

Matthew Walsh

September 1, 2005

Collocation and the learner of English

1. Introduction

Observations in the areas of theoretical and applied linguistics have identified lexis, the vessels that carry meaning, as having a major role in determining language patterns, a position more traditionally reserved for grammar. Conversely, these patterns or collocation often determine the meaning of the lexical items themselves, and these pre-assembled word combinations are employed throughout language to make communication more efficient. These insights present a challenge to a language pedagogy that takes a structural approach. Overlapping characteristics make classifying these predictable patterns difficult for linguistics and lexicography alike, and dealing with these patterns is a daunting task for the learner. Choices about how collocation should be treated in the classroom need to be made, but often relying on intuition fails to create an accurate picture of the extent that collocation exists in the real world. In this paper I will explain what is meant by collocation, examine collocations as they actually exist within an authentic text, discuss some methods used to focus on collocation, and offer some suggestions on how to use the collocations found within the text to help learners gain an awareness of the constraints on word-combinations in English.

2. Review of the literature: What is meant by the term ‘collocation’.

Any discussion of how to demonstrate the constraints on word-combinations in English, or ‘collocation’, assumes that we know what the word means. However, it is a slippery concept and is used to describe different things by different authors. A review of the history and usage of the term ‘collocation’ is necessary. There are two major senses in which it is used, - [a] that of collocation in a very broad sense -i.e. all patterns permeating language distinct from those determined by sentence or clause grammatical structure, and [b] that of specific predictable word combinations which do not appear to be determined solely by semantic restrictions but rather possess some degree of being: fixed, idiomatic, or ‘natural’ via linguistic convention. I will begin by reviewing the history of the attention given to collocation, highlight what this has meant for ELT and the learner, and proceed to explain some classifications of them I will later use to describe the collocation found in the text (see section 3).

2.1 Research on ‘collocation’: a view of language with implications for ELT and the learner

The term ‘collocation’ was first introduced by Firth (1951) as the level of meaning created in language independent of the ideational concepts of the individual words or the contextual level of meaning to be derived (pragmatics). Defining this level of meaning created by word combinations or patterns, where the pattern or ‘co-location’ with other words can actually create the meaning of a given word, he states “You shall know a word by the company it keeps” (Firth 1957). Another definition of the concept he offers is “collocations of a given word are statements of the habitual or customary places of that word” (Firth, 1968). Firth’s work triggered a heightened awareness of the role of lexis, or these ‘habits’ as meaning-creating over an ‘atomic’ view that tended to see language as individual words containing inherent meanings which simply fit into a structural or grammatical frame.

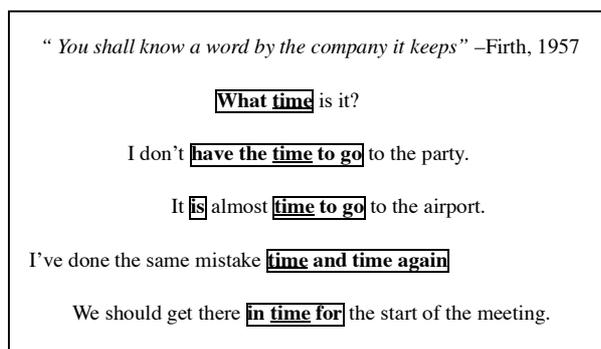


Fig 1

The word ‘*time*’ has many senses, but the surrounding words, or ‘collocation’ signals the sense being used. It could also be said that the sense in which ‘*time*’ is used determines the surrounding words.

Seminal work in what became known as the ‘Firthian’ tradition was continued by authors such as Halliday (1966), Sinclair (1966), coincided with pioneering work by linguists and lexicographers in East Europe such as Apresyan (1969), and forms the basis for recent views of language becoming more prevalent in ELT such as ‘The Lexical Approach’, of which collocation is of central importance (Lewis, 1993).

In line with the Firthian tradition, further awareness of collocation at an empirical level was brought to the forefront by analyzing language statistically via large computer-based corpora. Sinclair, who defines collocation as “the occurrence of two or more words within a short space of each other in a text” (1991) led the COBUILD corpus project which led to, among other

things, innovations in lexicography and dictionary design informed by corpus evidence in order to take more into account commonly found collocation and lexis behavior.

One of the lasting concepts to have grown out of this research is Sinclair's distinction between what he terms 'the idiom principle' and 'the open-choice principle'. 'The idiom principle' observes that language is largely systematic in that "... a language user has available to him or her a large number of semi-preconstructed phrases that constitute a single choice, even though they might appear to be analyzable into segments" (ibid). This is what is meant by collocation in the broad sense. 'The open-choice principle', in contrast, sees slots in a grammatical framework being filled more arbitrarily, thus a 'choice' of words being made at each turn limited only by grammaticality. Sinclair posits that users of a language use both principles, but rely on the open choice principle only when the idiom principle fails. This insight has serious implications for a pedagogy such as a structural approach that presents only 'the open choice principle' and focuses on grammar determined by sentence or phrase structure ignoring lexically determined behavior or restrictions. Using the word 'grammar' in a lexically centered sense, Michael Lewis (2000) observes that "every word has its own grammar ... (and) ...'knowing a word' involves knowing its grammar -the patterns in which it is regularly used." In this way, the 'collocationist' view of language sees "the predictability of the co-occurrence of words in the slots that constitute the underlying structural frame" (Moon, 1997), and with obvious relevance to the language learner, recognizes these combinations and/or 'chunks' as an important part of the mental lexicon. Furthermore, collocational competence is said to provide an efficiency of effort, or fluency in terms of both decoding and encoding language in that it allows learners to process and produce these chunks at a faster rate as well as convey more easily complex ideas that are often expressed lexically (Hill, 2000).

