

© Copyright Springer-Verlag Berlin.

In A. Mehler & R. Köhler eds (2007) *Aspects of Automatic Text Analysis*. Berlin: Springer. 233-53.

INFERRING MEANING: TEXT, TECHNOLOGY AND QUESTIONS OF INDUCTION

MICHAEL STUBBS

Corpus linguists, including lexicographers, use methods which are often called 'inductive'. That is, they study large corpora or large data sets (such as word-frequency lists) derived from these corpora, in order to identify patterns in the data. There is detailed discussion of a few statistical techniques (e.g. for identifying significant collocations), but little general discussion of the combination of automatic and intuitive methods which are used to make significant generalizations. It might be thought that, if linguists draw generalizations from large data sets, then they would generally agree about the resulting analyses, and certainly corpus work often reaches a remarkably large consensus across different studies. Findings from one corpus are regularly corroborated by studies of other independent corpora, and partly automated or computer-assisted analysis has led to major progress in the study of semantic and pragmatic data.

However, given the title of this book, I should say immediately that I will argue that there can be no entirely automatic semantic analysis. First, I discuss the historical and logical background to the concept of induction, from Francis Bacon in the 1600s, via David Hume in the 1700s, to Karl Popper in the 1980s. The broad consensus from this work is that induction does not exist, and that there are no automatic methods which can be used to infer reliable generalizations from repeated individual observations. Second, I discuss some problematic examples from corpus-based dictionaries in order to illustrate the uneasy balance in corpus lexicography between automatic and intuitive methods. As examples of the results of inductive inferences (in the rough sense), I discuss some definitions in modern corpus-based dictionaries, and the extent to which these definitions agree or disagree. It might be thought that they disagree surprisingly often in their definitions of individual words.

1. CONVENTIONS AND TERMINOLOGY

I use the following conventions. "Double quotes" are used for the meanings of words and phrases. 'Single quotes' are used for quotes from other authors. Upper case is used for lemmas (lexemes). *Italics* is used for word-forms. For example, the word forms *crony* and *cronies* are two realisations of the lemma CRONY. I

will also make a distinction which is very simple and could avoid much confusion, but which is only rarely made (though see Rundell 2001). The terms 'introspection' and 'intuition' are often used synonymously, but there is a clear distinction between using introspection as data (as neo-Chomskyan linguists have done since the 1950s) and using intuition to identify interesting problems and to analyse data. I assume that all corpus linguists reject introspection as the (only) source of data, but that none deny the essential role of intuition in formulating hypotheses and analysing data. One thing which becomes very clear in teaching students to interpret corpus data is that the ability to see patterns (e.g. in concordance lines) takes practice: recurrent patterns are not obvious, and recognizing them is more like a skill than knowledge. This already throws doubt on any mechanical view of induction. The same is presumably true in any observational science: chemistry students also have to be taught what it is important to observe.

2. AN INTRODUCTORY EXAMPLE

Since the historical part of the paper may seem rather far from concrete linguistic data, I will start with a brief linguistic example of the general problem. Since the late 1980s, corpus studies have reinstated observation as central to linguistics on a scale previously unimaginable. Linguists can now work with data in the form of tens or even hundreds of millions of running words sampled from many different speakers and writers) or with large data-sets which are derived from these corpora with minimal human intervention. Now, one might think that, if linguists use such data and methods - in particular concordance software which allows large numbers of examples to be brought together and studied - then they would infer the same generalizations from the data, with a high degree of consensus across different studies. Findings from one study predict similar findings in other independent corpora, and it is certainly true that that corpus studies are regularly corroborated in this way.

One very large set of parallel findings is available in the form of definitions of words in corpus-based dictionaries. However, if we compare the definitions in different dictionaries, which have been produced independently but with similar methods, we discover that the dictionaries sometimes differ in the meanings (perhaps especially the evaluative connotations) which they attribute to words. For example, here are the definitions of CRONY in four corpus-based dictionaries, all published in 1995.

crony. Your cronies are the friends who you spend a lot of time with, an informal word. *Daily he returned, tired and maudlin, from lunchtime drinking sessions with his business cronies.* (Cobuild.)

crony. (informal often derogatory) A close friend or companion. *He spends every evening drinking in the pub with his cronies.* (OALD.)

crony. (usually plural) One of a group of people, who spend a lot of time with each other and will usually help each other, even if this involves dishonesty. *Nixon gave political power to many of his political cronies.* (LDOCE.)

crony. (informal especially disapproving) A close friend or someone who works with a stated and usually dishonest person in authority. *The General and his cronies are now awaiting trial for drug-smuggling.* (CIDE.)

Although these definitions have much in common, they differ considerably in emphasis. Cobuild gives a neutral denotation as "friend", and notes that the usage is informal. The example of drinking companions perhaps implies something "disreputable" (but the word gets no "PRAGMATICS" label: see below). OALD uses a very similar example with drinking companions, but explicitly warns that the word is often (but not always?) derogatory. LDOCE goes further and gives "dishonesty" as part of the denotation(?), with a citation which refers to one of the most notoriously dishonest political figures of all time. CIDE also gives "dishonesty" as part of the denotation, and gives a citation involving a major crime. In summary, the definitions range from neutral "friend", with only an implicit hint of "disreputable activities", via an explicit warning that the word is "derogatory", to the statement that it implies "dishonesty", if not serious "criminality" and abuse of political power on a major scale.

Now, what is a foreign learner to make of this? Can the word be simply informal and casual, or is it insulting and therefore to be used only with great care? There is no clear dividing line between the connotations given by the dictionaries, but consider what would happen if these meanings were translated into German, and then back-translated into English. This is not an artificial example: I recently asked a lecture class of around seventy largely German-speaking university students of English if they knew the word CRONY. Only ten or so claimed to know it, so it is precisely the kind of word they might want to look up in a dictionary.

