-music-genre as dictionaries senses, and "dance round your handbag" as an only partially compositional collocation. Each exhibits lexical meaning which is not predictable from the base sense. So why do the dictionaries not list them? The short answer is frequency. Around 97% of *handbag* citations in the BNC are straightforward base sense uses. The music-genre sense is certainly rare, possibly already obsolete, and confined to a subculture. The collocation is partially compositional and occurs just twice in the corpus: for any single-volume dictionary, there will not be space for vast numbers of partially compositional collocations. Not only is a lexicographer "a lexicologist with a deadline" (Fillmore, 1988) but also a lexicologist with a page limit.[23]

5.2 Analytic definitions and entailments

The handbag-as-weapon sense is rather more common, and a further consideration comes into play. The denotations of base-sense *handbag* and handbag-as-(potential)-weapon are the same. Correspondingly, the lexical fact that there is a use of *handbag* in which it is conceptualised as a weapon does not render the LDOCE definition untrue. A lexicographer operating according to the classical approach whose goal was simply to provide necessary and sufficient conditions for identifying each word's denotation would say that the 'weapon' aspect of meaning was irrelevant to his or her task. A more pragmatic lexicographer might also follow this line, particularly since space is always at a premium.

The situation is a variant on autohyponymy (Cruse, 1986, pp 63–65), the phenomenon of

one sense being the genus of another sense of the same word. The prototypical example is *dog* (canine *vs.* male canine). *Dog* is a case where there clearly are distinct senses. For *knife* (weapon *vs.* cutlery *vs.* bladed object) Cruse (1995, pp 39–40) argues for "an intermediate status" between monosemy and polysemy, since, on the one hand, 'bladed-object' is a coherent category which covers the denotation, but on the other, in a scenario where there was a penknife but no cutlery knife at a table setting, one might reasonably say "I haven't got a knife". COBUILD2 distinguishes 'weapon' and 'cutlery' senses, while LDOCE3 provides a single, analytically adequate, 'bladed object' sense.

In a discussion of the polysemy of *sanction*, Kjellmer (1993) makes a related observation. His goal is to examine how language breakdown is avoided when a word has antagonistic readings. Nominal *sanction* is such a word: in "sanctions imposed on Iraq" the meaning is akin to punishment ('PUN') whereas in "the proposal was given official sanction" it is related to endorsement ('END'). A first response is that the context disambiguates - punishment, not support, is the sort of thing you "impose", whereas "give" implies, by default, a positively-evaluated thing given. Syntax is also a clue: the plural use is always PUN, whereas determinerless singular uses suggest END. Kjellmer then finds the following instances:

> The process of social control is operative insofar as sanction plays a part in the individual's behaviour, as well as in the group's behaviour. By means of this social control, deviance is either eliminated or somehow made compatible with the function of the social group.
> Historically, religion has also functioned as a tremendous engine of vindication, enforcement, sanction, and perpetuation of various other institutions.

Here the context does not particularly favour either reading against the other. In the second case, the co-ordination with both an END word (*vindication*) and a PUN one (*enforcement*) supports both readings simultaneously. How is this possible, given their antagonism? How come these uses do not result in ambiguity and the potential for misinterpretation? The answer seems to be that,

> we may operate, as readers or listeners, at a general, abstract level and take the word to mean 'control, authority' until the context specifies for us which type of control is intended, if indeed specification is intended. In other words, faced with

the dual semantic potentiality of the word, we normally stay at a higher level of abstraction, where the danger of ambiguity does not exist, until clearly invited to step down into specificity. (p 120)[24]

Citations where *sanction* is unspecified for either PUN or END are rare, and there is no case for including the unspecified 'control' sense in a dictionary.

The example demonstrates a relationship between a lexicographer's analytic defining strategy and the interpretation process. There are occasions where a 'lowest common denominator' of the usually distinct standard uses of a word will be the appropriate reading, in a process analogous to the way an analytically-inclined lexicographer might write a definition for a word like *charming* or *knife*, which would cover the word's uses in two or more distinct corpus clusters. Some dictionaries use nested entries as a means of representing meanings related in this way.

6   Implications for WSD

The argument so far exposes a lack of foundations to the concept of 'word sense'. But, a WSD researcher might say, "so what?" What are the implications for practical work in disambiguation?

