USING CORPORA FOR LEXICOGRAPHIC PURPOSES

Gwyneth Fox
Editorial Director
Cobuild
University of Birmingham

1. INTRODUCTION

Throughout the twentieth century linguists have been studying languages, analysing sentences, telling people how they work. Unfortunately, many of the statements made have not been based on evidence. Linguists have sometimes described how they think the language is used rather than how it actually is used. There are of course exceptions to this. The early 1960s saw the beginnings of the work of Quirk et al on the Survey of English Usage, a corpus of British English systematically designed and collected.

In the United States a one million word corpus of American English was collected by Francis and Kucera at Brown University. These, and other, corpora were used to provide evidence for statements about the language. But corpus linguistics did not really take off in a big way for a further twenty years.

In 1980 the second generation of machine-readable corpora was launched when the Cobuild project in lexical computing was set up at Birmingham University. This was a project jointly sponsored by the English Department in the university and William Collins the Publishers (now HarperCollins Publishers). The main aim of the project was to find out how people use English in the late twentieth century. To do this, evidence was needed, and so a corpus of British and American English was built, in the first instance consisting of about 7.3 million words but soon reaching approximately 20 million words.

These 20 million words were alphabetized and put into concordance format, so that researchers had access to all the examples of any particular word which appeared in the corpus. These concordances were then analysed in order to build up a picture of how a word was typically used, and anything that was considered relevant was stored in a database. This obviously included the word itself, all lemma forms, meanings, grammar and syntactic patterns, collocational information, synonyms and antonyms for each individual sense, register and style labels, pragmatic uses, and so

Revista Canaria de Estudios Ingleses, nº 25, 1992, págs. 45-53

on. Each statement that was made about the word was backed up by at least one example taken from the corpus.

When most of the words in the corpus had been analysed and entered into the database, work then started on the next phase of the project - extracting a dictionary for learners of English as a foreign language from all the statements in the database. This was completed by December 1986, and in May 1987 the Collins Cobuild English Language Dictionary was published.

Both Collins and the University felt that the project was a success, and it was decided to form a company, Cobuild Ltd, which would be jointly owned by the two original sponsors, in order to continue the work. This was done on 1st January 1987, with Cobuild Ltd housed at the University of Birmingham as an independent unit. Since that date, six dictionaries (including the BBC English Dictionary), two grammar books, a usage book, and a number of small guides have appeared. The corpus has grown from 20 million words to nearly 200 million words, and there are research and publication plans stretching into the 21st century.

2. LEXICAL EVIDENCE

The corpus is at the heart of the work done at Cobuild. The statements made in the dictionaries, grammars, and other reference books all derive from evidence of what is found in the corpus. So what is it that the corpus can give that intuition based on previous knowledge of the language cannot give?

In the first place, and most obvious, a corpus shows which words are frequently used and which are not. Most speakers of English know that the word 'the' is the most common word in English, with well over a million citations in the original 20 million word Cobuild corpus. How many speakers, though, know that the second most frequent word is 'of', with more than half a million citations? Most of the top 200 words are function words of this kind, with only 'said', 'time', 'way', and 'people' having much lexical content. The frequency of words drops dramatically, from 'the' with its million citations to 'thing' (number 172) with 10164, to 'set' (number 254) with 6246, to 'taciturn' (number 22801) with 29. And so on. So one reason for using a corpus is to find out which words are important for learners to know well - and these are obviously the more frequent words. There is no point investing a great deal of learning effort on an item which will only rarely be encountered.

When looking at an individual word, a corpus shows, not only which words are frequent, but which senses of the word are frequent and which are not. And the frequent senses are not always the same as those taught in EFL coursebooks. Take 'take', for example. Coursebooks deal mainly with 'take' meaning 'get from someone', as in 'Let me take your coat', and 'carry with you', as in 'Don't forget to take your umbrella'. As soon as you look at a corpus you realise that these senses can be found, but they are much less common than a delexical use of 'take' which is found in very few books - 'She bent down to take a look at the baby', 'I wish I could take a holiday in Italy too', 'He was too excited to take any notice'. In these examples

'take' has very little meaning at all, with most of the lexical burden being carried by the following noun phrase: what is important in the quoted sentences is 'look', 'holiday', and 'notice'. Another verb of this kind is 'give', as in 'We usually give a concert at Christmas', 'I shall give two examples relevant to the subject in hand', and 'He hoped he had given a good impression'. In none of these examples does 'give' have the meaning you might expect it to have, ie 'hand over (often as a present)'. All the examples quoted are taken from the corpus and thousands more examples of a similar kind could be given. It is obviously important for students to learn this use of 'take' and 'give', as it is the one they are most likely to come across outside the fairly rarefied vocabulary of the classroom.

