Lexical density in English

Leo Selivan looks at the influence of French and what it means for teachers.

Even though English belongs to the Germanic family of Indo-European languages, it can be argued that, if evaluated on the basis of its vocabulary, English is essentially a Romance language. Indeed the overwhelming number of borrowings from French and Latin by far outweighs the words of Germanic origin. This is due to the significant influence French and Latin (mostly through French) exerted on English after the Norman invasion in 1066, which shaped the language we know today.

This huge influx of French words in the thirteenth and the fourteenth century (Barber, 1993: p.145) greatly enriched English vocabulary with a number of Latin-derived words. As a result, what we have today is an extremely diverse, "fusion" language, the number of words in which, according to some claims, exceeds one million. How do we go about teaching such an enormous number of words?

Latinate and Germanic words

As a rule, English speakers have a lexical choice as regards *freedom* or *liberty*, *sight* or *vision*, *leave* or *desert*. We can *begin* or *commence* work and then *end* or *finish* the day. We can have *feelings* or *sensations*. We can be *flooded* with emails or *inundated* with them. In all of the above examples, the latter, Latinate synonym is often more elegant or sophisticated.

A trained eye can easily spot Latin-derived words in English as they tend to be longer and are often used in more formal or specific contexts. For example, Germanic *kill* and Latinate *assassinate*. Although in the source language (French) *assassiner* does not have any special connotation and merely means 'to kill', in English it has become

imbued with different undertones and is consequently used to refer to killings of presidents and other high-powered individuals.

Another characteristic of Latinate words is their more frequent use in writing, particularly in academic discourse, than in speech. For example, according to the British National Corpus (BNC), *rapid* is over 10 times more frequent in academic writing than in the spoken language, as opposed to its Germanic counterpart *quick* which is five times more common in speaking.

These features can be accounted for by the fact that French was the language of the aristocracy, whereas the Old English was mainly used by ordinary Anglo-Saxons. It may also be the case that the French words entered English later when all the essential meanings for basic communication had already been assigned to the Anglo-Saxon lexicon. When the vast number of French words penetrated the English language, a number of Old English (Anglo-Saxon) equivalents were displaced (Barber 1993: p.146). *Mountain* is an example of a French word whose Anglo-Saxon equivalent was lost. However most French loan words went on to co-exist with their Anglo-Saxon counterparts, substantially augmenting the English lexicon.

This resulted in a "hybrid" language with two distinct major sets of vocabulary. Together with contributions from the Scandinavian languages, a constant addition of coinages mostly related to technology and media (such as *dot-com* and *edutainment*) and new words continually borrowed from other languages, the English vocabulary is undoubtedly immense. By some estimates, the number of words in English exceeds one million and in June 2009 the Global Language Monitor controversially announced that English had passed the one million word mark. Although the announcement was met with criticism from some linguists (Shuester 2009), more modest estimates suggest that English is probably the richest language in the world with a vocabulary of more than 250,000 words. (Oxford Dictionaries, n.d.)

Learning vocabulary can then be a daunting task for speakers of other languages whose vocabulary is not as rich. Ironically, French itself, being the most significant source of the English vocabulary, has, according to some claims, fewer than 100,000 words (McCrum *et al*. 2002). Even German and Russian do not even come close to the stupendous size of English having fewer than 200,000 words, while Spanish and Japanese have around 230,000.

> ❝Learning vocabulary can then be a daunting task for speakers of other languages whose vocabulary is not as rich.❞

Consequently, learners are often faced with the dilemma of which word to choose. Not knowing the differences in usage and often very subtle nuances in connotation may often lead to errors in word choice. For example, Hebrew ("lehazmin") can be rendered into English in five different ways:

book *a room in a hotel*

invite *friends round for dinner*

reserve *a table (in a restaurant)*

order *food*

It's on me/**It's my treat**.

It is hardly surprising that Hebrew speakers often make mistakes when they make a lexical choice. This same is true of the speakers of other "economical" languages.

Lexical voids like this one exist, of course, in English too and are well-known to translators. The most popular example is probably English *know*, for which most European languages have two equivalents (*cf.* saber/conocer– *Sp.,* **savoir**/**connaître** *– Fr., kennen/wissen – Ger.).* Still, it would be reasonable to assume that such lexical voids would more often be found when translating from English into other languages.

Words and concepts

Since English is the richest language, English speakers can often express finer, subtler nuances of meaning by opting for one word or another. But does it mean that they have a richer conceptual system? Apart from being politically *in*correct in this post-modern era, even the post-Whorfian[1] linguists and psychologists assert that language does not influence thought. They do agree, however, that speakers of different languages use different conceptual processing when speaking (Slobin 2003: p.157). Even Pinker, normally an outspoken critic of the theory of linguistic relativity, concedes that "one's language does determine how one must conceptualise reality when one has to talk about it" (as cited in Gentner & Goldin Meadow 2002: p.8).

