Using Corpora to Help Teach Difficult-to-distinguish English Words

Dilin Liu

(University of Alabama, USA)


L2 synonyms (and other related words) that are undifferentiated (i.e., are translated as the same word) in learners’ L1 are among the most difficult lexical items for L2 learners to grasp. How to help students learn the appropriate use of these words has been a great challenge for L2 teachers. Drawing on recent research in corpus linguistics and corpus-based language teaching, this paper strives to show that corpus analysis can effectively differentiate synonyms and help L2 students successfully learn the appropriate use of the words in a synonym set. The paper begins with a brief overview of research on synonyms and their learning/use. Then using four sets of English synonyms/related words that are not differentiated in Korean, it discusses, with specific examples of learning activities and exercises, how corpus analysis and its results can be used to help learners differentiate these difficult words and grasp their usage patterns. The paper concludes with a summary of both the benefits and challenges of such a teaching approach and some strategies for dealing with the challenges.

**Key words:** corpora, corpus-based language teaching, synonyms, corpus analysis

1. INTRODUCTION

While learning vocabulary in an L2 is generally very challenging, some words are especially difficult, such as synonyms/near-synonyms1 (Hatch & Brown, 1995; Laufer, 1991; Martin, 1984; Nation, 2001). Synonyms are difficult because while they express essentially the same meaning, they do so in different manners, for different contexts, and/or

---

1 According to many linguists, true synonyms are extremely rare unless cross-dialect synonyms are considered (Edmonds & Hirst, 2002; Stubs, 2001; Taylor, 2003). In other words, all synonyms are really near-synonyms. For the sake of simplicity, however, we will use the only term “synonym” in the rest of the paper.
from different perspectives, i.e., they are often not entirely interchangeable (Edmonds & Hirst, 2002; Hatch & Brown; Liu, 2010a, 2013; Liu & Espino, 2012). Added to the difficulty is the fact that often some L2 synonyms are not lexicalized or differentiated in learners’ L1 (i.e., they are translated as the same word in their L1). For example, the English synonyms in each of the following two sets are often translated into the same word in Korean (also the same word in some other languages such as Chinese and Japanese): demand/request/require > 요구하다, and adverbs incorrectly/wrongly (also erroneously/mistakenly) > 틀리게. Besides synonyms, there are some other related-words in a target language that may also be undifferentiated in form in learners’ L1, e.g., both borrow and lend are translated as 대여 or 빌리다 in Korean, and both doubt and suspect are translated as 의심. Due to L1 interference/transfer, L2 learners often have difficulty correctly using these closely related words (Hatch & Brown, 1995; Jiang, 2000; Lee & Liu, 2009; Liu, 2011). Thus, how to help learners effectively grasp the use of these words has been an important but challenging task for L2 teachers. Recent research in corpus linguistics and corpus-based language teaching has, however, begun to show that corpus analysis can effectively differentiate synonyms and may help L2 learners better grasp them (Lee & Liu, 2009; Liu, 2010a, 2013; Liu & Espino, 2012; Stubbs, 2001; Tsui, 2004; Yeh, Liou, & Li, 2007). It is the purpose of this paper to discuss, based on research including publications by the author, how to use corpora to assist L2 students in grasping difficult-to-distinguish synonyms and other related words. The paper will begin with a brief overview of research (especially corpus-based studies) on synonyms and their use. Then it discusses, with specific examples of learning activities and exercises, how corpus analysis and its results can be used to help learners differentiate closely-related words and grasp their usage patterns. It concludes with a summary of both the benefits and challenges of such a teaching practice and some strategies for dealing with the challenges.

2. CORPUS ANALYSIS AS AN EFFECTIVE TOOL FOR SYNONYM STUDY: AN OVERVIEW

2.1. Synonymy and Corpus Analysis

Synonyms may differ in “any aspect of their meaning,” including “denotational,” “expressive,” and “structural” aspects (Edmonds & Hirst, 2002, p.109). The differences are often subtle and difficult to discern. Consequently, there was not much research progress made in the study of synonymy until the 1980s when experimental and corpus-based studies began yielding empirical evidence for the Firth/Halliday/Sinclair lexical semantic theory that the meaning of a word is largely shaped by its collocates (Firth, 1957; Halliday,
Using Corpora to Help Teach Difficult-to-distinguish English Words

1966, Sinclair, 1966, 1991), a theory arguably best articulated by Firth’s (1957, p. 11) now famous quote: we “know the word by the company it keeps.”