There are similar points to be made about many other words. The word 'like' is taught almost exclusively as a verb. But by far the most frequent use of 'like' is to say that two or more things are similar in some way: 'He looked like Tom Cruise', 'She's very like her sister', 'The lake was like a mirror', and so on. 'Thing' is usually taught as meaning 'object'. It does have that meaning, of course: 'chairs and tables and things like that'. But it is at least as frequently used as a preface in which you signal what you feel about something you are saying: 'The strange thing is that it doesn't taste fishy at all', 'The essential thing is: don't reply to him ', 'The important thing is that we love each other', 'The remarkable thing is that all the 58 countries agreed in principle to the proposal'. Again, this is a use that appears in few, if any, course books.

You notice other things (another typical use of the word 'thing') as you read the contexts in which a word has been used. We all know that there are some words - often insulting ones - that are used to refer to women: 'bitch', 'tart', 'butch', 'angel', and so on; and others that are used to refer to men: 'macho', 'hunky', 'effeminate', etc. But an examination of a corpus shows that there are other words which seem, for no obvious reason, to be typically used to refer to one or the other sex. In our original 20 million word corpus there are 29 examples of the word 'taciturn'. Of those 29 examples, only three refer to women. Why should this be so? Are women actually not so quiet that they seem unfriendly? Are they not perceived as being so? Is there another word that we use instead? (If so, I have not managed to find it.) Also, 'taciturn' seems rarely to be the only adjective used; a man is described as being 'reserved and taciturn', 'gloomy and taciturn', 'taciturn and unsmiling', 'taciturn and devoid of curiosity'. Thus, a picture of the person is built up, of which taciturnity is one vital element. This means that if learners of English want to use 'taciturn' in the way that is most typical of English users, they should use it to refer to men, and they should also use it with one or more other adjectives. It would have been difficult, if not impossible, to arrive at that picture of the word without the evidence of the corpus in front of you. You need to see numerous examples to be sure that what you are saying is representative of typical use.

There are many words which you think you know a lot about: their meanings, their word class, their syntactic patterns. But when you look at the examples, you discover something which is new to you. The phrasal verb 'set in' is an example of this in my own case. When looking at the lines, I realised for the first time that the