The apparent ubiquity of this predictability in language is supported by the observation that as much as 70% of adult native language may be formulaic (Altenberg 1990). The issue of whether or not native-likeness is or should be the most important goal for the language learner goes beyond the scope of this paper, but pertinent to some learners no doubt is the observation that often the failure to use native-like formulaic sequences (see below, 2.2), ultimately marks out the advanced L2 learner as non-native (Pawley and Syder, 1983). Similarly, 'the intermediate plateau' (Lewis, Morgan, 2000) has been identified as a prolonged stage in interlanguage development where learners can deploy approximated 'open principle' language to communicative ends, but appear stuck in terms of improving fluency because of their not possessing an awareness of, or perhaps an appropriate level of acquisition of collocation as *The*

Oxford Collocations dictionary for students of English defines it, “the way words combine in a language to produce natural-sounding speech and writing”.

2.2 Typologies of collocation: The broad sense.

In order to demonstrate the constraints on word-combinations in English to our learners, we need to determine what kinds of these patterns are more frequent in language, and thus relevant. In order to do that, we need to adapt some sort of system to classify them. Sinclair’s ‘idiom’, what is meant by collocation in the broad sense, can be classified in various ways and terms are abundant. Wray and Perkins (2000) identify as many as 42 separate terms used in the literature, including ‘collocation’ to describe what they term ‘formulaic sequences’. A lack of consensus would be expected because, as Carter (1998) observes, there are in fact no set lines dividing the characteristics of these language items, rather, each of them lie somewhere on a dual cline of relative fixedness and idiomacity (in Carter’s terms, ‘opacity’). In order to label collocation in the text described in section 3, one classification used is ‘fixed expressions’ (appendix 1), a term Carter (ibid) uses as a subset of collocation that encompasses 13 overlapping categories of idioms, stock phrases, discursal expressions, etcetera, but omits less idiomatic but fixed patterns such as ‘prefabricated routines’ (Bolinger, 1976) and more specific combinations often found in the collocations dictionaries described below (see 2.3), such as the BBI whose typology I will also borrow.

2.3 The narrow sense: typologies

What is meant by collocation in the narrow sense is perhaps ‘lexical collocation’, -word combinations made up of content words-, such as those found in *the LTP dictionary of selected collocations* (Hill, Lewis, 1997) which centers on a mere 5 different combinations of adjectives, nouns, verbs, and adverbs (see fig. 2 below). However, *The BBI combinatory dictionary of English* (Benson, Benson, and Ilson, 1986), includes ‘grammatical collocation’ (what Firth referred to as ‘colligation’) and uses an intricate system identifying 26 patterns of grammatical collocation and 7 categories of lexical collocation (appendix 2) defining collocation as “fixed, identifiable, non-idiomatic phrases and constructions”(ibid) It is worthy of mention that many other dictionaries concerned with collocation (in the broad sense) focus on, and make a distinction between idioms and phrasal verbs both of which are contained within Carter’s ‘fixed expressions’ but the BBI differs in the respect that its primary focus is with the non-idiomatic. It is because of this and it’s inclusion of grammatical collocation that its system of classification

along with Carter's, as well as information from a corpus, was used to identify and answer the question of which types of collocation are more frequent in authentic language. Any attempt to raise the learner's awareness of this level of language needs to take this into account in order to determine which types deserve attention in class. Some types of collocations were found to be more significant than others in the text used to demonstrate collocation to learners of English.

Fig. 2: Examples taken from the LTP dictionary of 'lexical collocation', or collocation in the narrow sense. For categories and examples from the BBI and Carter's 'fixed expressions' see the appendices.

Adj+Noun	<i>golden opportunity</i>
Verb+Noun	<i>accept responsibility</i>
Noun+Verb	<i>The gap widened.</i>
Adverb+Adj	<i>highly desirable</i>
Verb+Adverb	<i>discuss calmly</i>

3. The text: its origin

The text (appendix 3) was found on a digest of question and response exchanges from an internet forum called 'ChomskyChat'. In it, Massachusetts Institute of Technology linguistics professor Noam Chomsky of generative grammar fame acts in his perhaps more popularly well-known role of social and political activist centering mainly on his criticism and documentation of unjust foreign policy measures by the United States and other world powers. In the text, Chomsky answers an enquiry as how to effect change in the social consciousness about the issues he is famous for writing about.

3.1 The text: collocation within and analysis used

The text (appendix 3) consists of 259 words. Underlined and numbered within the text are 33 word combinations that amount to a total of 106 words or 41 percent of the text. The word combinations have then been classified in the table found in appendix 4 according to the aforementioned typologies of the BBI dictionary (appendix 2), and Carter's 'fixed expressions' (appendix 1)(see 2.2, 2.3). Although the collocations were initially identified intuitively, the table (appendix 4) provides corpus evidence, primarily of two types, mutual information (MI) and T-score, verifying that the word combinations were statistically significant and thus genuinely predictable collocations. Both MI and T-score numbers start with 0.0 as the random chance of words occurring near each other, but a high MI score is said to be more indicative of a unique collocation such as *totalitarian society* (number 7 in the text) where the collocational range of a word like *totalitarian* is assumed limited because it is rare, whereas a high T-score indicates significant collocations such as *free society* (number 5) where collocates are more common and thus have a wider range, but occur together frequently. Because MI and T-score give information about the occurrence of single words in the proximity of the node, 'fixed expressions' with more than 2 words were evidenced by the number of 'key word in context' (KWIC) concordance lines found with the same series of words used in the same sense. The corpus used, 'The Collins Wordbanks *Online* English Corpus' is composed of 56 million words of American and British contemporary written and spoken text, small enough that even a few KWIC lines occurring would denote significance. Examples of data and observations to be made from a corpus are illustrated below in Figure 3.

