Andrew Sheehan
CHILE

MAKING SENSE OF WORDS

FOR CENTURIES, LANGUAGES HAVE BEEN CHANGING, AND ENGLISH IS NO EXCEPTION. Samuel Johnson expected that his pioneering dictionary, published in 1755, "[S]hould fix our language, and put a stop to those alterations which time and chance have hitherto been suffered to make in it." In fact, English is changing faster than most languages. For teachers, the fact that English is constantly changing and evolving can be unsettling and sometimes even dispiriting. Thank goodness for good old reliable grammar! But of course, MAKING SENSE OF WORDS is what language teaching and learning is all about. A recent upsurge in the interest in, and importance of, vocabulary in ELT has prompted a host of new books on the subject, and the advent of corpus linguistics has added support to the movement. How far does this lexical revolution extend?

It depends on where in the world you are, geographically and chronologically. While methods and approaches may have come and gone for many ELT academics in their universities, in many parts of the world where English is taught, and where contact with developments in ELT has been limited, working at the chalk-face has remained the same for decades. In my experience, the Lexical Approach—or to be more specific, the research on which the hypothesis is based—has not reached many parts of the world, and the majority of the teachers I have been working with recently are not familiar with corpus linguistics.

This article will address the following questions:

- What is the current thinking on the role of vocabulary in ELT, and how does corpus linguistics support this?
- What does corpus linguistics tell us about lexis and the importance of lexical development?
- What are the implications for English language teaching and learning?
- How can we help students learn, store, and retrieve the words they need?

What is the current thinking?

Vocabulary has been the neglected Cinderella of language teaching; preference has always been, and still is, given to the two sisters Grammar and More Grammar. There are several reasons for the Cinderella status of vocabulary. First, there is the legacy of previous language-teaching methods, particularly the traditional Grammar Translation Method with its emphasis on the learning of rules and structures. As Brown (2000, 15) states: "[T]he Grammar Translation Method remarkably withstood attempts at the outset of the twentieth century to reform language teaching methodology, and to this day it remains a standard methodology for language teaching in educational institutions." The Audiolingual Method (ALM), with its emphasis on repetitive drills, did nothing to change the balance. Brown goes on to observe that a key feature of ALM is that "Vocabulary is strictly limited and learned in context" (2000, 74).

Many English language teachers like to stress grammar over vocabulary because grammar is a finite system, whereas vocabulary is not. The reasoning is that a language teacher—especially one who rarely reads in English and has no access to English-language newspapers— could not possibly keep track of even a fraction of the words the English language now contains, let alone its ever-expanding lexicon. Consider, for example, such recent additions as *greenhouse effect, global warming, hip hop, grunge, ethnic cleansing, cyberspace, CD-ROM, hacker,* and *embedded reporter.*

However, the argument in favor of placing greater weight on vocabulary is strong. Meara (1995) points out that knowing only 500 words is functionally useless. English learners with such a minimal vocabulary who try to process a text will encounter too many unfamiliar words, and frequently these are precisely the words that convey the meaning of the text. Consider, for example, the following:

> While Argentina was *celebrating* the *victory* of its team in the World Cup, the president and his *family* took the *opportunity* to go on *vacation.*

Given enough time, students reading this sentence who have a low level of reading vocabulary might recognize the italized words because they are cognates, although they could easily misinterpret the key signal word *while* at the beginning of the sentence. But the same students listening to a native speaker saying this sentence at normal speed are not likely to understand or recognize most of the words, except perhaps, the words *Argentina* and *World Cup.* Even the cognates are likely to become incomprehensible because the pronunciation of these words in English is completely different from that of, say, Spanish. Even given Nation's contention (1990) that learners need know only half as many words to understand spoken text as they need to understand written text —because of the usually greater lexical density of written text —listening, in my view, involves the additional problem of real time constraints on comprehension, which more than compensates for the discrepancy noted by Nation.