subject of the verb almost always refers to something unpleasant: panic sets in, decay sets in, the cold weather sets in, gloom and despondency set in. But never joy or happiness or sunshine or recovery. They could do, of course. The excitement of language is that you can play with it, you can make it do what you want it to do - but you do this only from the secure basis of real understanding. For learners things (that word again) are slightly different. If they are to use 'set in' in ways which native speakers will accept as normal, they need to know that it means something like: 'If something unpleasant sets in, it begins and seems likely to continue or develop for some time'. A concordance for 'set in' is given in Figure 1.

```
ed a lot, and nobody thought much about it. Then a reaction set in, and many parents decide that it was shameful. But t
the past, so presumably she would again, once normality had set in, and habit, and she was back in her family, sweeping
 out of hand. Though disillusionment with the Maharishi had set in, it was felt that the potentialities of India as a p
hat Christianity was tolerated by Abgar V. When persecution set in, it had to be removed. The real surprise is that, in
nterior of the trunk. Within a few hours rigor mortis would set in, jamming the corpse into its adopted position at the
er having left even for one moment the pattern her life was set in, seemed a mistake chosen by a madwoman. The violence
turnips. Harvesting Lift in the autumn before heavy frosts set in, top, put in little heaps covered with their own lea
live in the beam of its tedious, simpleminded glare. Doubts set in - is this why the natives barricade themselves behin
d the slippers off you the stiff socks. "Pongification might set in." This was a joke for the servitor. Who didn't respo
nder other names, and it was no wonder that disillusion had set in. Assassination was not far behind. The association b
 slant of the sun and a late afternoon depression begins to set in. At another time of day these mountain meadows would
ts zenith, had achieved little and a feeling of anti-climax set in. Attempts to repeat the occasion would have been fru
r them to find a roof to live under before the cold weather set in. But there was little chance of employ- ment for her
t the chance), and makes plenty of growth before the frosts set in. Frost may destroy very young wheat by dislodging th
xchange. By the time he had got it back in place, panic had set in. He had to get out of the doll. He just had to. Ther
 "Get firewood and coke in tonight in case this weather has set in. I'll listen to the forecest this evening and let yo
o return---a point at which spathy and emotional withdrawal set in. In short, the available evidence strongly suggests
tonous. Boredom, which I now saw was my industrial disease, set in. I was growing old. My accumulating discontent with
twentieth century, especially after World War I, a reaction set in. Several factors pushed it along. Pioneers in educat
em pass him in rank and privilege the bitterness started to set in. There was only one thing left to do, and that was j
inconvenience. Then the final stage of emotional exhaustion set in. The soldier seemed to lose the very will to live. H
ecome conscious there was something to examine, the rot had set in. The climactic moment of three years ago had been wh
ld also need a low tide--and it had to come before darkness set in. These two critical factors of moonlight and tide sh
ning openly to hope that a little infirmity of purpose will set in. The Opposition has succeeded superbly well in artic
nts of the long war - are going back to see where the rot set in. Who was to blame? President Johnson blamed the doub
ation worth while, whereat a desperate sense of frustration set in; after dreaming for so many centuries of travelling
othes, not old either, probably very good before sloppiness set in; good hat too, of an excellent felt, the band of whi
n that follows the death of any very successful author soon set in; it is still too early to say whether it will revive
eal outsider to report on the anarchy and ill-will that had set in between the British and the Jews. The editor gently
rong. It had nothing to do with the strike plans. The slump set in during the great hot summer of 1921. I remember it w

e of the slowdown of white blood cell production, infection sets in, and sores break out around the lips before spreadi
atistic. For at a certain point, a feeling of deep lethargy sets in, and our reaction is: "What's the use?" or "What ca
g as boredom with stereotyped thrillers and worn out comedy sets in, but the BBC provides for those who are privileged
le. She begins to breathe hard, mother is irritated, asthma sets in, mother apologizes and the "Asthma Game" now runs i
 private jet. I'd better say, before a universal prejudice sets in, that no American institution is worse understood a
n. When they die, this intake ceases, and radioactive decay sets in, the carbon 14 changing into other atoms at a preci
locks for the sheep on the fell tops before the real winter sets in - as it will. ((BEER AND SKITTLES NOW)) It's no won
ng comes up. He's only going to use it till the bad weather sets in." Mengele said, "I need to have the dates the men w
s at 11 pm. That, says Garry, is when the loneliness really sets in. But who wants to know? They've got their cars, the
 only increases the mother's uneasiness, so a vicious cycle sets in. If a mother understands this mechanism, she can us
es of snow. <P 198> May is the squishy month, when the thaw sets in. Summer bangs in with ninety degrees. Fall is a fou
 After a production has opened, yet another kind of tension sets in. The director feels the need to protect his work fr
and once this light is excluded a process of impoverishment sets in. Variety, which is not subject to numbering since i
```

Figure 1

A corpus often shows another area where dictionaries and other reference books get things wrong. Dictionaries give information about word families; for example, they say that 'lamely' and 'lameness' are related to 'lame' and they give the impression that these derived words are equally likely to be used with all senses of the headword. A corpus might disprove this. The adjective 'lame' has two main meanings: unable to walk properly, as in 'The illness left her permanently lame' and 'a lame horse'; weak or unconvincing, as in 'a lame excuse', 'these lame remarks'. An examination of the corpus evidence for 'lamely' shows that it refers to the second meaning only, and is almost always used with the verb 'say' or some other reporting verb: '"Well," Rudolph said lamely, "Good luck"'. It is important that learners know this, so that they do not say things like 'She walked lamely down the road' meaning

'She limped down the road', because this is likely to be misinterpreted by a native English speaker. 'Lameness', on the other hand, does seem to pick up on both meanings of 'lame'; you can deplore the lameness of someone's excuse and you can also be dismayed by the lameness of the horse you had hoped would win a race!

