3 Methods in corpus linguistics: Interpreting concordance lines

Introduction

This chapter is the first of two that present some of the most commonly used methods and approaches in corpus linguistics. Producing concordance lines is perhaps the most basic way of processing corpus information, and most corpus users rely heavily on concordances and their interpretation. This is particularly true for those who are using a corpus in day-to-day teaching or translation, for whom an intuitive response to data may be more immediately useful than a more statistical approach. For this reason, concordance lines have a chapter to themselves in this book.

This chapter offers a number of examples of corpus searches, each one illustrating one or more points about the methodology of finding and interpreting concordance lines, and about the kind of findings that emerge from such study. In this chapter, a number of topics are addressed, which might be summarised as follows:

1 What kind of searches are useful in finding out about how English works? Examples are given of searches for a single word-form (e.g. *point*), for a lemma (e.g. *CONDEMN*), or for a series of words (e.g. *on* ADJECTIVE *terms with*). Sometimes a search is designed to show not the words searched for but a concept that often co-occurs with it (e.g. *what would* co-occurring with expressions of hypotheticality). In some of the examples given, a series of searches has to be conducted, because a single search cannot give the required data in manageable form (e.g. a very frequent word such as *point*). In some cases, more context than single concordance lines is required to allow conclusions about regularity to be drawn (e.g. *I must admit*).

2 How can concordance lines be presented in an accessible way? Examples are given of sets of lines which are unsorted, and lines which are sorted alphabetically. Concordance lines are also shown in groups which have been selected and organised to illustrate a particular behaviour.

3 What are concordance lines useful for? Although this topic will be taken up in much greater depth in chapters 5 to 8, this chapter does suggest some of the main types of information that can be gained from concordance lines, and what this might be used for. The main use that is illustrated in this chapter is examining the meaning and behaviour of individual lexical items, and the pragmatic meaning of given phrases.

4 What do we see in concordance lines? or, Why are concordance lines worth looking at? These questions as they stand are perhaps a little disingenuous, and may be re-phrased as: What view of language does this chapter want to put across? and What general assumptions about language investigation are made in offering concordance lines as a source of information? The view of language taken in this chapter is in accord with that described in more depth in chapter 6, which prioritises lexis and stresses the association between pattern and meaning (both lexical and pragmatic). Examples are chosen that illustrate this view, illustrating general and detailed patterns of lexis, word meaning and pattern, and semantic prosody and pragmatic meaning. Additional assumptions are that observed language is a more valid object of study than intuition, that it matters how frequently a linguistic item occurs, and that a corpus can produce results that can be extrapolated to a more general category of language.

The chapter begins with an introductory section showing what concordance lines look like and how their presentation can influence their interpretation. The chapter then moves on to ask what is observable from concordance lines, and to show how searches can be adapted to cope with a lot of data. From there we look at information gained from more context than the simple concordance line, and probes are used. The final section summarises some of the issues raised in the chapter.

Concordances are an example of what I refer to as 'word-based' methods of investigating corpora. Chapter 4 will compare these methods with 'category-based' methods and consider the advantages and disadvantages of each.

Searches, concordance lines and their presentation

A concordancer is a program that searches a corpus for a selected word or phrase and presents every instance of that word or phrase in the centre of the computer screen, with the words that come before and after it to the left and right. The selected word, appearing in the centre of the screen, is known as the node word. Here are 15 randomly-selected concordance lines for the word *critical* (from the 1999 Bank of English corpus):

attack on ground targets will be **critical** to success in any Gulf conflict,
for parents is to children to be **critical**. Some ways to encourage
benefit by adopting a more self-**critical** and experimental approach.
These films have a self-**critical** subtext and a depth of
within Afghan society. Arney is **critical** of the lack of political
The Daily Telegraph remains **critical** of Syria and Iran. It says as on
guard who insists on a literary-**critical** seminar correct me if I'm wrong,
DeConcini's letter is even more **critical**. The committee said that even
claim to having helped, through his **critical** writing, to save figurative
this certainly might be a **critical** clue. Physical relaxation may just
But a more considered view, highly **critical** of Eden, was expressed among both
that colonial possessions were of **critical** importance for advanced or
a mental health professional is **critical** to the success of the therapy. In
nearly a century later. Led at this **critical** juncture by a man with vision and
the chief inspector's view may be **critical** in determining its future

It is helpful to sort these lines so that the lines that are like each
other in some way appear next to each other. Below are the same
lines sorted so that the words immediately before (to the left of)
critical are in alphabetical order. The words that the corpus user is
likely to focus on are printed in bold.

this certainly might be a **critical** clue. Physical relaxation may just
attack on ground targets will be **critical** to success in any Gulf conflict,
for parents is to children to be **critical**. Some ways to encourage
the chief inspector's view may be **critical** in determining its future
But a more considered view, highly **critical** of Eden, was expressed among both
claim to having helped, through his **critical** writing, to save figurative
within Afghan society. Arney is **critical** of the lack of political
of a mental health professional is **critical** to the success of the therapy. In
guard who insists on a literary-**critical** seminar correct me if I'm wrong,
DeConcini's letter is even more **critical**. The committee said that even
that colonial possessions were of **critical** importance for advanced or
THE DAILY TELEGRAPH remains **critical** of Syria and Iran. It says as on
benefit by adopting a more self-**critical** and experimental approach.
These films have a self-**critical** subtext and a depth of
nearly a century later. Led at this **critical** juncture by a man with vision and