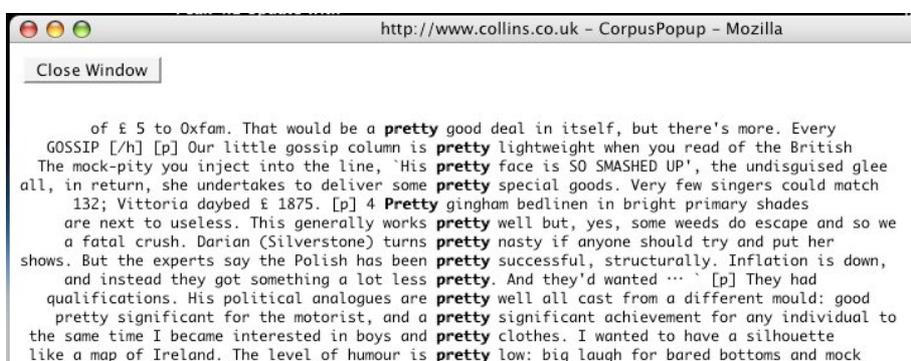
Fig. 3: Corpus data:

Collocate	Corpus Freq	Joint Freq	Significance
with	290716	3515	51.832029
a	973489	4229	42.270501
great	21745	1474	37.531564
to	1104731	3093	25.413031
good	42902	580	21.374691
big	17260	218	12.987454
new	61328	321	12.712065
how	50009	269	11.765287
pound	58153	280	11.449257
of	1100578	2138	10.049060
better	16584	140	9.701126
signed	2588	94	9.289510

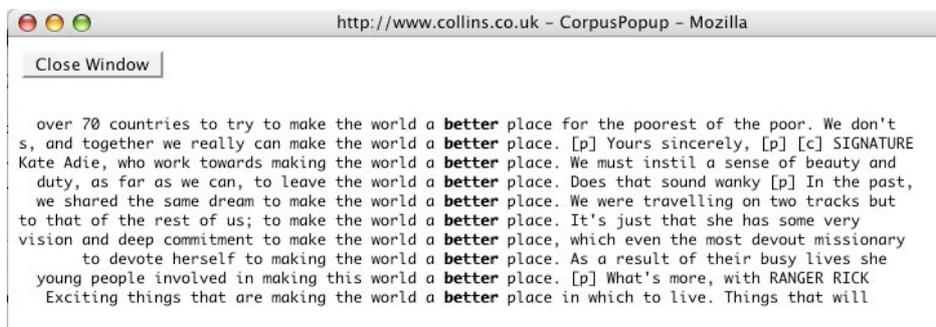
Collocate	Corpus Freq	Joint Freq	Significance
prisms	12	4	7.777122
brokering	14	4	7.554707
clinches	19	4	7.114091
finalize	16	3	6.946964
bofors	17	3	6.859493
clinching	57	9	6.699012
umbro	31	4	6.407752
clinch	176	22	6.361943
finalise	43	5	6.257596
persimmon	45	5	6.192002
finalised	82	8	6.004356

T-Score of 'deal'

MI-Score of 'deal'



KWIC sample concordance lines of 'pretty'



KWIC sample concordance lines of 'world+a+better+place'

- From the T-Score data of 'deal' it can be seen that *'to deal with'* and *'a great deal of'* are very common collocations, whereas the MI-Score reveals the more unique collocations *'to clinch a deal'*, or *'to finalize a deal'*.
- A KWIC sample of the node 'pretty' reveals that it is much more often used in the sense meaning *'fairly'* than its more prototypical sense, *'beautiful'*.
- The significant occurrence of *'to make the world a better place'* verifies it as a fixed expression.

3.2 Range and proportion of collocation types:

In section 4, suggestions will be made concerning how to go about using the text and all of the collocation it contains to direct learners' attention to the constraints on word-combinations in English as they exist in real language. The text, and assumably most authentic English, was found to contain pre-eminently three types of collocation according to the typologies adapted. Among other problems with the way collocation is often dealt with in the classroom discussed in section 4, this proportion is often not reflected and needs to be taken into account.

Adjective and noun combinations (BBI type L3)(appendix 2) were the most numerous type to be found with 8 instances of word combinations such as '*the ignorant masses*' (number 21), '*dedicated work*' (number 20) or '*thought control*' (number 4) which uses a noun as an adjective.

'Fixed expressions' such as '*it wouldn't do*' (number 10) idiomatically implying a supposed impropriety or unacceptability, '*And what's more*' (number 30), what Carter (1998) would classify as a stock phrase, or discorsal expression of the connective type (see appendix 1), or '*believe it or not*' (number 33), an 'allusion', or 'quotation' in the same typology were found 6 or 7 times in the text depending on if you count '*the left*' as a fixed expression because of its somewhat idiomatic usage of 'left' or accept that it only carries the meaning of an abstract noun referring to the group of people with certain political, or world-views when nominalized with '*the*'. An interesting item found in the previous 'allusion' category was, as it reads in the text, "*That's why the world continually gets to be a better place*" (number 32). I believe the choice of words here, specifically '*a better place*' about '*the world*' can be traced to a phrase any native speaker of English is familiar with (also used in the Micheal Jackson song), '*to make the world a better place*'. This is an example of parts of a fixed expression that when rearranged, still retain the phrase's idiomatic meaning. '*A better place*' in a purely literal sense could have a lot of meanings, but when '*the world*' becomes '*a better place*', it is safe to assume we mean more just, fair, or humane.

The next type of collocation most common in the text were the 6 created by a verb and preposition (BBI type G8-D)(appendix 2) such as *deal with* (number 1) (found twice), *instill in* (number 8), *immune to* (number 12), or *learn about*(number 23). It is worthy to note that these verb-particle constructions (which Chomsky, (1965) would label 'close construction') are very similar to phrasal verbs except that the term 'phrasal verb' is sometimes reserved for more idiomatic constructions that are thus treated as a separate headword in dictionaries (see 2.3). Phrasal verbs are notoriously difficult for the learner because they are often made with delexical

verbs and a preposition (e.g., ‘make up’) making meaning opaque, and involve complicated rules such as classes of separable and un-separable phrasal verbs, the former of which do not take a pronoun as an object unless separated. Furthermore, the same pair of morphological words can create different phrasal verbs with meanings only distinguishable by their being separated or not (e.g. *Things started to **look up** after I **looked** him **up**.*).

Other collocations found in the text, although numerous, were not as common as the three categories listed above. Two noun and preposition patterns (BBI type G1) (appendix 2), (*a*) *preface to* (number 3), and (*a*) *barrier to* (number 14) were in the text as were two verb-noun constructions (BBI type L1) (appendix 2), (*to be*) *kept secret* (number 18), and (*to*) *do something about it* (number 27), the latter far more idiomatic than the former and thus also a fixed expression. Beyond that, the remaining nine types of patterns found all occurred only once and for sake of brevity will not be mentioned individually, but can be examined in appendix 4.

