Dictionary word sense distinctions:
an enquiry into their nature

Adam Kilgarriff
School of Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH

Abstract

The word senses in a published dictionary are a valuable resource for natural language processing and textual criticism alike. In order that they can be further exploited, their nature must be better understood. Lexicographers have always had to decide where to say a word has one sense, where two. The two studies described here look into their grounds for making distinctions. The first develops a classification scheme to describe the commonly occurring distinction types. The second examines the task of matching the usages of a word from a corpus with the senses a dictionary provides. Finally, a view of the ontological status of dictionary word senses is presented.

Key words: Dictionary, lexicography, word senses, polysemy, homonymy, corpus

1 Introduction

The humanities, linguistics and natural language processing all share a concern for word meaning. In each, it is often essential to be able to say 'this word here means the same as that one there, but that third one means something different'. When a researcher is studying the ambiguities of a text, or the use of metaphor, or a pattern of correlation between form and meaning, the claims he or she makes will often include the observation that a word is being used in a particular way, perhaps in a hypothesis of the form,

'whenever such-and-such a word is used to mean x in the text, it carries with it an implication that y', or, 'all words for which the meaning follows pattern p can be used in syntactic construction q'. Then it will be necessary to ask whether a word is being used to mean x, or in a way that matches pattern p, in order to determine whether the usage is one that the hypothesis makes predictions about.

For natural language processing (NLP), the concern is blunter; how to represent word meanings, and how to access the right meaning for a particular input.

Published dictionaries provide a promising starting-point for all such enterprises. The best of them are the fruits of the cumulative wisdom of generations of lexicographers. They have always, for every word of the language, had to address the question, what should, and what should not, be treated as a distinct word sense. The sheer breadth of coverage makes them indispensible. But a central question still stands: what sorts of distinctions does the dictionary make? What rationale underlies them? How do lexicographers' distinctions relate to the concerns of NLP, linguistics and humanities research? The questions cannot be answered *a priori*. They demand a careful exercise in lexicology. Two pilot studies for such an exercise are presented below.

'Word sense' or 'sense' will be used to mean the different meanings of the word as presented in a particular dictionary, and a 'usage' to mean a particular occurrence of the word in a context.

As a general rule in English, words can be used to refer to either tokens or types. Thus the expression "Dog eats dog" could be said to have two, or three, words in it. This could lead to confusion in a paper such as this. So the word 'usage' will be reserved exclusively for tokens, making a comment such as 'the word has the same usage in sentence a and sentence b' (where a is not identical to b) contradictory. 'Usage-type', or simply 'use', will be used in such cases. A regular pattern, applying to a number of words, which relates one usage-type to another for each of the words will be called an 'alternation'.

The term 'sense selection' or 'resolution' is preferred to 'disambiguation', as this suggests there is an ambiguity to be resolved, which is often the issue at stake. In this paper we shall encounter many usages which are not ambiguous, even though the usage cannot be matched to one and only one sense. (An exception is in relation to NLP lexical

ambiguity research; here, the word is used as part of the name of the field.) Distinctions between vagueness and ambiguity, homonymy and polysemy, literal and figurative meaning and denotation and connotation are not discussed because, firstly, dictionary senses are the only data that are identifiable and countable. None of the other terms are sufficiently well-defined to support the kind of study undertaken. And secondly, lexicographers have already taken all of these (and many more) dimensions of variation in word meaning into account in arriving at the entries to be found in the published product, so, in studying the dictionary, we are in a position to test out ideas about homonymy and polysemy, figurative meaning and so forth. The dictionary provides the data, and it would be inappropriate to use such theoretical terms to play too great a role in our lexicological analysis. (These claims are elaborated and defended in section 6.)

2 Motivations: NLP, lexicography, statistical methods

2.1 NLP

Any NLP system which wants to arrive at a representation of the meaning of input sentences will need to look up the meanings of the words in the input in a lexicon. This sounds like one of the more straightforward aspects of NLP - and so it is, provided there is only one sense for the word given in the lexicon. If there is more than one, the system will have to make a choice. Where there are a few words in a system's lexicon which have perhaps two or three senses, it seems appropriate to develop an ambiguity resolution procedure for each of them. The context will generally force just one reading, so if an attempt to interpret the sentence is made for each sense, there will generally only be one interpretation which is plausible and coherent. However, in dictionaries very many words have multiple senses. The overall average is over two meanings per word in Longmans Dictionary of Contemporary English (LDOCE), and the situation gets drastically worse when more common words are considered. So it will often happen that a large number of words in a sentence are two-or-more way ambiguous, so the number of readings between which we must disambiguate grows exponentially. As Slator and Wilks (1987, p 4) say,

> A simple-seeming sentence like

There is a huge envelope of air around the surface of the earth.

. . . is a staggering 284,592-way ambiguous.

This demonstration simply serves to show that a parser for text, accessing a realistic machine-readable language resource like LDOCE, is faced with solving a large, and hard, problem.

There are various stages to the process of using a dictionary. A standard mode of dictionary use can be idealised as follows: first, find the entry or subentry about which information is sought. Then, access the details required. An NLP system analysing text uses its lexicon in a similar way. An entry for a word sense serves both to identify whether we are looking in the right place for the use of the word we are interested in, and also to provide the details we are seeking. This paper concerns only the first of these processes, so the only aspects of the adequacy or structure of entries to concern us are those that encode the distinctions between senses, or are relevant to the matching of usages to senses. Questions of whether sense distinctions are adequate or appropriate for any other task are outside the scope of the paper.

2.1.1 MRD research

Since Amsler's thesis (Amsler, 1980), NLP has been investigating and exploiting machine readable dictionaries (MRDs) as potential sources of large quantities of lexical information. Two collections of important papers are are Boguraev and Briscoe (1989) and the special edition of Computational Linguistics, Vol 13, parts 3-4 (1987). This work has made great progress in converting printers' tapes for dictionaries into systematic, structured, de-bugged lexical databases, with some of the information, particularly the syntax, directly available for use in NLP systems. Briscoe, Copestake, and Boguraev (1990) and Copestake and Briscoe (1991) have used analyses of MRD content as the empirical base for their theoretical studies of logical metonymy and sense extension. But when looking at meaning rather than syntax, the enterprise has encountered the problems posed by alternative word senses at every turn. Our lack of understanding of most of the kinds of relations existing between dictionary word senses remains a major obstacle to further

progress in MRD research.

2.1.2 The Bank Model and Its Shortcomings

There have been a number of substantial contributions from NLP to the task of automatic lexical disambiguation. They have generally operated with the 'Bank Model', which might be stated as follows:

> A word like *bank* presents a very clear case of a word with more than one meaning. It can mean the side of a river, or an institution which looks after your money for you. For any usage of the word as a noun, either a money bank or a river bank is being referred to, and the word always refers to one or the other, not both. When English speakers encounter the word in a discourse, they know instantly and effortlessly which meaning of the word applies. This knowledge is an important part of human competence in a language, and an NLP system, likewise, needs to be able to choose.
>
> The word *bank* has been used because, in it, the issues are clear to see, making it a good pedagogical example. It makes clear what sorts of issues we have to deal with in lexical disambiguation for NLP.

The first paragraph is unproblematic. But the second suggests that the kind of distinction between senses found in the *bank* case is the pervasive kind found on every page of the dictionary. This is an empirical claim, and to answer it we turn again to the question, 'what sorts of distinctions does the dictionary make?'

The position stated here is not to be found explicitly in the work of authors who have looked at lexical ambiguity, or word sense distinctions, from an NLP perspective. Yet the research in that tradition addresses only Bank Model ambiguity (see e. g. Wilks (1975), Small (1980), Hirst (1987), Cottrell (1989)). The task has been to determine, for a sentence containing a word with more than one sense in the lexicon, which sense the word is being used in. The system will have completed its task when one and only one of the senses has been identified as the active one. The research program has made important progress and clarified many questions. However, it has not been self-critical with respect to its idealisation of word sense distinctions. The 'matching' study presented in section 5 specifically addresses the empirical question, 'how much of the time does the

Bank Model hold?'