These four dictionary definitions are four little theories of meaning. They are generalizations from the corpus data (plus the intuitions of the lexicographers), which raise questions such as: Which theory is correct? Which is wrong? How can we test such semantic claims? What would be the evidence that one is wrong? Could we provide counter-examples? Can we at least defend a preference for one definition over another? Would this preference depend merely on the most frequent usage (implying that minority usages are wrong)? Are all four definitions correct, but for different speakers? Or are they all wrong, because each one is too narrow, and does not take into account variable usage across speakers? The last questions imply that even corpus-based dictionaries, which seem the ultimate example of linguistic **d**escription, are in fact **p**rescriptive, since they do not fully take into account variation in usage. And should foreign learners at least be warned of this variation, that many native speakers regard the word as insulting, and that it may be safest to avoid it altogether?

Chomsky (1957: 51) early dismissed any attempts at automatic discovery procedures as far too demanding and quite unworkable for syntax, and therefore presumably all the more so for semantics. I will also argue here that any completely automatic procedures are impossible, (and that claims by corpus

linguists in this direction are sometimes exaggerated). But it is worth bearing in mind that the scepticism of such procedures in the 1950s and 1960s was at least partly due to the very restricted amount of data which could be prepared in machine-readable form, and of the failure of early statistical models (which were often tarred with the brush of failed early attempts at machine translation).

So, corpus linguists - including lexicographers - look at large amounts of data, observe recurrent patterns, and use these observations as evidence of meanings. Such methods, which use large data-sets to infer patterns, are often referred to as inductive, and they would seem to provide a much firmer empirical basis for linguistic description than the small amounts of introspective data used in neo-Chomskyan work. However, they raise several questions which are largely unresolved (and indeed hardly discussed).

3. SOME TRADITIONAL DISTINCTIONS

I assume that no-one these days believes in automatic methods which can reliably lead, purely objectively, from (repeated) empirical observations to significant generalizations. Intuition and inspired guesswork are always involved in selecting the initial data (e.g. designing a corpus), deciding what problems to investigate, and identifying the interesting patterns. Since corpora are typically designed according to a sociolinguistic theory of language variation, theory is involved from the beginning. Thus, even though corpus linguists often talk of 'raw data', and even though they have methods of avoiding some of the assumptions of pre-corpus grammar, I assume also that no-one these days believes in the possibility of a neutral observation language. (But see <http://www.linguistlist.org/>, July 2002, for a debate between Mukherjee and Pullum: disputes between these approaches are by no means settled.) These questions have a history of at least 2,000 years, and I can here make only a few traditional distinctions as a preface to concrete corpus examples. Deduction and induction are often distinguished as follows.

A deductive argument starts from premises, and draws conclusions which must be true if the premises are true. Given two premises, 'all students like beer' and 'Bertha is a student', then it must follow that 'Bertha likes beer'. Deduction concerns the validity of the conclusion, given the truth of the premises (that all students really do like beer, and that Bertha really is a student). It can say nothing about how the premises are established, or whether they are well defined (e.g. does alcohol-free beer count as beer, or if Bertha is a mature part-time student, does she count as a valid case?) Deductive logic concerns only the validity of the argument which relates premises and conclusion.

An inductive argument starts from a number of specific observations (hopefully a large and representative number) and proposes a generalization which is true of similar cases. Thus if all the students we have seen like beer, then we have reason to believe that other students like beer. However, the observation does not lead logically to this conclusion, and indeed it is not clear how far we can extend the generalization: to all students, to most, or only to many? Other problems include

whether our sample was a good one: perhaps the students we observed were not typical.

Deductive reasoning takes place within a closed system, in the sense that all the information is already contained in the premises: implications are merely made explicit by argument. It studies ways in which sentences follow logically from other sentences, and thereby relates propositions (premises) to other propositions (conclusions). The conclusions would be true in all possible worlds, because they depend on the meaning of the words. Deduction can lead to new knowledge only in the sense of a new perspective on old knowledge which is already contained (implicitly) in the premises. However, one of its weaknesses is that it can tell us nothing about the truth of the premises.

Inductive reasoning claims to go beyond the particular starting point to a generalization about cases which we have not observed. It relates individual observations (facts?) to general statements. This is its strength: we have confidence in the starting point since we have observed these cases to be true, and it tells us something new by going from the particular to the general. However, this is also its weakness, since we cannot be certain about what we might see in the future. In addition, a generalization is not an explanation: it says we will observe more of the same, but does not explain why. (We might predict that the next student we meet also likes beer, but we do not know whether this is due to peer pressure, students having more money and spare time than is good for them, their needing alcohol due to depression brought on by too much work, or to depression brought on by too little work and subsequent fear of exams.)

Deduction and induction are often assumed to be symmetrically related. Deduction starts from premises, and goes from the general to the specific. Induction starts from observations, and goes from the specific to the general. However, this opposition is only apparent. Both deduction and induction assume reliable starting points: either self-evident premises or observed facts. Deductive logic simply assumes the truth of the premises, and treats only their consequences in a possible world. But inductive logic assumes the reliability of the initial observations in the real world, and since all observations are already interpretations and open to all kinds of potential errors, they can never be certain. If the initial evidence is unreliable, then the conclusions cannot be reliable (Popper 1983: 221-23).

When we talk of drawing generalizations from a finite sample of observations, we ought to distinguish between three rather different situations. First, if we have simply observed all the members of a group, then it is quite possible to summarize the observations quantitatively (e.g. 90 per cent of this class of 100 students like beer), and this is not open to any problems of generalizing to a larger population. Second, if we repeatedly observe some phenomenon (e.g. students drinking beer), we might think it likely that we will see further similar cases in future, without making any claims about numbers or proportions. That is, we will have precedents for such similar cases. (This is sometimes called 'eduction'.) Third - and this is the

difficult case which is usually meant by induction - we can observe only a (small) finite sample of a (very large) population, but we wish to make predictions about this large open-ended population.