The primary implication is that a task-independent set of word senses for a language is not a coherent concept. Word senses are simply undefined unless there is some underlying rationale for clustering, some context which classifies some distinctions as worth making and others as not worth making. For people, homonyms like *pike* are a limiting case: in almost any situation where a person considers it worth their while attending to a sentence containing *pike*, it is also worth their while making the fish/weapon distinction.

Lexicographers are aware of this: the senses they list are selected according to the editorial policy and anticipated users and uses of the particular dictionary they are writing. Until recently, WSD researchers have generally proceeded as if this was not the case: as if a single program —disambiguating, perhaps, in its English-language version, between the senses given

in some hybrid descendant of Merriam-Webster, LDOCE, COMLEX, Roget, OALDCE and WordNet —would be relevant to a wide range of NLP applications.

There is no reason to expect the same set of word senses to be relevant for different tasks.

The handbag data shows how various the non-standard uses of *handbag* are. These uses are sufficiently predictable or insufficiently frequent to be dictionary senses (in a dictionary such as LDOCE). They are licensed by a combination of linguistic principles, knowledge of collocations and lexico-syntactic contexts, and world knowledge. Only in a single case, the department store metonym, is there a plausible linguistic principle for extending the base meaning to render the non-standard use interpretable. The data suggest that little coverage will be gained by an NLP system exploiting generative principles which dictate meaning potential. The non-standard uses of words tend to have their own particular history, with one non-standard use often built on another, the connections being highly specific to a word or lexical field.

The handbag data also indicates how the corpus dictates the word senses. The BNC is designed to cover a wide range of standard English, so is consonant with a general purpose dictionary. The common uses in the one should be the senses in the other. But, were we to move to a music journalism corpus, the music-genre sense would be prominent. A 1990s music-journalism dictionary would include it.

The practical method to extend the coverage of NLP systems to non-standard uses is not to compute new meanings, but to list them. Verbal *handbag* can, if sufficiently frequent, be added to the lexicon as a synonym for *beat*; "WEAPON at NUMBER paces" as one for "have an argument". For the medium term future, the appropriate language-engineering response to a use of a word or phrase which the system needs to interpret but which it is currently misinterpreting because the word or phrase's use does not match that in the lexicon, is to add another lexical entry.[25]

The implications of the account for different varieties of NLP application are addressed in Kilgarriff (1997a, 1997b).

7   Conclusion

Following a description of the conflict between WSD and lexicological research, I examined the concept, 'word sense'. It was not found to be sufficiently well-defined to be a workable basic unit of meaning.

I then presented an account of word meaning in which 'word sense' or 'lexical unit' is not a basic unit. Rather, the basic units are occurrences of the word in context (operationalised as corpus citations). In the simplest case, corpus citations fall into one or more distinct clusters and each of these clusters, if large enough and distinct enough from other clusters, forms a distinct word sense. But many or most cases are not simple, and even for an apparently straightforward common noun with physical objects as denotation, *handbag*, there are a significant number of aberrant citations. The interactions between a word's uses and its senses were explored in some detail. The analysis also charted the potential for lexical creativity.

The implication for WSD is that word senses are only ever defined relative to a set of interests. The set of senses defined by a dictionary may or may not match the set that is relevant for an NLP application.

The scientific study of language should not include word senses as objects in its ontology. Where 'word senses' have a role to play in a scientific vocabulary, they are to be construed as abstractions over clusters of word usages. The non-technical term for ontological commitment is 'belief in', as in "I (don't) believe in ghosts/God/antimatter". One leading lexicographer doesn't believe in word senses. I don't believe in word senses, either.

Notes

[1] Yarowsky's work, this is just one of the options for providing seeds for the process.

[2] The MUC (Message Understanding Conference) is a series of US Government-funded, competitive, quantitatively-evaluated exercises in information extraction (MUC-5, 1994).

[3]The AAAI Spring Symposium on Representation and Acquisition of Lexical Information, Stanford, April 1995 and the ACL SIGLEX Workshop on The Breadth and Depth of Semantic Lexicons, Santa Cruz, June 1996.

[4]Sometimes not all the sense distinctions recognised in the dictionary are viewed as salient to the program. WSD researchers tend to be lumpers, not splitters (Dolan, 1994).