The weaknesses of the Bank Model have, of late, been noted by various researchers. Pustejovsky (1991), Briscoe et al. (1990), Copestake and Briscoe (1991), Fass (1991), Martin (1990), Wilensky (1991) and Kilgarriff (1992) all address words being used in ways which diverge from their core uses, where the word might receive two senses in the dictionary. In each case there is a close relation between the two types of use for the word. The distinction is not treated as a Bank Model distinction but as some form of mapping. A mapping is identified, a representation scheme built, and a processing strategy for interpreting the non-standard use as a function of the standard use discussed or developed. All this research bears witness to an awakening of NLP researchers to the complexity of lexical structure, and to the need for linguistic and lexical-semantic theory to guide the design of NLP systems.

The research presented in this paper is complementary to all this. Whereas those authors have identified an alternation and looked in detail at that, here we look across the dictionary (and gather corpus evidence) to get an overview of what patterns there are.

2.2 Lexical Computing

The computer presents new challenges and opportunities to lexicography. A large corpus provides the opportunity for lexicographers to examine a range of evidence for the behaviour of a word in a way not previously possible (Sinclair, 1987, 1991; Atkins, Kegl, & Levin, 1986, 1988). A concordancing program allows the lexicographer to look at all the contexts of occurrences for any word in their corpus, and, in situations and languages where corpora are available, this is coming to be the raw material from which all lexicographers work. However, the 'word' under examination is defined, in the first instance, as a sequence of letters. For common words, there will be hundreds of citations in a substantial corpus. The lexicographer would like to have them pre-sorted so he or she does not need to wade through hundreds of citations of, e.g., money *bank*s in order to find a population of river *bank*s to compare and contrast, to determine the characteristics of the river-bank usage-type.[1] So lexicography, like NLP, would like to be able to automate

the process of matching usages to a preliminary set of senses for the word —preliminary because one object of the exercise, for the lexicographer, would be to see whether the set of senses for the word needed revision. Atkins (1987) presents the case for unearthing 'semantic ID tags', the syntactic and lexical clues which serve to identify a usage as a usage in a particular sense, and by means of which automatic disambiguation will, in many cases, be possible. The two processes, identifying semantic ID tags and automatic classification, are interdependent: ID tags are needed for automatic classification, but a measure of automatic classification is needed to provide populations of usages to be examined (by lexicographers or automatically) for ID tags. Recent work on the various stages of this problem includes Church and Gale (1991), Hearst (1991) and Veronis and Ide (1990), and the topic is currently a central concern in lexical computing.

This leaves lexicography in a similar position to NLP. The goal is allocating usages to senses. Some usages will not fit. (This is a commonplace to lexicographers; Stock (1983), Atkins (1991). Dictionaries need only describe the recurring usage patterns for words so one-off unusual uses, which might trip up an NLP system, can be ignored, but where patterns of words defying classification recur, the fact will be part of the lexical knowledge that it is the job of the lexicographer to describe. Both are in need of a model of word meanings which makes sense of mismatches between usages and senses.

2.3 Statistics in literary criticism

Literary statistics usually involves counting words. Much of the time that word tokens are being counted, ideally it would be word sense tokens that were being counted. For exercises in identifying a genre, or an author's style, or contrasting periods or varieties of an author's work, word counts usually provide the data which is statistically analysed. But word counts are not sensitive to polysemy or figurative usages, and the accuracy of the analysis may be vitiated if words are frequently being used in ways that the experimenter had not taken into account. van Peer (1989) cites the difficulty in specifying what is, and what is not, figurative, and the concomitant difficulty in counting tokens of each, as considers it a central stumbling block for statistics in literary criticism. Choueka and Lusignan (1985, p 147) assert it to be the belief of "many workers in applied computational

linguistics and large text-processing projects", that

> ...lemmatization [e.g. syntactic and lexical disambiguation] is one of the most important and crucial steps in many non-trivial text-processing cycles ...

Choueka and Lusignan assume that disambiguation is an appropriate goal, in effect adopting the Bank Model:

> It is a matter of simple observation that with the exclusion of jokes, word games and maybe some rare instances of double meaning, occurrences will almost always have exactly one "meaning" (= lemmatization), and therefore will almost never be ambiguous. [2]

I suggest that, while 'lemmatization' is indeed a crucial process for many exercises in humanities computing, the 'simple observation' should not be taken on trust. This study puts it to the test, and finds that a model more sophisticated than the Bank Model will be needed for 'lemmatization' for humanities computing.

3 Lexicology: Preliminaries

3.1 Choosing a dictionary

The dictionary chosen for the studies was LDOCE, as it is an up-to-date dictionary compiled in the light of current ideas on lexicographic practice and, as a dictionary for people for whom English is not a first language, it is constrained to make many things explicit which a dictionary for English-first-language users would not need to. The first edition is available in a convenient, tidied-up, machine-readable form, with some additional information not available in the printed version, and has already been used for a variety of MRD and NLP projects. The revised and improved 1987 edition was used for these studies. It seems likely that the patterns of distinctions in LDOCE are broadly similar to those in other dictionaries, though comparative studies, particularly making comparisons with larger dictionaries such as the OED, are a subject for future research.

Arguably, the COBUILD dictionary would have been a more appropriate choice as it was explicitly based on corpus evidence. It would certainly have been a suitable dictionary

for the studies but I do not think it likely the results would have been markedly different. Any dictionary needs to choose a set of senses for a word and to express clearly what makes those senses different, and this task is central to the quality of the product. The important matter for this study was that these jobs were done well, rather than how they was done.

3.2 Limiting the Domain

To constrain and focus the topic, several kinds of meaning distinction will be excluded from the population of distinctions under scrutiny (inevitably, the boundaries will sometimes be hard to place). These are:

1. **Part-of-speech ambiguity**

   Where there are word senses falling into different parts-of-speech for the same spelt form, LDOCE gives them separate entries.[3] Our strategy shall be the same. While part-of-speech ambiguity is a major issue for NLP, other authors have addressed it. Here, we shall be concerned only with distinctions in word sense within the major word classes *n*, *v*, and *adj*.

   Distinctions between words which are morphological variants of each other will not be addressed.

2. **Set Phrases**

   Many senses of words are identified in the dictionary as occurring only in idioms, compounds or collocations. It is a large and difficult task for NLP to be able to detect these items while parsing. Issues concerning set phrases in the lexicon will only be considered where they cannot be separated from the tasks of specifying senses distinctions and allocating usages to senses (see also section 6.5.2).

3. **Semantically close words**

   There are clearly many similarities and distinctions to be drawn between synonyms or near-synonyms, antonyms, words in the same contrast set, derivational--morphological variants and so forth. They will not be considered.

3.3 Presentation of alternative word senses in LDOCE

How does LDOCE represent alternative word-senses? A thorough understanding of LDOCE's conventions for coding phenomena which might be classed as word sense distinctions is a prerequisite to both identifying the most suitable population of distinctions to study, and also to remaining sensitive to the assumptions implicit in that choice of population. Eight methods are identified. They are presented below, with the methods that draw the most vivid distinctions between usage-types presented first. Quotes are from the LDOCE's User Guide or 'Front Matter'.

1. Different entries.

   "... bank(1) and bank(3) are treated separately, even though they are both nouns, because there is no historical connection between the two words and their meanings are completely different." (page F17)

2. Numbered senses within an entry.

   This is the most common level at which meanings are explicitly distinguished.