If we adopt this view, we can make a case that English speakers do not have a richer conceptual system but rather English has too many words related to the same concept. In other words, English is a "densely populated" language with many words per concept, which results in many words having a more restricted or "narrow" use. This "narrowness"of its lexical semantics can explain why speakers of more economical languages often make mistakes when they have not fully mastered all the levels of word knowledge.

Synonymy and near– synonymy

The examples given earlier (*liberty* vs. *freedom*, *vision* vs. *sight*) are habitually treated as synonyms and often presented as such in English lessons. However one must remember that in most languages very few synonyms are actually pure, absolute synonyms. Most of the so-called synonyms are in fact near-synonyms. Since English is so lexically rich, the speaker can often choose between a few different words, for example *begin* and *start*. However, we *start a car* not **begin a car* and when we choose *desert* over *leave* we imply that leaving was done in violation of a duty or promise. Hence, word knowledge would be incomplete without knowing these nuances.

While there are unarguably synonyms within the Latinate set (e.g. *mount*/ *ascend*, *mansion*/*villa*) and within the Germanic (e.g. *go*/*walk*), the majority of synonyms in English are based on the Latinate/Germanic contrast. There are a number of factors which affect word choice, which we will look at below by examining one area where French exerted a considerable influence: verbs.

Phrasal verbs, which are generally of Germanic origin, are often taught alongside their so-called 'more formal' equivalents, which tend to be of Latin/ French origin.

set up	establish, found
find out	discover
put up (with)	tolerate; endure
take in	deceive
put out	extinguish
bring about	cause
look into	investigate

On the face of it, they seem synonymous; however there are a number of aspects of word meaning which need to be considered, such as collocation, register, semantic prosody and colligation.

Collocation

While English speakers may seem to exercise a lot of choice when selecting words, they are, in fact, restricted collocationally. According to the BNC, we tend to *establish relationships/ relations/links* while we *set up groups/ committees/businesses/companies.*

Likewise, the near-synonyms *look into* and *investigate* have different collocational fields. The former tends to be followed by: *matter/possibility/ problem/way (of doing something)* while the latter collocates with *allegation/ complaint/possibility/murder/effects.* These collocational limitations are probably the most significant constraints on word use, but there are other factors at play too.

Register, style and domain

Register refers here to the level of formality. The same word can be used differently or with a varying degree of frequency in different contexts. A quick search in the BNC shows that, for instance, *investigate* is more common in the academic domain and in the news. In contrast, *look into* is more frequent in fiction and in the spoken language, whereas the number of its occurrences in the academic sub-corpus is negligible.

---

1. Whorf is famous for his controversial and often misinterpreted theory of linguistic relativity which claims that language influences thought.

	Spoken	Fiction	News	Academic	Misc
investigate	15	21	85	96.6	58
look into	17.6	38	12	4	8.7

This observation can be extrapolated to all multi-part verbs in general: they tend to find preference in informal spoken discourse, while their one-word Latinate counterparts appear in writing, for example *put out* and *extinguish* (Celce-Murcia & Larsen-Freeman 1999).

Semantic prosody

This somewhat elusive term refers to the association a word has with particular meanings. If collocations are the lexical environment of a word, semantic prosody is its semantic environment. Needless to say, semantic environment will depend on the collocations as well as other surrounding words not readily distinguishable as collocations. For example, *put up with* tends to be found "in company of people" (*men*, *women*, *wife*) whereas we can *endure* not only people but various situations in life such as *pain*, *hardship* and *suffering*, but also time (days, months).

At first sight, semantic prosody may seem indistinguishable from collocation; one can only start to see it after observing a large number of examples. Learning collocations would often be based on pure memorization. By studying a number of individual instances (collocations), learners can be guided towards making a more abstract generalization about a word and that is semantic prosody, i.e. a sum of all the collocations of a given word.

Consider, for example, *cause* which tends to collocate with "negative" words such as *controversy*, *damage*, *problems*, *suffering*, *trouble* to name but a few. By pointing these out, learners gradually become aware that generally *cause* has a negative connotation. In contrast, *bring about* often does not have a negative semantic prosody.

Colligation

"Even semantic prosody […] is insufficient to account fully for how words are used" claims Hoey (2000: p.233). The last but not least restriction on word use is colligation, which refers to the grammatical context of a word or the grammatical function it prefers. For example, according to the Macmillan Dictionary, *taken in* is more common in the passive (e.g. *Don't be taken in by their promises.*), whereas its near-synonym *deceive* does not display such a tendency. The verb *to found* shows a similar preference for the passive voice (e.g *was founded in* (year)/*by* (person) while *set up* does not. It has to be noted that the use of *take in* or *found* in the active would not result in a grammatical error; it is merely a case of grammatical preference.

Teaching implications

As we have seen above, the enormous size and richness of the English lexicon, which stems largely from the huge influx of Norman French words, has resulted in the "narrowness" of the meaning and use of certain lexical items. Thus, full mastery of these items is contingent upon the knowledge of not only denotational meaning, but also collocations, colligation, register and often semantic prosody – important aspects of the word knowledge which are often overlooked. While not all the above mentioned aspects may always be equally important depending on the item, it is useful to point them out when dealing with near-synonyms.