Evidence from the field of corpus linguistics shows clearly that it is lexical competence, not the learning of grammatical structures, that must be the priority for language learners because lexical competence is at the heart of

communicative competence. Richards (2000, xi) states:

> Vocabulary and lexical units are at the heart of learning and communication. No amount of grammatical or other type of linguistic knowledge can be employed in communication or discourse without the mediation of vocabulary. Indeed, vocabulary and lexical expressions can sustain a great deal of rudimentary communication without much support from other aspects of the language system. Understanding of the nature and significance of vocabulary knowledge in a second language therefore needs to play a much more central role in the knowledge base of language teachers.

Implications for English language teachers and learners

Richard's insistence on the importance of vocabulary and lexical units has profound implications for English language teachers and learners. Six of them are discussed below:

1. What it means to know a word

Evidence suggests that language learners need to learn as many words as possible as soon as possible (initial 2000 word target, with 10,000 words as an ideal longer-term target). Several definitions have been proposed concerning what it means to know a word. I have adapted Ellis and Sinclair's (1989) list of criteria for knowing a word:

- To understand the word when it is written or spoken
- To recall it when you need it
- To use it with the correct meaning
- To use it in a grammatically correct way
- To pronounce it correctly
- To know which other words you can (and can not) use with it
- To spell it correctly
- To use it in the right situation
- To know if it has positive or negative connotations
- To know when (and when not) to use it

Of course, we know that all these cannot occur simultaneously. We know that learning and knowing words is an incremental process; it may take years of learning to fully know a word. A learner may learn the word *red* in terms of its spelling and pronunciation, and the learner may be able to apply it correctly when describing color. However, all the idiomatic expressions associated with red, including *in the red, to see red,* and *a red letter day,* may never be learned.

2. Recycling and revising words

We know that words should be recycled and revised as soon as possible after they are introduced; otherwise there is a tendency for them to be forgotten. There are several options for presenting and revising, or recycling, vocabulary. Nation (1990, 3–4) lists four typical ways, from "most indirect to most direct," which teachers may follow. Here, I have chosen McCarthy's (1990) categories as being particularly useful. According to this viewpoint, there are three main options:

1. By topic or theme, e.g., colors, rooms in a house, in the supermarket, on vacation, crime
2. By focusing on meaning, e.g., collocation, semantic sets, register, discourse analysis
3. By focusing on form, e.g., word formation, such as roots, suffixes, and prefixes; compounds; phrasal verbs

I shall introduce each of these, and give examples, in the *Three Options* section to follow. The examples are meant to be representative of each category only; there are many excellent books available that include a wide range of similar activities.

3. Teaching vocabulary systematically

Another implication from the findings of corpus linguistics is that vocabulary development will have to be given much more prominence in language teaching than it now gets. I believe that vocabulary development in the language classroom should be systematic. There is, however, disagreement on the extent to which vocabulary can or should be taught. Nation (1990, 1) opens his book with the question, "Should vocabulary be taught?" Until recently, the unstated assumption has been that learners must somehow learn vocabulary but that teachers should not really try to teach it, at least not systematically. This assumption was clearly revealed by Coe in his 1997 article, "Vocabulary must be learnt, not taught." Now, it seems, the introduction and development of lexis, defined here as the input, storage, and retrieval strategies for the development of an

appropriate mental lexicon, should probably receive a much higher and more explicit profile in the ELT classroom.

4. Learning the principles and techniques of vocabulary development

Teachers will have to gain expertise in vocabulary development principles and techniques so that they can provide appropriate introduction, storage, and retrieval activities for their students. There are many techniques, ranging from vocabulary notebooks and traditional word lists to mnemonic devices and word association methods. One technique, the keeping of well-organized vocabulary notebooks, should no longer be left to the discretion of the students, and perhaps instead should be mandatory. However, the system for organizing the notebooks should probably be the one that the individual student finds most effective and useful. Exactly what words should be introduced is not clear, and questions remain about whether the decision should be based on a list of the most frequent words in the English language (such as West's 1953—and now rather outdated—*General Service List of English Words),* the learners' immediate needs, or whatever the course book writer deems appropriate. My view is that emphasis should be placed on learners' needs.