The same thing is true of almost every adjective you examine. The Cobuild English Language Dictionary gives eight different adjective meanings for the word 'crisp', ranging from the most frequent meaning 'Fruit and vegetables that are crisp are fresh and have a firm texture, so that when you bite them they are hard and crunchy; used showing approval' to the least frequent meaning 'Behaviour that is crisp is cool, sharp, and unfriendly, and does not consider other people's feelings; used showing disapproval'. Note how the same word can in one meaning be appreciatory and in another derogatory. Although there are eight distinct meanings of the adjective, there is evidence for the adverb 'crisply' for only two of them - one of which is for the least frequent, derogatory, meaning quoted above. It is therefore misleading for students to be given the impression that every adjective meaning has a corresponding adverb.

By looking at word families it is often obvious that an adverb has senses that are not found in the related adjective. 'Broad' and 'broadly' have very little in common, except that someone with 'a broad smile' on her face could be described as 'smiling broadly'. Much more common, though, is the use of 'broadly' in sentences such as 'It was done broadly as planned' and 'Their views are broadly compatible', where it means roughly 'more or less'.

Another example of words which seem to be related is 'bare' and 'barely'. A quick glance at concordances for the two words show that they have no meanings in common. The most frequent sense of 'bare' is 'having no clothes on'; the most frequent sense of 'barely' is 'only just' or 'hardly at all'. (Notice how difficult it is to define 'barely'; no dictionaries succeed in getting it right!) A few lines from both 'bare' and 'barely' are given in Figure 2.

on her brown neck, and crumpled blue shorts. Her feet were bare, and her brown hair, very glossy, was caught up in a p
bby, nervous faces old before their time, the same rags and bare, black feet. The same smell of sweat and marijuana. Al
s to blossom. Why should those branches not remain for ever bare, the earth for ever hard and inhospitable? By what gra
he first two days were dreadful. The earth was scorched and bare, ugly grey dust covered everything and I got sick twic
life. Ripples of pleasure running down the length of those bare, white, rusted arms. "So you made your own candles," s
x children, and many rooms used for children are completely bare, without toys or equipment. Sonia Jackson explains in
in on his arm or hand or knee, any place where the skin was bare. His father, who was a bricklayer, said there must be
her tighten her raincoat about her. "Her arms and legs were bare. In our society it would be immodest, but Posy did not
is evening. She wore a yellow dress that left her shoulders bare. It was perfectly plain, managing, as if it had been i
investigate. He wondered if it mattered that his feet were bare; he was, in any case, naked inside his dressing-gown.
tep into my dungeon?" It was a small square room, white and bare above a line of brown panelling, windowless and lit by
ght either. The set, although it was rather impressive, was bare and bleak, and the audience had no idea what we were g
I plumped down my heavy paper bag and stood staring at the bare ankles of her sandalled feet, then at her silly face,
aching on his chastity. She reached out and gripped Bobby's bare arm and clung to him the way she had wanted to cling t
would upset such a lucid dear? Did I ever mention that her bare arm bore the 8 of vaccination? That I loved her hopele
irl, her long hair spread romantically over the pillow, one bare arm hanging to the floor. He sat in his pyjamas and lo
ear-old skinhead in monkey boots, with tattoos all over his bare arms, came lumbering down the stairs. He had been lurk
-cold stream; begging pennies for bottles of pop; of girls' bare arms and unripe cherries, green apples and liquid waln
and knees and dragged myself up on the bed beside her. Her bare arms felt unnaturally hot and I realized that I was sh
st have come tip-toeing up behind me - all at once I felt a bare arm sliding through mine, and one second later her fin
ace in your plaid skirt, my darling! The fragility of those bare arms of yours - how I longed to enfold them, all your
idence that might have brought them together. She threw her bare arms round his neck and gave him a long open-mouthed k
receive a public flogging of thirty-nine lashes across his bare back by his most recently wronged husband, according t
ing a masked and costumed figure brandishing a rod over the bare back of a screaming woman whom several other women hel
erless, without food, enduring the pitiless rain upon their bare backs with such humility and hopelessness. Tears spran
he scene outside might have come from a Christmas card. The bare black trees stood out starkly against their grey backg
a hard chair by the windows sat a girl in the sunlight. Her bare brown feet were planted side by side on rush matting.
some boys were in fact running the cinder track now, their bare chests, with their newly sprouting fleeces of hair,sli
y Eddie. Warm tears were dripping from her eyes and onto my bare chest. "Don't leave me, Ginny. Please don't. I couldn'
nment prices, cost around 150 rupees. This consisted of the bare essentials: eight pounds of rice per head, nine pounds
r actor playing John the Baptist in the school play. He had bare feet, wore jeans, a sort of smock, which was possibly
evening in a white peasant-style dress that reached to her bare feet, her long brown hair loose about her shoulders, h
e door. She stripped, and standing on the cold floor in her bare feet, looking at her reflection in the mirror above th
the house, padding happily through the dewy grass with her bare feet, her transparent cotton nightgown blowing around
uld not recommend walking down the streets of any city with bare feet, and shoes and boots do keep our feet warm in win
arkened by a suntan preparation, and he wore sandals on his bare feet - this in midJanuary with the temperature well be
urning, and the flames were moving closer and closer to her bare feet. In a few minutes they were going to reach her, a
to become a - what is it? - hippy, ja, and walk home in my bare feet." And she opened the door of the car. I leaned ac
and cooking. Her father was very tall, had a red beard and bare feet and swore at us, which was the only time he ever
e and red; a white blouse and red Paisley headscarf. By her bare feet lie a rush basket, sandals and a strew of books.
ing from a mattress as we entered the room and trampled the bare feet of the young girls as it lurched out through the
rinciple of the surgical mask operates. No face is seen. No bare feet pathetically poking out of a regulation gown. No
othing !" And I began to give out my howl. <P 19> A slap of bare feet slithered across the floor, and I heard sister Ma
instant and looked at the sleeping girls. She was safe! On bare feet she went to the washstand and took her tooth glas
en't any shoes!" she saw and protested, but not enough; his bare feet were no longer so incongruous as her clothed body

