

---

# 12

## How can a corpus be used to explore patterns?

*Susan Hunston*

---


### 1. What is a pattern?

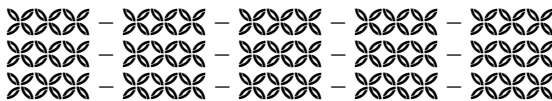
A pattern is essentially repetition. A symbol may occur once only:



but when it occurs twice or more it becomes a minimal pattern:



A combination of two symbols, such as  and  $-$ , may form a sequence that when repeated comprises a more noticeable pattern:



In language, pattern is observed when words, sounds, rhythms or structures are repeated. These are often observed in poems and nursery rhymes, such as this one quoted by Carter (2004: 1):

Hickory, dickory, dock  
The mouse ran up the clock.

There is considerable repetition even in these two lines: the same sounds and rhythm in *hickory* and *dickory*; the same end sounds in *dock* and *clock*; the same combination of weakly and strongly stressed syllables in *the mouse* and *ran up* and *the clock*. Carter goes on to note that creative repetition is more common than might be imagined in ordinary speech. What has been more frequently observed is that prepared oratory, such as in politicians' speeches, also tends to contain repetition (see for example Tannen 1989: 175). Here are some examples from a speech delivered by (at the time of writing) US presidential candidate Barack Obama:

I've gone to some of the best schools in America and lived in one of the world's poorest nations.

Repetition of the superlative and its co-text: *some of the best schools in America; one of the world's poorest nations.*

The church contains in full the kindness and cruelty, the fierce intelligence and the shocking ignorance, the struggles and successes, the love and yes, the bitterness and bias that make up the black experience in America.

Repetition of binomial opposites: *the kindness and cruelty; the intelligence and the ignorance; the struggles and successes; the love and the bitterness.*

The speech also contains a lengthy section in which a series of sentences begin with *This time we want to talk about:*

This time we want to talk about the crumbling schools ...

This time we want to talk about how the lines in the Emergency Room ...

This time we want to talk about the shuttered mills ...

This time we want to talk about the fact that the real problem ...

This time we want to talk about the men and women of every color and creed ...

We want to talk about how to bring them home from a war ...

In each of the examples above, it may be supposed that the pattern has been consciously designed. Words are chosen (or invented) in nursery rhymes to fit the sound and rhythm of the surrounding words. The writers of political speeches select the parallel words and structures for rhetorical effect.

It is apparent, however, that repetition, and therefore patterning, occurs in language without anyone having planned it. As an introduction to this chapter we will consider a few examples of this from the 450-million-word Bank of English corpus (owned jointly by the University of Birmingham and HarperCollins publishers; all examples in this chapter unless otherwise indicated come from this corpus).

The first example is the phrase *the black experience*, taken from Obama's speech quote above. If we search for *the* followed by *experience* we find that one set of words occurring frequently between those two refers to groups of people: *American, black, human, undergraduate, customer* and *British*. Replacing *the* by *a* or *an* gives very different results: in this case the word preceding *experience* usually expresses an affective response to the experience: *great, new, wonderful, bad, traumatic* and *pleasant*, for example. In short, there is a repeated phrase *the black experience*, but there is also a more frequently repeated pattern: '*the + group of people + experience*'.

The second example also comes from the speech above: *one of the best schools in America*. This represents a very common use of superlative adjectives (such as *best*) in English. Taking a random 1,000 instances of *best* followed by a noun, the word that occurs most frequently before it is *the*. The most frequent word before *the best* is *of*, and the most frequent word before *of the best* is *one*. In other words, '*one of the best + noun*' is a relatively frequent phrase. What is more, the most frequent single word following the noun is *in*, introducing prepositional phrases such as *in the country, in Europe, in the world*. If the exercise is repeated with adjectives occurring with *most*, the results are the same. The sequence '*one of the most + adjective + noun + in + time or place*' occurs

relatively frequently, and is one of the most typical environments in which the superlative occurs.

The final example interprets ‘pattern’ slightly differently, as a co-occurrence of a language form and a particular context. Conrad and Biber (2000: 63) note that stance adverbials of all kinds are more frequent in conversation than in either academic or newspaper prose. To a large extent, this difference is accounted for by the much greater number of epistemic adverbials in conversation, including items such as *of course*, *probably*, *actually*, *really* and *sort of* which are very much more frequent in conversation than in the written contexts. Such differences emerge also when the mode is the same. O’Keeffe *et al.* (2007: 208) notice considerable differences in word frequency between corpora of spoken conversation, business English and academic English. For example, although pronouns are frequently used in each of the corpora, *we* is significantly more frequent in the business corpus than in the other two.