One of the first synonym studies focusing on the contextual correlates of words was Miller and Charles (1991). In this study, the researchers had native English speakers rate the degree of similarity and substitutability between words in pairs embedded in sentences (contexts) taken from corpus data. Their results showed that “the more often two words can be substituted into the same contexts the more similar in meaning they are judged to be” (Miller & Charles, 1991, p. 1). Later, Church, Gale, Hanks, Hindle, and Moon (1994) also did a “lexical substitutability” study, which involved the synonyms ask for, request, and demand in corpus data, and their study produced essentially the same finding that the “textual substitutability” of multiple lexical items is a strong barometer of their semantic similarity (p. 169).

These early studies have later led to the development of an effective corpus-based approach to the study of synonyms and lexical semantics in general called the “behavioral profile” (BP) approach (Hanks, 1996). This approach examines the meanings and usage patterns of lexical items by concentrating on the distributional patterns of words, particularly their collocates. In arguably the first corpus-based BP study, Hanks (1996), using the British National Corpus (BNC) data, produced the BPs of several verbs including urge, incite, bother, and abandon. The BPs were based on the syntactical collocational patterns of these verbs, especially their subjects, objects, and modifying adverbs, and the frequencies with which the various patterns were used. Such BP information helped clearly define each verb’s primary and secondary meanings and differentiate it from its synonyms. In other words, the BP analysis helps provide valuable information that we may not be able to obtain otherwise. Since Hanks’ study, there have been quite a few sophisticated BP studies of synonymous verbs (e.g., Divjak & Gries, 2006) and adjectives (Gries & Otani, 2010; Liu, 2010a), as well as a couple on nouns (Janda & Solovyev, 2009; Liu, 2013) and one on adverbs (Liu & Espino, 2012).

While corpus-based studies on synonymous verbs concentrate largely on their subjects, objects, and adverbs, those on synonymous adjectives (Justeson & Katz, 1995; Liu, 2010a) focus mainly on the nouns they typically modify. This is because the types of nouns the adjectives typically modify can most effectively reveal the meanings of the adjectives, as evidenced Liu’s (2010a) study on the chief/main/major/primary/principal synonym set. By examining the types of nouns that the five adjectives frequently modify, Liu (2010a) was able to identify the major semantic and usage differences among the five adjectives. Although all of the five adjectives may modify abstract nouns (e.g., chief/main/major/primary/principal concerns/goals), they each also have unique types of nouns they modify. For example, while main is the only one that often modifies concrete nouns (e.g., main dish/gate/lobby), chief is employed mainly to modify position titles of power (chief
executive/justice/operating officer), and only primary is used to modify nouns that are first in order of a series (e.g., primary care/election/school). In order to determine and compare the degrees of importance the adjectives carry, Author also examined the frequencies the adjectives were used with the definite/indefinite (a/the) and singular/plural nouns. The examination reveals that main conveys the highest importance, while major expresses the lowest. Other important collocates for studying synonymous adjectives include the typical adverbs that modify a given adjective.

While the corpus-based studies on synonymous adjectives have focused mainly on the typical nouns adjectives modify, those on synonymous nouns (e.g., Liu, 2013) have concentrated on the typical modifiers (adjectives and infinitives) of the synonymous nouns in a set because the semantic types of the modifiers of nouns may help best uncover the semantic and usage differences among the synonymous nouns. For example, by examining the typical modifiers of two sets of synonymous nouns authority/power/right and duty/obligation/responsibility, Liu (2013) has successfully determined the semantic/usage differences among the nouns in each set. For example, authority/power/right differ in the source and nature of the power/right they typically refer to. In terms of source of power/right, although “law/constitution” may be a major source for all three nouns, “office” (an official position/institution) is a unique source for only authority/power (e.g., the authority/power to appoint/arrest/veto) whereas “birth” (natural prerogative) is the key source for right (e.g., the right to live/speak/vote). Concerning the nature of power/right, while authority and power are primarily “official,” right is essentially individual. As for the differences among duty/obligation/responsibility, while all three nouns can refer to things that one needs to do based on one’s job, membership in a community, and/or established morality, duty and obligation are much stronger than responsibility in terms of the expressed degree of commitment/dedication/necessity. More importantly, duty is sometimes used to refer to an undesirable thing that a person has to do based on his/her religious beliefs and/or assignment/job, etc. (e.g., sad duty and military/religious duty to kill), whereas obligation mostly means something one must do due to a mutual agreement and commitment.