4. SOME HISTORY

Modern ideas about inductive reasoning are often traced back to the early 1600s, when Francis Bacon argued that scientific progress must be based on systematic data collection and observation (though he himself admits that some of his main points had been made by Plato). Bacon rejected dogma and authority as sources of knowledge, and criticized deductive reasoning as being similar to spiders making webs of knowledge out of their own substance (Quinton 1980: 26, 55). In its stead, he proposed methodically recording observations, and then proceeding gradually and cumulatively towards general principles. He also clearly understood the difference between positive and negative observations, commenting that 'major est vis instantiae negativae' (the force of the negative instance is greater):

The human understanding [...] forces everything to add fresh support and confirmation; and although more cogent and abundant instances may exist to the **contrary**, yet either does not observe or despises them [...]. It was well answered by him who was shown in a temple the votive tablets by such as had escaped the peril of shipwreck, and was pressed as to whether he would then recognise the power of the gods, by an enquiry; 'But where are the portraits of those who have perished in spite of their vows?' [...] It is the peculiar and perpetual error of the human understanding to be more moved and excited by **affirmatives** than by **negatives**, whereas it ought duly and regularly to be impartial; nay, in establishing any true axiom, the **negative instance** is the **more powerful**. (Bacon *Novum Organum*, 1620, aphorism 46. Emphasis added.)

This is the point about the asymmetry of confirming and falsifying data that Popper much later built into a demarcation criterion for science. A confirming instance is just one more instance which neither proves, nor even makes more probable, a conclusion, since one single counter-example may falsify a hypothesis. One cannot prove that a statement is true (unless in the closed fashion of a simple deduction), but one can prove that a statement is false. (Though see Popper on why he is not a 'naive falsificationist': e.g. 1983: xxxiii.)

Not everyone has shared the view that Bacon's ideas were particularly original. In a very long, often very funny, review of an edition of Bacon's works (which is often a vitriolic attack on Bacon's character, rather than a review of the book), Lord Macaulay (1837) questions the originality of the concept of induction. Here is one of his more ironic statements:

The vulgar notion about Bacon we take to be this, that he invented a new method of arriving at truth, which method is called Induction. [...] The inductive method has been practised ever since the beginning of the world

by every human being. It is constantly practised by the most ignorant clown, by the most thoughtless schoolboy [...]. That method leads the clown to the conclusion that if he sows barley he shall not reap wheat. By that method the schoolboy learns that a cloudy day is the best for catching trout. [...] Not only is it not true that Bacon invented the inductive method; but it is not true that he was the first person who correctly analysed that method. (Macaulay 1837: pp.406-8).

Macaulay continues in this vein for another page or two. (Quinton 1980 provides a more balanced short account of Bacon's ideas.) Bacon is also standardly criticized for his naive faith in the possibility of reliable unbiased observation (he thought that we could start from our intuitions (NB!) that some observations and sense-perceptions were self-evident), and for his insistence on the careful and rather timid plodding accumulation of data, as opposed to the leaps of imagination and guesswork which lead to real progress in science (Popper 1983: 222-23).

In 1758, David Hume had already made the point that there is nothing new about the idea that we learn from experience: 'none but a fool or madman will ever pretend to dispute the authority of experience' and 'it is certain that the most ignorant and stupid peasants - nay infants, nay even brute beasts - improve by experience'. Hume admitted that we cannot avoid jumping to inductive conclusions. It is an unavoidable mental habit, and a perfectly reasonable thing to do, indeed often the only thing to do. But this is a matter of necessary everyday custom and habit, not of logic. He makes these comments in his famous discussion of different types of inference, where he distinguishes clearly between the psychological certainty which induction seems to bring, and the impossibility of inductive generalizations providing logical certainty, since any predictions about the future are open to potential counter-example. Any observations we have made were made in the past, but there is no logical reason to assume that the future will resemble the past, since there can always be new cases and new observations:

As to **past** Experience, it can be allowed to give direct and certain information of those precise objects only, and that precise period of time, which fell under its cognizance: but why this experience should be extended to **future** times, and to other objects, which for aught we know, may be only in appearance similar; this is the main question on which I would insist. [...] The consequence seems nowise necessary. [...] If we be, therefore, engaged in arguments to put trust in **past** experience, and make it the standard of our **future** judgement, these arguments must be **probable** only [...]. All our experimental conclusions proceed upon the supposition that the **future** will be conformable to the **past**. (Hume 1758: Section 4. Emphasis added.)

Hume continues in this vein for another page or two, making the same point several times in slightly different words, that all inferences from experience assume 'that the future will resemble the past', but that there is no proof that this will be so, since 'the course of nature may change'.

Hume has been read in two rather different ways. He is traditionally interpreted as simply rejecting induction as a rational procedure. Alternatively, he is interpreted as merely arguing that induction must be rejected as a rational procedure, only if reason is interpreted in very narrow deductive way. These different readings of Hume himself are not of direct concern in this article, but Noonan (1999: 116-31) summarizes the various positions and argues for this second view.