3. Subdivisions of numbered entries marked **a, b**, etc.

   The rationale for using subdivisions rather than main divisions is not clear. Subdivided word senses evidently are intended to be more closely related to each other than they are to other senses.

4. Bracketed optional part

   One sense is given by including the bracketed material, and another by excluding it.

   Examples:

   marquetry (the art of making) a type of pattern in wood ... .

   martini (a glass of) an alcoholic drink ...

   mazurka (a piece of quick lively music for) a Polish dance.

5. "(fig.)" in front of an example. According to the User's Guide,

> Some words are used in an imaginative or "figurative" way, to suggest a meaning that is not the literal meaning but has some similarities with it. If a word is often used like this, the examples will include a figurative use, and this is shown by the note (fig.) (page F36)

Example:

> **materialize 1** (fig.) *I'd arranged to meet him at seven, but he never materialized.*

There will always be a continuum between figurative uses of a word and distinct senses which have their origins in figurative usage, since at any point in time many word usage patterns will be in flux between extremes of originality and conventionality. Where a lexicographer draws the line between the two types of cases will always be somewhat arbitrary.

6. The main definition contains a disjunction.

Examples:

> **masked 2** by or for people wearing masks.
>
> **melody 1** a song or tune ...

A disjunction of this kind is not necessarily signalled by 'or'. It could be signalled by 'and', or by a comma. The semi-colon is used for the similar and occasionally overlapping purpose of providing an alternative statement of the definition, often a synonym or near-synonym.

7. Disjunction in the grammatical coding

Where a verb may be countable or non-count, or a verb transitive or intransitive, or otherwise occur as a member of more than one kind of minor word-class, a semi-colon in the grammatical code for the word or word sense will indicate the disjunction.

Example: **marriage** [C;U]

The square-bracketed code means the word can be a count or a non-count noun. The non-count, 'marriage is a good institution' reading of marriage might be taken as a distinct sense.

8. Brackets with a disjunction in them.

   As well as indicating an optional extra meaning, brackets sometimes give selection-restriction-like information on how a word is normally used, indicating what sort of an entity a noun or adjective is used to describe, or for a verb, what its subject or object is likely to be. When these brackets contain a disjunction of dissimilar kinds of entity, it could be said that two senses are being conveyed.

   Example: **meek** (of people or behaviour) ...

   This pattern is used particularly with the people/behaviour alternation. Other occasions when 'or' is used in brackets tend not to be disjunctions but indicators of a range of possibilities.

The next two sections present empirical studies which required a base population of word senses and word sense distinctions. The studies are of whatever distinctions the LDOCE lexicographers have drawn, and the base population will be the senses and distinctions that LDOCE not only makes but makes explicitly, using methods 1 to 3. The role of distinctions 4–8 will be taken up again in section 6.

4 Study 1: Classifying Sense Distinctions

It is very often the case that two dictionary senses for a word are closely related in meaning. For example *marriage* can be the event, as in "we're looking forward to the marriage", or the subsequent state, as in "they had a long and happy marriage". The distinction here is an example of what I shall call an MBT (for 'Must Be There'); a word sense distinction where the applicability of the one sense of the word to one aspect of a situation entails the applicability of the other sense to another aspect. The purpose of the study was to establish what, if any, were the commonly occurring patterns of distinctions between word senses.

Apresjan (1974) and Ostler and Atkins (1991) have catalogued recurring patterns, under the names of 'regular polysemy' and 'lexical implication rules'. Their lists are of well-defined alternations applying usually to relatively small numbers of words in a particular semantic field. The goal of this study was rather different: a classification scheme able to give some account of a high proportion of all sense distinctions.

I looked closely at 427 full entries, approximately 1% of the dictionary. The study was manual. An initial analysis is given below.

Full entries	427	
of which - capitalised		20
- compound/phrasal/hyphenated		37
leaving a base population of	370	
2 separate entries in same word class	6	
Entries divided into numbered meanings[1]	110	
of which - 2 meanings		75
- 3 meanings		17
- 4 meanings		10
- 5 or more meanings		8
nouns		63
verbs		21
adjectives		25
adverbs		1
Total of numbered word sense distinctions (minimum)[2]	187	
Total of numbered word sense distinctions (maximum)[3]	326	
Numbered entries subdivided using letters	10	
Brackets give another sense (as in 4 above)	23	
of which - 'cause to'		4
- 'too'		3
Another word sense given as "fig" example	10	

[1] Excluding numbered meanings where a phrase, idiom or collocation is given, also ignoring subdivisions by letter.

[2] E.g., where there are more than two senses, counting only one distinction per additional sense, so number of distinctions counted is one less than the number of senses.

[3] E.g., where there are more than two senses, counting all pairwise distinctions.

4.1 Clarifying Dictionary Definitions

There was a further task to be performed before a distinction could be classified. The entries had to be read closely and interpreted, in order to establish the exact nature of the distinction the lexicographer had in mind in deciding to identify two senses. His or her work in encoding the distinct senses and the distinctions between them in the highly constrained language of the dictionary entries had to be reversed (see section 6).

An LDOCE definition can be a very complex structure. In this exercise the distinctions under consideration were those between numbered senses. A numbered sense definition can have all or any of the following.

1. syntactic code
2. prepositions subcategorised for
3. definition-text or 'd-text' with
   (a) syntactic structure
   (b) semantic content
   (c) restriction on range of application
   (d) 'esp.' specifying a central or prototypical usage
4. examples with
   (a) syntactic structure
   (b) semantic content
   (c) collocations (sometimes bold typeface).

In interpreting sense definitions, many questions arise. Three were particularly salient to the current exercise: necessity, consistency, and centrality.

Firstly, of 1, 2, 3c, and 4a-c, it can sensibly be asked whether the specified feature is a necessary condition of the word being used in this sense, or merely typical.

Secondly, it can also be asked whether the information from the different aspects of the definitions is consistent. Where a word has a sense *a* and a sense *b* it could be that

examples and syntax suggest one dividing line between the usages to be classified as *a*s and *b*s, while the definition-text suggests another. An example: sense 2 of *application* is marked as non-count yet the d-text is 'the act of putting something to use', and acts are generally countable. In this case the potential inconsistency can be spotted when the definition is examined in isolation. The more common case is where the definition shows no inconsistency in isolation, but indicates that certain features, when taken together, identify a particular sense of the word, and another group of features indicate another sense; but then, on examination of citations, it becomes evident that many of the citations display some features from the one set and some from the other, so the clustering of features indicated in the dictionary does not tally with the evidence for how the word is used. The dictionary distinction is thus inconsistent with the citations.

Thirdly, we must ask, where a part of a sense definition does not present a necessary condition for the word being used in that sense, how much variation away from the form given in the dictionary is acceptable with the word still being properly classified as that sense? This problem is particularly clear in relation to examples. An example sentence, like a picture, suffers a shortcoming: it provides an instance, but does not provide any indication of how much variation away from the example presented is permissible, without the variant ceasing to exemplify what the example exemplifies. This was a pervasive problem. An example is appointment/2.

> **appointment 1** an arrangement for a meeting at an agreed time and place ... **2** [U] the agreement of a time and place for meeting: *He will only see you by appointment.*

It is not clear whether any usages of the word except those in the *by appointment* collocation are to be classified under sense 2. The intended distinction between senses 1 and 2 is clearly quite subtle. The easiest route to follow would be to equate sense 2 with *by appointment* exclusively, but if that had been what the lexicographer had intended, then he or she would, in principle, have made the fact explicit by putting the collocation in boldface at the head of the sense entry.

These questions often could not be resolved in isolation. The experimenter identifies a sense by contrasting its specifications with the specifications of the other senses. In most cases the contrasts will start to answer the questions, but it will take the evidence of

citations to remind the experimenter of the distinctions the lexicographer was attempting to capture, and to determine where the lexicographer's bundles of features seem not to belong together.