This organisation highlights the fact that *critical* often follows a
form of the verb *BE* (*be* or *is*) and sometimes follows a determiner (*a,
his* or *this*) in a noun group. It is sometimes used in compounds such
as *self-critical*, and it sometimes follows a grading adverb (*highly,
more*). Here are the same lines again, this time sorted so that the
words immediately after (to the right of) *critical* are in alphabetical
order. Again, the words the user is likely to focus on are in bold.

benefit by adopting a more self-**critical** and experimental approach.
this certainly might be a **critical** clue. Physical relaxation may just
that colonial possessions were of **critical** importance for advanced or
the chief inspector's view may be **critical** in determining its future
nearly a century later. Led at this **critical** juncture by a man with vision and

within Afghan society. Arney is **critical** of the lack of political
THE DAILY TELEGRAPH remains **critical** of Syria and Iran. It says as on
But a more considered view, highly **critical** of Eden, was expressed among both
guard who insists on a literary-**critical** seminar correct me if I'm wrong.
for parents is to children to be **critical**. Some ways to encourage
These films have a self-**critical** subtext and a depth of
DeConcini's letter is even more **critical**. The committee said that even
attack on ground targets will be **critical** to success in any Gulf conflict.
of a mental health professional is **critical** to the success of the therapy. In
claim to having helped, through his **critical** writing, to save figurative

Sorting the lines this way highlights the fact that *critical* is some-
times followed by *of, to* and *in*. A different meaning of *critical* is
associated with each preposition: *be critical of* is associated with the
'negative opinion' meaning of *critical*, whereas *be critical to* and *be
critical in* are associated with the 'important' meaning. The lines also
show, again, that *critical* is sometimes followed by a noun. The most
frequent meaning in that case (as in *critical clue, critical importance*
and *critical juncture*) is the 'important' meaning. (*Critical writing* is
another meaning.)

Even in these few lines, then, we begin to see something of how the
word *critical* behaves. It is both an attributive adjective (occurring
before a noun) and a predicative one (occurring after a link verb),
though on the evidence of these lines the predicative use is slightly
more frequent. The adjective is followed by certain prepositions, of
which *of* and *to* appear to be the most frequent. Each of the
behaviours of *critical* is associated more closely with one meaning
than another, according to the evidence of these lines. When used
attributively, *critical* is likely to mean 'important', as it does also in
the phrase *be critical to*. The meaning of 'negative opinion' is
associated with the phrase *be critical of*.

The lines for *critical* were selected using simply the word *critical* as
the search-word. Most concordancers allow the researcher to make
more specific searches, with various permutations. For example, you
may be able to search for a phrase, or for specific word-classes.
Below are some lines selected with the search '*on* followed by an
adjective, followed by *terms*, followed by *with*', which show the kind
of adjective that is used in this phrase. The lines have been sorted so
that the adjectives appear in alphabetical order.

family when she is known to be on bad terms with some of them. And it is hard
a case of erm trying to be on equal terms with everybody else. Yes.
older than she was, she felt on equal terms with him. Anyway, it's all
on they will have to beg on equal terms with the rest of the poor world for
team even if they're not on familiar terms with them. They've brought
Mr Botha is said to be on friendly terms with Mr Mandela. Nevertheless his
republics while staying on friendly terms with the resourceful but hard-

the Georgians, who were on friendly terms with the Germans, the Armenians had information. Moutet was on good terms with the porters and cleaners in though she wanted to live on good terms with the former Dutch colony, whose t write to me; I want to stay on good terms with my postman. Write to Pete any reason. He left the pub on good terms with everybody. He even borrowed dreams in which I'm on intimate terms with the Queen. Woody Allen is months in disguise and on intimate terms with nomads and pilgrims – not more important to be on reasonable terms with the Israelis than to

These adjectives can be grouped according to meaning: *familiar, friendly* and *intimate* indicate a degree of closeness; *good, reasonable* and *bad* indicate whether or not the two groups like each other; *equal* indicates a similarity in status. Thus, three aspects of a relationship – closeness, symmetry and affect (liking or disliking) – can be indicated by the phrase *on* + adjective + *terms* + *with*.

What is observable from concordance lines?

In this section the main types of observation that can be made about concordance lines are exemplified. They are presented under the following headings: observing the 'central and typical', observing meaning distinctions, observing meaning and pattern, and observing detail.

Observing the 'central and typical'

Linguistic description of the kind that has been traditional since the 1950s focuses on distinctions between what 'can' and 'cannot' be said in a particular language, with little regard for whether what is possible frequently or rarely occurs in practice. Corpora cannot be used to determine what is impossible in a language, as they do not offer negative evidence, and they cannot really even be used to determine what is possible, as a corpus may well contain utterances which any speaker of a language would reject as 'incorrect' (e.g. *I'm just sort of showing you perhaps some dishes which are more healthier than others*). In Swan's (1994) terms, there is no 'demarcation' between the correct and the incorrect. In place of demarcation, a corpus offers information that a native speaker cannot replicate: an indication of 'central and typical' usage (Hanks 1987: 124–125; Sinclair 1987a: 108, 114; Sinclair 1991: 17).