English language teachers need to familiarize themselves with the many excellent vocabulary development textbooks available today, most of which are filled with activities they can use to augment the activities they themselves devise. Teachers also need to encourage their students to take more responsibility for developing their own mental lexicon. As mentioned earlier, vocabulary notebooks, organized in whatever way works best for each student, should probably be considered essential items. At the moment, they tend to be used by some motivated learners but do not feature significantly in most ELT classrooms.

5. Learning the metalanguage of vocabulary

Teachers need to be aware of what is happening in the field of research and development in lexis (corpus linguistics), and they will need to grasp and use the necessary terminology to talk about vocabulary with their colleagues and, as needed, with their students. Most English language teachers know and use the metalanguage of grammar, such as *past perfect, first* and *second conditional, past participle,* and *irregular verb.* But few English language teachers are familiar with and understand the essential terms and concepts associated with corpus linguistics and lexis, such as *collocation, chunks of language, fixed expressions, sentence heads.* These concepts need to be as familiar to English language teachers as grammar is now. Teachers should familiarize themselves, for example, with these four major categories of lexical items outlined by Lewis (1993, 1996):

1. words, e.g., *push, exit, fruit*
   polywords, e.g., *by the way, on the other hand*
2. collocations or word partnerships, e.g., *an initial reaction, to assess the situation*
3. institutionalized utterances or fixed expressions, e.g., *I'll see what I can do, It's not the sort of thing you think will ever happen to you.*
4. sentence frames or heads, e.g., *Considerable research has been done in recent years on the question of…; At present, however, expert opinion remains divided; Some experts believe….* (from Lewis 1996, 10)

Lewis also suggests that there are two distinct modes of English, spoken and written, and that in many ways they represent two virtually different languages. Native speakers have a vast store of words in both spoken and written modes. As Lewis puts it, "It is now clear that students can usefully employ a repertoire of at least several hundred, if not many thousand, institutionalized expressions. Such expressions are central to effective spoken communication, both receptive and productive" (1996, 15).

Lewis argues strongly and convincingly that teachers should not be teaching traditional grammar but instead should be focusing on these chunks of language, that is the fixed expressions, of which there are thousands. One immediate consequence of corpus linguistics research is that the expression *chunks of language* and the need to recognize and teach language chunks has become part of ELT terminology, and up-to-date books on ELT methodology and vocabulary contain this term (e.g., Hedge 2000; Cameron 2001). Teachers may need to make a mental shift from thinking that language is *lexicalised grammar* to thinking of it as *grammaticalised lexis.*

Exercise 1
Word groups

Topic • *Activity 1*

Put these animals into groups:

sheep	dog	goat
lion	pig	zebra
horse	wolf	turtle
cow	camel	cat
crocodile	hamster	giraffe
tiger	elephant	guinea pig
bear	antelope	

6. Updating mental lexicons

Perhaps most daunting of all the implications is that English language teachers will need to update their own mental lexicons. Biber et al. (1999) in the *Longman Grammar of Written and Spoken English* (the title clearly reflects the current idea gained from corpus research that written and spoken English are different systems), examine language corpus in four areas, or *registers:* academic texts, newspapers, spoken texts, and fiction. Not surprisingly, nouns represent by far the most frequent lexical word class; every fourth word is a noun. Verbs are less frequent, occurring every tenth word, followed by adjectives and adverbs. Surprisingly, newspapers have by far the greatest range of lexis in terms of nouns, yet newspaper language is the least known and least understood among non-native English language teachers. Headlines such as POLICE CHIEF QUITS OVER PRESS SCAM and YARD IN GEMS SWOOP DRAMA leave most non-natives (and many natives) completely baffled. It is the almost exclusive use of the simple present form of verbs in headlines, associated with the often unfamiliar lexical items used by newspapers that creates confusion. Course books rarely, if ever, discuss this type of language, so EFL teachers and learners are rarely exposed to it. But they should be because newspapers keep pace with contemporary language usage, including newly coined words.