We have looked at phrasal verbs versus their Latinate one-word counterparts, often presented as synonyms. However, this phenomenon is pervasive in English and concerns other parts of speech too. What activities can be used to highlight these features in class?

Collocation forks

Present new items or elaborate on partially learnt items by using collocation forks providing the most common collocates. If necessary, both confusing words can be presented alongside each other.

	matter
look into (the)	possibility
	problem

	possibility
investigate (the)	murder
	effects

Look up twice

This activity was first proposed by Lewis. After finding the meaning of a word in a bilingual dictionary, students look it up in a monolingual dictionary to increase their depth of meaning. Learner dictionaries, such as Cambridge, Longman and Macmillan (all three are now available online) are ideal for this as they provide a lot of good natural examples and often highlight collocations.

« By studying a number of individual instances (collocations), learners can be guided towards making a more abstract generalization about a word and that is semantic prosody … »

Present a word

In her forthcoming book Penny Ur recommends getting students to make mini-presentations about a word including its origin and derivatives as well as frequency collocations and other important features we talked about (*All you know about a word*).

Another suggestion she makes is explicitly teaching students to distinguish between Germanic and Latinate words by using an etymological dictionary or comparing different texts, e.g. spoken and academic, where the proportion of Latinate words would be higher.

If learners can be trained to spot Latinate words in English, it will immediately unlock at least one of the aspects of the word knowledge they need to master – register. As we have seen above, Latinate words tend to occur more in formal contexts and written discourse.

Corpus and concordances

Get your students to look up near-synonyms in the corpus (for example, BNC or Corpus of American Contemporary English – COCA) or study concordance lines. just-the-word.com is a user-friendly corpus-based website which allows you to easily key in a word and see the list of its collocations.

Teaching affixes

This follows on the previous suggestion of explicit teaching of Latinate vs German words. Students will be able to spot the difference between Latin and German words if they are taught basic Latin-derived prefixes such as *ex-*, *inter-*, suffixes *-tion*, *-ment* or basic Latin roots like *dic/dict, scrib/script,* etc.

Conclusion

No definition or translation can fully account for how words are used. Caution must be taken when treating multi-part verbs as synonymous with their one-word counterparts, because such practice ignores various aspects of meaning and usage that are essential for the learner to know. This is also true when dealing with other confusing word pairs or near-synonyms.

Teachers would help their students tremendously by exposing them to a lot of examples, patterns and concordance lines of any lexical item taught, and by inviting them to explore by themselves different nuances associated with the word use. Vocabulary should be taught in collocational context and learners should be made aware of connotation and colligation, a practice which should also be judiciously supplemented by more explicit vocabulary teaching techniques. This can help learners not only master the difference between confusing and seemingly synonymous words but also bridge the all-too-wide gap between receptive and productive knowledge.

I would like to thank Penny Ur for critically reviewing the first draft of this article.

REFERENCES

Barber, C. (1993) *The English Language*. Cambridge: Cambridge University Press

Celce-Murcia, M. & Larsen-Freeman, D. (1999) *The Grammar Book: An ESL/EFL Teacher's Course*, 2nd Edition. Boston: Thomson Heinle

Davies, M. (2004-). BYU-BNC: The British National Corpus. Available at http://corpus.byu.edu/bnc

Gentner, D. & Goldin-Meadow, S. (2002) Whither Whorf. In Gentner, D. & Goldin-Meadow, S. (eds.), *Language in Mind: Advances in the Study of Language and Thought*. (pp. 3-14). Cambridge: MIT Press

How many words are there in the English language? (n.d.). *Oxford Dictionaries*. Retrieved from http://oxforddictionaries.com/page/howmanywords

Hoey, M. (2000) A World Beyond Collocation: New Perspectives on Vocabulary Teaching. In Lewis, M. (ed.), *Teaching Collocation: Further Developments in the Lexical Approach.* Hove: Thomson Heinle

McCrum, R., Cran, W. & MacNeil, R. (2002) *The Story of English*. London: Penguin

Slobin, D. I. (2003) Language and thought online: Cognitive consequences of linguistic relativity. In Gentner, D. & Goldin-Meadow, S. (eds.), *Language in Mind: Advances in the Study of Language and Thought.* (pp.157-191). Cambridge: MIT Press

Schuessler, J. (2009, June 13) Keeping It Real on Dictionary Row. *The New York Times*. Retrieved from http://www.nytimes.com/2009/06/14/weekinreview/14shuessler.html?_r=2

Ur, P. (2012) *Vocabulary activities.* Cambridge: Cambridge University Press

Welcome to the World of Global English. (2010, February 24) *The Global Language Monitor*. Retrieved from http://www.languagemonitor.com/

Leo Selivan

Leo Selivan is a teacher/teacher trainer and materials developer with the British Council in Tel Aviv. He has taught in Cyprus, Turkey and Russia. He has a BA in PR and Languages, CELTA and an M.Ed in TESOL. His key interests are vocabulary development and using films in the classroom. In addition to writing materials for the TeachingEnglish website, he maintains his own blog Leoxicon.