4.2 Classification Scheme

Distinctions have been classified according to the following categories:

4.2.1 Generalising Metaphors (Genmets)

The distinction is between one specific sense, which is unequivocally the right word for the specific situation, and a less specific sense which shares some features of the specific sense but can be applied in a wider range of situations. The less specific sense has in several cases acquired positive or negative connotations, while the more specific one has more of a factual flavour. Amongst the verbs, a bracketed note stating the type of entity that the more specific sense applied to was common.

Examples:

> **martyr** *n* **1** someone who is put to death or suffers for their beliefs ... **2** someone who gives up their own wishes or suffers ... **3** someone who suffers something they cannot avoid ...
>
> **maul** *v* **1** (esp of animals) to hurt badly ... **2** to handle roughly ...
>
> **meaty** *adj* **1** full of meat **2** full of valuable ideas.
>
> **benefit** *n* **1** [U] anything that brings help, advantage or profit ... **2** [C; U] money provided by the government to people who need it ... **3** [C] an event, esp. a theatrical performance, to raise money for some person or special purpose ...

For many genmets, the denotation of the specific sense was a subset of the denotation of the general sense. Hence any martyr/1 is a martyr/2 (and also a martyr/3, if it is allowed that moral imperatives cannot be avoided). A benefit/2 or benefit/3 is always a benefit/1. The subset relation between denotations does not, however, have any straightforward implications for the relationship between the usage-patterns for the two senses.

In particular, it is not valid to say that "My benefit is £22.04" is unspecific between benefit/1 and benefit/2 on the grounds that, if the money was a benefit/1, it was also a benefit/2. While it is true that the money was a benefit/1, this is not what the sentence asserts, and, had the assertion been that the money was a benefit/1, the structure of the sentence would have been quite different, perhaps "The benefit to me was £22.04".

4.2.2 Must-Be-Theres (MBTs)

MBT distinctions are those where, if the situation is such that the one sense can be applied, then it is a logical consequence that the other can also be applied, though to some other aspect of the same reality. If the situation is thought of as described by a schema with slots and fillers, an MBT distinction is one which attends to different slots of the same schema. The presence or existence of the filler of the one slot entails the presence or existence of the filler for the other.

Examples:

> **marriage** *n* **1** the union ... **2** the state of being married.
>
> **matricide** *n* **1** the murder of one's mother **2** a person guilty of this crime
>
> **marry** *v* **1** to take (a person) in marriage ... **2** to perform the ceremony ...
>
> **married** *adj* **1** having a husband or wife ... **2** having as a husband/wife

4.2.3 Domain shift

It is in the nature of words that, in learning and using them, we apply them to situations beyond where we have previously applied them. The point at which two situations in which the word is applied are sufficiently far removed from each other to constitute distinct senses will never be well-defined. A semantic type distinction is one where the two situations of usage are just sufficiently far apart so that the lexicographers have decided two distinct senses are called for, even though one might say, 'but the word has only been adapted as far as it had to be, given the different entity or situation to be described'.

Examples:

mastery *n* **1** full power to control or defeat something **2** great skill or knowledge in a particular subject or activity

marshal *v* **1** to arrange (esp. facts) in good or effective order **2** to lead or show (a person) ceremonially or carefully to the correct place

mellow *adj* **s1** (of fruit and wine) sweet and ripe ... **2** (of a colour) soft ... **3** (of people and behaviour) wise and gentle ... 4 (feeling) pleasantly calm and friendly ...

4.2.4 Natural and Social Kinds

The distinctions classified under this heading are those where, owing to a non-linguistic fact, the entities or situations identified by the different word senses have distinct denotata, and although the denotata have many attributes in common, they will always remain different classes of things. Membership of both classes is possible, in the social if not the biological cases, but is then co-incidental.

Examples:

marrow s1 also **bone marrow** ... **2** also **vegetable marrow** ...

marshal s1 an officer of the highest rank in certain armies and airforces ... **2** an official in charge of making arrangements for an important public or royal ceremony or event **3** an official in charge of making arrangements for a race ... 4 (in the US) **a** an official who carries out the judgements in a court of law ... **b** a chief officer of a police or fire- fighting force

It is to be noted that in some of these cases, a dictionary definition is less a definition, than a pointer to a class of entities. LDOCE often uses pictures, illustrating a typical class member. A complex set of facts, some of which may not be known to science, serve to define the class, and no more than a reference to these facts will be found in the dictionary.

4.2.5 Relations between types

The sample contained only noun 'kind' distinctions. It may be that this is true of the language at large. Kind distinctions are closely related to domain-shifts, and it seems likely that verbs and adjectives are treated as domain-shifts where a noun would be a kind distinction. The 'kind' senses of *marshal* do, after all, shift in domain between army, airforce, royal or public ceremonies, races, courts of law, police and fire-fighting forces. But whereas any of the senses of *marshal* can (in a naive view of the semantics of nouns) be taken to denote a set of entities, denotations for verb and adjective senses are not readily seen in this way, so 'kind' seems an inappropriate label.

A distinction could be both a 'kind' and a domain-shift, but it could not be a kind and a genmet, since the general sense would not be a natural or social kind contrasting with the specific sense. Nor could it be a domain-shift or kind and an MBT, since the MBT relationship is specifically not one between distinct domains or sets of entities. Genmets and domain-shifts were compatible.

4.3 A Worked Example: *mess*

mess *n* **1** [S;U] (a state of) untidiness or dirt; dirty material: *This room's* **in a mess.** | *There's a lot of mess to clear up.* | *What an awful mess!* **2** [S] *infml* a situation full of difficulty and disorder; trouble: *The company's affairs are in a terrible mess.* | *That's another fine mess you've got us into!* **3** [C *usu. sing.*] *infml* someone or something untidy, disordered etc.: *You look a mess — you can't go to the office like that.* | *That report you did's a real mess — do it again!* **4** [C] a room in which members of the armed forces eat together: *the officers' mess* **5** [C;U] *euph* a quantity of animal FAECES (= solid waste material): *The dog made a mess on the carpet.*

The grammatical codes, in square brackets, are: 'S' for senses used only in the singular; 'C' (the default) for countables, where the sense can be modified by a number and can occur in the plural; 'U' for uncountable (no modification by numbers and no plural) and ';' for disjunction.

The relationship between 1 and 2 could be a domain-shift, from the 'physical' to the 'mental', or a genmet, particularly since the 'physical-to-mental' route is a particularly

common one for metaphors to take. It could also be interpreted as both, where sense 2 is a generalised metaphor, based on sense 1, but, while generalised, is not entirely general and in fact applied to a different domain to sense 1 so is a domain-shift.

Sense 3 appears, at first sight, to be an MBT related to sense 1, with a relationship analogous to that holding between count and non-count senses throughout the language (e.g. *whisky, cheese, cake*). For there to be 'untidiness' (one disjunct for sense 1) there must be 'someone or something untidy' (one disjunct for sense 3). Considering the sense 1 examples, "This room's in a mess/1" implies "This room's a mess/3", and "There's a lot of mess/1 to clear up" and "What an awful mess/1" both imply "The place is a mess/3". The first example from sense 3 fits the theory, but the second, "That report's a real mess — do it again" relates to a 'generalised metaphor' sense closer to 2 than 1.[4] We may also note that the two sense 3 examples both have the form

REFERRING-EXP COPULA INDEF-ARTICLE (OPTIONAL MODIFIER) mess

and are both being used to tell someone that something is not of the expected standard. This suggests that sense 3 relates specifically to this usage pattern which, introspection confirms, is one with something of a distinctive feel to it. But if that is so, then none of the distinctions involving sense 3 are essentially MBTs because what sets sense 3 apart is syntax and pragmatics rather than semantics.