## 2. Why are patterns difficult to spot?

This section will consider why individual language users and researchers may find it difficult to identify patterns. I will suggest three reasons, each of which approaches the problem from a slightly different perspective.

The first reason: repetition in naturally occurring conversation is transient, fleeting; it may have no perceptible effect, or its effects may not be ascribed to the repetition itself. In other words, it is not noticed. Carter (2004: 2) quotes an example from a corpus of spoken English where the following interaction takes place:

A: Yes, he must have a bob or two.

B: Whatever he does he makes money out of it, just like that.

C: Bob’s your uncle.

As Carter points out, the word *bob* is used twice, once by speaker A in the phrase *a bob or two* (‘quite a lot of money’) and once by speaker C in the phrase *Bob’s your uncle* (‘it happens without effort’); in addition, speaker B repeats the idea in speaker A’s utterance without repeating the actual words. Carter notes that it is difficult to say why speaker C says what he/she does, and whether the repetition is conscious or unconscious, and what effect it has on the other speakers. It is only very recently that examples such as this in ordinary conversation have been identified and commented upon. Carter’s point is that repetitions of this kind are creative rather than conventional and that they deserve the same kind of attention as creativity in literature or art.

The second reason: speakers of a language have relatively untuned intuitions about frequency, and about frequency of co-occurrence in particular. This is a point made by Sinclair (1991: 39) as an argument in favour of using a corpus as a source of evidence about collocations and other kinds of patterning.

Intuitions often come into play when a user of a language encounters something that sounds unusual. Here is an example from a student for whom English is a second language. She writes: ‘Three verbs were found frequent in both corpora.’ This use of the passive *were found* followed by the adjective *frequent* strikes me as odd: intuitively I would re-write this as ‘Three verbs were found *to be* frequent in both corpora’ with the explanation that ‘find something + adjective’ is used when the adjective is an evaluative one

(*strange, interesting* or *exciting*, for example) but not when the adjective does not indicate subjective judgement. A corpus search corroborates this to a certain extent. Examples of FIND followed by *them* and an adjective suggest that adjectives such as *interesting, attractive, difficult, useful* and *boring* are particularly frequent in this environment. On the other hand, searching for 'FIND + *them* + *to be* + adjective' shows both 'objective' and 'subjective' adjectives, with examples such as *she looked inside all of them one by one, but found them to be clean* as well as *he explored it and found that to be nonsense*. It turns out, however, that when the verb FIND is used in the passive, and followed by an adjective, a small range of subjective adjectives such as *helpful* and *useful* do occur, but much more frequent are the adjectives *guilty* and *innocent*, as well as *alive, unconscious, safe* and *naked*. When the adjective is subjective (*useful*) the subject of the clause indicates an inanimate object; when the adjective is objective (*alive*) the subject of the clause indicates a human being. My intuition that *be found frequent* is an unusual usage is correct, but the reasoning I used intuitively to explain this turns out to be totally incorrect.

The third reason: patterning involves the repetition of 'things', but those 'things' may be of many different kinds. Two or more words may frequently co-occur, as in *was found guilty*, but we may also regard adjectives such as *dead, alive, unconscious* and the phrase *in a coma* as belonging to a single category, so that *was found dead/ alive/ unconscious/ in a coma* would represent a single pattern. Similarly, a sequence of individual words such as *it is ironic that* may be considered representative of a more general pattern which begins with *it*, ends with a *that*-clause (with or without *that*), and contains a link verb that might be BE but could just as well be SEEM or LOOK and an adjective that indicates an evaluation of a situation. Examples of this pattern, illustrating the degree of variation possible, would include:

It is not surprising that ...  
 It seems very peculiar that ...  
 It is absolutely right that ...  
 It looks very unlikely that ...  
 It's obvious he doesn't want ...