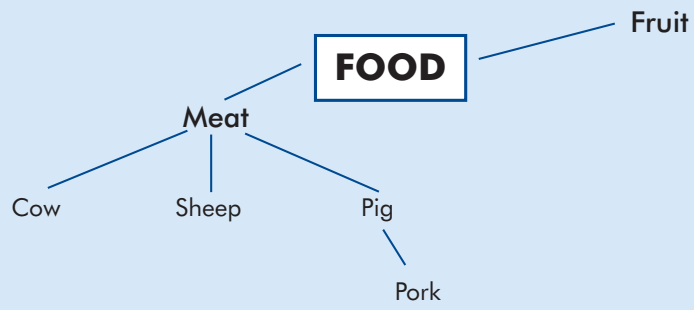
Helping our students learn, store, and retrieve the words they need: Three options

As indicated earlier, there are several options for presenting and revising or recycling vocabulary. Unfortunately, some tradi-

Exercise 2
Word webs

Topic • *Activity 2*

Word Webs



tional course books do not go much beyond topic or theme when dealing with vocabulary development. McCarthy's (1990) three main options are described in this section.

Option 1: By topic or theme

This is the way most writers introduce vocabulary in course books, and it is a logical way to introduce the many important lexical sets that make up the bulk of learners' early lexicons.

Topic activity 1: Word groups

There is no right or wrong answer for this task; the animals may be grouped in a variety of ways, for example, herbivores /carnivores/ omnivores or pets/domestic/wild. See Exercise 1.

Topic activity 2: Word webs

The teacher begins by writing the topic FOOD in large letters in the middle of the board. S/he then adds two categories of food: *Meat* and *Fruit.* S/he then extends the Meat category by adding *Pig, Cow,* and *Sheep*. Finally, the subcategory *Pork* is added. The board now looks like Exercise 2.

The teacher then explains that there are more categories that can be added (for example, Vegetables) and many more words for each category. Students, working in groups of two or three copy the web onto a large piece of paper and are given a time limit to extend the web as far as they can. The webs are then displayed and compared. Similar webs can be done using topics such as Clothes, Rooms in a House, Animals, or Transport.

Option 2: Focus on meaning

With the renewed interest in lexical development, the importance of focusing student attention on meaning has increased significantly. Better, more up-to-date course books now mention the word *collocation* and include activities focused on meaning, which is the second option to be considered here. In this option, we would look at collocations within a specific context. We could examine items such as *final whistle, defending champions, off-side rule, penalty kick,* which would typically all come from a newspaper article about a soccer game. Or we could examine lexical sets that reflect semantic fields (for example, words that describe size, such as: *large, enormous, big, gigantic, vast,* and *huge)* and discuss the way they are used and how they do or do not collocate with other words. So, with the students we might discuss why it is acceptable to say *a large dog,* but not *a vast dog,* and we might work together to place the words on a continuum based on size.

Focus on meaning activity 1

Which adjectives can be combined with which nouns in Exercise 3? This activity nicely points to the different meanings words may take when collocating with other

Focus on Meaning • *Activity 1*

Which adjectives can be combined with which nouns?

Adjectives	Nouns
hot or mild	sea
hot or cold	cheese
sweet or dry	curry
sweet or sour	bed
strong or weak	water
strong or mild	wine
rough or calm	cigarettes
rough or smooth	tea
hard or soft	exam
hard or easy	grapes
	skin
	surface

Adapted from Redman, Ellis, and Viney (1996)

Exercise 3
Adjective and noun combinations

Exercise 4
"Normal" and "extreme"

Focus on Meaning • *Activity 2*

Complete the table with the appropriate "normal" adjective. The first one has been done as an example.

"Normal" word	"Extreme" word
hot	boiling
_____	enormous
_____	delicious
_____	tiny
_____	exhausted
_____	freezing
_____	awful
_____	filthy
_____	ancient
_____	wonderful

Adapted from Redman, Ellis, and Viney (1996)

words and the inherent dangers of teaching synonyms and antonyms too freely, for example *rough sea* and *calm sea; rough surface* and *smooth surface).*

In the *Dictionary of Selected Collocations,* a book made possible because of evidence obtained from corpus linguistics research, Hill and Lewis (1997, 6) identify the five most important kinds of collocations as:

1. adjective + noun, e.g., *fatal accident, golden opportunity*
2. verb + noun, e.g., *accept responsibility, undermine (my) self confidence*
3. noun + verb, e.g., *the gap widened, a fight broke out*
4. adverb + adjective, e.g., *highly desirable, potentially embarrassing*
5. verb + adverb, e.g., *discuss calmly, lead eventually to*

Focus on meaning activity 2

Complete Exercise 4 with the appropriate "normal" adjective. The first one has been done as an example.