The terms 'central' and 'typical' are often used synonymously, but it is useful also to make a distinction between them in order to illustrate different types of centrality/typicality. 'Typical' might be used to describe the most frequent meanings or collocates or phraseology of an individual word or phrase. To illustrate typicality,

here are ten randomly selected concordance lines for the phrase *recipe for*:

```
1      and making merry are a sure-fire recipe for long-term damage to the mind
2    Tuigamala and you could have the recipe for a World Cup surprise or two.
3         On the face of it, it should be a recipe for easy success. A female singer
4    Civil Aviation Authority are not a recipe for failure but a time-bomb for
5     not as a dangerous Treasury-view recipe for prolonged slump, but as a
6            rally in Edinburgh: 'It's a recipe for constitutional chaos. The
7       massacre was no part of their recipe for government. Because of this
8   demand for efficiency savings was a recipe for disaster in education.
9      to good material, and you have a recipe for serious success. But it's not a
10     or cucumber. A favourite Welsh recipe for salmon trout (or sewin, as it
```

From the evidence of these lines, the typical meaning of *recipe for* is metaphoric rather than literal (only line 10 is an exception to this). Furthermore, the nouns following *for* are slightly more frequently negative (*damage* 1, *failure* 4, *slump* 5, *chaos* 6, *disaster* 8) than they are positive (*surprise* 2, *success* 3 and 9) or neutral (*government* 7). When the phrase is metaphoric, it most frequently follows the verb *BE* and the determiner *a* (lines 1, 3, 4, 6, 8); most exceptions to this (lines 2, 7, 9) are positive or neutral rather than negative. Thus, although the phrase *recipe for* has a range of meanings, collocates and grammatical co-texts, its typical use is in the sequence 'something is a recipe for something bad'. A typical example would be, *Eating, drinking and making merry are a sure-fire recipe for long-term damage to the mind and body* (line 1). This example does not show all the ways that the phrase can be used, but it combines all the most frequent features.

The concept of 'centrality' can be applied to categories of things rather than to individual words. For example, the present progressive in English can indicate the present (e.g. *she's working at the moment*), the future (e.g. *she's taking an exam tomorrow*), or no specific time (*she's always making mistakes*). As Mindt (2000: 262) shows, however, the future and 'timeless' uses are relatively infrequent, making the reference to present time the central use of the present progressive. A central adjective might be one that occurs both attributively (*rich man*) and predicatively (*he was rich*), whereas an adjective such as *asleep* occurs only predicatively, and so is not central. In the *recipe for* example, words such as *damage* and *failure* form the class of nouns that centrally follow the phrase, whereas *government* is less central.

Although speakers of a language may have intuitions about typicality, these intuitions do not always accord with evidence of frequency. Barlow (1996) and Shortall (1999) use the term 'prototypical' to indicate a usage which is commonly felt to be typical but

which is not necessarily most frequent. They suggest that English teaching course books tend to present usage which is prototypical in this sense but not typical in the sense of 'most frequently occurring'. One example of this is comparatives. A prototypical (if out of date) example might be *The USSR is larger than China* (Hsia et al 1989: 178), in which the comparative adjective *larger* is followed by *than*. From a sample of 100 lines of *larger* from the Bank of English, however, only 17 included *than*. In most lines, the adjective is followed by a noun, as in *a much larger plan* or *their larger but poorer northern neighbours*, where the comparison is implicit. A truly typical example would be as in the following:

In some cultures such as Britain and America, women architects and engineers are seldom found. Yet in others, such as Egypt and Eastern Europe, they form a far larger proportion of these professions. (British non-fiction book)

Here the two parts of the comparison are present, but not in the same sentence.

As noted in chapter 2, Barlow (1996) has made similar observations with relation to the use of reflexive pronouns such as *herself*, noting that coursebook writers often present these pronouns contrastively, setting *be proud of oneself* in opposition to *be proud of one's child*, for example. Hsia et al (1989: 222) ask students to produce sentences such as *I saw myself in the mirror*; *He hit himself with the hammer*; *We dried ourselves with the towel*. Barlow notes that in fact the reflexives have phraseologies which are quite distinct from those associated with other pronouns. For example, the most frequently used verb is *FIND*, and *found myself by the sea* has a very different meaning from *found him by the sea*. In addition, the other verbs with which reflexives co-occur most frequently are those indicating thoughts and speech, such as *SEE*, *IMAGINE*, *VISUAL-ISE*, *CONSIDER* and *ASK* (Barlow 1996: 9), rather than the verbs of physical action represented by *he hit himself* and so on.

The psychologically prototypical is not necessarily to be ignored in language teaching. Learners may find it difficult to grasp the meaning of comparatives unless they are first presented with examples in which two items are compared explicitly. Also, usages which are non-central and non-typical do still exist and cannot be omitted from the learner's repertoire. Knowing what is central and typical in frequency terms, however, can indicate what the bulk of examples that a learner is exposed to should be like.

