Watch Out! The Corpus, Verb Usage, and the Non-Native Teacher of English

Stephen Pihlaja

Niigata Meikun High School

Abstract

For non-native teachers of English (especially in Japanese high schools), questions of appropriate English usage and possibility are constant. Although corpora usage has widely been discussed as a useful tool for EFL students, this paper discusses the possibility of corpus-usage for non-native EFL instructors in answering questions of usage and helping prepare material for classrooms. Using the online British National Corpus, a simple step-by-step methodology for teachers searching words with shared meaning and usage was presented. Analysis of the search results shows that although corpus results are often complex and require time to sufficiently analyze, corpus studies can help clarify usage and meaning questions. The paper discusses the problems of motivation with using the corpus and discusses possibilities for improving teacher motivation for using corpus studies.

Introduction

For non-native EFL teachers (especially in Japanese high schools), one of the most difficult aspects of language learning and teaching is words that share parts of meaning, but differ greatly in usage. Theoretically, a corpus should be able to help non-native EFL teachers find examples that can be used to fashion working 'rules' of usage apart from traditional textbooks and dictionary definitions. This paper will explore the possibility of using a corpus (The British National Corpus) to help non-native EFL high school teachers delineate proper usage by presenting naturally occurring examples. It will also discuss whether corpora can serve as a helpful pedagogic resource for non-native English teachers when trying to better understand the nuance of a given lexical item. Finally, the paper will briefly discuss ways to motive Japanese EFL instructors to use corpora, both for enriching their own learning experience and that of students.

Review of Literature

Usage of corpora for EFL purposes has centered namely on the how it can be used in the classroom and how to create teaching methodologies that are more 'organic' than typical grammar-based models. Aarts writes, 'Ideally, the intuition-based grammar, through its confrontation with corpus data, becomes an *observation based* grammar' (Aijmer, 1991). A descriptive, observation-based grammar is valued because it describes grammar as it is, rather than how grammar should be. Data-Driven Learning (DDL) is one of the embodiments of

descriptive grammar centered teaching approaches. DDL methodology uses corpus studies and authentic texts to answer questions posed by a student or researcher. One of the many rewards this type of learning offers is answering questions that researchers and students have not asked and discovering new questions when the data one is presented with does not match what was expected. This process creates a kind of 'serendipity learning' that can be extremely rewarding (Flowerdew, 1996).

In DDL methodologies, the corpus becomes a kind of 'informant' for a learner rather than an instructor (Johns, 1991). For EFL students, the corpus can be used to answer common questions such as "What's the difference between . . .?" and "Can I say this?" Hadley (2002) notes, "A DDL Approach suggests a move away from unnatural, 'simplified' textbook English, and allow for greater use of authentic materials." In Japan, a popular English teaching television program employs a character named *Corpusie* (*kopasu-cun*) who presents corpus lines of lexis to viewers and was also the main character of a corpus reference book Corpus Practice Register (*koopasu renshucho*). Corpora has been shown to be effective at all levels, including 'low-achievement' EFL learners (Tian, 2006).

This DDL approach may be effective in the classroom at many levels, but in the Japanese high school context, where non-native speakers of English are teaching English, corpora can be an effective tool for not only non-native EFL students, but also non-native EFL teachers in answering some of the most common questions regarding questions of usage. Using a corpus to help answer the ever-present "Can I say this?" question might prove extremely useful for helping non-native EFL teachers achieve independence when addressing problems that come up in English study.

Presentation of Search Methodology

This paper will present a simple, easy-to-follow methodology for EFL teachers seeking to use corpora to investigate words with overlapping meanings or usages. The method follows these steps:

1. Clearly state question to be investigated through corpus study
2. Collect corpus data
3. Manage data and seek out unknown lexis and structures
4. Categorize usage of words
5. Note collocations
6. Contrast results of words within a given study

7. Analyze and theorize

8. Test theory through practice and further research

For this research, the British National Corpus was employed, using only the free, online version (http://www.natcorp.ox.ac.uk/). The question addressed is a common one among many EFL learners and non-native teachers of English: what are the appropriate usage for the words *look*, *see*, and *watch*? How do they differ in usage and how do usages overlap? The key to this corpus studies for non-native English teachers is to first clearly state the question they have about a given lexical item or set of lexical items. As corpus material can be overwhelming, it is key to start out with a clear goal rather than simply fishing for answers.