3.3 Conclusions to be drawn

The analysis of the text reveals the fact that some types of collocation seem to appear more frequently in authentic language than others, namely adjective+noun, fixed expressions with an element of idiomacity, and verb-particle constructions. Although one could expect the proportions to vary slightly depending on factors such as mode, register, or genre, it is safe to assume they probably remain significant throughout all language and thus any discussion of how to focus on collocation to demonstrate the constraints on word-combinations in English should reflect this reality.

4. Discussion on how to focus on collocation: strengths and weaknesses of procedures typically taken.

The gist of my argument is that, in several ways, intuition fails the language teacher attempting to focus on collocation and attempting to present collocation out of context in an item-by-item fashion carries false expectations and does not give the learner an accurate picture of the extent that collocation permeates language. Instead, a general awareness can be facilitated by drawing attention to patterns that exist in texts learners have already dealt with.

4.1 Which collocations should be focused on?

The findings in section three have implications concerning which types of collocation deserve attention. Let's start by examining how some authors make that choice. It is noted by Woolard (2000) that words with high lexical content will have less collocational partners than superordinates of the same word, and thus more attention needs to be given to words with a wider collocational range. In terms of frequency in the text analyzed above, this holds true with a mere 4 out of 31 identified collocations having a higher MI than T-score, the MI score identifying these 'humdinger' combinations made from infrequent words that Woolard is probably referring to. (see 3.1). An implication could be that while these unique collocations may be easy to identify for the teacher as restricted, they may also be easier for the learner to identify as bound because of their relative rarity in input, but even more so, may be less applicable in use because of this rarity. Within these collocations made from frequent and perhaps core words, Woolard suggests attention be given primarily to 'unexpected collocations'. Along the same lines, Hill (2000) suggests primary attention be given to 'medium-strength collocations', those made up of common words (he gives the example of choices of delexical verbs. e.g. 'do, have, make, take, put', to verbalize nouns) because, "the main thrust of classroom vocabulary teaching at intermediate level and above should be to increase students' collocational competence with their basic vocabulary...".

Although valid to some extent, these suggestions pose several dangers. One is that even with this focus on 'unexpected', 'medium-strength collocations' made from common words, a disparity between intuition and reality arises. Case in point, idiomatic fixed expressions such as those found to be the second most common in the text (see 3.2) seem to be absent from their recommendations. By expanding the definition of collocation to include fixed expressions and using texts that learners have already had contact with to focus on the specific types of collocation found most common in language, learners would get a more accurate picture of the patterning in language determined by lexis.

Another difficulty is that these suggestions could be interpreted as a recommendation to focus on these types of collocations in an explicit item-by-item fashion (as Hill does). Time constraints make any item by item approach unrewarding since applying a 'collocationist' view to language reveals that "there are many more –tens of thousands of more- individual items to be learned than language teaching has ever recognized" (Lewis, 1996). Also, the very existence of the 'intermediate plateau' (Lewis, Morgan, 2000) (see 2.1), or the fact that "even very advanced learners often make inappropriate or unacceptable collocations" (McCarthy, 1990) is

evidence that collocation may resist explicit instruction. Furthermore, not limited to collocation, the overall value of explicit teaching in terms of acquisition is seriously doubtful given insights into second language acquisition as is epitomized by ‘The Natural Approach’ (Krashen, Terrel, 1983).

Instead, the goal of activities aimed at collocation should be that of a ‘consciousness-raising’ aimed at “sensitizing learners to the general difficulties involved (*which*) may help them to understand (*these principles*) in the future.” (Gairns, Redman, 1986), and encouraging learners to notice all such patterns in input rather than making an attempt to present just a few specific instances out of the tens of thousands that exist, with the false expectation that learners will retain them.

4.2 Grids

Grids, such as the one below (fig.4) (Rudska, et al, 1981) are another way of aiming at creating such an awareness but cannot be custom made for items encountered in texts, lack a context which may provide other collocates (e.g. *It’s* a lovely day.), and while other usages are briefly mentioned below the grid, can easily be contrived to the point of using words in infrequent senses to make them fit. This difficulty perhaps arises from the fact that “a coherent and meaningful group of collocations is often difficult to organize and it may be better to teach (them) as they arise” (Gairns, Redman, 1986). As an example, a corpus search of ‘*pretty*’ (used in the grid below) gives 75% of usages meaning ‘*fairly*’. While it may be true that most people would reply ‘*beautiful*’ if asked what ‘*pretty*’ meant, learners would not be exposed to the more frequent sense from the grid. We need to work from the data rather than attempt to create it to fit unnatural specifications.

There may also be a danger in such grids of an ‘over-focus’ that doesn’t achieve the desired general overall awareness, but gives the impression that there is something special and important about the particular items. While questions about word selection from students are rare, this writer has experienced countless learners stopping mid-sentence to ask which way of saying ‘*beautiful*’ would be appropriate for their purposes, perhaps because this example is so commonly used to display selection restriction.

	woman	man	child	dog	bird	flower	weather	landscape	view	day	village	house	furniture	bed	picture	dress	present	voice	proposal
beautiful	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
lovely	+		+	(+)		+	+		+	+	+	+	+	(+)	+	+	+	+	
pretty	+		+		+	+			+		+	+		+	+	+			
charming	+	+	+								+	+						+	
attractive	+	+								+	+					+		+	+
good-looking	+	+	+	+															
handsome	+	+																+	

In speech, **beautiful**, **lovely**, **charming** and **attractive** are often used for situations in which their real meaning would be too strong, in order to express enthusiasm.

EXAMPLES

The walls were covered with a most { **beautiful**
lovely
charming
attractive } wall paper.

I'll come to see you about seven – will you be there? **Beautiful** – okay – see you later.

She does really **lovely** things for people like bringing them their favourite flowers on their birthday.