Observing pattern in this set of examples involves perceiving a similarity between items that have no surface similarity. A more extreme example is illustrated by the following concordance lines:

I'm male I can learn how women react to women's texts, as opposed to platform. And I wonder how did you react to that as Republicans? Do you have them it's just a question of how they react to it. < p > Some teams suffer with went to a cattery to check how he'd react to my cat.' Beverley thinks the and I couldn't predict how he would react to my presence. < p > Cain, it's cold But it's not clear how Ankara would react to a request because any such move knows us knows exactly how we would react to the of thing our manager was then watched how they behaved and reacted to her. A video of each child was and first asked her how people were reacting to the suspension, given the test the waters to see how the press reacts to an idea of some sort of militar

These lines were obtained by a search for REACT + *to* preceded by the word *how*. The lines might be said to have more in common than the search terms, however. Looking at the items preceding *how*, we might identify a general sense of 'finding out' and 'knowing'

or ‘not knowing’: *I can learn how, I wonder how, it just a question of how, to check how, I couldn’t predict how, it’s not clear how, knows exactly how, watched how, asked her how and to see how*. A sub-set of these, identified by the presence of *would*, indicates a hypothetical situation. ‘Seeing’ a pattern such as this one is not as simple a matter as, for example, recognising the frequency of *to* following REACT. The pattern is realised by a very diverse set of items, and is difficult to discover automatically, that is, without a human observer. For the same reason it is more difficult to quantify. We might also feel that whereas *react to* might be said to indicate a particular sense of REACT (as opposed to *react on* as in *lichens react on rocks*), the items preceding *how ... react to* occur simply because this is the kind of thing people often need to say about reactions (they are unpredictable). There is no question of ‘(in)correct English’ here. Furthermore, the identification of the ‘finding out + how someone reacts to something’ pattern is very much open to debate – many people might deny that this pattern exists at all.

### 3. Concordances and how to read them

The identification of pattern in a corpus implies a connection between theory, method and technique. It implies a theory of meaning in language that stresses that meaning is discovered in language situated in context, not in words in isolation (Sinclair 2004; Teubert 2005). Such a theory both requires and is borne out by a methodology that prioritises the search for repetition and co-occurrence. The technique often used to carry out such a search is the ‘key word in context’ (KWIC), or the concordance line.

Examples of concordance lines have been given in other chapters in this volume (see in particular chapters by Scott, Greaves and Warren). Essentially they consist of a node word or phrase with a small amount of context (measured in characters) to the left and the right. In most concordancing programs the amount of context can be increased on demand; showing a line longer than can be accommodated on a standard screen or page without wrapping allows more context to be observed but reduces the visual impact of any repetition. Most programs, too, allow the concordance lines to be manipulated in ways that make repetition visually salient. These techniques range from being able to sort lines so that the word(s) before or after the node occur in alphabetical order to using colour to identify word class (see Scott, this volume).

To illustrate this, here is an example of twenty-five randomly selected concordance lines with the word *view* as the search word and the node.

the success of their treatment. This view of clients as experts on matters  
 s numbers, most elephant specialists view culling and the ivory trade as two  
 from any abstract ethical point of view – Judeo-Christian, whatever – from any  
 on. < p > I have to take as positive a view as possible and I would like to  
 in Capital and Class (1986) . 10 This view was expounded by RIIA prominents  
 lays the foundation of a broader view. It connects the idea of democracy  
 time and stood looking at a fantastic view across the mountains whilst I waited  
 < p > Others take a more sympathetic view. Bishop Skinner offer special  
 Black himself did not have a clear view of the age at which a child could  
 things that they didn’ t need. But the view that design is dead portrays the  
 medieval scholars. Yet this was the view that all major Christian theologians  
 got stuck in the 1960s, with a world view in which collectivism was compulsory

eventually takes place. However, the view from the Moroccan side of the wall efforts. And I think there's a shared view here that we would hope the historians by and large take a dim view of my grandfather's role and dome - is never obscured from view by new developments. < p > Regina unions represent a radical left-wing view to students and many have embarked for all time. < p > Hidden from view in the Staffordshire Moorlands, you get them together for a long view back. Stack up the fact that he's which reflected a market-frame view of the state's responsibility to differently, for instance, do women view rock? Who do they think are its from the applicant. It took the view that the risk of his association the proceedings had reinforced his view in the report that Mr Irving has biological and chemical weapons. This view was then supported by a senior texts but always from a point of view. If you look at his project closely

Sorting these lines alphabetically to the left of *view* focuses our attention on prepositions occurring before *view* and in particular the phrases *hidden/ obscured from view* and *point of view*:

dome - is never obscured from view by new developments. < p > Regina for all time. < p > Hidden from view in the Staffordshire Moorlands, from any abstract ethical point of view - Judeo-Christian, whatever - from any texts but always from a point of view. If you look at his project closely

It also brings together, and so makes noticeable, those lines where *view* is preceded by *the* or *this*. *The view* focuses attention on the phrase *the view that* and incidentally suggests that *the view that* may be used frequently (in three of the four instances cited here) in a contrastive statement (*But the view that ... Yet this was the view that ...*), though more evidence would be needed to be sure of this:

things that they didn't need. But the view that design is dead portrays the medieval scholars. Yet this was the view that all major Christian theologians eventually takes place. However, the view from the Moroccan side of the wall from the applicant. It took the view that the risk of his association