In the 1960s, these points from Bacon and Hume were developed by Popper (1983), who refers (p.62) rather ironically to Hume's 'problem of tomorrow', which he regards as a simple philosophical muddle. First, it is rather likely that the future will not resemble the past. Second, Hume thought there was a paradox in saying both that the laws of nature may change, and also that the laws of nature are just these things which we think can never change. Popper points out that there is no paradox: it simply means that we formulated the laws wrongly in the first place, and shows again that our theories are always open to correction and counter-example. Popper (e.g. 1983: 11-158) agrees with Hume (pp.31-32) that there are countless regularities in nature on which we rely in practice, but that we cannot logically reason from singular observations to general laws of nature. This landed Hume in what he saw as another clash between the invalidity of induction and the principle of empiricism. Since Hume was unwilling to abandon empiricism, he concluded pessimistically that we have to rely on habit, but cannot rely on reason, and this drove him into an irrationalist position. Popper (pp.32-33) accepts both the argument against induction and the principle of empiricism (that theories are accepted or rejected on the basis of observational evidence), but changes the role which observation plays. Observation is essential, but it cannot prove that a theory is true, only that it is false. Popper argues at length that there is no such thing as induction, since hypotheses are always provisional conjectures (a) which are influenced by prior knowledge and expectations, and (b) which may turn out to be false due to refutation by counter-example.

So there are in fact three stages which should be distinguished. The problem is how we get (1) from exploratory data analysis (2) to hypotheses (3) which we can test. First, we explore a mass of messy data (e.g. a large corpus plus associated concordances, word lists, statistics on frequent collocations, etc) and these facts somehow suggest a theory. But how we arrive at the theory (argues Popper) is irrelevant to its possible truth. In our search for patterns we certainly get ideas from observations, but we never draw true inductive inferences, since we always start from expectations. Second, we formulate generalizations and hypotheses (e.g. dictionary entries about word meaning). These hypotheses do not emerge from pure logic, since they also depend on the categories which we use to classify and interpret the world (e.g. a distinction between denotation and connotation) and on our assumptions (e.g. that words have relatively stable meanings in a speech community). Third comes a process of formulating and testing consequences from these hypotheses. With reference to dictionaries, I am not sure if such testing is ever carried out systematically, or indeed whether it could be carried out in practice across large comprehensive dictionaries.

In summary, I think there is nowadays general acceptance of Popper's view that there is no such thing as pure induction. McGuire (1999: 399) provides a useful summary: knowledge is always an underrepresentation (since there is always selective attention to data), a misrepresentation (since it is influenced by the knower), and an overrepresentation (since it is based on inferences which go beyond the given data).

5. SOME LEXICOGRAPHIC EXAMPLES

These questions of research method are unresolved despite some 400 years of intense discussion. Phrased rather negatively, it would be valuable if corpus linguists were at least more aware of these questions. Words such as *deduction* and *induction* do not appear in several widely-used introductions to corpus methods (including my own, I must admit: Stubbs 1996, 2001). Phrased positively, corpus linguists could use their unique combination of very large data sets, computer-assisted quantitative methods and human intuition to make some conceptual progress on the problem. (Relevant methods, especially cyclic procedures of data analysis, have been described by Sinclair (1991), Sinclair, Mason et al (1998), and Barnbrook (2002).) So, from a rather abstract historical discussion, I return now to the concrete questions I raised at the beginning.

Below I will be using the codings for pragmatic connotations used by the Cobuild (1995) dictionary. In an excellent article, Channell (2000), the linguist who developed the pragmatic coding framework for the dictionary, discusses the methods used to discover these connotations, and provides clear examples of how the lexicographers worked in practice. The essential method involved using concordance lines to display recurrent patterns in the use of a given word. Channell then illustrates how statements about evaluative connotations can be 'based in systematic observation', which makes it possible to 'produc[e] a sound description' (p.39). Concordance data provide facts which are not accessible to introspection, and are 'not visible from the study of single examples' (p.40). The main part of the article makes no simplistic claims about automatic methods (and does not use the term 'induction' at all). However, the summary section at the end makes claims which cannot be taken literally: '*Without recourse to intuitions*, quantitative data show clear evidence of where there is an evaluative polarity to an item' (p.54, emphasis added). This implies more automatism than is possible, and a more guarded statement would be more accurate: that the concordance software can be instructed to find the appropriate data (possibly with further help from the kind of software described by Sinclair et al 1998), but that these data still require the lexicographers' intuition to extract the significant patterns.

Given the broad historical discussion so far, the following section may seem disappointingly modest, but it does ask a specific question: How far do dictionaries agree in their definitions of words? Lexicographers have large corpora and associated data-sets, plus the hermeneutic procedures described by Channell (2000). Do these procedures lead to consistent results? Inter-subjective

agreement would not of course prove that the analysis is correct: the analysts may all have been misled in the same way, Hume would point out that 'the course of nature may change', and Popper would point out that corroboration does not even increase the probability of a generalization being true. But disagreement would point to a potential problem.

We have no independent statements of what the meaning of a word is, but we can compare definitions of the same words in the four highly comparable dictionaries used above for comparing the definitions of CRONY. They are all corpus-based, all intended for advanced foreign learners, and all published in 1995:

CIDE: Cambridge International Dictionary, based on the Cambridge Language Survey corpus of 100 million words

Cobuild: Collins Cobuild English Dictionary, 2nd ed, based on the Bank of English corpus of over 200 million words

LDOCE: Longman Dictionary of Contemporary English, 3rd ed, based on the Longman Corpus Network and the British National Corpus of 140 million words

OALD: Oxford Advanced Learner's Dictionary, 5th ed, based on the British National Corpus of 100 million words, and an American English Corpus of 40 million words.

The remainder of the paper discusses cases where these dictionaries disagree in their analyses of individual words, and therefore discusses the uneasy balance in corpus studies between automatic and intuitive methods.

