Results of a Corpus Study of LOOK, SEE, and WATCH

Gregory C. Anthony
Hachinohe University

The near synonymy of the verbs LOOK, SEE, and WATCH consistently cause problems for EFL teachers and students. How can we identify the various senses of these verbs? Which senses are the most common in actual language use? How should vocabulary items be presented in the classroom? These questions motivated an exploratory corpus study of these three verbs. After introducing some of the basic terms and theories of corpus studies, this paper indentifies the most common unique collocations and colligations that pair with these verbs. Through an analysis of the varied phraseologies that result, we see evidence for the argument that the majority of language is constructed of fixed phrases. As these phrases seemingly acquire meaning from the co-text in which they are found, I argue that, similarly in the classroom, such words should be also presented in context, along with their common word pairings.

　動詞LOOKの類義語であるSEEおよびWATCHは、EFLの教師および生徒にとって、常にまぎらわしい語彙であり、問題を引き起こす。どのようにこの単語を認識できるようになるのか。どの単語が適切で、使い分けできるようになるのか。どのように教室で言葉を提示し紹介すべきか。これらの疑問に答えるため、3つの動詞に関してコーパスで検索した。ここでは、コーパス研究に関連する基本的な用語や理論を紹介した後、この3つの動詞と最も頻繁に用いられる連語(collocation)と類連結語(colligation)を確認する。様々な語法(phraseologies)の分析を通して、多くの言語が定型文で構成されているという主張を見ることができる。これらの句は、一見、身近な言葉から意味を獲得することから、同様に教室内でも単語だけではなく、一般的な言葉と組み合わせ、文脈で紹介すべきと考えられる。

Most **EFL** teachers in Japan find that there are groups of verbs which consistently cause problems for their students. One such group is *LOOK, SEE*, and *WATCH*. While the three verbs obviously represent very similar actions, their near synonymy is often a bane to EFL students. I attempted to help my own students by accessing my native speaker intuition to provide some simple definitions to differentiate the literal uses of these verbs. However, it was soon apparent that my makeshift definitions were of little use to my students, as they related only to specific senses of each word. In actual language use, there are a deluge of other senses, metaphors, and idiomatic expressions employing these same words. A quick glance at a Longman corpus-based dictionary provides 50 varied entries for *look*, over 60 for *see*, and 17 for *watch* (Bullon, 2003). Some examples are:

- *Do you see what I am saying?*
- *Been there, seen that.*

- *He looked tired.*
- *I'm looking forward to the concert.*
- *Watch out!*
- *Can you watch the kids tonight?*

My inquiry led me to corpus linguistics, the study of language structure and use employing large collections of authentic texts stored on computer databases for quick analysis. John Sinclair, one of the pioneers of computer-based corpus linguistics, maintains that language is understood through the lexico-grammatical patterns found in real world use. Sinclair (1991) argues that native speaker intuition is unreliable because it lacks the "proof" of language use. He states that "actual usage plays a very minor role in one's consciousness of language" and that our intuition is "largely ideas about language rather than facts on it" (p. 39). In particular, native speaker intuition can be considered unreliable in judgments about collocations, frequency, pragmatic meaning, and phraseology (Hunston, Laviosa, & Groom, 2005, pp. 109-110). Such claims motivated me to investigate the following questions:

- How can we identify the linguistic "facts" about these three verbs?
- Which senses of these verbs are the most common?
- How should vocabulary items be presented to students?

Considering the shortcomings of native speaker intuition, I decided to take advantage of the "proof" of actual language use available through a 450 million word corpus (HarperCollins Publishers & University of Birmingham, n.d.), examining and comparing the use of the verbs LOOK, SEE, and WATCH. My study focused on identifying common collocations as well as the major lexico-grammatical patterns of each verb.

Although LOOK, SEE, and WATCH are considered near synonyms, a corpus examination of these verbs reveals that they have much less in common than would be expected. Hunston

(2002) states that "the meaning of a word is closely associated with its co-text" (p. 46). This study exemplifies how the various senses identified from concordance lines are in fact defined by the items that the verbs collocate with.

After reviewing some basic terms and theories relevant to corpus studies, I identify the major unique collocates of each verb as revealed by the corpus. From these specific pairings, I then classify the phraseologies that appear in the concordance line results. Finally, based on my analysis, I raise the broader implications for the teaching of vocabulary in the classroom.