Option 3: Focus on form

The final option is to focus on form. Students who have some knowledge of suffixes and prefixes can often work out for themselves the meanings of words. Prefixes are particularly important because, generally speaking, they change the meaning of a word. Thus, students can learn that adding the prefixes *im-, un-* and *in-* produces the opposite meaning of the word to which they are attached *(im*possible, *un*satisfactory, *in*expensive), which can provide them with a useful strategy.

Focus on form activity 1

Which of the words in Exercise 5 can be combined with *-less* and *-ful?*

Focus on form activity 2

Most suffixes change the category of a word, for example, from a verb to a noun. This activity is based on a humorous song, *When You're Old and Gray,* by Tom Lehrer. Lehrer is known for the black humor in his songs, and this one makes repeated use of the rhyme of the *-ility* suffix (in many cases, transforming an adjective into a noun).

Since I still appreciate you
Let's find love while we may,
Because I know I'll hate you
When you're old and gray.
So say you love me here and now
I'll make the most of that,
Say you love me and trust me
For I know you'll disgust me
When you're old and getting fat.
An awful debility
A lessened utility
A loss of mobility
Is a strong possibility.
In all probability
I'll lose my virility

Focus on Form • *Activity 1*

Which of these words can be combined with *-less* and *-ful*?

use	_____
home	_____
end	_____
harm	_____
tact	_____
care	_____
thought	_____
taste	_____
pain	_____
hope	_____

Adapted from Redman, Ellis, and Viney (1996)

And you your fertility
And desirability.
And this liability
Of total sterility
Will lead to hostility
And a sense of futility.
So let's act with agility
While we still have facility
For we'll soon reach senility
And lose the ability.
Your teeth will start to go, dear
Your waist will start to spread
In twenty years or so, dear
I'll wish that you were dead.
I'll never love you then at all
The way I do today,
So please remember
When I leave in December
I told you so in May.

From Tom Lehrer Revisited, recorded in 1959 for Lehrer Records, re-released on CD in 1990 by Reprise Records, 9-26203-2

Conclusion

In that 1755 dictionary, Samuel Johnson also admitted that neither he, nor anyone, "shall imagine that his dictionary can embalm his language, and secure it from corruption and decay." While we shouldn't equate normal language change with corruption and decay, as language teachers, we must accept the fact of change in the subject we teach. We must also accept the challenge of staying abreast of the kinds of changes occurring in English, which are most apparent in its lexicon. Corpus linguistics, the study and analysis of large collections of written and spoken text, has contributed immeasurably to our understanding of how English is actually used. A renewed focus on teaching vocabulary and lexical items in English can help our students be more successful in learning, storing, and retrieving the words they need.

Exercise 5
-less and -ful word combinations

References

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman grammar of spoken and written English.* Harlow, UK: Longman Pearson.

Brown, H.D. 2000. *Principles of language learning and teaching.* 4th ed. Longman: Pearson Education NY.

Cameron, L. 2001. *Teaching languages to young learners.* Cambridge: Cambridge University Press.

Coe, N. 1997. Vocabulary must be learnt, not taught. *Modern English Teacher* 6 (3): 47–48.

Ellis, G. and B. Sinclair. 1989. *Learning to learn English.* Cambridge: Cambridge University Press.

Hedge, T. 2000. *Teaching and learning in the language classroom.* Oxford: Oxford University Press.

Hill, J. and M. Lewis. 1997. *Dictionary of selected collocations.* Hove, UK: Language Teaching Publications.

Lewis, M. 1993. *The lexical approach.,* Hove, UK: Language Teaching Publications.

———. 1996. *Implications of a lexical view of language* in J. and D. Willis *Challenge and Change in Language Teaching.* Oxford: Heinemann.