In chapter 6 the notion of typicality will be related to the idiom principle and to the reduction of ambiguity in English. Ambiguous utterances, such as *Time flies*, are sometimes used to illustrate

parsing techniques, or to make jokes. (It can mean either 'we perceive time to pass quickly' or 'use a stop-watch to time how quickly flies move'.) In practice, however, the first meaning is the most typical. Of 52 instances in the Bank of English, 51 had the temporal meaning, and the exception was part of a discussion about parsing. A speaker of English hearing *Time flies* does not need to work out which meaning is the most likely in a given context: typicality encourages the hearer to assume only one meaning unless the context makes that impossible.

Observing meaning distinctions

Many words have meanings that are similar, and yet the words are not able to be substituted one for the other. Dictionaries, which deal with words separately rather than comparatively, can be of little help, but observing typical usages of near-synonyms can clarify differences in meaning. Corpus investigations that distinguish between near-synonyms include those by Kennedy (1991; and see chapter 1) and Partington (1998: 33–46).

Partington's study focuses on what he calls 'semi-grammatical' words, that is, words which by themselves carry only a general meaning. The examples he gives are the intensifying adjectives *sheer*, *pure*, *complete*, *utter* and *absolute*. He points out that dictionaries tend to define these words in similar ways, and even give them as synonyms of each other. The *Collins COBUILD English Dictionary* (CCED), for example, suggests that *complete* and *pure* are synonyms of *sheer*; the *Longman Dictionary of Contemporary English* (LDOCE) gives *pure* as a synonym of *sheer*; the earlier *Collins COBUILD English Language Dictionary* (CCELD) gives *absolute* as a superordinate of *sheer*. In spite of this apparent similarity in meaning, the typical collocates of each adjective differ to quite a considerable degree. For example, *sheer* is used with nouns of degree or magnitude (*sheer weight*, *sheer number*), often in the pattern *the sheer* N *of* N (*the sheer weight of noise*, *the sheer scale of the shelling*). The other adjectives do not collocate with these nouns. In addition, *sheer* alone is often used in expressions indicating causality (*through sheer insistence*; *out of sheer cussedness*; *by sheer hard work*; *because of sheer hard work*; *his sheer integrity got him though*; *his enthusiasm and sheer hard work meant that things moved quickly*) (1998: 36). *Complete* is used with nouns indicating absence (*complete ban*), change (*complete revamping*), and destruction (*complete collapse*), often in the pattern *the complete* N *of* N (*the complete absence of perspective*) (1998: 43). *Absolute* is used with what Partington

(1998: 45–46) calls 'hyperbolic' nouns, such as *chaos, disgrace, genius, godsend*, whereas the other adjectives are not.

Observing meaning and pattern

One theme that has emerged from all the examples above is that the meaning of a word is closely associated with its co-text. That is, although ambiguity is possible, for the most part the meanings of words are distinguished by the patterns or phraseologies in which they typically occur. To illustrate this, it is common to divide concordance lines into sets, each set exemplifying one meaning. The first example given here is of a noun: *initiative*. This noun has three distinct meanings, shown here in three sets of concordance lines.

Set 1

       a practising GP. 'That makes this initiative attractive to all parties. We all
    said it will announce a new Punjab initiative in the next few days to deal with
          Giuliani's latest law-and-order initiative is aimed at New York's drink –
       a need for a new national initiative on immunization – one that
    capital that he can use. And every initiative that a president undertakes,
    continual latecomers as part of an initiative that has produced dramatic
             education. He said the initiative was aimed at remedying Britain's

Here *initiative* is a count noun meaning 'something that someone (usually a government agency or other institution) starts to try to solve a problem'.

Set 2

       I would be expected to take the initiative and tell players to sit back and
    should refrain from taking the initiative in these matters. According to so
          only if France takes the initiative is there a chance of success at
    elsewhere, women have seized the initiative. We've come a long way from the
          has his best chance to seize the initiative – or at least go some way down
    pocket before saying, 'Losing the initiative can be disastrous.' It never
       then veto, thereby losing the initiative in its efforts to isolate Iraq.

Here the phrase *the initiative* is used with verbs meaning 'take' or 'lose'. 'Take the initiative' means 'start something and so gain an advantage over a competitor' while 'lose the initiative' means 'fail to start something and so allow a competitor to gain an advantage'.

Set 3

    that the president must show more initiative and leadership on domestic
       think and not to act on their own initiative. Only the elite are encouraged to
             had the imagination or initiative to read books that were not on
    cult. Greater enthusiasm and more initiative, together with the expansion of
    encourage them to use their own initiative when learning. The modern
          the road and showed a lot of initiative'. 'She used to sit on my lounge
    failing, however, was his lack of initiative and leadership in dealing with

Here the meaning is 'the quality of being able to do things without

being told'. Only the possessive (e.g. *their, his*) is used as a determiner; in most cases there is no determiner.

In short, the word *initiative* does not have one meaning but several. Distinguishing between the meanings is a matter of distinguishing between patterns of usage. *An initiative* or *a government initiative* indicates the first meaning. *TAKE/LOSE the initiative* indicates the second meaning. *A lot of initiative* indicates the third meaning. Meaning and phraseology are indistinguishable, and the concordance lines show both.

The second example is a verb: *CONDEMN*. Again, there are several different meanings, and the concordance lines are grouped so that each set illustrates one meaning.