Bacon and eggs for breakfast! **Lovely!**

Fig. 4: collocational grid from "The words you need" (Rudska, et al, 1981)

4.3 Other general suggestions found in the literature.

Something as simple as teaching learners the word 'collocation' can be useful (Conzett, 2000), and conveying the concept would be necessary before any consciousness-raising activities. Because the term, even in the narrow sense, covers such a wide range of constructions, it would be difficult for learners to understand what was being focused on in such an activity. Encouraging students to keep notebooks journaling encountered collocations or fixed expressions can be productive if students are willing and time permits, and for the teacher, making a habit of eliciting or providing a few common collocations and related words when offering clarification of any vocabulary item in the classroom is something of primary importance (Lewis, Morgan, 2000).

4.4 Dictionaries and learner data-first analysis.

Enabling learner independence by directing learners to specialized dictionaries with collocation information such as the ones mentioned earlier (see 2.3), or instructing students to 'look it up twice' (Conzett, 2000), once for meaning in a bilingual dictionary and once in such a dictionary is helpful, but encouraging students to notice patterns in texts dealt with in class as well as the example sentences in traditional dictionaries can foster the necessary initial sensitivity needed to see such patterns in input. Going even further by teaching students how to use the KWIC function of computer based corpora to research word groupings on their own is another classroom procedure to be utilized (Woolard, 2002). Data-Driven Learning (DDL)

proposed by Tim Johns (1991), while involving a certain amount of pre-arrangement of language data by the instructor, uses KWIC concordance lines to cultivate collocation awareness with the goal of “stimulating enquiry and speculation on the part of the learner, and helping the learner also to see patterning in the target language and to form generalizations to account for that patterning” (ibid). The data-first manner in which learners are set to the task of looking for patterning in language in DDL is ingenious, but the use of concordance lines for language learning can be troublesome because they are very difficult to decipher for all but the most advanced non-native speaker. Clues to the context, such as the purpose or type of text, what precedes it, etcetera, are all stripped bare and the reader is left with a single line of text, often with words cut off midway at the beginning and the end. The authenticity itself can also be a problem because the corpora often contain highly specialized technical or academic texts. This is another reason I recommend using texts previously dealt with in class.

A simple but interesting mixture of reference material training, and a data-first analysis of language can be seen in the activity below aimed at collocational awareness (Yoshimura, 2004). Assuming that the dictionaries available would be of the traditional type, learners would be working together to pull collocates in a data-first manner from the dictionary example sentences and the teacher-provided usages, all of which are created to be comprehensible. A fault with this exercise may be its narrow focus. Collocation patterns can be found surrounding many words in almost any text and this needs to be perceived for awareness to be heightened. Also assuming learners would understand what they were to be looking for; ‘collocates’, assumes they understand the concept, in which case they would already have an awareness and already be looking for them in input. This would to some extent make the exercise redundant except to examine ‘behave’ using various dictionaries.

<p>ACTIVITY 2</p> <ol style="list-style-type: none">1. Provide students with several examples with ‘behave’.2. Make several groups of three or four. Give each group different monolingual/bilingual dictionaries and ask them to find information about what types of collocates ‘behave’ has.3. Discuss the findings with other groups.
--

fig. 5: From Yoshimura Yuka (2004) “How to use a corpus to teach collocation”

5. How to use the collocations within the text to foster an awareness of the constraints of word combinations in English.

The collocations found in the text examined (see 3.1, 3.2), could be used in such a data-first way to demonstrate collocational patterns of a wider variety and surrounding a more assorted selection of words. Intentionally, a focus would be put on the types of collocations found to be prevalent in language in section 3. To avoid the pitfalls found in the ways collocation is often dealt with in the language classroom (see section 4), attention would be pulled to collocation as it actually exists in a text the learners are familiar with. This would provide learners the chance to view the extent that this level of language exists with the hopes they might notice it in the future. Below is an outline of how to use the text and the collocations within to achieve these aims.

5.1 Procedures for the activity aimed at collocation awareness (appendix 5).

The actual consciousness-raising activity (appendix 5) is preceded by a meaning based activity which will not be specified, but is necessary to ensure that attention reserves can be devoted to collocation at this stage. The teacher should first explain the concept of collocation, and the types to be focused on. For this purpose, there is a box at the top of the sheet using collocation found within the text to illustrate the types. Justification for the choice of types can be found in section 3. The main activity, to be done in groups or pairs, is a gap-fill with a select number of each type of collocation containing blanks. The less dominant words of the collocations are chosen to be filled in. The learners then make a second pass of the text attempting to identify remaining similar patterns. The learners are allowed to use dictionaries if they wish to look for patterns in example sentences within definitions.

Recycling texts in this way to create a focus on collocations avoids the pitfalls outlined in section 4 of relying on intuition to determine what kinds of collocations are frequent, the unrealistic expectations of an item-by-item approach, or contrived exercises which paint a misleading picture of collocation. It achieves the aim of demonstrating how abundant fixed lexical patterns are, which types are frequent and encourages learners to look for them in the future. Creating activities like this, the language teacher can use any text to demonstrate the constraints on word-combinations in English in the true way it is used in language.

6. Conclusion.

Developing an awareness of collocation as an important level of language, although a significant hurdle for the language learner, is necessary because of its efficient meaning-creating function and abundance in language. For the language teacher, research indicates that a shift in attention from the structural to the lexical is in order. Classifications, as well as the methods dictionaries employ to handle these items or patterns vary, and attempts guided by intuition to artificially create exercises aimed at them can often be revealed as misleading or misrepresentative when compared to real language. Such exercises ignore a much more rich and accurate resource, all language the learner processes for meaning in class. An attempt to tackle these items one-by-one would be not be effective because of the sheer number of them, but consciousness-raising activities based on texts they have come in contact with, broadly aimed at sensitizing learners to frequently found types of patterns or constraints, will provide valuable clues to be used in further input as the process of acquisition advances.