6. THE FREQUENCY OF DISAGREEMENTS

I will start with the simplifying assumption that dictionaries may tend to agree most easily over the denotations of words, but less easily over their connotations, (though I will also question below whether this distinction can be maintained). All(?) dictionaries use labels of one kind or another (such as "informal" and "derogatory") for evaluative connotations, but these labels are notoriously unstandardized. Cobuild (1995) attempts to systematize such description by placing the label PRAGMATICS next to words which have features of usage which need to be specially signalled. This label is used for several rather distinct purposes, including conversational markers (e.g. *anyway*), and words where the core semantic meaning already denotes something good or bad (e.g. *dreadful*). However, the dictionary also uses a range of syntactic forms (discussed in detail by Barnbrook 2002), in order to explicitly mark evaluative connotations and speaker attitudes with phrases such as 'if you say x, you want to emphasize it', or 'if you say x, you approve/disapprove of it', or 'if you say x, it is because you are irritated'. If we take words labelled in this way, and compare their definitions in the four dictionaries, we would have an initial rough sample for discussion. So, from these four corpus-based dictionaries, a sample of words and phrases was selected as follows:

- (1) I started on Cobuild page 5, and took the first word (i.e. word sense) with the PRAGMATICS label, where the entry has a further explicit evaluative descriptor as defined above.
- (2) I selected words in this way every hundred pages (pages 5, 105, 205, etc). If there was no such word on the page, I went to the next page.
- (3) I compared the definitions in the four dictionaries. (It was of course possible that the given sense was not listed at all in a dictionary.)

Given my starting point, all entries selected from Cobuild have, by definition, an explicit evaluative label. However, I do not distinguish further between the dictionaries below. (And the main comparisons are amongst only the 1995 printed versions of the four dictionaries. Different printed and CD-ROM editions of the dictionaries, between 1995 and 2000, often have distinct differences of emphasis in their definitions and give different citation examples.) Here are some brief comments on each of the words in this small sample.

1. page 6: **absolute** (sense 2)

All four dictionaries explicitly label this word as "emphatic" or "expressing a strong opinion".

One dictionary adds that it "emphasizes your opinion [...] especially when you think [that something is] very bad, stupid", etc (*absolute disgrace*). But the other three dictionaries give a mixture of negative, neutral and positive citations (*absolute nonsense*, *absolute minimum*, *absolute trust*).

2. page 106: (do something) **behind someone's back** (*back* sense 10)

All four dictionaries give the denotation of "doing something without a person's knowledge or agreement".

But two of the dictionaries add that it is "disapproving" (e.g. *saying nasty things behind his back*).

3. page 205: **brood** (sense 2)

All four dictionaries give the denotation of "a family of young children".

One dictionary adds: "when you want to emphasize that there are a lot of them". The other three dictionaries label the usage "jocular" or "humorous".

(For what it is worth, my intuitive judgement is that it could be rather risky and potentially insulting to use the word "humorously".)

4. page 305: **cohort** (sense 1)

All four dictionaries give the denotation of a person's companions or supporters, and explicitly label it as "disapproving" or "derogatory" (i.e. rather similar to *cronies*). Three of the dictionaries label this usage as especially American, or give a citation which implies an American usage.

5. page 405: **cut something out** (sense 4)

All four dictionaries agree that if you tell someone to *cut it out*, it is because their behaviour is "annoying" and/or you are "irritated".

6. page 505: **would not dream of** (*dream* sense 9)

All four dictionaries either label this explicitly as "emphasizing" that the speaker would never do something because they think it is (morally) wrong or unpleasant, or imply this in the citation.

It is not strictly within my comparisons, however another dictionary gives no such restriction: *I wouldn't dream of going without you*. This seems more accurate than the four dictionaries under comparison.

7. page 605: **far be it from me** (*far* sense 17)

Two dictionaries give a neutral gloss: "I certainly would not want to do this". The other two agree that the speaker is about to criticize someone, but one thinks that the speaker wants to appear hostile, whereas the second thinks that s/he wants to pretend to agree.

8. page 705: **gerrymandering**

All four dictionaries agree that the word denotes "altering political boundaries to give advantage to a political party". All use the word "unfair", which implies speaker attitude, and two dictionaries explicitly label it "disapproving" or "derogatory".

9. page 805: **hole** (sense 5)

For the sense of referring to a place (e.g. usually where someone lives) as a *hole*, all four dictionaries label the usage as "unpleasant" and "informal".

10. page 905: **not a jot** (*jot* sense 2)

Two dictionaries label the phrase "old-fashioned".

One labels it "informal" (which seems logically inconsistent with "old-fashioned"), and one has neither of these labels.

11. page 1007: **on the make** (*make* 3 sense 9)

Three dictionaries label the phrase "disapproving".

One of these goes further and labels it as possibly implying "illegal and immoral" methods. But one dictionary labels the phrase merely "informal".

12. page 1106: **nepotism**

All four dictionaries agree on the denotation: using power to gain advantage for friends or relatives.

Two dictionaries label this practice "unfair" and as signalling the speaker's "disapproval". The other two do not use any such evaluative descriptors: everything is left to implication (see below).

13. page 1205: **party politics** (sense 2)

Here the differences between the dictionaries are not entirely distinguishable from the issue of whether a word has distinct senses: a literal denotation (= "relating to political parties") and an extended usage which is an "accusation" and which "criticizes" people for doing or saying something which they do not believe.

Only one dictionary gives these attitudinal labels.

Two dictionaries give only the first literal sense. One gives only the second sense. One gives both.

14. page 1306: **principled**

One dictionary labels the word "approving"; the other three dictionaries imply this with phrases such as "honest and moral", or "esp good".

Only one dictionary labels it "formal".

15. page 1405: **repetitious**

All four dictionaries label the word "disapproving" or give citations with the word *boring*.

16. page 1505: **self-important**

(This item occurred by chance in my small sample: it is one of the examples discussed in detail by Channell 2000.)

Three dictionaries label the word "disapproving". The fourth implies this in the citation (*a self-important, pompous little man*).