[5]Cruse identifies two major varieties of modulation, of which highlighting is one.

[6]See Zwicky and Sadock (1975) for a fuller discussion of the terms and their sources.

[7]Also related to this distinction is the polysemy/homonymy distinction: when do we have two distinct words, and when, one word with two meanings? Most commentators agree that there is a gradation between the two, with the distinction being of limited theoretical interest. For some purposes, the distinction may be more useful than the vagueness/ambiguity one (Krovetz, 1996). In practice, similar difficulties arise in distinguishing homonymy from polysemy, as in distinguishing vagueness from ambiguity.

[8] The examples are taken by comparing four state-of-the-art English learners' dictionaries (LDOCE, 1995; OALDCE5, 1995; COBUILD, 1995; CIDE, 1995) and finding words where the lexicographers in one team made one decision regarding what the distinct word senses were, whereas those in another made another. This immediately has the effect of introducing various factors which have not been considered in earlier theoretical discussions.

[9]For many putatively ambiguous reading-pairs, there are intermediate cases. A sprouting bean, or one bought for planting, is intermediate between FOOD and PLANT. But the possibility of intermediate cases does not preclude ambiguity: whether two readings of a word are completely disjoint, permitting no intermediate cases, is a different question to whether a word is ambiguous. This imposes a further constraint on ambiguity tests. A speaker might say, "Ellen and Harry must have bought the same kind of *bean*, unless, say, Ellen bought plants and Harry bought beans sold at the supermarket but which he was intending to plant". We should not infer that *bean* is vague. Rather, we must insist that both of the crossed readings are prototypical. (There are of course further difficulties in making this constraint precise).

[10]Eight out of ten informants found the related sentence, "I loved and married her", odd. The two who found it acceptable were reading *and* as an indicator of temporal sequence.

[11]This is the 'lexical decision' task in a mixed, visual and auditory procedure. It is one of a variety of versions of semantic priming experiments. The basic effect is robust across a number of experimental strategies.

[12]**Lumping** is considering two slightly different patterns of usage as a single meaning. **Splitting** is the converse: dividing or separating them into different meanings.

[13]The format of the dictionary has remained fairly stable since Dr. Johnson's day. The reasons for the format, and the reasons it has proved so resistant to change and innovation, are explored at length in (Nunberg, 1994). In short, the development of printed discourse, particularly the new periodicals, in England in the early part of the eighteenth century brought about a re-evaluation of the nature of meaning. No longer could it be assumed that a disagreement or confusion about a word's meaning could be settled face-to-face, and it seemed at the time that the new discourse would only be secure if there was some mutually acceptable authority on what words meant. The resolution to the crisis came in the form of Johnson's Dictionary. Thus, from its inception, the modern dictionary has had a crucial symbolic role: it represents a methodology for resolving questions of meaning. Hence "the dictionary", with its implications of unique reference and authority (cf. "the Bible") (Leech, 1981). Further evidence for this position is to be found in McArthur (1987), for whom the "religious or quasi-religious tinge" (p 38) to reference materials is an enduring theme in their history; Summers (1988), whose research into dictionary use found that "settl[ing] family arguments" was one of its major uses (p 114, cited in Bejoint (1994), p 151); and Moon (1989) who catalogues the use of the UAD (Unidentified Authorising Dictionary) from newspapers letters pages to restaurant advertising materials (pp 60–64).

The implications for ambiguity are this: to solve disputes about meaning, a dictionary must be, above all, clear. It must draw a line around a meaning, so that a use can be classified as on one side of the line or the other. A dictionary which dwells on marginal or vague uses of a word, or which presents its meaning as context-dependent or variable or flexible, will be of little use for purposes of settling arguments. The pressure from this quarter is for the dictionary to present a set of discrete, non-overlapping meanings for a word, each defined by the necessary and sufficient conditions for its application —whatever the facts of the word's usage.

[14]By 'concordance' I mean a display which presents a line of context for each occurrence of the word under scrutiny in the corpus, with all occurrences of the key word aligned. Fuller details are, of course, system specific, but it has rapidly become evident that this kind of display is the basic requirement for any corpus lexicography system.