'criticise'
    should be welcomed rather than condemned and the most enlightened – not
             s action, saying it totally condemned any steps violating the right of
    a Western-style democracy, which he condemned as 'bourgeois liberalism'. But
    The Fable of the Bees was tried and condemned as a public nuisance by the Grand
       leader, Jean-Marie le Pen while condemning the desecration as ignoble
       involved in the talks yesterday, condemned the incident as 'the dangerous

'pass sentence'
    one that sixteen years later would condemn him to death on even more
    people who a few days before had condemned him to death. Ironically, the
    secret recesses of his own mind and condemned them to deportation. Even if he

'sentenced to death'
    advisory committee, although the condemned man had no right that he should
       Then he would turn to face the condemned man: 'You who have sinned against
    the death sentences of many of the condemned prisoners, it urges that the

'make something bad happen'
          in restricting uniforms and condemning people to wear tattered clothes
    Many fear Fuerstenberg's past will condemn it to being a ghost town, trapped
       Four decades ago, Asia seemed condemned to poverty. Half of Japan's
    an in-person this decade and who is condemned to the social wilderness. Our

As expected, each meaning is clearly associated with a particular pattern. The 'criticise' meaning, for example, is associated with the patterns 'condemn something' and 'condemn something *as* something'. On the other hand, 'condemn someone *to* something' is used with both the 'make something bad happen' and the 'pass sentence' meanings, suggesting that one may be a metaphoric extension of the other. The pattern 'condemn someone *to* do something' (or the passive 'be condemned to do something') always indicates that the 'something' is bad or undesirable. This is true even when 'something' is not bad on the surface. In the following example, *a life of nursery nursing* may not seem to be a terrible fate to many people, but the phrase *condemn her to* indicates that Elaine would see it in this way.

In Britain, despite a bad choice of school – 'my parents didn't know the system' – and a careers teacher who tried to condemn her to a life of nursery nursing. Elaine got together enough qualifications to go to teacher training college.

In each of the examples above, the corpus search specified an individual word. The association between meaning and pattern of use was illustrated by words with several meanings. The other side of the coin – that words with similar meanings tend to share patterns – can be illustrated by searching for a phrase and noting the words that frequently occur within it. This will be illustrated with the phrase or pattern '*it* followed by *BE* followed by an adjective, followed by a that-clause', as in *it is apparent that fate intervened*. Below are 45 lines, randomly selected, that illustrate the pattern.

> Constable Jones, Mr Casey said. It is apparent that fate intervened th
> Elegance being a key factor, it was appropriate that Joanne
> evening of a change of plan. It was clear that Mr Chernomyrdin did
> I can tell you.' From his tone it was clear that Dick Ryle had had
> pessimism and quiescence. It is clear that the Nobel Laureate's
> to discredit UN forces. By now it was clear that both the Croats and
> in Mecca to perform the haj. It is clear that the revolution in mas
> Union of Students said that it is crucial that universities give a
> o add the fateful words: 'But it is essential that we end it in such
> dgate Valley Country Park, and it was fitting that two of the city's
> t result might become positive. It is important that social workers
> health clinic, or any hospital. It is important that the woman
> ewn up is not so important; but it is important that North Korea should
> impossible. At the same time it is inevitable that those at home,
> rs to Stick Letters. I suppose it was inevitable that this passionate
> im to act so out of character? It was ironic that Penelope's insistence
> co-operation, however he said it was likely that Germany would have t
> acing three different products, it is likely that management will be
> s and the quality of your life. It is likely that you will exhibit all
> lethal muscle wasting disease. It is likely that any child with this
> joyment of this exquisite poem, it is necessary that the reader should
> a camera attached and working. It is obvious that the chance of a
> zing in wonderment about them. It was obvious that they believed they
> h it had been contracted here, it was obvious that that was not what
> And we asked them and it was overwhelming that the the
> nt parts of every relationship. It is possible that your partner's mil
> emplate for building proteins – it is possible that the FraX protein h
> xchange transactions in London. It is revealing that the Socialists who
> st the end of the year. Indeed it was significant that the Jakarta
> euthanasia is a guess at best. It is surprising that, following an
> others. Some doctors have said it is suspicious that the pills named
> r as simple as presented. While it is true that vertical integration
> in the Old Testament, because it is true that it has been used for
> the PPI after Sturzo's exile, it was true that he had 'no experience
> Europeans are two-thirds wrong. It is true that the Turks are Muslims,
> e will require a high-wire act. It is true that Malcolm Rifkind, the

> e were absolutely amazed and if it is true that St George's Hall has
> him, but he couldn't help it. It was typical that Robyn would have
> said: 'In a civilised society, it is unacceptable that women are
> in this Year of Remembrance. It is unfortunate that the article made
> i-government protests. He said it was unfortunate that a number of
> re foreign policy picture, that it is unfortunate that we cannot have
> its policy on public debate. It is unlikely that the second and third
> recent work has suggested that it is unlikely that family boundary
> to testify. In retrospect, it is unlikely that a US court would

The adjectives that appear in these lines are: *apparent, appropriate, clear, crucial, essential, fitting, important, inevitable, ironic, likely, necessary, obvious, overwhelming, possible, revealing, significant, necessary, obvious, overwhelming, possible, revealing, significant, surprising, suspicious, true, typical, unacceptable, unfortunate* and *unlikely*. Some of these adjectives occur several times, others only once or twice: a larger sample of concordance lines would show more examples of all these adjectives, together with others. As noted by Francis (1993; see also Francis et al 1998: 480–484), the adjectives all indicate an evaluation or judgement, specifically judgements of likelihood (*inevitable, likely, possible, unlikely*), clarity (*apparent, clear, obvious*), necessity (*crucial, essential, important, necessary*), significance (*revealing, significant*), goodness or badness (*appropriate, fitting, suspicious, unacceptable, unfortunate*), as well as a few other kinds of judgement (*ironic, overwhelming, surprising, true, typical*).