17. page 1605: **spendthrift**

All dictionaries give denotations such as "spending money wastefully or extravagantly". Two dictionaries label it "disapproving". The other two use words such as "careless" and "wastes money", and seem to assume that the evaluative connotations of these words are clear.

18. page 1705: **talk down to**

All four dictionaries agree on the denotation: talking to someone as if they are not very intelligent. One dictionary explicitly labels the phrase "disapproving". The other three dictionaries imply this with phrases such as "too simple".

19. page 1805: **two-dimensional** (sense 2)

One dictionary labels the phrase "critical"; a second labels it "disapproving"; a third implies this with "not very interesting"; the fourth implies it even less directly: "does not seem real".

20. page 1905: **whatever you say** (*whatever* sense 7)

Two dictionaries agree that this means "you do not believe or accept what someone has said" or "do not really agree" with someone. The other two dictionaries have no entry for the phrase.

7. INTERPRETATION

Dictionaries can differ widely in how they represent word-meanings, including how they divide word-meanings into different senses, whether they present senses as separate or as specialized cases of a more general meaning, and so on. (See Kilgarriff 1993, 1997, who concludes 'I don't believe in word senses'.)

My topic here is more specific: the extent of agreement over attitudinal meanings. First, dictionaries differ in the labels they use, simply because there are no standard terms for presenting pragmatic information, but only a rather small and crude set (e.g. "emphatic" and "formal"), which are not always clearly distinguishable (e.g. "disapproving" and "derogatory"). Second, dictionary entries differ in whether connotations are explicitly labelled, or only implicitly encoded in value-loaded words. In examples above, definitions use words such as *unfair*, which certainly implies disapproval. Third, there are cases which Channell (1995) argues are more serious than this type of implication. Consider the definition in one dictionary (Channell's point, but my example) of *nepotism* as "the practice of giving the best jobs to members of your family when you are in a position of power". The lexicographer presumably takes it for granted that this practice is a bad thing, and has (covertly) encoded something that s/he disapproves of. However, there is no evaluative label, the definition contains positive words (*best*, *power*), and some readers might regard it as behaviour which is completely rational and only to be expected. In another case, as a citation for *principled*, one dictionary gives *a principled stand against federalism*, which seems to imply that federalism is a bad thing (and that the lexicographer may be sharing particularly British prejudices against this form of government)!

This set of twenty definitions is a very small, and clearly not a random, sample. Given my starting point, all the Cobuild definitions obviously contain explicit

descriptors of speaker attitude. Nevertheless, this small sample shows a surprising number of differences between the dictionaries. Depending on how strictly one interprets the conventions used by the four dictionaries, there is good agreement across all four of them in perhaps half of the cases. There is one case where two dictionaries did not have the word or phrase at all, and in the other ten or so cases, there is either distinctly different information (e.g. the same phrase is labelled "old-fashioned" and "informal"), or at least distinctly different emphases in the connotations given (e.g. a speaker wants to appear "hostile" or wants to "pretend to agree"). How might one explain these differences?

(1) Perhaps the lexicographers looked at different corpora, which were not entirely comparable samples, and drew legitimate generalizations from different data. In the case of relatively infrequent words and their even less frequent combinations in longer phrases (e.g. *brood?*), perhaps there were not enough examples even in a large corpus to allow valid generalizations. This would throw doubt on claims that the dictionaries are based on representative samples of language use. This explanation is, however, not very convincing, since all four dictionaries are based on large mixed corpora.

(2) A second explanation might be that native speakers do not always agree in their use of words, (which would explain why corpus samples differ), and that the search for meanings shared across a discourse community is misconceived. This could imply in turn that language use is more variable than is admitted in dictionaries of general English. This explanation sounds superficially similar to the frequently heard excuse in neo-Chomskyan linguistics of why native speaker intuitions often fail to agree: 'it's grammatical in my dialect'. However, the reasons for this variability in word meanings may be more interesting: I return to this below.

(3) This could lead to a third, methodological, problem. If analyses are based on highly variable data, they should logically be formulated as probabilistic statements. First, dictionaries only seldom adopt this strategy (cf Cobuild 1995 on *foreigner*: "some people believe this word is slightly offensive"). Second, it is difficult to see how such definitions can be refutable in any clear way, since the concept of individual counter-examples does not apply. And, third, if variable data are given a categorial description, this implies that the dictionary is prescribing one usage as correct. If there is a difference between this meaning and my usage, does this imply that I am using the word wrongly?

Barnbrook (2002: 38-39) summarizes some of the main issues. Is meaning in the mind of the lexicographer or in the usage of the speech community? Do lexicographers decide, on the basis of their native speaker competence, what the meaning of a word is, and then search for corpus examples which illustrate this meaning? (Is the lexicographer the source of the meaning?) Or do they discover the meaning in the corpus data (induce it from the data)? In what sense can semantic information be '**derived**' from '**reliable** sources' and '**based directly** on representative corpus data' (Barnbrook 2002: 46, emphasis added)? In addition,

does a word have a 'correct' meaning, which can be illustrated by such data and then recorded in dictionaries?

8. A FINAL EXAMPLE: SIZE OF CONTEXT

There is a fourth possible explanation of disagreements between dictionaries. Here is a final illustration of the hermeneutic method, but also of a problem, which is simple in this case, but unresolved in general: how much context (co-text) is relevant to deciding meaning, especially connotation? Here are some examples (from the 100-million word British National Corpus) of the word *horde*, followed by *of* and a noun phrase. At first sight, there seem to be examples of both neutral and negatively evaluated uses, respectively:

horde of children; horde of courtiers; horde of souls; horde of tiny crablets; horde of young girls; horde of young men; horde of volunteers

horde of disturbed bats; horde of goblins; horde of hooligans; horde of the damned from hell; horde of troublesome workmen

However, collocates in a larger context reveal the apparently neutral and positive examples in (a) as often very negatively evaluated indeed:

horde of children: 'He left it [his carriage] [...] where it attracted the interest of a **horde of children** [...] Garbage was piled high in corners, and Maggie watched Sarah stepping carefully so as not to tread in the filth.'