[15]The interactions between the lexicographers' clusters and the automatic clusters produced for Information Retrieval purposes (Schütze and Pederson, 1995), and the potential for automating some of the clustering that

the lexicographer performs, are subjects of current research.

[16]A psycholinguistic investigation along these lines is presented in Jorgensen (1990).

[17]I do not dwell on cases of simple similarity, where there is a straightforward match between corpus lines, or between a corpus line and a word's core meaning. While it is a major language-engineering problem to operationalise even 'simple similarity', it is not a problematic matter, either theoretically or for lexicographers or other human language users.

[18]For the BNC see `http://info.ox.ac.uk/bnc`. Counts were: *handbag* 609, *handbags* 103, *handbagging* 1, *handbagged* 2.

[19]This turns out to be a (sexist and homophobic) in-group joke, as well as a case of both metonymy and of a distinct idiomatic use of the word. Interestingly, in the text, "the bouncing handbag" succeeds in referring, even though the idiom is not known to the addressee, as is made explicit in the text.

[20]In Kilgarriff (1992), in the context of an analysis of polysemy, I call the first four knowledge types HOMONYMY, ALTERNATION, ANALOGY and COLLOCATION. (Taxonomy is addressed separately.)

[21]This is presented as a quotation of a football manager's spoken comment; quite whether it is verbatim, or the Daily Telegraph journalist's paraphrase, we shall never know.

[22]Thanks to Simon Shurville for sharing his expertise.

[23]It is an interesting question, touched on in Kilgarriff (1993) but worthy of a much fuller investigation, what the percentage of 'anomalous' uses might be for various classes of words. One would expect the figures to be highly corpus-dependent. A large proportion of the BNC is material written by novelists and journalists —who earn their living, in some measure, through their skills in the original and engaging use of language. (The music-genre use of *handbag* probably first occurred in advertising material, probably the most fecund discourse of all.) Also one might expect spoken material to have a higher proportion of set phrases, owing to the time constraints on the production of spoken language.

[24]Kjellmer implies that the further specification is a temporal process, there being a time in the interpretation process when the lexical meaning of the word is accessed but specified for 'control' but not for either PUN or END. I see no grounds for inferring the temporal process from the logical structure.

[25]A well-organised, hierarchical lexicon will mean that this need not introduce redundancy into the lexicon.

References

Apresjan, Juri D. 1974. Regular polysemy. *Linguistics*, 142:5–32.

Atkins, Sue. 1993. Tools for computer-aided lexicography: the Hector project. In *Papers in Computational Lexicography: COMPLEX '93*, Budapest.

Ayto, John R. 1983. On specifying meaning. In R. R. K. Hartmann, editor, *Lexicography: Principles and Practice*. Academic Press, London, pages 89–98.

Béjoint, Henri. 1994. *Tradition and Innovation in Modern English Dictionaries*. OUP, Oxford.

Briscoe, Edward J., Ann A. Copestake, and Branimir K. Boguraev. 1990. Enjoy the paper: Lexical semantics via lexicology. In *COLING 90*, volume 2, pages 42–47, Helsinki.

Brown, Peter, Stephen Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991.

Word sense disambiguation using statistical methods. In *Annual Meeting of the Association of Computational Linguistics*, Berkeley, California.

CIDE, 1995. *Cambridge International Dictionary of English*. CUP, Cambridge, England.

Clear, Jeremy. 1994. I can't see the sense in a large corpus. In Ferenc Kiefer, Gabor Kiss, and Julia Pajzs, editors, *Papers in Computational Lexicography: COMPLEX '94*, pages 33–48, Budapest.

COBUILD, 1995. *The Collins COBUILD English Language Dictionary. 2nd Edition.* Edited by John McH. Sinclair *et al.* London.

Copestake, Ann A. and Edward J. Briscoe. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics*, forthcoming.

CorpusBench. 1993. *CorpusBench Manual*. Textware A/S, Copenhagen, Denmark.

Cottrell, Garrison W. 1989. *A Connectionist Approach to Word Sense Disambiguation*. Pitman, London.

Cruse, Alan. 1995. Polysemy and related phenomena from a cognitive linguistic viewpoint. In Patrick St. Dizier and Evelyne Viegas, editors, *Computational Lexical Semantics*. CUP, Cambridge, England, pages 33–49.