Section 4 bears no synchronic relation to any of the other senses, so no distinctions involving it fit any of our categories. Sense 5 is either a domain-shift from sense 1 or genmet, where sense 5 is the specific and sense 1, the general sense.

The outcome for *mess* was that the distinction between senses 1 and 2 was classified as a domain-shift, that between 1 and 5 as a genmet, and no others received classifications.

4.4 Results

	Nouns	Verbs	Adjectives	Totals
Genmets	10	6	3	19
MBTs	14	6	2	22
Domain-shifts	6	13	7	26
Kinds	11	-	-	11
	41	25	12	78

The classification scheme is neither exclusive, nor exhaustive, nor final. As the discussion of *mess* revealed, some distinctions seem to fit more than one classification. Some fit none, either because there is no interesting synchronic relation, as in relation to *officers' mess*, or because the distinction is one that does not relate to the repertoire we have to choose between, like the animal/food one for *chicken, lamb* etc.[5], or because, like the mess/1-mess/3 one, the salient factors were syntactic or pragmatic rather than semantic. Since the primary object of the study was simply to identify major types, all of these were left unclassified. Just 78 of the distinctions —42% on the lower count of distinctions, 24% on the higher — received a classification. While very preliminary, and limited in its concerns to the semantic aspects of distinctions, the study does indicate the repeated patterns that are to be found in the distinctions between senses in a published dictionary. It provides a vocabulary for distinguishing some of the very different kinds of phenomena which all fall under the umbrella of word sense distinctions.

5 Study 2: Matching Usages to Senses

A second study used a corpus, and experimented with matching corpus usages to dictionary senses. The motivation was to see how widely the Bank Model has application. A sample of words was selected on the basis of frequency alone. For each, a set of citations was gathered from the LOB corpus. For each citation, an assessment was made as to whether it fitted one, none, or several of the LDOCE senses for the word.

For the purposes of this study, a 'usage' for a word was specified by a citation of

about eighty characters with the word under scrutiny in the middle. Thus the only available context was verbal and was given in the (on average) seven preceding and seven following words. The notion of 'usage' was thus a severely limited one, and different results might have been obtained had more context been taken into consideration. Both a longer citation and other details such as author, title, genre and source could have been examined. This was not done because, firstly, this was an initial study, serving to indicate the kinds of relations existing between usages and dictionary senses rather than making strong quantitative claims. Secondly, in some cases it seemed very unlikely that any amount of context would have disambiguated. I hope the examples below and in the appendix will convince the reader of the validity of this claim, *contra* Choueka and Lusignan (1985, *op cit.*). Thirdly, for automatic sense resolution of any kind, the context available to the system will generally be limited. There may be knowledge of the domain, genre and source but the major part of the evidence will be the immediately preceding and following text. An automatic sense-resolution system will be working with a similar account of 'context' to that considered in this study.

5.1 Identifying the sample

The source of usages was the LOB corpus. The sample of words to be investigated was arrived at in the following way. Very common words were excluded because they tend to have very large numbers of senses and to present complex and difficult cases. For this study, simpler cases were to be examined. Low frequency words were excluded because it would not be possible to see any patterns emerge unless there was a reasonable number of usages to be examined. A range meeting these constraints was 26-29. So the initial sample was chosen by taking all those words which had between 26 and 29 occurrences in the first half of the LOB corpus. Half of these, a sample of 154, were taken for further analysis.[6] From this set the following were removed; prepositions and adverbs (there would only have been two of these, not enough to make any general comments), proper names, adjectives relating to countries ('dutch', 'greek'), titles ('earl', 'congress'), and non-base forms of words ('cutting', 'created', 'directors') or forms which were base but where a non-base form occurs much more frequently than the base form. The size of the

filtered sample was 83.

The exercise was, in outline, to determine the following. For each word, for each usage of that word, which, if any, of the word's senses did it fit? Where each of the usages of a word could be straightforwardly matched against one and only one sense, the limited evidence allowed that the word fitted the Bank Model. Where they did not, the word fell outside the scope of the Bank Model. It was hoped that the misfit cases might be revealing of the sorts of difficulties that lay in store for any new model.

Semantic rather than syntactic distinctions are our concern. For those words that were used as more than one part of speech, the nominal, adjectival and verbal uses were treated as separate sets, and the problem was to classify the nominal usages according to the nominal word senses, the verbal according to the verbal and the adjectival according to the adjectival. Cases where the usage could not be readily classified as nominal, verbal or adjectival were not common.

The exercise had two stages: clarifying the dictionary definitions, as discussed in section 4.1, and determining which of the senses a usage fits. Here there were several possibilities:

1. Exactly one sense fitted.

2. More than one sense was near the meaning in the citation. In these cases the definitions were examined for clues as to how the lexicographer would have intended the usage to be classified; this generally involved a close examination of how well the syntax and meaning of the citation matched that of the dictionary examples for each candidate sense. Sometimes a unique sense was established as a 'best fit', sometimes not.

3. Two or more non-exclusive senses applied, both making different contributions to the word's contribution to the citation's meaning.

4. A usage was indeterminate between different senses. For example, the usage of 'guest' in

   > the opportunity of showing a guest something of ourselves ...

is indeterminate between guest/1, 'a person who stays in someone's home ...' and guest/2, 'a person who is invited out and paid for ...'. This kind of indeterminacy occurred particularly where the use of the word was non-referential.

5. A usage was not covered by any of the senses, perhaps because it was an unusual figurative use of the word, or a rare use, or simply because the lexicographer had left something out.

6. It seemed that the word was being used in one and only one sense but there was insufficient context to determine which.

5.2 Two worked examples: *image* and *exercise*

There follow two accounts of how usages were found not to fit one and only one of the senses. First:

image *n* **1** [C(**of**)] a picture formed in the mind: *She had a clear image of how she would look in twenty years time.* **2** [C] a picture formed of an object in front of a mirror or LENS, such as the picture formed on the film inside a camera or one's REFLECTION in a mirror **3** [C] the general opinion about a person, organization, etc., that has been formed or intentionally created in people's minds: *The government will have to improve its image if it wants to win the next general election.* | *The company tries to project an image of being innovative and progressive.* **4** [(*the*)S(**of**)] a copy: *He's the (very) image of his father.* **5** [*the*+S+**of**] a phrase giving an idea of something in a poetical form, esp. a METAPHOR or SIMILE **6** *old use* likeness; form: *According to the bible, man was made in the image of God.* —*see also* MIRROR IMAGE, SPITTING IMAGE

The citations included the following.

> of the Garonne, which becomes an unforgettable image. This is a very individual film, mannered,

Here the 'image' is an image/1, a picture in someone's mind (probably the author's; possibly the whole cinema-audience's). It is also the image/2 produced on a screen by means of projection equipment from the image/2 on the photographic film.

Regarding image/3, if we look only at the d-text we may think that the 'image' in the text is the opinion that film-going people will have of the Garonne (etc.) which has been formed intentionally by the film-maker. However the examples indicate that the lexicographers probably did not have this sort of thing in mind. On looking further at the d-text we shall conclude that the category of 'person, organization etc.' alluded to as what the image/3 might be of, is a category that probably does not stretch as far as whatever it was about the Garonne that the image was of. (If the film was made by the Garonne tourist board, then it is likely that more context would have supported image/3.) Image/4 does not seem a likely candidate. We do not have the *the image of* expression which is typical of this sense[7], nor does there seem to be any copying, in any very obvious sense, going on.