The next step is to gather corpus lines by searching the given word and limiting examples to verb forms. In this study, searches for each word were done twice, resulting in 100 random lines of corpus. Corpus lines (see Appendices) were not re-arranged and appear as the BNC presented them, the first search constituting lines 1-50 and the second constituting lines 51-100. Ambiguous lines were not replaced. Whether or not 100 lines are needed for a given question is up to the teacher and time constraints. As the goal is to move towards an understanding of the authentic use of the word rather than having a complete scientific picture of all uses of the verb, the number of corpus lines should be kept manageable.

While this study will involve 100 examples that have not been manipulated in any way, for teachers doing searches, it is inevitable that difficult examples with unknown lexical items are bound to occur. Rather than be bogged down with examples that are confusing, it may be more advantageous for teachers to throw out lines they find confusing or involve lexical items they don't understand, unless the items occur frequently. In this way, the information is kept manageable and not unnecessarily complex.

Finally, before starting, one should consider the goals of corpus studies. When looking at corpus data (as with looking at any authentic language information), clean, simple answers to questions are unattainable. The goal of corpus studies for the non-native teacher of English should be to aid theorizations about language that can be tested again and again as new data appears. The corpus helps give a picture of language as it is and, therefore, helps teachers understand how words are used. But this hardly a complete picture and should only be viewed as a step in research, not the end point. Users of the corpus are not writing dictionaries, they are trying to understand in small ways how words are used. The goals of study should reflect this modest outlook. In the corpus study presented in the paper, the goal is to set forth

working theories about usage that can then be applied to new information and contrasted with other available resources.

Corpus Study Data

LOOK

After physically collecting the data, a researcher must categorize usages. This is clearly the most time-consuming step of corpus studies. Teachers might want to limit their studies to 30 to 50 lines to help ease the time commitment. Teachers should also take care to note any frequent collocations, especially pronouns following verbs as these set phrases often have very different meanings.

The temptation of corpus research is to look immediately at the lines that seem to be the most clearly related to the question being asked; that is to say, to look immediately at the entries related to the meaning *to perceive visually*. These entries are obviously important, but the results of the categorizing usages of '*look*' from the search yields the first substantial finding in the study. As the primary question of this study was how '*look*' differs from '*see*' and '*watch*,' it is important to not only look at uses of the '*look*' to mean '*perceive visually*', but other uses of the word and how the two uses might be related or not. As for '*look*,' one can observe uses meaning '*to perceive visually*' and '*to consider*'. There are also a number of lines that might prove to be difficult in that they should fit on one of these categories, but they seem to denote both 'consideration' and '*visual perception*.'

Corpus lines of '*look*'
29. I look at these girls.
41. Well to be quite , to be quite honestly sweety, I mean, I look at him and I think I'm not unfond of him
48. and he said crikey look at this, what am I gonna do here and he sort of had to fill quite a big gap in, but he sort of pushed it and filled it and then the wallpaper goes up so you can just see it, a little bit, what's a name now, just in the hall there's a bit
54. Look at Paul,'; he said.
63. Must be pretty tough for them when they look at other lasses.';
93. Look at these hairstyles!
94. Just look at the headline!
100. And look at my nails Marg!

In all of these instances, although an argument might be made that one line or another clearly means either only '*perceive visually*' or only '*consider*,' there seems to be some overlap between these two usages. In classic dictionary definitions, this relationship might be difficult

to see as examples and definitions are labeled separately. The corpus study allows teachers to see that, in authentic texts, these definitions might not be as clearly differentiated. By seeing the blurring between the usages, one is able to see that '*looking*' and '*considering*' are related and their meanings can be embedded in each other.