In corpus-based research on synonymous adverbs, the focus has been on the verbs and the adjectives that the synonymous adverbs in a set each typically modify, as well as the positions in a sentence the adverbs each usually appear in. This is because while the verbs and adjectives that adverbs modify often provide valuable information about the meanings and usage patterns of adverbs, the sentential positions they typically appear in are also very helpful information for differentiating adverbs due to the fact that adverbs may appear in different positions in a sentence and that the different positions an adverb takes may result in different meanings, as can be seen in the following two examples:
Using Corpora to Help Teach Difficult-to-distinguish English Words

31

a. *Strangely,* he talked a lot at the meeting.
b. He talked *strangely* at the meeting.

Because of the sentential position variation of *strangely,* the meanings of the two utterances differ: sentence A means that it was strange that "he" talked a lot while B means "he" talked in a strange fashion. Liu and Espino's (2012) study on *actually/genuinely/really/truly* demonstrates that it is an effective approach in studying synonymous adverbs to focus on their verb/adjective collocates and sentential positions. The results of the study reveal that whereas all of the four adverbs may express/emphasize the meaning of reality/truth, they differ noticeably in usage. *Actually* is typically used to highlight a fact by contradicting what has been said or believed to be and it is often used clause-initially (e.g., "He said he didn’t know it; actually, he did know it"). In contrast, *really,* the most frequently used adverb in the set, is used mainly as an emphasis/intensifier (e.g., "She is really good"). *Genuinely/truly,* the most infrequent ones in the set, are used mainly to modify adjectives and verbs of attitude/emotion to help stress the truthfulness of the attitude/emotion in question (e.g., "Tom *genuinely/truly* loves Mary").

As we conclude the discussion in this section, it is imperative to note that while corpus analyses can help us effectively differentiate synonyms, the task is often a complex one. The person who does the corpus query and analysis will often have to try not only different types of information to examine but also different ways/perspectives at looking at the query results. As Hanks (1996, p. 96) states,

But the [usage] patterns do not spring, untouched by human hand, fully fledged from the corpus. They have to be teased out, often painstakingly and slowly. Procedures have to be developed for distinguishing relevant features from noise. Appropriate levels of generalization have to be chosen at every step.

In short, querying and analyzing corpus data for language usage patterns is a complex and challenging endeavor.

2.2. L2 Synonym Learning, Use, and Teaching

All of the aforementioned corpus-based studies are related to L1 or native-language speakers' use of synonyms. Few studies have examined the use and learning of synonyms in L2. Lee and Liu (2009), Liu and Zhong (2012), Tsui (2004), and Yeh et al. (2007) appear to have been the only ones. Using both corpus and elicited data (fill-in missing words with the right synonyms), Liu and Zhong (2012) investigated intermediate and
advanced Chinese EFL learners’ use of four sets of synonyms against native English speakers’ use of them (authority/power/right; duty/obligation/responsibility, actually/genuinely/really/truly, and erroneously/incorrectly/mistakenly/wrongly). The results of both the corpus and elicited data analyses show that EFL learners, even advanced ones, have serious difficulty knowing which synonym to use in a given context. More importantly, it is found that not knowing the right choice in such a context essentially means not knowing the typical collocates of a given synonym. For example, many Chinese EFL learners did not know that native English speakers typically say fatherly/motherly/religious/sad duty (rather than fatherly/motherly/sad obligation/responsibility). This finding highlights the importance of learning the typical collocates of a word in L2 synonym learning.

Using corpus-search generated findings (in the form of lists or concordance lines), Lee and Liu’s (2009) and Yeh et al.‘s (2007) studies each investigated the usefulness of having EFL learners focus on the typical collocates of synonyms in understanding/learning the different usage patterns of the target synonyms. They gave their students tests and questionnaires to ascertain the effectiveness of such a learning approach. The results of both studies show that such an approach helped students successfully grasp the target synonyms and that students found the approach very useful. Tsui (2004) showed how corpus analysis helped nonnative English-speaker teachers in Hong Kong effectively differentiate synonyms such as high/tall and learn idiomatic lexical collocations. In fact, there have also been a few other studies showing the effectiveness of using concordances in helping EFL learners grasp word collocations in general (not just synonyms), including verb + noun and verb + preposition collocations (Chan & Liou, 2005; Sun & Wang, 2003; Liu, 2010b). The effectiveness of corpus-based/driven language learning also results from the discovering learning opportunities and the ample language input it provides to learners (Aston, 2001; Liu, 2010b, 2011; Liu & Jiang, 2009).