Sense 5 would seem to fit except that, in this case, the *the image of* expression is required[8] so sense 5, which from d-text alone might seem the best fit, is ruled out. The d-text for image/6, 'likeness; form', is so open as to seem quite plausible as a match for the text (in which the *image* could be exchanged for *form* without dramatically changing the meaning), and the example sentence does nothing to discourage this interpretation. It is only the style label 'old usage', which might deter us from allowing that the text uses image/6, since there seems nothing old-fashioned about this citation. Subentry order also weighs against image/6. LDOCE aims to list more standard senses before less standard ones, and this seems a standard use for the word, so if this were an example of sense 6, we would not expect to find sense 6 listed so late in the entry.

Senses 4, 5 and 6 can then be rejected (though sense 5, only unwillingly; the d-text did seem the most natural match for our citation, but the grammar code disallowed it. We may wonder if we have here an inconsistency, as described in section 4.2 above. The dictionary indicates that a usage which matches the d-text of image/5 will only be found in the expression, *the image of*, yet this citation seems to cut across that supposed correlation[9].) Sense 3 can be set to one side on the basis that we have no evidence that the film was made by the Garonne tourist board (or similar). But we would rather not have to make a choice between senses 1 and 2. They are not mutually exclusive. The usage makes reference to both the projected image/2 and the images/1 that the projected

images/2 caused in people's minds, and to make a choice would be to reject half the story.

Second:

exercise *n* **1** [C;U] (a) use of any part of the body or mind so as to strengthen or improve it: *if you don't take/get more exercise you'll get fat.* | *She does exercises to strengthen her voice.* **2** [C] a question or set of questions to be answered by a student for practice: *Look at Exercise 17 in your book.* **3** [C] a set of actions carried out by soldiers, naval ships etc., in time of peace to practise fighting: *The soldiers are here for a NATO exercise.* **4** [S(**in**)] any set of actions, esp. when expected to have a particular effect: *Getting this report done in such a short time was quite a difficult exercise.* | *After the President's embarrassing remark, his staff had to stage an exercise in damage limitation* (=try to limit the damage he had done) **5** [S;U(**of**)]*fml* the use of a (stated) power or right: *Expelling him from the club was a legitimate exercise of the committee's authority.*

A usage which defies classification as one and only one sense is

but at best only portions of the exercise can be significant artistic expression -

Syntactic clues offer no assistance. All senses admit a singular form with article *the* and nothing subcategorised for. Sense 1 will be appropriate provided that a goal of the exercise was to strengthen and improve the faculty used in undertaking it, but in this citation no such goal is mentioned, nor would it necessarily be even if we had access to much fuller context. Exercise/2 could well apply, since the exercise may very well have been one set for a student for practice, for example in an art class. Setting aside the possibility that the form of art under consideration is the symmetry and elegance of a military attack, exercise/3 can be rejected. Sense 4 is so broad that the usage certainly could be said to fit, but it could be that this very general sense should be set aside on the basis that there is a more specific sense available. Sense 5 can be discarded with some confidence on the grounds that 'artistic expression' and 'the exercise of a power or right' are unlikely bedfellows, that there is no 'stated' power or right, and that there is no *of* following *exercise*.

So in this case we have a hierarchy of specificity. If sense 2 is appropriate then sense 1 could be set aside, although in a sense it did apply, since its applicability was already

implicit in the more specific sense 2. A larger context might and might not resolve whether sense 2 applies. Likewise, if sense 1 applies, then sense 4 need not be invoked since it is already implicit in sense 1 —but we do not know, on the basis of the citation, whether sense 1 does apply.

The examples should not be read as a criticism of the dictionary. Dictionaries are not written with a view to the task of assigning usages of a word to one and only one of the senses for the word that the dictionary provides. It is not a problem for a person, be they a native or a non-native speaker of the language, if more than one dictionary sense is able to resolve their uncertainty about what a word is contributing to the meaning of a sentence where it puzzles them. A user need only read an entry up to the point where their puzzlement is resolved[10]. Even if a word sense coming later in the entry would have fitted the usage equally well or better, such a user has no need of that further information. Hence it is not surprising, and no criticism of lexicographers, if usages can often not be assigned one and only one word sense. There is no practical or theoretical reason why it should be possible.

5.3 Results

Of a sample size of 83 words, 14 had just one sense for each part of speech they featured in, leaving 69 for which there was potentially a choice to be made. For 60 of these 69 words, there was at least one usage which could not with any confidence be classified into one rather than an other of the senses. Thus the sense selection task presented in the experiment could sometimes not be resolved to a single sense for 87% of words where the possibility arose.[11]

Sceptics and advocates of the Bank Model may argue that another researcher would have succeeded in classifying all or nearly all usages. The appendix presents definitions and concordance lines which, in this experimenter's opinion, could not be satisfactorily classified as one and only one sense. The challenge for the sceptic is then to identify how each of the usages in the appendix should be treated.

The usages which could not be assigned one and only one sense could have been classified according to the reasons they defied classification. This was not done in this

case because the range of explanations was arrived at only in the course of conducting the experiment. For a further experiment it would be a worthwhile exercise.

5.4 Observations

The first point to make is that the exercise was, much of the time, hard. According to the Bank Model, people select senses instantly and effortlessly. For the sample of words chosen here, the experimenter was frequently toiling laboriously.

The task was hard in the cases where more than one of the dictionary senses was near the usage in the citation. The dictionary provides only a set of clues to the nature of the senses that the lexicographer was intending to discriminate. Identifying the divisions that the lexicographer saw in the conceptual space of usages of a word is a matter of reconstruction, and the citations are essential to the task, so as experimenter I was working at clarifying the sense distinctions throughout the process of classifying usages. It was not possible to work with an unchanging conception of the distinctions. Each time a new citation neither clearly fitted one and only one sense, nor replicated a pattern already seen, a re-evaluation of the sense distinctions for the word was inevitable. Examples of awkward words were *apply*, *image*, and *design*. Each has a collection of closely related senses many of which are fairly abstract.

The issues encountered varied greatly from word to word. Words where a sense could always be selected, and selected easily, included *absence*, where sense 1 is a specified person's absence, while sense 2 is the non-existence or lack of something, and *capable*, where all but one usage were in the *capable of* construction, which is specified as capable/1.

Football and *chapel* were words which could often not be classified owing to lack of context. The distinctions under consideration are between soccer and American football, and chapels within churches, stand-alone chapels, and chapels which are rooms for worship within houses or institutions. It was in these kinds of cases that non-linguistic knowledge played the clearest role. Knowing the corpus is of British English, and that soccer is more popular than American football in Spain, I am confident (though not certain) that in

> I watched Spanish soldiers playing football and called up a vision of serried lines

soccer is the game in question (see *chapel* in the appendix for more examples). The *football* and *chapel* distinctions are clearly both 'kinds', and the response that more context was needed was strongly associated with kinds.

In general, dictionary entries displayed an inverse correlation between the specificity of the d-text and the number of examples. Matter/5 has d-text, 'things of a particular kind or for a particular purpose'. Enlightenment only arrives with the examples

> *I must take some* **reading matter** (= books, magazines etc.) *for the journey.* | *advertising matter* | *vegetable matter* | *waste matter*

(See also discussions of mess/3 and exercise/4 above.) Where the d-text only very loosely constrains where the sense might be suitable, the sense tends to have examples to clarify where the sense is used, which may be characterised by syntax, collocates or pragmatics. Other senses, like the 'American' sense of *football* have a specific meaning made clear in the d-text and examples are neither needed nor given. We might contrast 'd-text-heavy' and 'example-heavy' items. In general the d-text-heavy ones were easier to work with. Semantic contrasts played a large role, and they were relatively likely to receive a classification in the first study. Interpreting the example-heavy ones required the incorporation of evidence from a number of examples and from syntax codes, and was generally harder. This was not a result of a lexicographer's decision to convey information through examples rather than d-text, but rather a fact about the word sense and its usage pattern: a different kind of information needed conveying.