A variation on this approach is the concept of 'frames'. Frames are sequences of, usually, three words in which the first and last are fixed but the middle word is not. Renouf and Sinclair (1991), for example, list the following frames from a small corpus (10 million words of written English; 1 million words of spoken English):

a . . . of e.g. *a lot of, a kind of, a number of, a couple, of, a matter of, a sort of, a series of;*

an . . . of e.g. *an act of, an example of, an average of, an expression of, an air of, an element of;*

be . . . to e.g. *be able to, be allowed to, be expected to, be said to, be put to, be made to;*

too . . . to e.g. *too late to, too much to, too young to, too easy to, too small to, too close to;*

for . . . of e.g. *for most of, for all of, for one of, for fear of, for both of, for some of, for lack of;*

had . . . of e.g. *had enough of, had plenty of, had thought of, had heard of, had one of, had died of;*

many . . . of e.g. *many thousands of, many years of, many kinds of, many parts of, many millions of.*

Each of these frames is frequent in the corpus used, much more frequent than any one of the triplets formed by it. For example, in Renouf and Sinclair's corpus there are 3,830 occurrences of *a . . . of* but only 1,322 occurrences of the most frequent triplet with this frame – *a lot of* – and only 174 occurrences of the twentieth most frequent triplet – *a quarter of*. In other words, the frame as a whole has a numerical significance to the corpus which far outweighs the significance of any one of the triplets. Not only that, but the frame is significant to each word which occurs as the middle item in it. Of all the instances of *lot* in the corpus, 53% of them are in the triplet *a lot of*.

When the triplets formed by each frame are examined, it is found that the middle words are not a random selection, but belong to particular meaning groups. For example, the words that are found most frequently with *many . . . of* in Renouf and Sinclair's corpus fall into these groups:

words indicating numbers: e.g. *thousands, millions, hundreds*
words indicating a type or aspect: e.g. *kinds, ways, aspects, types, varieties*
words indicating a length of time: e.g. *years, hours*
words indicating a group of people or things: e.g. *members, examples, species.*

Frames are particularly useful because programs can be written to identify them automatically, without the researcher knowing or guessing in advance what they will be. They show part of what is typical in a corpus and, because they incorporate variation, they are much more frequent than fixed phrases are. Frames and patterns are an alternative to the very general statements made by most grammars and the very specific statements that can be made about the collocations of each individual word in a language. Because of this, they might be a good starting point for organising the way that language is presented to learners. Frames are particularly useful when looking at a specialised corpus, and can be used as the basis for investigating the language of a discipline (Luzon Marco 2000). Chapter 6 considers this point further. Corpus-based work on frames has been carried out on languages other than English; Butler (1998), for example, considers frames in Spanish.

Observing detail

So far we have seen how concordances can be used to give very general ideas about the ways that words behave and the meanings

that can be associated with patterns. Any work with concordances, however, tends to lead to more specific observations about the behaviour of individual words. For example, the nouns *advice* and ANSWER, among others, are often followed by *as to* and a clause beginning with a wh- word. In the case of each of these nouns, however, more detail can be added to this patterning. For example, *advice as to* often follows a verb indicating 'getting', 'giving', 'wanting' or 'offering'. The following are ten randomly selected concordance lines, with the relevant words underlined.

women, had surreptitiously asked his advice as to whether the Orthodox Church
    meeting was invalid. I'm seeking advice as to my remedies in relation to
        Instead of providing medical advice as to why this may have occurred,
Metropolitan Opera, in order to get advice as to where I should be sent in
            Erm I'd like some advice as to how Key Stage Four would er
    though there was no practicable advice as to how this could be done
    of food allergies you should take advice as to the best way of altering your
with water and said I would need his advice as to how to cut around his ears
picnics, petty economies, wholesome advice as to how to succeed in life, and
Youthfully dispensing criticism and advice as to how they could and should do

ANSWER *as to* tends to follow the same kind of verb, underlined below. In addition, it often follows a phrase indicating either that a clear answer is not available, or that to give a clear answer is difficult or unexpected, as shown in bold below.