Collocations play an important role in any corpus study. In this case, it is clear that collocations with pronouns (e.g. '*look out*,' '*look at*,' '*look for*,' etc.) were extremely important in determining the usage and meaning of '*look*.' In the categorization stage of corpus research, these verbs should be grouped together by meaning, and further analyzed based on the importance of collocating words to the study. In this case, as the study pertains specifically to the usage of the verbs as '*perceive visually*', collocations play a less central role.

Although not necessarily an issue of collocation, it might be worth noting (especially in this search) what sort of things are the objects of the verb '*look*' (i.e. '*sketches*', '*galaxies*,' etc.) If the researcher is able to note commonalities among the object (for example, they are all inanimate), this information might be helpful for determining difference between the three verbs.

SEE

Having noticed that overlapping meanings in the study of '*look*' helped point towards the meaning of '*look*,' the same tendency of overlapping might be investigated in the search of '*see*'. Again, starting with categorization, all words were labeled by usage before any analysis is done, and all frequent collocations are noted.

Unfortunately, in the search of the BNC, '*see*' in is used to mean '*to refer to*' in well over half of the corpus lines that we were presented with, often appearing in parenthesis and instructing the reader to refer to another part of the same article, another text, a graph, and so on. Perusing the corpus lines, it becomes clear that the search of '*see*' seems to have been negatively affected by the high concentration of academic and technical texts in the BNC. Kilgarriff (2003) notes that this problem occurs with rare words or rare meanings of common words, but it can observed that even common meanings might be excluded in a small search. For example, this search yielded no meanings of the verb '*see*' as '*to have a romantic relationship*,' which one might expect to observe.

The problem of '*representativeness*' is abundantly clear. If this search were done using a corpus with a higher concentration of spoken texts or fiction text, one would likely see few or

no occurrences of this usage. As the BNC is 90% written texts and only 25% of those are '*imaginative*' texts (Kennedy, 1997), it is clear why '*refer to*' is so prominent. Given a higher percentage of spoken texts, it is likely that an increase in usages like '*to meet*' might occur. In this sense, native intuition is still an incredibly valuable tool when viewing corpus lines. All of this should lead us to a very careful conclusion that in a corpus that has a high concentration of a specific type of writing (i.e. academic and technical writing), we are likely to find much different answers to our question of usage than if we search a different kind of corpus.

Although this information is not completely useless ('*refer to*' is certainly a legitimate use of the word '*see*'), it might be advisable to move onto the next categorized set of sentences. At just over 20% of the lines, 'understand' is the second most common usage of '*see*.' This meaning seemed to occur most frequently in the phrase '*I see*' or '*You see.*'

As with '*look*,' it seems that the definitions of see as '*perceive visually*' and '*understand*' are not always separated perfectly into one meaning or another. Line 37 (labeled as '*to observe*') is a good example of this problem. '*Sometime and see what the facilities we've got there.*' In this case, certainly a kind of visual perception will be included, but the speaker seems to be emphasizing observation over perception. On the other end of the spectrum, there are lines like 3 (labeled as '*perceive visually*'): '*Climbing the steps through the central arch to the altar-like plinth, I see that someone has left a pot of red begonias.*' Here the speaker is perceiving '*a pot of red begonias*' with their eyes, but also deducing that '*someone has left*' them: a kind of observation.

Unlike '*look*,' collocations seemed less important to the meaning of '*see*.' There were very few examples when the pronoun dictated the meaning. The usage of '*perceive visually*' occurred relatively infrequently (3, 6, 48, 57, and 69) as did the meaning of see as '*to meet*' (only three occurrences) which might be cause for another search, especially if the researcher notes a significant difference in the kind of things that are the objects of these verbs. In the case of see, the objects observed seem to have be more varied than with look (feet, data, sunrise). Whether or not this difference holds up when confronted with more data will require additional searches. Moreover, if additional searches uncover only more unclear data, this too might cause the researcher to be more careful in theorizing about any findings.