Finally, the frequency of *This view ...* suggests that *view* is often used to refer back and to summarise a segment of the preceding discourse ('encapsulation' in Sinclair's 2004 terminology):

the success of their treatment. This view of clients as experts on matters in Capital and Class (1986). 10 This view was expounded by RIIA prominents biological and chemical weapons. This view was then supported by a senior

Sorting the lines to the right additionally draws attention to phrases where *view* is followed by *of* and is preceded by an adjective:

the success of their treatment. This view of clients as experts on matters Black himself did not have a clear view of the age at which a child could historians by and large take a dim view of my grandfather's role and which reflected a market-frame view of the state's responsibility to

Marking word class in addition separates those instances where *view* is a verb from the majority where it is a noun:

s numbers, most elephant specialists view culling and the ivory trade as two differently, for instance, do women view rock? Who do they think are its

Expanding the concordance lines allows ambiguities to be resolved. For example, expanding the lines where *view* is followed by *that* shows that there are two distinct patterns – *view* + appositive clause and *view* + relative clause:

### *Appositive clause*

But the view that design is dead portrays the same lack of perspective for which the 1980s themselves are so famous.

### *Relative clause*

Yet this was the view that all major Christian theologians insisted on – and many still do today.

What should be obvious, however, is that the concordancing programs only find and organise the data. Interpretation is a human activity. We now consider what skills are needed to find pattern in concordance lines.

The first skilled activity is formulating the search that will produce the concordance lines (see chapters by Tribble, Evison, Scott, Chambers, Sripicharn, Walter). Too general a search leads to too much ‘noise’; too specific a search leads to important information being overlooked. For example, searching for *it is surprising that* will yield plenty of concordance lines (176 in the Bank of English, for example). However, performing the same search but allowing for an additional word between *is* and *surprising* yields many more lines (almost ten times as many). The concordance lines then show that only about 10 per cent are of the form *it is surprising that*; a further 20 per cent are *it is hardly surprising that*, 65 per cent are *it is not surprising that*, and there are additional less frequent occurrences of *it is scarcely surprising that* and *it is perhaps/therefore/somewhat surprising that*.

Once concordance lines are obtained, they must be interpreted. As noted above, observing pattern involves identifying similarity and forming notional categories. It also involves ignoring distractors, that is, separating what might be important from what is unlikely to be so. (Note, though, that it is impossible to be precise about what is and is not patterned – what is overlooked by one observer might be noticed by another.) These points will be illustrated with a small set of concordance lines obtained by searching for the lemma REACT. The result of this search gives examples of all wordforms in the lemma: *react*, *reacted*, *reacting*, *reacts*. Twenty random lines have been selected and then sorted so that the words following the node word (REACT) are in alphabetical order.

1           could not believe the way Vieira reacted after he was dismissed. The  
2           at all. When asked today how they’ d react if the White House sent them a ne  
3           step, which will enable viewers to react immediately to what they have see  
4           two-thirds of the radical pairs reacting (in a field of typically only

5 any more, I don't know how he would react. Is there any point in making  
 6 growth because stock markets could react.' Mr Visco said stock markets in  
 7 police officer at Selhurst Park reacted similarly to the Cantona incident  
 8 mail, in New York, Adrian Clark reacted to Simon Hoggart's discussion of  
 9 market has come, and how people will react to it. The best seats and places  
 10 strength of a substance and the body reacts to fight off any diseases which  
 11 from the air and induce them to react to form harmless gases. Last  
 12 is the poster!' Herzen was reacting to a swelling trade in images. T  
 13 efforts you may find the magician reacting too early or late. Also bear in  
 14 conference was to see how he would react when asked questions by journalists  
 15 protect. How is management likely to react when a group threatens to quit?  
 16 twenty-year-old son felt free to react with such ferocity indicates that  
 17 position sulfur and nitrogen oxides react with atmospheric water vapor to  
 18 above such common tasks, refusing to react with the molecular masses. < p > Bu  
 19 Commentators and crowd alike reacted with astonishment when Lara  
 20 during the investigation. They reacted with anger and said: The finding

Which of these lines might be grouped together to illustrate the 'same pattern'? Looking only at what follows the node word, we might observe the following:

- a full stop (lines 5 and 6);
- the word *to* (lines 8, 9, 10, 11, 12);
- the word *when* (lines 14 and 15);
- the word *with* (lines 16, 17, 18, 19, 20).