1   Picasso. There is no definitive answer as to who was the greatest," said
2   readers who can give us a persuasive answer as to why the narrator is male or
3   I want to wait fourteen days for an answer as to whether they can get a
4   that lawyers could not give a clear answer as to what would happen if the Bil
5   means. There can only be one answer as to where Croatia could possibly
6   waving his hands to demand an answer as to whether Labour would join a
7   Well, I can't give you an exact answer as to what he has. Reporter 5:
8   But no one really has a concrete answer as to why the incidence is
9   will one day present us with the answer as to what chemical change takes
10   seemed disturbed. He even had an answer as to why no scientific publicatio
11   reason why there must be a single answer as to the relative effects of
12   not only for a cure, but also for answers as to why we suffer from such big
13   searchers hope to find some of the answers as to why the virus is
14   President, to demand of him clear answers as to how the plan stood. Leonida
15   re hampered by the lack of medical answers as to why many miscarriages

For some lines, a short concordance line is not sufficient to show the negativity and it is necessary to obtain more co-text. Two examples from these lines are:

3 The public don't want to wait fourteen days for an answer as to whether they can get a mortgage on a house.

9 It may well be that science or medical research will one day present us with the answer as to what chemical change takes place in the body of the alcoholic . . . But, to date, no answer has been forthcoming.

This suggests that *answer as to* is part of a phrase which might be expressed as '[negative] [clear] *answer as to* [wh-word]'. The precise wording is not fixed, but it is constrained within limits. As we shall see in chapter 6, a great deal of English is made up of sequences of this type, and it raises a problem as to how the phraseology is to be described in a useful way. The sequence '[negative] [clear] *answer as to* [wh-word]' may well apply to only this noun, which makes it useless as a generalisation. A more general expression '[get] NOUN *as to* [wh-word]' could be applied to *advice* and *answer*, but would miss some of the essential detail.

Coping with a lot of data: using phraseology

One of the problems brought about by the increasing size of corpora is that searches for frequent words will yield too much data to be interpretable in the form of concordance lines. A corpus-user can probably cope with looking at, at any one time, about 100 lines for general patterns and about 30 lines for detailed patterns. If the word under investigation is a frequent one, such a small number of lines will not show all the patterning, and slightly different methods may be employed. Sinclair (1999) advocates selecting 30 random lines, and noting the patterns in them, then selecting a different 30, noting the new patterns, then another 30 and so on, until further selections of 30 lines no longer yield anything new. An adaptation of this method is 'hypothesis testing', in which a small selection of lines is used as a basis for a set of hypotheses about patterns. Other searches are then employed to test those hypotheses and form new ones. This method will be demonstrated using two frequent nouns: SUG-GESTION and *point*.

Suggestion

To begin, here are 20 random concordance lines for the noun SUGGESTION, sorted one to the right of the node-word:

insistence always that it was his suggestion; and if it turned our to be a
Island, and Lady Stallard. Suggestions are being sought by the Histor
there's never a shortage of suggestions as to what's on in town and we
and controversial field. The Suggestions for Further Reading list severa
and find a centre that will hold. Suggestions for Discussion. Why does Russell
style book awaits the liveliest suggestion for the name of the place and the
reasons', although there were suggestions he was unhappy with the safety
to the House of Lords; there is a suggestion legal advisers to the Guinness
he was hoping I would fail. My suggestions never got past his desk. He
Hailsham said. Responding to the suggestion of removing the remaining

letter and style. Erm the suggestion of how you write it past
signed.' Andrew was angered at suggestions that he resigned because he had
George Hannaford rejected a suggestion that the Government delay the
QUEENSLAND Softball officials' suggestion that Ballymore should be the
of an ankle injury. There were suggestions that midfielder Todd Viney would
party programme, and agreed to his suggestion that the party should hold a
since the war, has rejected suggestions that there should be a
nor did I ever hear the slightest suggestion that we should do otherwise
have been interrogated. Another suggestion that something was going on. The
has never been any suggestion that his action is anything other

These lines show that SUGGESTION is frequently followed by a finite clause, either introduced by *that* or not. It is also followed by *as to*, *for* and *of*. To investigate this behaviour further, another 50 lines are selected, but this time the lines in which SUGGESTION is followed by a finite clause are deleted (as no further evidence is needed for this), as are lines in which SUGGESTION simply behaves as an ordinary noun, as in *My suggestions never got past his desk*. The remaining lines are as follows:

to us. One key reflection: many suggestions for classic Hollywood and
and find a centre that will hold. Suggestions for Discussion. Why does Russe
Hailsham said. Responding to the suggestion of removing the remaining
advertising. Send information and suggestions of relevant book/articles to
quite the contrary. So no suggestion of a coup? The word was floating
dispute at Wapping. There were suggestions of organised violence on the
became the Royal Arsenal at the suggestion of George III, when he paid a
a helping hand? The panel's suggestion of giving the authorities power
hour, if you like, so there's no suggestion of office harassment. Presumably
said. He smiled at her with no suggestion of patronization, no sense of
himself from the cautious suggestion of the Treasury's post-crash
travel is bearing fruit. Our suggestion to run weekday trains every 20
down the dust; there was even a suggestion to pipe seawater from Brighton
General Shintaro Abe. The clear suggestion was that they had decided to

These lines confirm that SUGGESTION is frequently followed by *of*. There are two lines in which it is followed by *for*, but no lines in which it is followed by *as to*. On the other hand, a new pattern seems to be emerging, in which the noun is followed by a to-infinitive clause (*a suggestion to pipe seawater*). There is also a variation of the that-clause, in which the clause follows a link verb (*The clear suggestion was that they had decided . . .*). The behaviours of SUGGESTION for which there has not been much evidence so far can now be tested using searches for individual combinations of words: SUGGESTION *for*, SUGGESTION *to*, and SUGGESTION *as to*.