What the Corpus Can Show Us

An electronic corpus provides quick and detailed statistics of words, based on massive volumes of annotated authentic texts. In particular, this study took advantage of the particular strengths of a text corpus in identifying frequency, collocation and colligation, as well as phraseology.

Frequency is a calculation of how many particular word types, or total differing word-forms appear in the corpus. *Collocation* in a corpus study refers to a statistical measurement of how often particular lexical items occur with each other. Conversely, *colligation* refers to a strong statistical connection between a lexical and grammatical item. In this paper, I identify such constructions and the unique meanings that the pairings create in relation to LOOK, SEE, and WATCH.

In dealing with collocation and colligation, it is appropriate to consider two models of language interpretation: the open-choice principle and the idiom principle. According to Sinclair (1991), the two theories are meant to support each other, one answering questions about meanings in language that the other cannot as "no single principle has been advanced which accounts for the evidence in a satisfactory way" (p. 109). The open-choice principle falls more in line with the traditional understanding of language construction,

with language as a series of slots that may be filled with almost any word, phrase, or clause, limited only by grammatical correctness. The idiom principle suggests that language is available as preset phrases to choose from in creating meaning. Most language theories consider the open-choice principle to be the primary model of construction, with fixed phrases being the exception. However, Sinclair argues that the opposite is true, claiming that the majority of actual language use is not so flexible.

Phraseology deals with the patterns that particular words find themselves produced in, and emphasizes the affiliation between pattern and meaning, identifiable in concordance lines. Colligation comes into play here, as grammatical items that frequently appear with lexical items can be said to form patterns of phraseology. Hunston (2001) also argues this as she demonstrates "the extent to which the phraseology of the text is the outcome of colligation" (p.18). I show examples of the idiom principle at work in phraseology when we look at the more common phrases that emerge from this study below.

Method

For this study I utilized the Bank of English (BoE), a 450 million word general corpus comprised of 20 smaller specialized corpora (HarperCollins Publishers & University of Birmingham, n.d.). This study looks at the 50 top collocates for LOOK, SEE, and WATCH, ranked by t-scores. T-scores rank the strength of a collocation based on the calculated frequency between the target word, or node, and potential collocates.

As the verbs in this particular study have four to five separate word forms each, the ability to search for the headword, or lemma, in the BoE is invaluable in this case. I set the default search span of up four words before and after the node word, as Sinclair argues that a 4:4 span is most appropriate for collocational studies (Sinclair, 1991, p. 175).

The BoE holds 1,115,929 total matching lines for these three lemma combined. It is quite interesting to notice the disproportionate ratio of roughly 9:13:2 for LOOK, SEE, and WATCH, respectively in each survey (see Table 1). Although a collocate with a t-score of 2.0 or higher is considered significant (Hunston, 2002), given the frequency with which these three verbs appear in the BoE, the lists used here consist of items with t-scores much higher than 2.0.

Table 1. Comparative Corpus Line Results for LOOK, SEE, and WATCH

Corpus	LOOK	SEE	WATCH
All Bank of English	408,069	614,804	93,056
US and UK Books	96,193	152,969	19,895
US and UK Spoken	28,241	53,040	3,109

Although my initial analysis was based on results from the entire BoE, in order to offer a more delimited scope of findings, two additional sub-corpora analyses were done for each verb using (a) the UK and US books corpora (76, 859 lines), and (b) the UK and US spoken corpora (22,283 lines). Comparative results allow us to narrow the focus to the most common unique collocates of each word that appear in all three studies. In order to identify some ways these three verbs differ in use, this paper examines the major collocates that are (a) unique to each verb, and (b) shared across the findings from all three corpora studies.

Results

Major Collocations and Colligations of LOOK

Table 2 shows the most frequent positions of the top five lexical collocates that are shared across the three corpus searches for LOOK: *forward, after, good, around*, and *now*. Given that these collocates are not shared with the other verbs, and given their high t-scores, these collocations are indicative of certain uses of LOOK that set it apart from those of SEE and WATCH.

Table 2. Top Lexical Collocates of LOOK

collocate	most frequent position	frequency at this position	t-score at this position
forward	N+1	12,785	112.55
after	N+1	13,868	112.50
good	N+1	4,257	60.09
around	N+1	3,782	58.18
now	N-1	3,387	48.45

Although any given collocate may appear in multiple positions in relation to the node word, I have only listed the collocational position with the highest t-score here. Interestingly enough, within a span of 4:4 from the node, *forward, after,* and *around* appear in only the N+1 position, directly after LOOK, with *now* only showing up in N-1, or directly before LOOK (*good* appears in two separate positions). This positional exclusivity allows us to make firm judgments about the lexical items: *LOOK forward, LOOK after, LOOK good, LOOK around*, and *now LOOK*.