of public rental housing. In fact, the list of authors who barely, or sometimes never, mention - let alone deal adequa
rkable and illuminating confession Lucas did not listen, or barely, since he was going on to explain that he had also r
- which seemed incredible in a county whose population was barely 3 percent Black. "Do you have any experience in havi
ightly swathed waist and hips; and the whole, though short, barely 5"8", certainly quite twenty years of age, certainly
ies have shut down this year, and the rest are operating at barely 68 per cent of capacity. Americans have shifted away
PECTED TO FEATURE IN SEVERAL EVENTS WAS DALEY THOMPSON; BUT BARELY A MONTH AFTER DECLINING TO COMPETE IN THE EUROPA CUP
the terrace of the Continentsl waiting for the bar to open, barely able to light their own cigarettes until it did. The
n excelled. At ground level he seemed gangling and awkward, barely able to control his great limbs whereas on rock he c
gedy as close friends of Steve's, and still in shock, still barely able to believe he was dead, we wrote these articles
st twice as big as this one," said Brody. His <P 226> mind, barely able to comprehend the fish he had seen, could not g
rehead. The Negro got up at the count of eight, staggering, barely able to get his guard up and Thomas, his face a bloo
o, subaqueous, a watery world into which the visitor swims, barely able to discern the tall break-front containing tatt
here were you, God damn it?" And again, in that soft voice, barely above a whisper, she answered: "I can't tell you, Th
sight of the long-preserved footprints of young children in barely accessible nooks of extensive cave galleries is part
elf (not the same as selfconsciousness) is so pervasive, he barely acknowledges the presence of other actors. He plays
address, to outwit unconsciousness.) When he got in, Rhoda barely acknowledged his return: she went on shredding lettu
before boarding an airplane. Yet this all came to pass with barely a demurring voice. Protests and campaigns on the par
o spend the night there and carry on next day, but they had barely adequate bivouac kit, and a line of high-flying clou
, particularly for families with children, incomes that are barely adequate to meet their needs at a level that is cons
to Botswana to land soon at the capital, Gaberone, which is barely a dozen miles from the border, when we flew out of a
m, lying alongside a narrow dirt road, topped a slight rise barely a half-mile from <P 68> the sea. He hadn't
gravity. The scene had changed since the party had arrived barely an hour ago. Though the stars, and the half-earth, w
, she said with a sincerity he couldn't doubt - in fact she barely avoided turning it into <P 291> him into what became
n them: a flimsy one it is true, and Ash, for his part, was barely aware of it. But because of those years in England,
an Afrikaner Nationalist. His objection was overruled, and barely a week later, during the screening of a film at the
as to satisfy demands which even a <P 84> few years ago had barely been articulated. In Europe the main pressure for ch
cycle, and my muscles weren't only unknown - they and I had barely been introduced. I had managed to start in the morni
he had collapsed at the court of his revered Rabbi and had barely been revived. There had been a great commotion among
nalism still retains its dynamic, and social revolution has barely begun. Nationalism and Communism have therefore been
usandth of what was yet to come; the story of this star had barely begun. The pod had ceased to roll; the great red sun
you, Johnny," he said, his voice slurring so that he could barely be understood. "I'm a better singer than you now. I
now felt in her still flat one? Whatever was in there would barely be visible through a strong lens, yet Lynn knew she
you will find them, in the crude state rough and sometimes barely blushed with colour or light - and in the cut and po
earable. He seems to be so busy on his own projects that he barely comes home to sleep. From what I can tell, especiall
re "Bamuthi was satisfied that they had enough wood. He had barely come to this conclusion when from the direction of t
olonel, holding the memorandum in his fingers as if he were barely conscious of it as he talked, said with what he obvi
ly, as if sucking on a wellrubbed briar pipe. "Ego-trippers barely constrained megalomaniacs and blundering oppressors
man jovially, "and get to know the animals a bit." He could barely control the smirk on the other side of his face. My
ed his way out over the river, and his insides frothed with barely controlled terror. But he kept going. He even forced
room. Greene put his drink down on the table and said with barely controlled fury, "What's this I hear the Corleone Fa
t our sleeping bags and discover that the pine needles just barely cover what must be many feet of soft spongy dust. I'
was successfully recommended. The pay was criminally poor, barely covered the money she spent on clothes, and nothing
women buyers among their best customers. My widow's pension barely covers the essentials after the inflation of the las
knife and fork neatly and without haste on the steak he had barely cut into yet; he even folded his napkin and pushed i
r aspect of the age structure of the housing stock which is barely dealt with in the literature is the impact upon the