(A second example of *horde of children* collocates with *awfully crowded*. An example of the plural *hordes of children* collocates with *a slum*.)

horde of courtiers: '[T]he scene was one of frenetic confusion, servants scurrying around, shouting and gesticulating [...] the situation was not improved by a **horde of courtiers** standing around also issuing their instructions to a vast army of retainers [...].'

horde of souls: '[A] vast **horde of souls** were rumbling towards heaven. There were whole companies of white-trash, clean for the first time in their whole lives, and bands of black niggers in white robes, and battalions of freaks and lunatics shouting and clapping and leaping like frogs'.

(These three examples are all from fiction: the first two from novels by Pamela Pope and Doherty Crown, the third from a famous short story by Flannery O'Connor.) Similarly, the *horde of tiny crablets* is described as having an extravagant and wasteful breeding strategy; the *horde of young girls* is mobbing a limousine carrying a pop group; the *horde of young men* is a critical reference to the speaker's rivals for the attentions of a young lady; and the *horde of volunteers* refers to volunteer soldiers in an incident in Scottish history.

In this case, the four dictionaries agree on the unpleasant connotations:

horde: A large, noisy and excited crowd. *A horde of students on bikes made crossing the road difficult.* (CIDE)

horde: (sometimes derogatory) A very large group, especially of people; a huge crowd. *Fans descended on the concert hall in their hordes.* (OALD)

horde: A large crowd moving in a noisy uncontrolled way. *Hordes of people milling around the station.* (LDOCE)

horde: If you describe a crowd of people as a horde, you mean that the crowd is very large and excited and, often, rather frightening. ... *a horde of people was screaming for tickets.* (Cobuild)

The four definitions, plus the citations, differ in emphasis, ranging from a crowd which is "excited" (not necessarily a bad thing), "noisy" (sounds like a nuisance), *milling around* (implies aimless, useless activity), "uncontrolled" (might be dangerous), to causing "difficulties" and "frightening". That is, they differ in how explicit the connotations are made and how strong they are claimed to be. In terms of methodology, we do not know exactly how much context the lexicographers have used in phrasing their definitions. More fundamentally, we do not know when connotations should be made part of the denotation, or when they should be generated by inference from common-sense knowledge (that excited, aimless and uncontrolled crowds can become dangerous).

9. A FINAL ATTEMPT AT EXPLANATION

Given practical constraints (including restrictions of space, the conservativeness of their users, and ultimately commercial pressures) dictionaries are forced to present their definitions as though words have definitive meanings and/or distinct senses. This is, however, a misleading model of semantics. Sampson (2001: 180-207) presents a detailed account of why this is so, and why linguistic and encyclopedic knowledge cannot be neatly separated. Due to cultural changes and the need for new concepts, the meanings of words change, many words encode concepts which have been recently institutionalized in our culture, and due to their different experiences and cultural beliefs, speakers simply reach different conclusions about word meanings. Meanings are therefore unpredictable and creative and, in this sense, Hume is right: all our observations were made in the past, but there is no logical reason to assume that the future will resemble the past.

Perhaps the problem of recording attitudinal meanings is not then, after all, separable from the question of how many distinct senses of a word a dictionary should list. Pustejovsky (1991) also argues that linguistic and encyclopedic knowledge cannot be neatly separated, and formalizes a position which seems close to Sampson's. He argues that words do not have a fixed number of distinct senses, but that these senses can be generated by a fixed number of rules which operate in context. For example, we know that *Susan has finished her book* can be interpreted in two ways ("finished reading" or "finished writing"). This does not

mean that *finish* has two senses, but that we know things about books (and about Susan, for example, whether she is an avid reader or a budding author).

Take again the case of CRONY. Speakers draw inferences about the meanings of words, and these inferences are based partly on what they know about people's behaviour (or possibly men's behaviour: do women have cronies?). If people (men?) are friends, then they spend time together, often go drinking together, form alliances, acquire commitments to each other, and may then be led into supporting dubious, and sometimes criminal, activities. That is, the word makes sense only within larger cultural frames, which generate implications. The same points hold for words and phrases such as *brood* (how large is a large family?), *nepotism* (when does helping one's relatives become illegal favouritism?), or *party politics* (when does working for one's own party turn into ignoring the general good?). Since the meanings depend on assumed common sense knowledge of cultural schemas, and on shared evaluations of what is mildly disapproved of versus regarded as immoral or illegal, they cannot have fixed meanings.

These observations start to explain why it is particularly difficult to handle pragmatic meaning in dictionary entries, but they do not solve the problem. Indeed, they produce two different problems. (1) When do such inferences become conventionalized? When do connotations become part of the denotation, and therefore when do words such as CRONY become insulting independent of context? (2) If words make sense only within larger cultural frames, then this implies that their meanings should be represented within the kind of frame semantics which has long been recommended by Fillmore (e.g. Fillmore & Atkins 1994).

10. CONCLUSIONS

Despite his overall argument for 'empirical linguistics', Sampson's (2001: 181, 206) conclusion about semantics is pessimistic and uncompromising:

[O]ne might expect to find [...] areas which cannot be treated scientifically at all. The outstanding example [...] is word meaning. [...] [A]nalysis of word meaning cannot be part of empirical science. [...] Word meanings are not among the phenomena which can be covered by empirical, predictive scientific theories.