Table 3 identifies the unique grammatical collocates, or colligates, of LOOK. Despite the high t-scores, these grammatical words tend to be words that are common colligates of many other words. As such, there are comparatively fewer grammatical colligates of LOOK that are not shared with SEE and WATCH.

Table 3. Top Shared Grammatical Colligates of LOOK

collocate	most frequent position	frequency at this position	t-score at this position
at	N+1	86,521	286.93
for	N+1	40,439	183.12
into	N+1	5,016	62.230

There are a few items that are common across the three corpora studies: *at, for* and *into*. Within the top collocates of LOOK we find three prepositions that all occur after the node word within a 4:4 span. Each is most common in the N+1 position, resulting in the phrasal verbs *LOOK at, LOOK for,* and *LOOK into*.

Major Collocations and Colligations of SEE

The most common of the three verbs in this study is SEE. Comparatively speaking, SEE represents the greatest number of matching lines in all three corpus studies, and correspondingly, the largest number of dictionary entries. Given this, within the top collocates here, it is not unexpected that we find lower t-scores (as compared to the studies of LOOK and WATCH), as well as items that are more difficult to place in common groupings.

Table 4. Top Shared Lexical Collocates of SEE

collocate	most frequent position	frequency at this position	t-score at this position
never	N-1	9,103	91.86
ever	N-1	6,120	75.77
coming	N+2	957	27.27

As shown in Table 4, only three lexical collocates are common across all three corpus studies for SEE: *never, ever*, and *coming*. However, frequency data and high t-scores show us that these are strong collocates. *Never* and *ever* both take position in front of the node word, with *ever* being unique to the N-1 position. *Coming* appears in multiple positions, but only after the node.

The common colligates of SEE, as shown in Table 5, tend to be much more varied in their positioning as most of these items are found in multiple locations around the node. With this variation and the large number of corpus results in general, each colligate has a comparatively lower t-score compared to the other two verbs' top colligates.

Table 5. Top Shared Grammatical Colligates of SEE

collocate	most frequent position	frequency at this position	t-score at this position
can	N-1	17,355	122.08
could	N-1	13,020	107.15
how	N+1	13,434	110.93
d	N-1	3,079	47.56
don('t)	N-2	6,050	71.20
couldn('t)	N-2	1,948	42.22

We immediately see that many of the top colligates can be identified with the common trait of modality of ability or permission: *can, could*, and *couldn('t)*. A quick look at representative concordance lines shows that the ''d contraction is also a modal collocate representing items such as *I'd (*I would*), he'd (*he could*)*, and *they'd (*they should*)*.

Major Collocations and Colligations of WATCH

With comparatively lower frequency across all three of the corpus analyses, it can be argued that WATCH is a much more specialized, and a more meaning-heavy, or lexicalized, word than LOOK or SEE. We can see from the listing of collocates in Table 6 that, despite the more restricted uses of WATCH, the t-scores are comparatively lower than those for top collocates of LOOK and SEE. This is due to the multiplicity of collocates that, while varied, can be placed into fewer common groupings.

Table 6. Top Shared Lexical Collocates of WATCH

collocate	most frequent position	frequency at this position	t-score at this position
television	N+1	1211	34.50
closely	N-1	417	20.25
play	N+2	596	23.10
sit	N-2	330	17.80
go	N+2	428	17.23
film	N+2	334	17.24

Beyond this list, many of the top collocates of WATCH identify varied terms for the same noun, such as *television, telly*, and *tv*. In actuality, if we counted all of these terms synonymously,

we would have a much higher t-score for this item. We also find many related items such as *football, programmes, news, films, videos,* and *movies,* all appearing directly after the node. All of these collocates are used in a similar way, identifying what is arguably one of the major contemporary senses of WATCH.