Figure 2

3. GRAMMATICAL EVIDENCE

It is also possible to study grammar from a corpus by looking at all the words which behave in similar ways. This again can yield surprising results. In the classroom learners are often taught that most verbs are either transitive or intransitive, with a few being both. The truth is that most verbs can be both transitive and intransitive, and that there are very few which are only transitive or intransitive all the time. Depending on the context, and the amount of shared knowledge there is between us and our audience, we as speakers or writers choose how much information we want to give, so we can say 'I never drink' or 'I never drink whisky', 'I cook every evening' or 'I cook a big meal every evening', 'I drive to work most days' or 'I drive my daughter's car to work most days'. Even verbs of behaviour, which are typically used intransitively, can have a cognate object when you want to specify precisely: 'She smiled at him' or 'She smiled her most beguiling smile', 'He coughed continuously' or 'He coughed a very painful cough'. And verbs which are so general that they normally need an object to specify the topic can be used without one when the object is obvious: 'He teases me mercilessly' or 'He likes to tease', where 'people in general' is understood. What is important here is the speaker's intention; he or she chooses how much to say, and that choice influences the decision to use a transitive or an intransitive clause. Verbs are not intrinsically transitive; it is the intentions of the speaker that are all-important.

Coursebooks teach that there is a choice to be made between the active and the passive voice. But looking at the evidence from verbs which behave in similar ways, we see that there is often a three-way rather than a two-way choice. We can use an active verb with an object 'A little girl opened the door', a passive voice 'The door was opened by a little girl in a green dress', or an intransitive verb 'The door opened'. If we choose the first option, we have to say who opened the door; if we choose the second, we can say who did it, but we do not need to; if we choose the third, we cannot say who did it. So once again we can choose how much we want to say, and that choice conditions the type of clause we use. There are over 800 verbs which Cobuild has isolated as behaving in this way, and has called ergative verbs.