Looking at the same evidence but in a more linguistically informed way we might express this somewhat differently:

- a subordinating conjunction (*after* in line 1, *if* in line 2 and *when* in lines 14 and 15), noting that in each case the subordinate clause indicates an event that is the stimulus of the reaction
- a preposition (*to* in lines 8, 9 and 12, *with* in lines 16–20);
- an adverb (lines 3 and 7), noting that in each case the adverb is followed by the preposition *to*;
- a to-infinitive clause (lines 10 and 11), noting that the clause indicates a reason for the reaction.

A further observation might be that *react with* in lines 17 and 18 works differently from *react / reacted with* in lines 16, 19 and 20. The question that might prompt line 17 or 18 is something like 'what does the object / substance (not) react with?' whereas the question prompting the other lines is something like 'how did the person / people react?' Putting all this together suggests that a maximum of eight different patterns might be identified in these lines:

#### 1. REACT followed by a subordinate clause indicating stimulus:

1 could not believe the way Vieira reacted after he was dismissed. The  
 2 at all. When asked today how they'd react if the White House sent them a ne



14 conference was to see how he would react when asked questions by journalis  
15 protect. How is management likely to react when a group threat-ens to quit?

2. REACT followed by the preposition *to*:

8 mail, in New York, Adrian Clark reacted to Simon Hoggart' s discussion of  
9 market has come, and how people will react to it. The best seats and places  
12 is the poster!' Herzen was reacting to a swelling trade in images. T

3. REACT followed by an adverb and then by the preposition *to*:

3 step, which will enable viewers to react immediately to what they have see  
7 police officer at Selhurst Park reacted similarly to the Cantona inciden

4. REACT followed by a to-infinitive clause indicating consequence:

10 strength of a substance and the body reacts to fight off any diseases which  
11 from the air and induce them to react to form harmless gases. Last

5. REACT followed by the preposition *with* answering the question 'how?'

16 twenty-year-old son felt free to react with such ferocity indicates that  
19 Commentators and crowd alike reacted with astonishment when Lara  
20 during the investigation. They reacted with anger and said: The findin

6. REACT followed by the preposition *with* answering the question 'what?'

17 eposition sulfur and nitrogen oxides react with atmospheric water vapor to  
18 above such common tasks, refusing to react with the molecular masses. < p > Bu

7. REACT followed by a full stop:

5 any more, I don' t know how he would react. Is there any point in making  
6 growth because stock markets could react.' Mr Visco said stock markets in

8. Other lines:

4 two-thirds of the radical pairs reacting (in a field of typically only  
13 efforts you may find the magician reacting too early or late. Also bear in

Some observers might wish to amalgamate some of these groups. For example, it might be argued that group 3 is simply a variant of group 2 – that the presence or absence of the adverb does not affect the pattern of 'REACT + *to* + noun group'. It is possible too to join group 6 with groups 2 and 3 because in each case the prepositional phrase is obligatory. Others might argue that groups 1, 4 and 7 should be conflated because in each case REACT is the end of a clause. Still others would want to add group 5 to those because, it could be argued, in those lines the prepositional phrase beginning with *with* adds only peripheral information. Discounting the 'other lines' (group 8), this would yield only two groups:

## (A) REACT coming at the possible end of a clause:

1           could not believe the way Vieira reacted after he was dismissed. The  
 2    at all. When asked today how they' d react if the White House sent them a ne  
 5    any more, I don' t know how he would react. Is there any point in making  
 6    growth because stock markets could react.' Mr Visco said stock markets in  
 10   strength of a substance and the body reacts to fight off any diseases which  
 11        from the air and induce them to react to form harmless gases. Last  
 14    conference was to see how he would react when asked questions by journalis  
 15   protect. How is management likely to react when a group threat-ens to quit?  
 16        twenty-year-old son felt free to react with such ferocity indicates that  
 19        Commentators and crowd alike reacted with astonishment when Lara  
 20        during the investigation. They reacted with anger and said: The findin

(B) REACT followed by the preposition *to* or *with* as a necessary part of the clause:

3    step, which will enable viewers to react immediately to what they have see  
 7        police officer at Selhurst Park reacted similarly to the Cantona inciden  
 8        mail, in New York, Adrian Clark reacted to Simon Hoggart' s discussion of  
 9    market has come, and how people will react to it. The best seats and places  
 12        is the poster!' Herzen was reacting to a swelling trade in images. T  
 17    eposition sulfur and nitrogen oxides react with atmospheric water vapor to  
 18    above such common tasks, refusing to react with the molecular masses. < p > Bu