Before we move onto the next section on how to use corpus data in learning/teaching synonyms and other closely-related words, it is imperative to note that identifying the collocation information that differentiates synonyms is a challenging and sometimes labor-intensive task. Furthermore, sometimes, the analysis of the immediate collocates of synonyms may not yield the discerning information needed; in such a case, it requires a scrutiny of larger contextual information to truly understand the semantic differences among some synonyms because their immediate collocates do not reveal much about their differences. Let us look at the case of surprise and amaze, two synonymous verbs that are often translated into the same word in Korean. While both verbs share the meaning of causing someone to feel surprised (e.g., “Tom has been known as a quiet and obedient employee; his vocal criticism of his boss has surprised/amazed everyone in the company”), they differ in that amaze may often conveys the sense of wonder/amazement, e.g., “The
Using Corpora to Help Teach Difficult-to-distinguish English Words

figure-skater’s spectacular high flying triple axel jumps amazed not only the audience but also the judges.” The verb surprise would not be appropriate in the latter sentence because it does not possess the sense of wonder/amazement. However, this subtle difference may be difficult for Korean (and some other Asian) EFL learners due to their L1 interference. For example, several of my former students from Japan, South Korea, and Taiwan wrote me something like the following in their post-graduation thank-you emails: “We’re all surprised you were able to answer all of our questions.” Though the statement was meant as a compliment, it actually was not. With the word “surprised,” the statement actually meant that the writers all had thought I was an incompetent teacher and then was surprised to find out that I was able to answer their questions.

The two verbs could not be differentiated by an examination of their typical (subject/object) collocates because these collocates are essentially identical for both verbs with many being pronouns (e.g., It/you/he/she surprised/amazed me/him/her/us). Of course, there is some collocation information that may help differentiate surprise and amaze to a certain degree, but such information requires close analyses to obtain. For example, a scrutiny of the nouns after amazed/surprised at will show that amazed at is sometimes followed by very positive noun phrases, such as achievement/creativity, showing again that amaze may express the sense of wonder/amazement. Also a look at the typical adverbs that modify the two verbs will show that surprise is frequently modified by the positive adverb pleasantly but almost never modified by negative adverbs like unpleasantly, a fact that suggests that the act of surprise is typically viewed either negatively or in natural fashion. Otherwise, there would not be the need to use pleasantly to distinguish a specific surprise act from the other mostly negative or neutral ones. These two pieces of discerning information are not easy to find and are not sufficient to definitively differentiate the two verbs. Thus, we may need to examine the larger discourse context of each token of the two verbs in use by reading the entire sentence and sometimes the previous and following sentences. This way, we will find that, unlike surprise, amaze is quite frequently used in the positive sense of wonder/amazement.

3. TEACHING DIFFICULT-TO-DIFFERENTIATE WORDS WITH CORPUS DATA/ANALYSIS: SOME EXAMPLES

3.1. What Corpus and Query/Analysis Procedures to Use and Some General Principles

To discuss how to use corpus-based activities to teach difficult-to-distinguish synonyms and other closely-related words, we must first address some basic issues, such as which
corpus/corpora to use, what corpus query procedures/techniques to employ, and how to conduct corpus analysis. In terms of which corpus to use, given our need to identify the typical usage patterns of English synonyms, the corpus we use has to be large so as to ensure the reliability and validity of the data. This is because, when other variables are constant, the larger a corpus is, the more representative/reliable its data will be. Currently, the free online 450 million-word Corpus of Contemporary American English (COCA) provided by Mark Davies of Brigham Young University is an excellent choice. Besides being large and freely accessible, COCA boasts systematically-selected data and consists of five sub-corpora that represent most of the major registers/genres including speaking, fiction, newspaper, magazine, and academic writing. Furthermore, COCA is equipped with a powerful and user-friendly search engine that offers a variety of useful query functions which can yield various types of meaningful information. For example, one can effectively and efficiently query for collocates by part of speech using the “part of speech” searching codes it provides, e.g., querying for all the nouns after the adjectives before the noun authority (for information about the query codes and functions, read the introduction information on the COCA interface/webpage). Also and importantly, a query of the collocates of a word/structure may generate not only the frequency but also the Mutual Information (MI) score for each collocate. MI (developed/introduced by Church & Hanks, 1990) is a statistical procedure that determines how likely two words may co-occur by comparing "the probability of observing x [word] and y [word] together (the joint probability) with the probabilities of observing x and y independently (chance)" (p. 23). The ability of COCA to generate both frequency and MI information for lexical collocates is especially helpful for researching and teaching synonymous and other closely-related words. Specific examples of collocation query and analysis procedures will be provided below.