WATCH is represented by a large number of unique lexical collocates but has very few unique grammatical colligates. This statistic tells us something about the restricted ways in which WATCH is used. The top colligates of WATCH include items such as *the, my,* and *his.* Being common collocates to many words in general, most of these items tell us little about the specific senses of WATCH. However, WATCH does share a single unique colligate across all three corpus studies: *while. While* most frequently appears in the N-1 position with a t-score of 18.95 and exemplified in 466 concordance lines as *while watching.*

Identifying Major Phraseologies of the Three Verbs

The lexical items produced by the major collocates and colligates of LOOK, SEE, and WATCH create units of meaning that differ from the literal meanings of the separate items that form them. These phraseologies help identify unique senses of each of the three verbs. The key concept is that we define the meanings of these items based on the words they collocate with and the new patterns that these collocations create. Hunston (2002) similarly states, "for the most part the meanings of words are distinguished by the patterns or phraseologies in which they typically occur" (p. 46).

I have chosen to look at the common patterns identified by one of the top lexical collocates and one of the top grammatical colligates for each of the three verbs, listed in Table 7. By analyzing concordance line results for each pattern, I have determined the major phraseologies that each of these patterns define. For

each separate sense, I have also included a concordance line example from the Bank of English as a reference. In this part of my study, I found that even within particular phraseologies, there are still distinct and restricted *senses* of the newly-formed item, which supports the argument of the idiom principle.

Table 7. Major Phraseologies of LOOK, SEE, and WATCH

collocates	colligates
LOOK forward	LOOK into
never SEE	can SEE
closely WATCH	while WATCH

Major Phraseologies of LOOK

My study identified *forward* as a major collocate and *into* as a major colligate of LOOK. Both of these words appeared exclusively in the N+1 position, which tells us that these are both very common and uniquely defined uses. A study of random concordance lines shows us that the fixed expression *LOOK forward* is actually representative of a number of similar, yet perceptibly differing senses. The four main senses, each shown with a concordance line example, are:

1. To look ahead of one's physical location (literal meaning): *a crook of the road, and **looking forward** again, beheld the figure of a man,*

2. To contemplate future events or outcomes (but not necessarily hoping, acting, or expecting a positive outcome): *you now have in place, if you **look forward** 15 to 20 years, do you believe*

3. To hope for and/or act towards positive future outcomes: *the Guardian, is very much **looking forward** and stressing the positive. As*

4. To anticipate or expect a positive future event or outcome:
   *than its rivals, the PUP **looks forward** confidently to electoral victory,*

Furthermore, the following senses were found to match the phrase *LOOK into*.

1. To peer into a physical object using one's eyes (literal meaning):
   *of your favourite bands is on. You **look into** the view finder, see the #*

2. To think about, try to understand, or attempt to realize someone's feelings, way of thinking, or position:
   *abroad. Let him go to Poland. Let him **look into** the eyes of the Polish people*

3. To contemplate or attempt to predict future events or outcomes:
   *Today we are attempting to **look into** the uncertain future of the*

4. To investigate or research an issue extensively:
   *up an independent judicial inquiry to **look into** allegations that the British*

Again, as with *LOOK forward*, we can see representations of a literal meaning, and three separate, more figurative meanings.

Major Phraseologies of SEE

As described earlier, I determined the major collocates and colligates of SEE, identifying the phrases *never see* and *can see*. Here, using concordance lines, we can recognize the differing senses of these phrases, starting first with *never see*:

1. Not having the experience of seeing (literal meaning) someone or something :
   *can draw Bronze Age houses he has **never seen** and talk to the animals. In*

2. Not having the experience of visiting a place
   *akluyt, a supremely industrious man, **never saw** America. Nevertheless, his*

3. To not have been witness to or to not have realized something or act:
   *reminded me of a correlation I've **never seen** mentioned in any article on*

4. Not being of a group that is commonly exposed to or witness to certain conditions:
   *the Earth's, so some crater floors **never see** sunlight. Although the recent*

Similarly, here are the identifiable senses of the colligate *can SEE*, as defined by concordance lines:

1. The ability to visually perceive something (literal meaning):
   *ATM and anyone waiting, so no one **can see** your PIN number or you*

2. The ability to visit someone in person:
   *of the billed amount. You **can see** any doctor or go to any hospital you*

3. The ability to comprehend or understand:
   *traumagenic dynamics. Using it we **can see** the homologous relationship of*

4. To realize something about someone or something:
   *may be quite simple. <p> We **can see** the deep roots of marital conflict*

Perhaps expectedly, these colligates of SEE that have a much higher frequency and t-score provide us with a greater number of variant senses using the same phraseology.