Meara, P. 1995. The importance of an early emphasis on L2 vocabulary. *The Language Teacher,* 19, 2: 8–10.

McCarthy, M. 1990. *Vocabulary.* Oxford: Oxford University Press.

McCarthy, M. and F. O'Dell. 1999. *English vocabulary in use elementary.* Cambridge: Cambridge University Press.

Nation, P. 1990. *Teaching and learning vocabulary.* New York: Newbury House.

Redman, S., R. Ellis, and B. Viney. 1996. *A way with words resource pack 1.* Cambridge: Cambridge University Press.

Richards, J. 2000. Series Editor's Preface. In N. Schmitt *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

West, M. 1953. *A general service list of English words.* London: Longman, Green and Co.

ANDREW SHEEHAN is a member of the Chilean Ministry of Education's new English Project team; one of his main responsibilities is the coordination of the professional development of English language teachers.

2000 Words: The magic number

For many years, the question "How many words do we need and, therefore, how many words do language learners need?" has been central to the debate about the role of vocabulary in language teaching and learning. The figure 2000 as the basic number of words needed has been around for a while. West (1953) had 2000 "head" words in his list. In their *Introduction to the student,* McCarthy and O'Dell (1999, 4) suggest that "…to speak and write English in normal situations you need at least 1-2000 words." The authors said this based on evidence provided by corpus linguistics research.

Analysis of corpora tells us about word *frequency* and *text-coverage.* Not surprisingly, we could predict that the most frequently occurring words in English are words such as *the, of, and, to, a, in,* and *that.* These are *functional* words, which in themselves carry no meaning. By analysis, we know that the three most frequent words in English *(the, I, you* in spoken English, and *the, to, and* in written English) represent 11.5% of all word *tokens,* or occurrences, in texts. We know that the top 100 words represent 44% of texts.

Perhaps the most significant figure, however, is the one corresponding to the most frequent 2000 words. The top 2000 words account for about 80% of texts. In other words, a learner who knows the most frequent 2000 words will be able to understand about 80% of a text (or, to put it another way, one in five words, or 20%, will be unknown). From this evidence, we can surmise that 2000 words is the absolute minimum a language learner needs—the survival level—in order to be able to process a text. Any fewer, and the unknown gaps in the text will be too many to en-able the learner to deduce meaning from context. This minimal 2000 figure is critical; there is only a 5% increase for the next 2000 words (up to 4000), and even less for each subsequent 2000 words (the figures are, approximately: 4000 = 88%, 6000 = 91%, 8000 = 93%). Analysis also shows that "knowing" 10,000 words means that 93% of a text will be understood, and this could be recommended as the next ideal target for a proficient language learner.

Important Terminology

Corpus (pl corpora): a large collection of spoken or written text, nowadays stored on a computer. Examples of corpora include:

The *British National Corpus* with over 100 million words (90 million written and 10 million spoken)

The *COBUILD Bank of English Corpus* with over 300 million words (spoken and written)

The *Cambridge International Corpus* with over 100 million words (spoken and written)

The *Longman Written American Corpus* with over 200 million words of written American English

Corpus linguistics: the study and analysis of these corpora of written and spoken text. Corpora-based research began in the first part of the 20th century. During the 1960s computers began to play a vital role. Today, with the use of powerful computers containing vast memories, high speed data-processing capability, and employing powerful new programs (such as *concordancers),* we know a great deal about the frequency of words used in English; how many words are needed to understand a particular text; which words tend to co-occur; and how words are used and which meanings are associated with them. All major dictionary publishers now rely on corpora stored electronically via computers.

Words/lexical items: Linguists prefer to use the term lexical items for what we commonly think of as words for reasons that become clear when examining the following expressions:

Saw is one word, but at least three lexical items.

Take off is two words, but several lexical items.

To put up with and *to get along with* are multi-word verbs, but each represents one lexical item.

To be taken for a ride can be understood literally, or it can be interpreted idiomatically to mean *to be cheated or tricked.*

A *Black Hole* is a lexical item in which two words, which mean entirely different things when isolated, have a special meaning when combined.