It would be more helpful to distinguish between different stages of research. It may be misguided to try and 'establish a rigorous scientific analysis of word meanings' (Sampson 2001: 197). Nevertheless, there are ways of collecting empirical observational data on meaning, even if there can be no automatic analysis of the data. There are automatic procedures which can select data and put them into a convenient form for the human analyst, but the interpretation of these data can be only partly automated or computer-assisted, since it requires the intuition and experience of the lexicographers.

In a fascinating book, Macfarlane and Martin (2002) argue that a specific technological discovery decisively influenced what we often call the Scientific Revolution, that is, the development of modern scientific methods and findings from the 1600s. This was the discovery of how glass could be used to make scientific instruments. Their most telling points are (1) that, without glass, we would have no test-tubes, retorts, thermometers or barometers, and that many forms of experiment and measurement would be impossible, and (2) that, without glass, we would have no lenses, and therefore no magnifying glasses, microscopes, telescopes or cameras, and no instruments for observing small things, distant things and fast-moving things. As a consequence, micro-biology, astronomy and many other disciplines could simply not have developed. In other words, there was a close relation between the development of a material technology, a scientific method and a whole mode of thought. Glass instruments made it possible to collect many new observations., and the authors point out that the telescope and the microscope were invented in the late 1500s, only a few years before Francis Bacon was making his points about empirical methods: his *New Atlantis* was written around 1610. They carefully point out that glass is an enabling, and probably necessary, technology for the development of science: as they say, it is the only substance which directly influences how humans see the world, and reveals things which were previously 'invisible to the naked eye' (p.81). However, they emphasize that it was not a sufficient cause and there were many other factors involved in the development of a sceptical scientific method.

How does this relate to corpus linguistics? Computer-readable corpora and access software also allow linguists to see things that they have never seen before. They are no longer restricted to observing their own individual introspections or short individual texts ('the extent of language that can comfortably be accommodated on the average blackboard': Phillips 1989: 8), but can now observe large-scale patterns of language behaviour across large text-collections, which are evidence of the mental lexicon of thousands of speakers across a speech community. Technologies alter what can be observed, suggest problems, make scholars satisfied with particular answers, and therefore alter descriptions and theories. For linguistics, this was true of the invention of written language and of tape-recorders. Computers are particularly good at repetitive tasks, and it is types of repetition which are particularly significant in corpus semantics. Since corpora plus software are now one of the 'technologies of the mind' (Macfarlane & Martin 2002: 31) of contemporary linguistics, linguists need to be clear about the relations between these technologies and their observations, generalizations and theories.

ACKNOWLEDGEMENTS

For help with data collection I am grateful to Bettina Starcke. For discussion of some of the data and of issues relating to practical dictionary-making, I am grateful to Joanna Channell (one of the Assistant Editors responsible for the

Pragmatics coding in Cobuild 1995) and Gwynneth Fox (Editorial Director of Cobuild 1995).

REFERENCES

- Bacon, Francis (1620) *Novum Organum*. (Translation by B. Montague, Philadelphia: Parry & MacMillan, 1854.)
- Barnbrook, G. (2002) *Defining Language*. Amsterdam: Benjamins.
- Channell, J. (1995) Cultural assumptions and shared knowledge in English dictionaries. Unpublished lecture, University of Trier.
- Channell, J. (2000) Corpus-based analysis of evaluative lexis. In S. Hunston & G. Thompson eds *Evaluation in Text*. Oxford: Oxford University Press. 38-55.
- Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.
- CIDE (1995). *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- Cobuild (1995). *Collins COBUILD English Dictionary*. London: HarperCollins.
- Fillmore, C. J. & Atkins, B. T. S. (1994). Starting where the dictionaries stop: the challenge of corpus lexicography. In B. T. S. Atkins & A. Zampolli (eds.) *Computational Approaches to the Lexicon*. Oxford: Oxford University Press. 349-93.
- Hume, David (1758) *An Enquiry Concerning Human Understanding*. Edinburgh. (Harvard Classics, Collier & Son, 1910.)
- Kilgarriff, A. (1993) Dictionary word sense distinctions. *Computers and the Humanities*, 26: 365-87.
- Kilgarriff, A. (1997) "I don't believe in word senses". *Computers and the Humanities*, 31: 91-113.
- LDOCE (1995). *Longman Dictionary of Contemporary English*. London: Longman.
- Macaulay, Lord (Thomas Babington) (1837) Lord Bacon. In *Critical and Historical Essays*. London: Longmans, Green & Co. 1878. 280-429.
- Macfarlane, A. & Martin, G. (2002) *The Glass Bathyscaphe*. London: Profile.
- McGuire, W. J. (1999) *Constructing Social Psychology*. Cambridge: Cambridge University Press.
- Noonan, H. W. (1999) *Hume on Knowledge*. London: Routledge.
- OALD (1995). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Phillips, M. (1989) *Lexical Structure of Text*. Discourse Analysis Monograph 12. English Language Research, University of Birmingham.
- Popper, K. R. (1983) *Realism and the Aim of Science*. London: Routledge.
- Pustejovsky, J. (1991) The generative lexicon. *Computational Linguistics*, 17, 4: 409-41.
- Quinton, A. (1980) *Francis Bacon*. Oxford: Oxford University Press.
- Rundell, M. (2001) "Corpora, evidence and intuition". *Corpora List*, 1 November 2001. <http://helmer.hit.uib.no/2001-4/0080.htm>.
- Sampson, G. (2001) *Empirical Linguistics*. London: Continuum.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J., Mason, O., Ball, J. & Barnbrook, G. (1998) Language independent statistical software for corpus exploration. *Computers and the Humanities*, 31: 229-55.
- Stubbs, M. (1996) *Text and Corpus Analysis*. Oxford: Blackwell.
- Stubbs, M (2001) *Words and Phrases*. Oxford: Blackwell.