5.4.1 The Bigger the Better?

It might be argued that LDOCE is a small dictionary, and it simply did not make sufficiently fine distinctions for each usage to find its true sense; a larger dictionary would allow the process to be more precise. My experience indicates the opposite. In general, the more possibilities there are to choose between, the more evaluating of different evidence and assessing of contrasting pairs is required, and the classification task becomes ever more difficult. Where LDOCE gave only one sense for a word, no difficulties were encountered![12]

One important exception was collocational information. Where the pattern of words in the text matched, exactly or very closely, a pattern of words found in an example or as a subentry, the usage could be classified directly. Where there were many collocations, clearly presented, the task was greatly assisted, and was turned into a process which could readily be automated.

6 The ideal lexicographer and the essential word sense

'Word sense' is a useful concept, and the dictionary provides a huge store of data about them. However, the fact that they are useful and we intend to exploit them does not provide them with a philosophical pedigree. Questions such as 'What are they? Where does one end and the next begin?' remain unanswered. 'No entity without identity' runs Quine's test, and without identity conditions for word senses the concept remains hazardously ill-defined.

An idealisation of lexicographic practice provides a working definition of a word sense, as follows. We assume that for each word, the lexicographer

1. gathers a corpus of citations for the word;
2. divides the citations up into clusters, so that, as far as possible, all the members of each cluster have more in common with any other member of that cluster, than with any member of any other cluster;
3. for each cluster, works out what it is that makes its members belong together;
4. takes these conclusions and codes them in the highly constrained language of a dictionary definition.

The process is an idealisation of what actually happens in dictionary-making, displayed to expose 'the central core of the lexicographer's art' (Krishnamurthy, 1987, p 75). Now that extensive corpora are available to lexicographers (at least in English and some other languages), lexicography is moving towards the idealisation. Its merit is that it enables us to say that there was, in principle, a process of clustering usages, performed

by a lexicographer. The lexicographer was probably not explicitly aware of the criteria according to which he or she clustered at the time, and stage 3 is a fallible *post hoc* attempt to make the criteria explicit. Yet it is those criteria which determine the senses that eventually appear in the dictionary. They are a result of that process. But they are a result at several removes, and with each of these removes comes the possibility of confusion or error.

The idealisation is of use for our search for the nature of word senses. We should like to know what they are, and where one ends and the next begins. The idealisation points us towards the criteria the lexicographer was using for his or her clustering, because, however quirky they may have been, they are the data that the published form of the dictionary is attempting to communicate. They answer, as well as anything can, the Quinean test. 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.

6.1 Sufficient Frequency, Insufficient Predictability

To say that senses correspond to clusters of usages invites the question, what holds the clusters together? Our response is: a usage-type should be listed in a dictionary as a distinct sense if it is sufficiently frequent and insufficiently predictable, and these are the criteria that lexicographers, more or less consistently and consciously, apply.

First, frequency. Words which are used only very rarely are omitted from any dictionary but one on the scale of the OED. The same consideration applies to word senses. A usage-type must be of sufficient frequency, for it to earn its place a dictionary.

The second is predictability. Where a usage-type is entirely predictable, as for example a 'type' usage-type for *word* is entirely predictable from a 'token' usage-type, the matter will go unremarked in the dictionary. The overhead of putting in the extra information for every word where it applied would be vast, and users will in any case know and be able to apply the rule. Since both options are available for most nouns, it is likely that the genus term for one noun which has both options will be another which has both options. Indeed it will probably be difficult for the lexicographer to write easy-to-read

entries which specify token-only sense or a type-only sense. It would also appear that alternations as predictable as this are not language-specific. There is no need to spell out that *word* has the two senses in a dictionary for non-native speakers because their first language, whatever it is, will have both token and type readings for its translation of *word*. The traditional linguistic account of what goes in the lexicon is 'whatever is irregular about the words of the language'. Regular or predictable usage-types need not be listed.

Predictability, like frequency, is a matter of degree for lexicographers. While the type/token distinction is highly predictable, the person/behaviour alternation is less so. Both people and behaviour can be *meek* or *gentle* but it is in general only people that are *uncomplaining*. As we saw in section 3.3, LDOCE has a range of means at its disposal for conveying alternative usage-types. Different techniques are appropriate to different degrees of predictability and, less directly, frequency. The high frequency, low predictability items are worthy of the space allocation of a distinct entry or distinct numbered sense. Where usages are less frequent or more predictable they are less worthy of space.

Where they are more predictable, the full expressive power of methods 1-3 is not needed. The existence of both usage-types can be confirmed by a much simpler device. Where alternation patterns occur fairly frequently in the language, for the user it is likely he or she could predict that the word will follow the pattern, and for the lexicographer it is worthwhile developing a device which will do the job economically, avoiding undue repetition or the stating of the obvious. The device is explained in the front matter for anyone who does in fact find it perplexing. Thus 'bracketed optional part' (method 4) is used for a range of common alternations. Drink-words, for example, had definitions which started with '(a glass of)'. With regard to method 8, any word used to describe people might also be used to describe behaviour, and vice versa, but which ones are? That *meek*

> (of people or behaviour) gentle and uncomplaining; accepting others' actions and opinions without argument

follows this pattern would not be known by someone who did not know the word, but from the remainder of the definition, it is likely. The '(of people or behaviour)' label

means, specifically, 'yes it does'.

7 Conclusion

Dictionaries are treasure houses of data on the uses words have. They are also our best starting point for all questions regarding word sense distinctions, in NLP, the humanities or lexicography. But to reveal the dictionary's treasures in a systematic way is no simple task. The variety is bewildering and defies any easy classification. The prior task is lexicology: to look closely at dictionary senses, alongside corpus evidence, and to determine what is there.

The two studies indicate that this is possible and fruitful, and there is much more there yet to be revealed. The evidence from both is that word sense distinctions form a highly heterogeneous set. The Bank Model applies to some, but not others. Sometimes two senses of a word are mutually exclusive, but more often they are not, and for some usages, both senses contribute different elements to the meaning. There are usages where a distinction is irrelevant, and the context simply does not specify one sense rather than the other. Sometimes the key to a sense distinction lies in the semantics, at others, in the syntax, collocates, or pragmatics. Often, the senses as identified in the dictionary identify points on a continuum of possibilities for how the word is used and dictionary senses might equally have been written which divided up the space differently. Different words, more often than not, do not range along the same dimensions, but define their own, so the lexicographer needs to express the dimension of variation before stating how the word ranges along it.

In this great web, the two studies identify some design. The first provides categories for common kinds of distinction, while dispelling optimistic illusions that word sense distinctions will ever succumb to a neat classification scheme. The second confronts the distinctions in the dictionary with citations from the corpus, and identifies a range of reasons why usages sometimes do not match one and only one sense. In 'practical' uses of language, as exemplified in NLP, the specificity of the meaning of a word tends to be the focus of attention. For poetic and literary uses the ways in which they may be

interpreted and reinterpreted and carry layers of allusions, their 'openness', is often of greater interest. The studies cast a little light on how it is that words can at one moment relate complex descriptions of specifics, and at another, shed their specificity and float free on a sea of possibilities.

Acknowledgements

I would like to thank Gerald Gazdar, Michael Rundell and the anonymous reviewers for comments on earlier drafts, Longman for their co-operation and permission to quote from LDOCE, and SERC for financial support.

Notes

[1]Syntactic 'tagging', so that verbal and nominal *bank* can be separated, and morphological processing, so that *bank, banks, banking* and *banked* can be seen alongside each other, are in themselves important areas of research but are not our concern here.

[2]The quotation proceeds: "True enough, a sentence context may be sometimes insufficient for the correct lemmatization of an occurrence, as in the example: "He dropped the bat on the floor" (Is it $BAT1$ or $BAT2$?). The whole paragraph or the whole text — or maybe even some extra textual information — has to be taken into account in these cases, but this is not the point. The important thing is that any occurrence has one intended meaning (excluding again the cases mentioned above), and this is the lemmatization we have to find." The issue of length of context is taken up in section 5.