Works Cited:

- Alexander, R.J. (1984) Fixed expressions in English: Reference books and the teacher In: *English Language Teaching Journal* , vol 38, no.2, pp.127-34
- Altenberg, B, 1990. Speech as linear composition. IN: Caie, G. Hastrup, K., Jakobsen, A.L., Neilsen, J.E., Sevaldsen, J., Spect, H. and Zettersten, A. (Eds.) Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1. Department of English, University of Copenhagen, pp. 133-143
- Apresyan, Y.D., Melcuk, I.A. and Zolkovsky, A.K. (1969) Semantics and Lexicography: Towards a new type of unilingual dictionary, IN: Keifer, F. (Ed.)(1969) pp. 1-33
- Benson, M., Benson, E. & Ilson, R. (1986) The BBI Combinatory Dictionary of English, John Benjamins Publishing Company, Amsterdam
- Bolinger, D. (1976) 'Meaning and Memory', IN:*Forum Linguisticum*, vol.1, no.1
- Carter, R. (1987) *Vocabulary: Applied Linguistic Perspectives*, Routledge.
- Chomsky, N. (1965) *Aspects of the theory of syntax*, MIT Press, Cambridge
- Conzett, J. (2000) "Integrating collocation into a reading and writing course" IN: *Teaching Collocation: Further developments in the lexical approach*, Michael Lewis (Ed.) Heinle, Boston
- Firth, J. R. (1951) "Modes of Meaning" , in Firth, J. R. (ed.) (1957) *Papers in Linguistics*, Oxford University Press.
- Papers of J.R. Firth 1952-59. Bloomington:Indiana University Press, 1-32
- Firth, John R., 1957. *Modes of meaning*. Oxford: Oxford University Press, pp.190-215
- Firth, J. 1968 A Synopsis of Linguistic Theory 1930-1955. In: Palmer (ed) *Selected Papers of J.R. Firth 1952-59*, Longman, Harlow.
- Gairns, R. Redman, S. (1986) *Working with Words: A guide to teaching and learning vocabulary*, CUP
- Halliday, M.A.K.. (1966) "Lexis as a Linguistic Level", IN: Bazell, C. E., Catford, J.C., Halliday, M.A.K. & Robins, R.H. *In Memory of J.R. Firth*, Longman
- Hill, J (2000) "Revising priorities; from grammatical failure to collocational success" IN: *Teaching Collocation: Further developments in the lexical approach*, Michael Lewis (Ed.) Heinle, Boston
- Hill, J, Lewis, M (Ed.) (1997) *LTP Dictionary of Selected Collocations*, Language Teaching Publications, Hove

- Johns, T.F. (1991) 'Should You Be Persuaded: Two Samples of Data-Driven Learning Materials',
IN: Birmingham University English Language Research Journal 4 pp1-16
- Krashen, S. & Terrell, T.D. (1983), *The Natural Approach*, Pergamon
- Lewis, Michael (1993) *The Lexical Approach: The State of ELT and a Way Forward*, Heinle,
Boston, MA
- Lewis, Micheal (1996) "Implications of a lexical view of language" IN: *Challenge and Change
in Language Teaching*, Jane Willis & Dave Willis (Eds.), Macmillian Heinemann,
Oxford
- Lewis, Michael (1997) *Implementing The Lexical Approach: Putting theory into practice*,
Heinle, Boston.
- Lewis, Michael (2000) *Teaching Collocation: Further Developments in the Lexical Approach*,
Heinle, Boston.
- Lewis, Morgan (2000) "There is nothing as practical as a good theory" IN: *Teaching
Collocation: Further developments in the lexical approach*, Michael Lewis (Ed.)
Heinle, Boston
- McCarthy, M. (1990) *Vocabulary*, Oxford University Press.
- Moon, R. (1997) "Vocabulary connections: multi-word items in English" IN: *Vocabulary*,
Schmidt, N. and McCarthy, M. CUP
- Pawley, A., Syder, F.H., (1983) Two puzzles for linguistic theory: nativelike selection and
nativelike fluency. IN: Richards, J.C., Schmidt, R.W. (Eds.), *Language and
Communication*. Longman, New York pp. 191-226
- Rudska, B., Channel, J., Putseys, Y., and Ostyn, P., *The Words You Need*, Macmillian, 1981
- Sinclair, J. McH. (1966) "Beginning the Study of Lexis" , IN: Bazell, C.E., Catford, J.C.,
Sinclair, J.McH. (1991) *Corpus Concordance Collocation*, Oxford University Press.
- Wollard, G. Collocation – encouraging learner independence, IN: *Teaching Collocation: Further
developments in the lexical approach*, Michael Lewis (Ed.) Heinle, Boston
- Wray, A., Perkins, M. (2000) "The functions of formulaic language: an integrated model" IN:
Language and Communication 20, pp.1-28
- Yoshimura, Y.. (2004) "How to use a corpus to teach collocation" Presentation at the 第1回英
語コーパス学会東支部研究談話会 9/11/04

Appendix 1: types of fixed expression (From Carter, 1987 after Alexander, 1984)

I Idioms

(i) irreversible binomials: *spick and span; dead drunk*

(ii) full idioms: *run up (a bill); tear off (rush away)*

(iii) semi-idioms: *beefy-looking; dead drunk*

II Proverbs: *a watched pot never boils*

III Stock phrases: *When all is said and done; a recipe for disaster*

IV Catchphrases: *That's another fine mess you got me into*

V Allusions/Quotations: *You've never had it so good; We are not amused*

VI Idiomatic similes: *as sober as a judge; as old as the hills*

VII Discoursal Expressions

(i) social formulae/clichés: *How do you do? Long time, no see.*

(ii) connectives; structuring devices: *Mark my words; Once upon a time.*

(iii) conversational gambits: *We'll now take questions from the floor; Guess what?*

(iv) stylistic formulae: *Further to my letter of 11nt inst., Ladies and gentleman.*

(v) Stereotypes: *We're just good friends; I thought you'd never ask.*

Appendix 2: BBI typology of collocations

(From Benson, M., Benson, E. & Ilson, R. (1986))

Grammatical Collocations

G1 = **noun + preposition**: *apathy towards* (minus 'of' constructions)

G2 = **noun + to-inf**: *to have the right to do it* (minus 'to' meaning 'in order to' constructions)

G3 = **noun + that-clause**: *an agreement that she would do it* (minus clause constructions where 'that' can be replaced by 'which', or that-clause modifying nouns that are objects of prepositions: (It was by chance that...)