First, SUGGESTION *for* occurs over 1,000 times in the Bank of English corpus, confirming the hypothesis that *for* is a significant part of the way SUGGESTION behaves. Below is a random selection from these lines for illustration:

Thames TV has rejected this suggestion for its annual amateur tournament
for the arms clampdown and the suggestion for the meeting came from the
There have been numerous suggestions for clarifying decision-making
the home throughout the year, with suggestions for scented indoor plants,
He was also doubtful of Italian suggestions for merging the organisation
polish a telescope mirror, and add suggestions for building an observatory from
a variety of imaginal devices and suggestions for pain control. Before you
repair strategies? What other suggestions for repairing a relationship by
confusing passages, and offer suggestions for revision. Together, through
union and in come tentative suggestions for institutional reforms to

We might further note that *suggestions* is more frequent than *suggestion* in this pattern, and that *for* is as likely to be followed by a present participle (*for repairing a relationship*) as by a noun (*for pain control*).

Next, a search for *SUGGESTION to* followed by the base form of a verb yields just over 200 lines. These lines need interpreting, however. In some cases, the to-infinitive clause is dependent on a word other than *SUGGESTION*. For example, *suggestion* sometimes follows *HAVE* (e.g. *I have a simple suggestion to make*) and exemplifies one of the ways that *HAVE* behaves (e.g. *have a book to read*). In terms of the specific behaviour of *SUGGESTION*, then, these lines can be discounted. Of the remaining lines, some are like these:

NTUC after he accepted the NTUC's suggestion to contest in the presidential
of my friends the post-hypnotic suggestion to freeze at my command.
officials say they oppose American suggestions to include European countries
Hongkong Bank has blithely ignored suggestions to release it, which seems a

In these lines, the to-infinitive clause explains what the suggestion is. For example, someone in the NTUC said, 'I think you should contest in the presidential election', or something like that. Therefore, the to-infinitive clause is an important part of the behaviour of *SUGGESTION*. In the following lines, however, the meaning is different, and the to-infinitive clause means something like 'in order to':

the rescue with 30 sanity-saving suggestions to make this summer a holiday
listener may also have ideas and suggestions to help you get started
turned down some of the dafter suggestions to save coal

These clauses do not tell us what the suggestion is, but what it does. For example, no-one said 'I think you should make this summer a holiday to remember', or anything like that. In fact, we do not know (yet) what the thirty suggestions are, only what they are designed to achieve. The conclusion is that *SUGGESTION* does have a pattern in which it is followed by a to-infinitive, but that pattern is not exemplified by all 200 lines of *SUGGESTION to*.

Finally, there are almost seventy examples in the Bank of English of *SUGGESTION* followed by *as to*. In nearly every case, *as to* is followed by a clause beginning with a wh-word (*what, where, how, why, when, who, which*), as in these examples:

had asked for constructive suggestions as to how the tax could be
There have been suggestions as to how this difficulty might
of unemployment, while giving suggestions as to how to deal with the
fracas and the navel. Any suggestions as to what prevented this from
Croll asked Salter if he had any suggestions as to where Brady might be
this power play to counter Chris's suggestion as to why the car isn't starting

Thus, concordance lines can be selected to test hypotheses, so that the corpus-user does not need to examine every one of thousands of lines to obtain a reasonably accurate picture of how a word behaves. The same method will be used to describe the behaviour of an even more frequent noun: *point*.

Point

The noun *point* is an extremely frequent word in English: the Bank of English corpus has almost 100,000 instances. (It is also a word that, like many other frequent words, finds its way into very few lists of vocabulary.) Below are a mere 20 randomly selected concordance lines to illustrate the wide variety of ways the word is used. The phraseology of *point* is highlighted by bold type.

Kinnock as the 'big idea' selling **point** of Labour's next election
They told him they were on the **point** of passing a vote of no confidence
has attacked the document **point** by **point**. In part it says Primac has
investing in the burgeoning Kangaroo **Point** residential precinct. Mr Dick
If not be entirely finished at this **point**, and therefore be prepared for
you. Thank you for making the **point**. Nice to talk to you. Bye.
Mm. There's no **point** in us trying to make a cooker in
the military Commanders made the **point** that there was a limit to the civil
Commonwealth.' To underline that **point**, he said that France was ready to
lish the black comedy. Even at the **point** at which the Entrance Hall was
rivals. On the face of it, he has a **point**: it is difficult to find, difficult
in fact, is at perhaps the weakest **point** in its 10-year history in office.
u appear to have reached a turning-**point** in your life, but that implies that
for the lorry. Fortunately at this **point** a group of children came walking
your training skill from the group **point** of view. There will also be open-
me, but that wasn't the **point**. The **point** was to make him yearn for her, to
d charge a 'reasonable' fee. A key **point** in the debate will be determining
greater Cincinnati area prove the **point**. In rich Indian Hill, for example,
f the 19th century, resigning on a **point** of principle in 1867, there have
successive win. But Barcelona, a **point** behind, showed the kind of form

Each of the phraseologies noticed here can be used as the starting point for other searches. The search may be based on what comes