There are of course intermediate positions – it is possible to make three or four groups here as well as eight or two. The point is that no one grouping is absolutely right or wrong; all the groupings use formal information (that is, information based on the form of words) but also linguistic interpretation (distinguishing between the preposition *to* and the *to*-infinitive, for example, or between the two uses of *with*). A smaller number of groups tends to give a limited amount of information – the division into groups A and B, for example, tells us very little except that REACT may occur with *to* and *with* or may not. On the other hand, division in many groups runs the danger of masking genuine similarities – placing the lines with adverbs into a different group from those without tends to hide the importance of the link between REACT and *to*.

In addition, of course, quite different groups can be made if different aspects of the concordance lines are brought into account. For example, all the lines in which the subject of REACT is non-intentional (not a thinking human being or animal) can be grouped together, giving a set comprising these lines (in line 18 the subject is *Gold*):

4           two-thirds of the radical pairs reacting (in a field of typically only  
 6    growth because stock markets could react.' Mr Visco said stock markets in  
 10   strength of a substance and the body reacts to fight off any diseases which  
 11        from the air and induce them to react to form harmless gases. Last  
 18    above such common tasks, refusing to react with the molecular masses. < p > Bu

It is noticeable that this set includes all the lines from the original where REACT is followed by a *to*-infinitive clause and the only line where it is followed by *with* giving essential information.

To summarise: observing pattern in concordance lines essentially involves grouping those lines together. In most examples, several alternative groupings could be proposed, each highlighting different kinds of information. There is no objectively correct grouping, although some will be more useful for particular purposes than others. Although the presence of individual words may provide help in grouping, usually a wider context and more interpretation is needed to form groups (that is, to identify patterns) that might be thought to be appropriate.

#### 4. Short cuts and how to find them

There seems to be a limit to the amount of data a human observer can usefully process using concordance lines. (It is partly for this reason that some people advocate limiting the size of a corpus, so that no search will produce more concordance lines than can reasonably be studied.) Most concordancing programs will allow a limited number of lines to be selected from a larger set; in effect every 'nth' line is displayed, giving a random sample (see Evison, this volume). If the corpus is large, and/or if the word is one that occurs frequently, however, a sample that can reasonably be dealt with may be too small in relative terms to reliably demonstrate patterns. For example, in the 450-million-word Bank of English there are 705,866 instances of the noun *time*. Looking at all those concordance lines would be a daunting task. If 100 lines are selected, some patterns start to be revealed, such as *the first time* (four occurrences), *at the same time* (four occurrences), *at the time* (five), *at the time of* (three), *it is time to* (three), *by the time* (four) and so on. Some patterns which occur in this sample only once, however, include: *by this time* (one), *for a time* (one), *it's just a matter of time* (one), *it wasn't the first time that* (one). By contrast, the phrase 'it + BE + negative + *the first time* + clause' (as in *it wasn't the first time that*) occurs in the whole Bank of English 564 times. In short, obtaining a manageable sample of concordance lines for a very frequent word can make it difficult to observe patterns reliably. An alternative method of finding patterns needs to be found, or rather, an alternative way of focusing searches so that a smaller number of lines with a smaller range of patterns is identified.

Fortunately, because the starting point for identifying patterns is linguistic form, the collocates of a given word can be used to narrow down the search. Many concordancers allow two-or-three-word phrases (n-grams) containing a given word to be identified. For example, the most frequently occurring three-word phrases including the word *time* are:

the first time  
 at the time  
 the same time  
 a long time  
 by the time  
 all the time  
 at a time  
 the time of  
 a time when  
 of the time

Each of these can then be the starting point for further, more manageable, searches. For example, *a time when* occurs in longer frequent phrases such as *at a time when, there was a time when, to a time when, there will come a time when*. Random concordance lines based on the phrase *to a time when* show further patterning:

< p > If Brett wants to go back to a time when you could get the shit kicked a common impulse to 'hark back to a time when Christmas was simpler, more day United States, going back to a time when the southwestern states were a gratification of having children to a time when they would have a better shot many of which speak directly to a time when things were fragile and belongs to an earlier era - to a time when current-account imbalances, not victory will provide a flashback to a time when the US was at the height of its Dream' address, looking forward to a time when racial prejudice no longer and we hope and look forward to a time when this technique, along with it is possible to see forward to a time when ARM will prosper. < p > Indeed, set back the opening 30 minutes to a time when generally sufficient liquidity injections. He immediately went to a time when he was only about five years

From these lines it appears that the word *to* links to phrases indicating a figurative movement in time, either backwards (*go back to, hark back to, provide a flashback to, set something back to*) or forwards (*look forward to, see forward to*). The clause beginning with *when* indicates a situation contrasting with the present one: *Christmas was simpler; racial prejudice no longer existed; they would have a better shot at supporting them*. The contrast is often an evaluative one, that is, things in the past or future are either better (*Christmas was simpler, more authentic*) or worse (*things were fragile and uncertain*) than the present. From a simple three-word phrase, therefore, we have established a much longer pattern that might be represented as 'reference to past or future time + *to a time when* + evaluatively contrasting situation'.