Major Phraseologies of WATCH

As previously argued, the comparatively lower frequency of WATCH in the corpus provides us with a much less varied, more specialized, and more lexicalized, set of collocates. There was only one identifiable sense of *closely WATCH*:

1. To maintain observation of a process, group, person, or specific action, with an anticipation that there may be a problem or a negative outcome:
    *an American journalist who has been **closely watching** the border story. The*

There are no literal senses of observing a process, group, person, or action from a close physical proximity here. All of the concordance examples found were understood from their metaphorical sense.

Finally, some senses of the pairing *while WATCH* are as follows:

1. To simultaneously perform other activities while watching an activity, event, TV program, or other action in progress (literal meaning):
    *doing homework and eating **while watching** TV. There is research to suggest*
2. To be witness to, or to realize a consistency or change in a state of affairs:
    *fight to adopt Christabel **while watching** the adoption system make error after*

Both of the senses here refer to performing (or not performing) an action while simultaneously witnessing or being aware of some sort of action in process, either in a literal or metaphorical sense. It is not surprising to find fewer differentiated senses of *while WATCH,* as I have argued that the uses of WATCH in general are much more restricted than LOOK or SEE.

In examining these major phraseologies and the varied senses that each has shown, it is now much easier to see how it is the linguistic behavior of the patterned items that define them. They create new meanings that, in many instances, have little resemblance to the semantic meanings of the individual items they are comprised of.

Discussion

Having looked at these three verbs more in depth, it is now apparent that corpus studies have much insight to offer language learners. I will discuss three of the larger implications here.

The ways in which these verbs collocate with items unique to each of them brings us back to Sinclair's idiom principle. At least from the few examples that I have identified, with certain collocations being preferred among each of the verbs, the idiom principle does in fact seem to be the default model of language construction. I have also established that some items can be much more lexicalized than others, as WATCH in this study features far fewer unique collocates, making it less "collocationally flexible" than LOOK or SEE. Understanding language in these terms helps distinguish the ways in which these verbs are not as synonymously interchangeable as language users may be inclined to believe. As such, it would seem pertinent that teachers present words along with their co-text, to provide notions of meaning through use.

As is indentified in the phraseologies, collocational and colligational pairings often create new and unique meanings, such as *LOOK forward*, *never SEE*, and *while WATCH*. While some of these word pairings may be understood either literally or figuratively, I have found here that the metaphorical senses are often more common. In each of these instances, identifying the appropriate sense of a word is dependent upon observing its

phraseology within context. Given that words acquire meaning in this way, in the classroom it would seem rather inadequate to teach words in isolation, in word lists or on flash cards, where they are separated from their common pairings and removed from any context of use.

In my initial attempt to help my students better understand the uses of LOOK, SEE, and WATCH, I relied on my native speaker intuition to provide simple definitions for these three verbs. Although these simple definitions were not entirely incorrect, my intuition alone was ultimately ineffective at helping my students grasp the multitude of senses and uses that each of these verbs retain. It is quite apparent that the corpus can be invaluable tool for language teachers in preparing their students for real world language use by providing statistics and authentic examples in context.

Conclusion

Although the scope of this exploratory study was limited by the sheer frequency of the words it set out to analyze, it succeeded in illustrating how the meanings of words are indeed defined by their co-text. By analyzing the proof of language use found in modern computerized corpora, it is obvious that EFL teachers could be better at enabling their students by presenting language in terms of the company that words keep and the varied meanings that these associations cultivate.

Bio Data

Gregory C. Anthony is currently a lecturer at Hachinohe University in Aomori Prefecture. He has taught English in Japan since 1996 and his research interests include metaphor, computer-assisted language learning, and teaching English to young learners.

References

Bullon, S. (Ed.). (2003). *Longman dictionary of contemporary English: The living dictionary*. (3rd ed.). Harlow, UK: Pearson Education Limited.

HarperCollins Publishers, & University of Birmingham. (n.d.). Bank of English. Retrieved from http://www.titania.bham.ac.uk

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. (2001). Colligation, lexis, pattern, and text. In M. Scott & G. Thompson (Eds.), *Patterns of text: In honour of Michael Hoey* (pp. 13-34). Amsterdam: Benjamins.

Hunston, S., Laviosa, S., & Groom, N. (2005). *Corpus linguistics*. Birmingham, UK: Centre for English Language Studies at the University of Birmingham.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.