G4 = **preposition + noun**: *by accident, in advance*

G5 = **adjective + preposition/prepositional phrase**: *angry at everyone, hungry for, fond of* (some cannot stand alone)

G6 = **predicate adjective + to-inf**: *It was necessary to work, She is ready to go.* (minus {too adj to inf}: *too tired to work*, and 'to' following past participle in a passive construction: *was proofread to eliminate errors, however, amazed to see him would be pure adjectival use of past participle so a collocate*)

G7 = **adjective + that-clause**: *afraid that she would fail the exam*

G8 = **verb patterns (Nineteen patterns described below)**

A = **verbs allowing dative movement transformation**

He sent the book to him, He sent him the book

B = **transitive verbs not allowing dative movement transformation**

*They described the book to her, *They described her the book*

C = **transitive verbs + for** (allowing dative movement transformation)

She bought a shirt for her husband, She bought her husband a shirt

D = **verb + preposition**

Adhere to the plan, catch up to his friend (minus 'by':*by train*, other free combinations such as 'in': *walk in the park*)('to' overlaps with type B)

E = **verb + to-inf.**

Began to speak, offered to help

F = **verb + inf** (without to)

we must work, he had better go (modals)

G = verb + verb-ing

They enjoy watching tv, I miss going to work (overlaps with E: began reading)

H = tr. verb + object+to-inf.

Asked me to come, forced them to leave (most can be passivized)

I = tr. verb + object+inf. (without to)

She heard them leave, we let the children go to the park (most cannot be passivized)

J = verb + verb-ing

I caught them stealing apples, she heard them leaving (overlaps with I) (most can be passivized)

K = verb + possessive+gerund

Please excuse my coming late, I remember Bill('s) making that mistake.(some NS' differ in preference)

L = verb + that-clause

They admitted that they were wrong, she believed that her sister would come (some verbs always take a noun/pronoun object before the that clause: They convinced us that he was innocent.) (some allow insertion of 'the fact': he mentioned the fact that he was guilty.)

M = tr. verb + direct object + to be +adj/past participle/(pro)noun

We consider her to be very capable, The court declared the law to be a violation of civil rights.

N= tr. verb + direct object+adjective/past participle/(pro)noun

She dyed her hair red, we found it interesting

O = tr. verb + two objects (not normally used in a prepositional phrase with to or for)

The teacher asked the pupil a question, We bet him ten pounds.

P = verb + (obligatory) adverbial (but not a particle)

He carried himself with dignity.

Q = verb + interrogative word

He asked how to do it, she could not decide whether to begin

R = it + tr. verb + to-inf/that-clause

It surprised me to hear that. It puzzled me that he never answered the phone,

S = intr. verb + predicate noun/adjective

He was a teacher, he became an engineer

Lexical collocations

L1 = verb (meaning creation, activation) + noun/pronoun/prepositional phrase

Come to an agreement, make an impression, draw a line vs. form a line: collocate changes the meaning.

L2 = verb (meaning eradication, nullification) + noun

Reject an appeal, reverse a decision, repeal a law

L3 = adjective + noun

strong tea, weak tea, A chronic alcoholic, a crushing defeat (noun used as adj: house arrest)

L4 = noun + verb (the verb names an action characteristic of the person or thing designated by it)

Bees buzz, bombs explode, blood circulates

L5 = noun + noun: units (usually noun + of + noun)

A herd of buffalo, a pack of whales

L6 = adverb + adjective

Strictly accurate, deeply absorbed, hopelessly addicted

L7 = verb + adverb (not adverbial particle)

Apologize humbly, amuse thoroughly

Appendix 3:

Text from 'A selection of Chomsky's posts from the ChomskyChat Forum' {<http://www.zmag.org/chomsky/other/chomchataarch.htm>},

Title heading: On "The left is weak here..."

Question to Noam Chomsky from one of the forum participant:

"You write: "I never really believed how highly indoctrinated the left wing is, until election to youth rep. I would like to know what you attribute this to, and if you can suggest any possible ways to help the people I know find roles as active and dissenting members of society. I have had trouble getting my 'radical' and 'crazy' ideas heard..."

Chomsky's reply:

1 I suspect you are mostly dealing with¹ people who've had what's called "a good
2 education²." These are the people Orwell wrote about in his preface to³ "Animal Farm" -
3 - which was not published: it was about thought control⁴ in free societies⁵. He concluded
4 that⁶ the effects were not unlike those of the totalitarian society⁷ he was satirizing, but
5 the mechanisms were different. One is a good education, which instills in⁸ its victims
6 the understanding that⁹ there are some things it just "wouldn't do¹⁰ to say" -- or think, we
7 may add. The left¹¹ is not immune to the effects of¹² a good education, though for the
8 left, what it wouldn't do to say, or think, sometimes is different. I don't want to
9 romanticize. For people who lack the benefits of¹³ a "good education," there are many
10 other barriers to¹⁴ understanding: the need to put food on the table¹⁵ so one's kids don't
11 starve, for example¹⁶. How does one deal with all such matters? If there's a magic key¹⁷,
12 it's been kept secret¹⁸ for thousands of years¹⁹. The only way that is known is dedicated
13 work²⁰, not to "teach the ignorant masses²¹," but to join with²² people to try to learn
14 about²³ what the world is like, and then to act on²⁴ (largely shared²⁵) moral instincts²⁶ to
15 do something about it²⁷. Sounds trite²⁸? It should: it is. But it's also about as much as can
16 be said²⁹, I think. And what's more³⁰, it often works. That's why³¹ the world continually
17 gets to be a better place³² (believe it or not³³).