Using collocation gradually to refine a search for pattern does not depend solely on the identification of n-grams. A technique of 'accumulative collocation' (Bang 2009) can be used to perform recursive searches that gradually refine what is observed. This was demonstrated above in the discussion of *some of the best schools in America*. The example given below partially replicates one given in Hunston (2008) and starts with the word-form *distinguishing*. The most frequent adjacent-word collocates of *distinguishing* is *between*, so the string *distinguishing between* is then taken as the starting point for a further search. The most frequent adjacent collocates of *distinguishing between* is *of*. Taking *of distinguishing between* as the node, the words which most frequently precede this string are: *way, capable, importance, difficulty, means, incapable, task, point, method* and *ways*. These might be divided into three sets: (1) *way, means, method, ways, task*; (2) *capable, difficulty, incapable*; (3) *importance, point*. What is important here, in terms of method, is that none of these words, individually, is very frequent. For example, the string *ways of distinguishing between* occurs only twice in the Bank of English, so techniques that identify frequent n-grams are unlikely to find it. This point will be taken up again in the next section.

Before going any further, we might note that other prepositions also occur frequently before *distinguishing between*. These are: *in, by* and *for*. Repeating the exercise above for each of these expands the three sets, as shown in Table 12.1.

Each set can then be taken as the starting point for further searches.

**Table 12.1** Collocates of ‘distinguish between’ divided according to meaning

‘method’	‘able/unable’	‘importance’
Way/s of	Capable of	Importance of
Means of	Difficulty of	Point of
Method of	Incapable of	Important for
Task of	Difficulty/ies in	
Done by	Problems in	
Achieved by		
Basis for		
Method for		

## 5. A note on frequency

This chapter has made some assumptions about the notion of ‘frequency’; some of these will now be made explicit.

Statements about frequency are always relative (although figures can be attached to them). For example, describing a word as ‘frequent’ means that it is frequent compared with other words. In a corpus of spoken British English (one of the spoken components of the Bank of English), for instance, the word *time* occurs just under 1,670 times per million words, whereas the word *circumstance* occurs only twice per million words. It is reasonable to say, therefore, that *time* is frequent in this corpus and that *circumstance* is less frequent. Programs such as the Keywords function in *WordSmith Tools* (Scott 1999) identify significant differences in the frequency of a given word in two corpora; the usefulness of this is discussed in Scott (this volume).

Frequency of co-occurrence is also relative. Much of the discussion in the previous section relied on a calculation of the words that co-occurred most frequently with a node word. There are statistical computations that measure the significance of co-occurrence (this is explained in Scott, this volume), but in this chapter ‘raw frequency’ of co-occurrence is all that has been used. Raw frequency may be an absolute concept (for example, in the spoken corpus mentioned above *at the time* occurs over 100 times per million words), but the concept of accumulated collocation assumes that a sequence or phrase can be identified as patterned not because it occurs frequently as such but because each element in the phrase co-occurs relatively frequently with the cumulative other elements. For example, although *way of distinguishing between* is relatively infrequent (occurring only 0.1 times per million in the Bank of English corpus), as a phrase it can be built up by taking *distinguishing* as the node and then progressively identifying the most frequent adjacent collocate (*distinguishing between*, then *of distinguishing between*, then *way of distinguishing between*).