Cruse, D. A. 1986. *Lexical Semantics*. CUP, Cambridge, England.

Dolan, William B. 1994. Word sense ambiguation: clustering related senses. In *COLING 94*, Tokyo.

Fillmore, Charles. 1988. Keynote lecture, british assn. applied linguistics conference, exeter, September.

Fillmore, Charles J. and Beryl T. S. Atkins. 1992. Towards a frame-based lexicon: the semantics of RISK and its neighbours. In Adrienne Lehrer and Eva Kittay, editors, *Frames, Fields and Contrasts*. Lawrence Erlbaum, New Jersey, pages 75–102.

Gale, William, Kenneth Church, and David Yarowsky. 1992. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings, 30th ACL*, pages 249–156.

Gale, William, Kenneth Church, and David Yarowsky. 1993. A method for disambiguating word senses in a large corpus. *Computers and the Humanities*, 26(1–2):415–439.

Geeraerts, Dirk. 1990. The lexicographical treatment of prototypical polysemy. In Savas L. Tsohatzidis, editor, *Meanings and Prototypes: Studies in Linguistic Classification*. Routledge, London, pages 195–210.

Geeraerts, Dirk. 1993. Vagueness's puzzles, polysemy's vagueness. *Cognitive Linguistics*, 4(3):223–272.

Guthrie, Joe A., Louise Guthrie, Yorick Wilks, and Homa Aidinejad. 1991. Subject-dependent co-occurrence and word sense disambiguation. In *Proc. 29th Annual Meeting of the Association for Computational Linguistics*, pages 00–00, Berkeley.

Guthrie, Louise, Brian M. Slator, Yorick Wilks, and Rebecca Bruce. 1990. Is there content in empty heads? In *COLING 90*, volume 3, pages 138–143, Helsinki.

Hanks, Patrick. 1994. Linguistic norms and pragmatic exploitations or, why lexicographers need prototype theory, and vice versa. In Ferenc Kiefer, Gabor Kiss, and Julia Pajzs, editors, *Papers in Computational Lexicography: COMPLEX '94*, pages 89–113, Budapest.

Hearst, Marti A. 1991. Noun homograph disambiguation using local context in large text corpora. In *Using Corpora: Proc. Seventh Ann. Conf. of the UW Centre for the New OED*, pages 1–22, Waterloo, Canada.

Jensen, Karen and Jean-Louis Binot. 1987. Disambiguating prepositional phrase attachment by using on-line dictionary definitions. *Computational Linguistics*, 13:251–260.

Jorgensen, Julia C. 1990. The psychological reality of word senses. *Journal of Psycholinguistic Research*, 19(3):167–190.

Kilgarriff, Adam. 1992. *Polysemy.* Ph.D. thesis, University of Sussex, CSRP 261, School of Cognitive and Computing Sciences.

Kilgarriff, Adam. 1993. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, 26(1–2):365–387.

Kilgarriff, Adam. 1997a. Foreground and background lexicons and word sense disambiguation for information extraction. In *Proc. Workshop on Lexicon Driven Information Extraction*, Frascati, Italy, July.

Kilgarriff, Adam. 1997b. What is word sense disambiguation good for? In *Proc. Natural Language Processing in the Pacific Rim (NLPRS '97)*, Phuket, Thailand, December.

Kilgarriff, Adam and Gerald Gazdar. 1995. Polysemous relations. In Frank R. Palmer, editor, *Grammar and Meaning: Essays in Honour of Sir John Lyons*. CUP, Cambridge, England, pages 1–25.

Kjellmer, Göran. 1993. Multiple meaning and interpretation: the case of *sanction. Zeitschrift für Anglistik und Amerikanistik*, 41(2):115–123.

Krovetz, Robert. 1996. Surprises under the hood: An investigation of word meanings and information retrieval. *Computational Linguistics*, Special Issue on Word Sense Disambiguation. submitted.

Lakoff, George. 1987. *Women, Fire and Dangerous Things.* University of Chicago Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By.* University of Chicago Press.

Landes, Shari, Claudia Leacock, and Randee Tengi. 1996. Building semantic concordances. In Christiane Fellbaum, editor, *WordNet: An Electronic Lexical Database and Some of its Applications*. MIT Press, Cambridge, Mass. forthcoming.