When we look closely at the verbs which are typically used in this way, we find that many of them fall into semantic groups. For example, there is a group of verbs to do with cooking: bake, boil, cook, fry, melt, roast, simmer, stew, and so on; another group to do with change of state of some kind: begin, end, start, stop, change, grow, increase, etc; yet another to do with vehicles: back, crash, drive, sail, etc. All of these can be used with this three-way choice, although some are more frequently found in one pattern rather than the others. The fact that these words have both semantic and grammatical similarities is worth pointing out to students, who can then learn them as a group rather than as individual lexical items which take individual grammatical patterns.

There are many grammatical points thrown up by a corpus. For example, most grammar books say that when you have a reported clause starting with 'that' you can put it in or leave it out: 'She said that she was tired' or 'She said she was tired'. And

you can. However, the less extra meaning the verb has, the more likely you are to leave it out. So the statement is true for 'say' or 'think' or 'know'. But verbs which show your attitude to what you are saying, or which show the way in which you say it, are much more likely to have a 'that' at the beginning of the clause: 'The Prime Minister conceded that a mistake had been made', 'The girl announced defiantly that she was leaving', 'He murmured that he loved her'. It is possible to omit 'that'; it is much more common to use it.

The frequency of grammatical structures is relevant to teaching; there is little point in teaching something that is rare and unlikely to be heard. Take the different ways of expressing the future in English. Workers on The Survey of English Usage counted all instances of the future being used; researchers at Cobuild have not yet had the time to do that exhaustively, but when spot checks are made Cobuild data bears out what was found on the Survey: the 'will' form of the future is found in over 50 per cent of cases; the 'going to' form is next, with about 15 per cent; followed by the simple present and finally the present continuous. This shows that the most useful form to teach first is the 'will' form, and that the present continuous can be left until much later. Interestingly, too, you can always use 'will', but not the others, which all have various constraints. So why bother to teach them at an early stage?

At Cobuild we are just starting to do statistical work of this kind, and I am sure that the results of this work will lead to suggestions for the order of teaching structures in the classroom, with a concentration on those structures which can be most generally used in the widest variety of contexts. One of the areas we have recently been studying is the statistical relationship of positive to negative clauses. This consistently comes out at approximately 9 positive clauses to 1 negative clause. This again is relevant both to the classroom and to reference books. Examples given in both are normally positive examples, except where negatives are specifically being dealt with. This means that students do not get sufficient exposure to negative environments compared with what they will find outside the classroom. The more work of this kind that is done, the more typical will be the language taught to learners.

4. EXAMPLES FROM A CORPUS

All learners' dictionaries and grammars give examples.

When we were planning the Cobuild English Dictionary we took it for granted that examples would be given for most, if not all, of the senses we had isolated. The use of examples forms an important part of the learning of the word. The learner needs both an explanation of what the word means, and one or two examples of the word in use. This should help to reinforce the meaning - not by being a reformulation of the definition, but by showing the word used in a typical grammatical structure, with typical collocation, and in an appropriate context.

Having decided that there would be exemplification, we then realised that all the examples should come straight from the corpus. All Cobuild work is based on evidence derived from real language, and the whole issue of 'naturalness' is extremely

relevant here. Made-up examples may be well-formed grammatically, but they are often seen to be overloaded with information, when compared with ones which have occurred naturally. That is because naturally occurring language does not consist of neat little chunks which act as props to definitions; real language is untidy, with ragged edges, and has about it a lack of explicitness which is sometimes at odds with the perceived needs of a lexicographer.

It is not always easy to choose a good example. A concordance line, as can be seen in Figures 1 and 2, is a bit of a text, thus a part of something which is much bigger than it is. It relies on the text for its meaning. When you take a little bit out of a text, you take it from its setting, and occasionally the results seem bizarre, to say the least. It is important, though, to persevere. And fortunately, the bigger the corpus, the more chance there is of finding examples which are both natural and meaningful to a learner.

5. CONCLUSION

There is no doubt that corpus linguistics and corpus-based lexicography are here to stay. More and more computational linguists are gathering corpora for the expressed intention of studying a language - whether they are interested in it as a whole or are interested in individual genres such as business English, airline English, or legal English.

Corpora are being built in many countries, and it will soon be possible to compare language corpora, to find out what are the real similarities and differences between languages. This will lead to the use of two corpora for bilingual lexicography, and yet another milestone in our description of language will have been reached.