It has further been assumed that the elements that are identified as frequent may not have a formal similarity. To illustrate this I shall take a sample sentence which uses the verb REACT that was discussed in some detail above:

It is somewhat ironic how people in general react to the circumstances in their lives.  
([http://ezinearticles.com/?expert=Kenia\\_Morales](http://ezinearticles.com/?expert=Kenia_Morales) (accessed 11 August 2008))

In all probability this sentence is unique. It certainly occurs only once in a search of the world-wide web. On the other hand, the sentence does ‘feel’ familiar, because it is composed of elements that are highly predictable because they are part of the patterning

of English. Not all this patterning relies on repeated words, however. For example, although *it is somewhat ironic* occurs less than 0.1 times per million in the Bank of English, *it* followed by some form of *be* or *seem* followed optionally by an adverb and finally by *ironic* occurs nearly three times per million words, accounting for 28 per cent of the total instances of *ironic*. Similarly, a sequence *it* followed by some form of *be* followed by an adjective followed by *how* occurs about four times per million words, and the same sequence with any one of *what*, *where*, *when*, *why*, *who*, *if* or *how* as the final item occurs almost fourteen times per million words. It might be argued, then, that what is perceived as ‘pattern’ is not *it is somewhat ironic how* but ‘*it* + link verb + *ironic*’ and ‘*it* + link verb + adjective + wh-word’.

It was noted above that REACT *to* frequently follows expressions of ‘not knowing’ or of ‘finding out’. This is not the case in the sample sentence, however. To discover whether REACT *to* frequently follows the ‘*it* + adjective’ pattern we can ask whether specifying a sequence of ‘adjective + *how* + someone reacts to something’ reveals an alternative pattern into which the sample sentence would fit. A search for this sequence gives these concordance lines:

rock. It's amazing how ordinary folk react to a fiddle in the safety of a  
 nd it will be critical how consumers react to the new tax system.'  
 an never be certain how someone will react to something which may have been  
 PLO. It is not clear how Israel will react to the advisory council. Israel TV  
 It is less clear how consumers will react to the growing financial turmoil.  
 It's not clear how Moscow will react to this latest move. The position  
 But it's not clear how Ankara would react to a request because any such move  
 It isn't clear how Americans would react to the new IRA proposals because  
 of erm It's interesting how people react to something like this. Okay. < M01 >  
 < p > It was unclear how Cigar would react to a wet and sticky racing surface  
 for his reply, yet unsure how she'd react to it. May we know how she died,

The dominant pattern here is indeed an anticipatory *it* and an evaluative adjective. This might be further subdivided into cases where the adjective indicates lack of clarity (*never certain, not clear*) and the reaction is hypothetical (*will/would react*) – these examples have something in common with the ‘not knowing’ examples given above – and a smaller set where the adjective indicates other kinds of evaluation (*amazing, critical* and *interesting*) and the reaction is actual.

In short, the sample sentence exemplifies both pattern and non-pattern. At quite an abstract level, the sequence ‘evaluation of reaction + reaction + stimulus to reaction’ recurs, though it is less frequent than other sequences such as ‘not knowing/finding out + hypothetical reaction + stimulus’. In grammatical terms, the sequence ‘*it is* adjective *how*’ occurs relatively frequently before REACT. In more specific lexical terms, that sequence with the adjective *ironic* is relatively infrequent, though by no means unknown. At every stage of this investigation varying concepts of what constitutes frequency have been used.

## Further reading

Hunston, S. and Francis, G. (1999) *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins. (This book takes a rather more restricted view of ‘pattern’ than

- the one discussed in this chapter, but it goes into quite a lot of detail about the relationship between pattern and meaning.)
- Partington, A. (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins. (This book is rich in examples of lexical patterning and how concepts such as this might be applied.)
- Sinclair, J. (2004) *Trust the Text*. London: Routledge. (This book sets the concept of pattern within a more general theory of language.)
- Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. New York: Blackwell. (Along with Sinclair 2004, see above, this develops a theory of language based on patterns.)
- Tognini Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins. (Another book with a lot of examples in both English and Italian.)

## References

- Bang, M. (2009) 'The Representation of Foreign Countries in the US Press', unpublished PhD thesis, University of Birmingham.
- Carter, R. A. (2004) *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Conrad, S. and Biber, D. (2000) 'Adverbial Marking of Stance in Speech and Writing', in S. Hunston and G. Thompson (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, pp. 57–73.
- Hunston, S. (2008) 'Starting with the Small Words: Patterns, Lexis and Semantic Sequences', *International Journal of Corpus Linguistics* 13: 271–95.
- Obama, B. (2008) <http://my.barackobama.com/page/content/hisownwords> (accessed 7 September 2008).
- O’Keeffe, A., McCarthy, M. J. and Carter, R. A. (2007) *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Scott, M. (1999) *WordSmith Tools version 3*. Oxford: Oxford University Press.
- Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- (2004) *Trust the Text*. London: Routledge.
- Tannen, D. (1989) *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.
- Teubert, W. (2005) 'My Version of Corpus Linguistics', *International Journal of Corpus Linguistics* 10: 1–13.