WATCH

Of all three of these verbs, 'watch' is the only one to produce a majority of lines meaning 'perceive visually' at 30 lines. Given its frequent occurrence, one is able to get better understanding of this usage than in the other searches. Also, as seen in the previous two searches, two meanings of 'watch' were closely related: 'perceive visually' and 'observe.' The close relationship to the 'observation' usage seems to show that attentiveness is often implied in 'watching.' Line 63, for example: 'Watch any good swing closely and you will notice that the left forearm rotates clockwise on the backswing.' Although this is clearly referring to a visual perception, a kind of attentive observation is also implied. The phrase 'watch this space' which occurs rather often and refers to watching an advertising space for changes and information. It shows the close relationship between visual perception and observation.

As with 'see' and 'look,' the collocations of pronouns with the verb seemed to play an important role in the meaning of the verb. Some collocations and meanings also overlapped (e.g. 'look out' and 'watch out').

Although the objects that were observed in the 'see' and 'look' searches were relatively similar, the objects of 'watch' were, in many cases, very different. In many cases of this meaning, the thing that is being perceived visually is moving or changing. For example, line 3: 'I turn and watch them go' and line 9 'Sometimes they'd stand it up and watch it fall over.' and line 44, 'Watch my juggling.' A common collocation with 'watch' was 'television' or its derivative. In every instance of the collocation with 'watch television,' one can observe that the speaker is watching a program on the television, which is obviously moving and changing.

Theorizing

Taking the information gathered from the corpus study, the researcher can now begin to theorize about the differences in usage between the three verbs. First, here is the information that was gathered about the usages of these verbs to mean 'perceive visually.'

LOOK	Close relationship to meaning '*to consider*'
	Pronoun collocation often dictated meaning
	Objects of '*look*' were inanimate objects
SEE	Search was dominated by '*refer to*' meaning
	Close relationship to meaning '*to understand*'
	Pronoun collocations largely absent
	Pronoun collocation often dictated meaning
	Objects of '*see*' were varied
WATCH	Close relationship to meaning '*to observe*'
	Pronoun collocation often dictated meaning
	Objects of '*watch*' were animate

From this information, the following theories can be put forth:

・Pronoun collocation is frequent with '*look*' and '*watch*' and often helps explicate the meaning. Conversely, pronoun collocation does not seem to influence '*see*' in the same way.

・Relationship to a second usage was important in all three cases and may be helpful in deciding which verb to use in a given situation.

・While it seems that '*look*' is used when the object is inanimate, the object of '*watch*' is often animate. The objects of '*see*' were mixed, although the majority were inanimate.

These theories are not meant to be concrete definitions, but rather road maps for analyzing further research or examples that one might discover in usage. The corpus is by no means the definitive authority on language. Moreover, unless one looks at every instance of the word, there is still the chance that data might not be representative of even the corpus one is studying. What one does gain, however, in any corpus search, is an appreciation for the word as it is, rather than the word as it should be or as it is conjectured by a textbook or reference book author. Using all of these resources in concert with one another is bound to give the teacher the most well rounded picture English usage.

Implications

This corpus study has made two points very clear. First, corpus study is not very clean; that is to say, the question posed at the beginning of the study was not answered definitively through corpus study. Although the information gleaned may be useful for exposing researchers to language information, it is not nearly as easy and convenient as simply looking the words up in the dictionary. With a multitude of simpler reference books and textbooks available that can clearly explain these differences, use of the corpus seemed to be relegated to an interesting possibility, but not pragmatically applicable. Why would a teacher take several hours to research a subject when the answer is seemingly available in a reference book?

Looking at this corpus data, the value of corpus study is abundantly clear. Although reference books, dictionaries, and textbooks might give simple answers to questions, language is not a simple matter. The corpus data presented in this study shows that usage of words is diverse and complicated. Moreover, looking at corpus data, one can see usages of words that are ambiguous, but that, in their ambiguity, give hints to usage of the word. For example, being able to observe these overlapping areas can be crucial in understanding usage. By looking past the obvious examples and considering corpus lines, very useful information can be found resting beneath the surface.