LDOCE, 1987. *Longman Dictionary of Contemporary English, New Edition.* Edited by Della Summers. Harlow.

LDOCE, 1995. *Longman Dictionary of Contemporary English, 3rd Edition.* Edited by Della Summers. Harlow.

Leech, Geoffrey. 1981. *Semantics.* CUP, Cambridge, England.

Lesk, Michael E. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proc. 1986 SIGDOC Conference,* Toronto, Canada.

Levin, Beth. 1993. *English Verb Classes and Alternations.* University of Chicago Press.

Levin, Beth and Malka Rappoport Hovav. 1991. Wiping the slate clean: A lexical semantic exploration. *Cognition*, 41:123–151.

Levin, Beth, Grace Song, and B. T. S. Atkins. 1997. Making sense of corpus data: a case study of verbs of sound. *International Journal of Corpus Linguistics*, 2(1):23–64.

McArthur, Tom. 1987. *Worlds of reference.* CUP, Cambridge, England.

McRoy, Susan W. 1992. Using multiple knowledge sources for word sense discrimination. *Computational Linguistics*, 18(1).

Miller, George. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography (special issue)*, 3(4):235–312.

Moon, Rosamund. 1989. Objective or objectionable? ideological aspects of dictionaries. *English Language Research*, 3: Language and Ideology:59–94.

MUC-5. 1994. *Proc. Message Understanding Conference.* DARPA.

Nunberg, Geoffrey. 1978. *The Pragmatics of Reference.* University of Indiana Linguistics Club, Bloomington, Indiana.

Nunberg, Geoffrey. 1994. The once and future dictionary. Presentation at *The Future of the Dictionary* Workshop, Uriage-les-Bains, France, October.

OALDCE5, 1995. *Oxford Advanced Learner's Dictionary of Current English, Fifth Edition.* OUP, Oxford.

Ostler, Nicholas and B. T. S. Atkins. 1991. Predictable meaning shift: some linguistic properties of lexical implication rules. In James Pustejovsky and Sabine Bergler, editors, *Lexical semantics and knowledge representation: ACL SIGLEX Workshop*, Berkeley, California.

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

Quine, W. v. O. 1969. Speaking of objects. In *Ontological Relativity*. Columbia University Press, New York, pages 1–25.

Schulze, Bruno and Oliver Christ, 1994. *The IMS Corpus Workbench.* Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.

Schütze, Hinrich and Jan O. Pederson. 1995. Information retrieval based on word senses. In *Proceedings, ACM Special Interest Group on Information retrieval.*

Sinclair, John M., editor. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing.* Collins, London.

Slator, Brian M. 1988. Lexical semantics and a preference semantics parser. Technical Report MCCS-88-16, Computing Research Laboratory, New Mexico State University, New Mexico.

Sparck Jones, Karen. 1986. *Synonymy and semantic classification.* Edinburgh University Press, Edinburgh.

Stock, Penelope F. 1983. Polysemy. In *Proc. Exeter Lexicography Conference*, pages 131–140.

Summers, Della. 1988. The role of dictionaries in language learning. In R. A. Carter and M. McCarthy, editors, *Vocabulary and Language Teaching*. Longman, London, pages 111–125.

Sweetser, Eve. 1990. *From etymology to pragmatics : metaphorical and cultural aspects of semantic structure.* CUP, Cambridge, England.

Taylor, John. 1989. *Linguistic Categorization: prototypes in linguistic theory.* OUP, Oxford.

Veronis, Jean and Nancy M. Ide. 1990. Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In *COLING 90*, volume 2, pages 389–394, Helsinki.

Williams, John N. 1992. Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*, 21:193–218.

Yarowsky, David. 1992. Word-sense disambiguation using statistical models of roget's categories trained on large corpora. In *COLING 92*, Nantes.

Yarowsky, David. 1995. Unsupervised word sense disambiguation rivalling supervised methods. In *ACL 95*, pages 189–196, MIT.

Zgusta, Ladislav. 1971. *Manual of Lexicography.* Mouton, The Hague.

Zwicky, Arnold M. and J. M. Sadock. 1975. Ambiguity tests and how to fail them. *Syntax and Semantics*, 4:1–36.