[3]Unless there is no meaning difference, as between nominal and adjectival readings for colour words, and then a bold type subentry is given.

[4]At this point we may query whether there are grounds for saying "What an awful mess!" and "This is another fine mess you've got me into" are particularly sense 1 and sense 2 usages, as the dictionary lays down (not that it is necessarily a flaw of the dictionary if they are not; see section 5.2).

[5]Alternations such as animal/food have been identified in Apresjan's and Atkins's work. An extension of this study would combine their more narrowly defined alternations and our wider-ranging ones, and would also integrate syntactically marked alternations such as the 'diathesis alternations' examined by Levin (1991).

[6]The reason half the corpus, and half the sample, were taken was so that if a model were to be built on the basis of the study of the sample in relation to the usages in the first half of the corpus, then the untouched halves would provide an environment for testing the model.

[7]As conveyed by the LDOCE grammar code [(*the*)S(**of**)].

[8]As conveyed by the LDOCE grammar code [*the*+S+**of**].

[9]We may also be tempted to ask the deeper question, what grounds does the lexicographer ever have for claiming that a sense is always used in a particular syntactic context? Where a comparable claim is made about a word (in all its senses) the lexicographer's evidence would be that all the corpus citations for that word were in that syntactic context. But to make the equivalent claim about one of a number of senses for a word, the lexicographer would have to present as evidence a corpus with all occurrences of the word under consideration sense-disambiguated. Yet such corpora do not exist and, as the example demonstrates, for a word such as *image*, it is highly implausible that one ever could. So there is reason to suspect that 'obligatory' grammar codes on word senses might often be better interpretated as 'expected' grammar codes.

[10]Where the user is using the dictionary for some purpose other than looking up words to determine their meaning, as when a non-native speaker uses it for generation, rather different considerations apply.

[11]No statistics are presented in relation to numbers of usages which could not be resolved because, firstly, the sample sizes were too small, and secondly, the question addressed is, for each word, a qualitative one; does the evidence permit, or does it not permit, that the word fits the Bank Model. Moreover statistics based on usages are highly susceptible to distortion, owing to the great contrasts between frequencies of different words, senses and usage-types.

[12]There could of course have been occasions where none of the LDOCE senses fitted. None of these occurred. This was in part because LDOCE is a well-researched dictionary and the sample sizes were small, but also because there is usually potential for re-interpreting a 'misfit' usage as a non-standard but acceptable use of the one sense. Only when there are two (or more) senses, so the question, 'is the misfit nearer the first or the second?' arises, does the misfit defy classification as one and only one sense. MacWhinney (1989)'s Competition Model of classification offers a theoretical account of the process. This treatment tallies with the problem faced by NLP. Where a word has

only one sense, the NLP system is not faced with the kind of problem the study addressed, even if, occasionally, the word is used in a way that only bears a remote relation to the dictionary entry.

Reference

Amsler, R. A. (1980). *The Structure of the Merriam-Webster Pocket Dictionary.* Ph.D. thesis, University of Texas at Austin.

Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, *142*, 5–32.

Atkins, B. T. S., Kegl, J., & Levin, B. (1986). Explicit and implicit information in dictionaries. In *Advances in Lexicography: Proc. Second Ann. Conf. of the UW Centre for the New OED*, pp. 45–65 Waterloo, Canada.

Atkins, B. T. S., Kegl, J., & Levin, B. (1988). Anatomy of a verb entry: from linguistic theory to lexicographic practice. *International Journal of Lexicography*, *1*(2), 84–126.

Atkins, B. T. S. (1987). Semantic ID tags: corpus evidence for dictionary senses. In *Proc. Third Ann. Conf. of the UW Centre for the New OED*, pp. 17–36 Waterloo, Canada.

Atkins, B. T. S. (1991). Building a lexicon: the contribution of lexicography. *International Journal of Lexicography*, *4*(3), 167–204.

Boguraev, B. K. & Briscoe, E. J. (Eds.). (1989). *Computational Lexicography for Natural Language Processing.* Longman, Harlow.

Briscoe, E. J., Copestake, A. A., & Boguraev, B. K. (1990). Enjoy the paper: lexical semantics via lexicology. In *COLING 90*, Vol. 2, pp. 42–47 Helsinki.

Choueka, Y. & Lusignan, S. (1985). Disambiguation by short contexts. *Computers and the Humanities*, *19*, 147–157.

Church, K. & Gale, W. (1991). Concordances from parallel text. In *Using Corpora: Proc. Seventh Ann. Conf. of the UW Centre for the New OED*, pp. 40–62 Waterloo, Canada.

COBUILD (1987). *The Collins COBUILD English Language Dictionary*. London.

Copestake, A. A. & Briscoe, E. J. (1991). Lexical operations in a unification-based framework. In Pustejovsky, J. & Bergler, S. (Eds.), *Lexical semantics and knowledge representation: ACL SIGLEX Workshop* Berkeley, California.

Cottrell, G. W. (1989). *A Connectionist Approach to Word Sense Disambiguation*. Pitman, London.

Fass, D. (1991). MET*: a method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, *17*(1), 49–90.

Hearst, M. 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*, pp. 1–22 Waterloo, Canada.

Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity*. CUP, Cambridge.

Kilgarriff, A. (1992). Inheriting polysemy. In *Proc. Second Seminar on Computational Lexical Semantics*, pp. 198–211 IRIT, Toulouse, France.

Krishnamurthy, R. (1987). The process of compilation. In Sinclair, J. M. (Ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing*, chap. 3. Collins, London.

LDOCE (1987). *Longman Dictionary of Contemporary English, New Edition*. Longman, Harlow.

Levin, B. (1991). Building a lexicon: the contribution of linguistics. *International Journal of Lexicography*, *4*(3), 205–226.

LOB. *The Lancaster Oslo Bergen Corpus.* Norwegian Computing Centre for the Humanities, Bergen, Norway.

MacWhinney, B. (1989). Competition and lexical categorization. In Corrigan, R., Eckman, F., & Noonan, M. (Eds.), *Linguistic Categorization*, pp. 195–241. John Benjamins, Amsterdam.

Martin, J. H. (1990). *A Computational Model of Metaphor Interpretation.* Academic Press, Boston.

Ostler, N. & Atkins, B. T. S. (1991). Predictable meaning shift: some linguistic properties of lexical implication rules. In Pustejovsky, J. & Bergler, S. (Eds.), *Lexical semantics and knowledge representation: ACL SIGLEX Workshop* Berkeley, California.

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

Sinclair, J. M. (Ed.). (1987). *Looking Up: An Account of the COBUILD Project in Lexical Computing.* Collins, London.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation.* OUP, Oxford.

Slator, B. M. & Wilks, Y. A. (1987). Towards semantic structures from dictionary entries. Tech. rep. MCCS-87-96, Computing Research Laboratory, New Mexico State University, New Mexico.

Small, S. L. (1980). *Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding.* Ph.D. thesis, Department of Computer Science, University of Maryland, Maryland.

Stock, P. F. (1983). Polysemy. In *Proc. Exeter Lexicography Conference*, pp. 131–140.

van Peer, W. (1989). Quantitative studies of literature: a critique and an outlook. *Computers and the Humanities*, *23*, 301–307.

Veronis, J. & Ide, N. M. (1990). Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In *COLING 90*, Vol. 2, pp. 389–394 Helsinki.

Wilensky, R. (1991). Extending the lexicon by exploiting subregularities. Tech. rep. UCB/CSD91/618, Computer Science Division, University of California, Berkeley.

Wilks, Y. A. (1975). A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence*, *6*, 53–74.