If the goal is observation-based grammar, it seems that first teachers need to be convinced that observation-based grammar is more authentic (and better) than inauthentic textbook models. The question is whether or not teachers should take the time for corpus studies. Particularly in high school contexts, where textbooks often work backwards from grammar rules that often present awkward, uncomfortable English sentences, teachers must be convinced that perfectly behaving grammar structures are inferior to authentic usage and communication models. For teachers, however, although the problem is often understood, with pressures from school boards, parents, principals, and vice-principals to produce college entrance exam passing students, the corpus may seem like an unaffordable luxury.

A shift in thinking about English is desperately needed (and occurring slowly) in the Japanese jr. high and high school EFL pedagogy. The change, it seems, needs to come from individual non-native EFL teachers who understand the value and necessity of corpus studies, use it effectively to answer their own questions and pass the information on to their students. The teachers that take initiative to incorporate corpus studies, not only into their own study, but also into their classes and subtly change the way their classes are run.

Given the pleasure derived from finding out information for oneself rather than simply being spoon-fed grammatical formulas, teachers who have an appetite for English knowledge can be encouraged to see the power of the corpus as new tool. This does not replace the native speaker or teacher whose answers to questions of '*Can I say X?*' may still prove to be the most accurate. Working with this tool, teachers (both native and non-native) will be able energize and empower students toward the goal of becoming descriptive linguists.

Final Thoughts
The corpus is not able to answer questions quickly and easily, especially when compared to traditional grammar texts and guides. In the end, however, the results are far more likely to

shed light on real usage of words and meanings.   Study of the corpus alone will not improve proficiency in language and it is not a magic solution to any problem. It is, however, one of the most useful and practical tools non-native EFL instructors have in seeing how the English language actually works.   This is the most powerful asset of the corpus: simply showing language as it is. The challenge is to convince high school teachers that the time and effort is worth it. For the Japanese high school system, engaging and using this information will be crucial for teachers, and subsequently students, becoming communicators of English.

References

Aijmer, Karin and Altenbrg, Bengt. (1991) English Corpus Linguistics. UK: Longman Group UK Ltd.

Hadley, Gregory. (2002) Sensing the winds of change: An introduction to data-driven learning. RELC Journal 33(2), 99-124.

Hunston, Susan and Francis, Gill. (1998) Verbs observed: A corpus-driven pedagogic grammar. Applied Linguistics, 19(1), 45-72.

Flowerdew, John. (1996) Concordancing in language learning. In Pennington, Martha C. (ed.) The Power of CALL (pp. 97-113). Houston: Athelstan.

Johns, T. (2006) Homepage. Retreived November 8th, 2006 from http://web.bham.ac.uk/

Johns, T. F. (1991) Should you be persuaded: Two samples of data-driven learning materials. English Language Research Journal, 4, 1-16.

Kennedy, Graeme. (1998) An Introduction to Corpus Linguistics. London: Pearson Education Ltd.

Kilgarriff, Adam and Grefenstette, Gregory. (2003) Web as corpus. Computational Linguistics, V(N), 1-15.

Meyer, C. F. (2002) English Corpus Linguistics. Cambridge: Cambridge University Press.

Owen, C. (1996) Do concordances require to be consulted? ELTJ 50, 219-224.

Pihlaja, S. (2006) Survey Results. Unpublished survey results, Niigata City, Japan.

Reppen, R. (2001) Review of MonoConc Pro and WordSmith tools. Language Learning and Technology, 5(3), 32-36.

Tian, S. P. (2004). Using corpora concordancing to assist low-achievement EFL students. Paper presented at the 4th International Conference on ELT in China, Beijing, China. Retrieved November 8th, 2006 from http://www.elt-china.org/lw/pdf/TianShiauping.pdf

Yukio, T. (2004) Corpus Practice Register (*koopasu renshucho*).Tokyo: NHK.

Stephen Pihlaja is an American English teacher from Chicago, IL. He has been teaching English in Japan for over three years and holds a master's degree in Applied Linguistics through the University of Birmingham (UK). He serves in an Assistant Language Teacher (ALT) capacity at a large private high school in Niigata City, Japan.