Appendix 4: Collocations from Chomsky text with typology and corpus information

(See below for guide to entries)

<i>number in text</i>	<i>Phrase</i>	<i>Line in text</i>	<i>BBI type</i>	<i>Carter type</i>	<i>MI</i>	<i>T-score/KWIC information</i>
	(to) deal with	1	G8,D	x	x	51.83 (deal to with)
2	a good education	1-2	L3	x	x	7.02 (education to good)
3	(a) preface to	2	G1	x	1.45(#20)	4.62 (#1) (preface to to)
4	thought control	3	L3,	x	x	x/ 4 KWIC same sense
5	free society	3	L3	x	x	4.3 society to free
6	concluded that	3-4	G8,L	x	x	18.8(#1) concluded to that
7	(a) totalitarian society	4	L3	x	5.34 from totalitarian	2.58 totalitarian to society
8	(to) instill in	5	G8,D	x	1.3 from instill (#8)	1.6 (#7) instill to in
9	the understanding that	6	G3	x	x	8.08 (#6) (understanding to that)
10	It wouldn't do	6	x	Full-idiom	x	x
11	the left	7	x	Full-idiom	x	27.07(left to the)/ 7/40 KWIC same sense
12	immune to the effects of	7	G8,D G5	x	3.4 (immune to effect)	2.2 (immune to effect) 3 kwic, none the same
13	(to) lack the benefits of	9	G8,S	x	x-	x/ 4 kwic, 1 in exactly the same sense
14	(a) barrier to	10	G1	x	1.26(barrier+IN)	3.14(barrier to to), 5.11(barrier+IN)(#2)

15	to put food on the table	10	G8,B	Semi-idiom, Stock phrase	x	5.1(food to table) 9 KWIC, all the same sense
16	for example	11	G4	Stock phrase, Discoursal Expression (connective)	3.4(example to for)	73.1(example to for)
17	a magic key	11	L3	Allusion	x	-One KWIC, same sense
18	(to be) kept secret	12	L1	x	4.9(#6) (secret+VERB to keep)	5. 55(#4) (secret+VERB to keep) 8.1(keep to secret)
19	thousands of years	12	L5	x	x	9.8 (years to thousands)
20	dedicated work	12-13	L3	x	x	x
21	the ignorant masses	13	L3	x	x	x/ 1 KWIC same sense
22	(to) join with	13	G8,D	x	6. 03(Join+IN)	7.8(Join+IN)
23	(to) learn about	13-14	G8,D	x	x	17.87(learn to about)
24	(to) act on	14	G8,D	x	x	9.02(act+IN)
25	largely shared	14	L6	x	x	x
26	moral instincts	14	L3	x	x	x
27	(to) do something about it	15	L1	Semi idiom	x	41.5 (something to do) 40 plus KWIC, same sense
28	(to) sound trite	15	G8,S	x	7.5 (#6) trite to sounds	2.6 (#2) trite to sounds
29	as much as can be said	15-16	x	Stock phrase	x	x/ 1 KWIC used ironically meaning 'ugly'. 4 'as much as can be' with different verbs

30	And what's more...	16	x	Stock phrase, Discoursal Expression (connective)	x	x/- 1 KWIC same sense.
31	That's why...	16	x	x	x	35.22 (why to that) 30/40 KWIC used in same sense.
32	(to)(make) the world a better place (Basis for 'gets to be a better place)	17	G8,N	Allusion	5.76(#2) ('a better place' to world)	4.04(#3)('a better place' to world), 13/40 KWIC 'a better place' used with 'world'
33	Believe it or not	17	x	allusion	x	8.3(or),10.6(not) from 'believe it'. 40 KWIC, same sense

Key to table:

[x] = No applicable classification (BBI or Carter), or corpus data not available (collocates were limited to first 100).

[(#2), (#3), etc.] = Rank of significant collocate. Provided when MI or T-score were low in comparison.

[(deal to with)] indicates that *deal* was the node and *with* was the significant collocate found.

[(x/ 4 KWIC same sense)] indicates that no T-Score data was found, but 4 concordance lines with the feature were found.

[30/40 KWIC same sense] indicates that of 40 concordance lines found with the feature, 30 used it in the same way.

[IN,VERB] indicate part of speech tags specified when a deeper probe was made. IN=preposition, VERB= verb

Note: The corpus used, The Collins Wordbanks *Online* English corpus is composed of 56 million words of contemporary written and spoken text, relatively small by modern standards. Also, collocates listed for MI or T-score are limited to 100, and concordance lines around a node are limited to 40. Although generally sufficient, limitations were apparent when an obvious collocation such as *thought control* (number 4) could not be found in MI or T-score using either of the words as a node, but had to be probed for directly by doing a KWIC search on those two words in that order.

Appendix 5: Worksheet for consciousness-raising activity on collocation.

Worksheet 2: Collocations

<p>Adjectives + nouns:</p> <p>'She has <u>a good education</u>' sounds natural but 'she has <u>a nice education</u>' sounds funny.</p> <p style="text-align: center;">Why?</p>	<p>Fixed expressions:</p> <p>'To make the world a better place' has a special meaning, but if we say 'to change the world into a better place', or 'to make the world a better location' it sounds funny.</p> <p style="text-align: center;">Why?</p>	<p>Verb-particle constructions:</p> <p>We can say 'talk <u>with</u> people' or 'talk <u>to</u> people'. However, we can only say 'deal <u>with</u> people', not 'deal <u>to</u> people'.</p> <p style="text-align: center;">Why?</p>
--	---	--

① In the text that we have just read, Can you remember the words in the blanks?

The 'collocation' makes the words correct.

② How many other 'collocations' can you find?

Hint: all together, there are 33, one is used twice and one is used 4 times.

Which group can find the most?

(Using dictionaries is OK!)

I suspect you are mostly dealing people who've had what's called "a good education." These are the people Orwell wrote about in his preface to "Animal Farm" -- which was not published: it was about control in free societies. He concluded that the effects were not unlike those of the totalitarian society he was satirizing, but the mechanisms were different. One is a good education, which instills in its victims the understanding that there are some things it just "wouldn't to say" -- or think, we may add. The left is not immune to the effects of a good education, though for the left, what it wouldn't do to say, or think, sometimes is different. I don't want to romanticize. For people who lack the benefits of a "good education," there are many other barriers to understanding: the need to put food on the table so one's kids don't starve, for example. How does one deal all such matters? If there's a magic key, it's been kept secret for thousands of years. The only way that is known is work, not to "teach the ignorant , but to join with people to try to learn what the world is like, and then to act (largely shared) moral instincts to do something about it. Sounds trite? It should: it is. But it's also about as much as can be said, I think. And what's , it often works. That's why the world continually gets to be a better (it or not).