Given the complexity and difficulty involved in corpus query and analysis of the usage patterns of difficult-to-distinguish words as has been shown above, it is necessary to note, before we proceed further, that teachers should usually do a corpus query/analysis of the semantic/usage patterns of the synonyms they plan to teach before class, so they can be well prepared. Research (Tsui, 2004) has found that corpus analysis is especially helpful for teachers in helping them become prepared for teaching difficult language usage issues. Another important point to bear in mind is that generally we should not engage low/intermediate-level students in direct corpus queries and analyses, especially not in the initial stage. Instead, teachers can use the results of the corpus queries and analyses they did to develop various learning materials/activities to help students more effectively learn the target language usages. Four specific examples are given below to illustrate how corpus analyses and results can be used in teaching difficult-to-distinguish words, including three sets of synonyms (incorrectly/wrongly; important/significant; demand/request/require)
and one set of closely related words (doubt/suspect). These sets are chosen because the words in each set are often difficult for Korean and some other Asian EFL learners to distinguish and learn.

3.2. Teaching Synonymous Adverbs: incorrectly and wrongly

3.2.1. Corpus queries and analyses

We begin with incorrectly/wrongly because, of the words in each of the four sets, these two are the easiest to differentiate. A simple, straightforward query of the verbs they each typically modify will clearly reveal their semantic/usage differences. We do not need to query for their typical adjective collocates because these adverbs do not modify adjectives (i.e., no one says *incorrectly/wrongly good/large). To query for the verb collocates of each adverb, simply type, in the search string space, “incorrectly [vv*]” or “wrongly [vv*]” where [vv*] stands for all the lexical verbs and their forms. Figures 1 and 2 are screenshots of the queries and results for each of the two adverbs. The verbs they each modify are listed in order of the frequency (shown in the frequency column). Because I selected “lemma” instead of “word” for the display of the results, the verbs (also the adverb) are listed in their basic form (infinitive form for verbs), but they each include all of its other tense forms, e.g., [accuse] stands for accuse/accuses/accusing/accused. The “All” column displays the total frequency of the verb (in all its forms) in COCA; the “%” column lists what percentage each “incorrectly + verb” collocation accounts in the total frequency of the verb; the MI column reports the MI scores. The higher an MI is, the stronger the tendency for the words to collocate is. A comparison of the results between the two adverbs shows clearly that the verbs they each typically modify differ noticeably. While both are used to modify identified/assumed, wrongly is the only one that modifies accused/convicted/imprisoned, verbs that deal with law or justice. In fact, the unusual strong association of wrongly with these law-related verbs is evidenced not only by their high frequencies but also their very high MI scores: the MI scores for wrongly and the three verbs are the highest (all above 11.58), much higher than its MI scores with any of the other verbs. This unique usage of wrongly indicates it carries the meaning of unjustly, unfairly, and/or unethically/immorally, a meaning that incorrectly (as well as all the other synonyms in the set such as erroneously/mistakenly) does not possess.
It is imperative to note that, in many cases, to truly understand the semantic differences between two synonyms, we will need to read and analyze the concordance lines of the result tokens (the actual sentences in which the synonym is used), a practice we will discuss below in due course. To help students understand the identified difference between *wrongly* and *incorrectly*, the teacher can do the following depending on the students’ English proficiency level and some other factors. If the students are at the upper-intermediate level or above and if the classroom is equipped with internet access, then the
teacher can do the queries with the class or have students do guided queries by themselves (if each student has a computer). Then ask students to compare the results and identify the difference, but the teacher should be ready to provide any necessary assistance/guidance. For students of other levels, the teacher can give students the query results (the frequency lists in a printout or on a screen) directly and then do the analysis with them. In addition to working on identifying the semantic/usage differences based on corpus query results, the teacher can and should also use the corpus data and analysis results to develop additional learning materials/activities to help students fully grasp the two adverbs. Three types of exercises are included below as examples of possible learning activities or exercises. All the sentences in the exercises, except for those erroneous ones in the error identification and correction activity, are taken or adapted from COCA. The erroneous ones were actual sentences produced by EFL/ESL speakers. The reason for using sentences from COCA is to ensure the authenticity and the meaningfulness of the teaching materials.

3.2.2. Sample learning activities

(1) **Exercise 1**: Decide whether *incorrectly* or *wrongly* fills in each blank better semantically; write *either* if you believe either adverb works equally well.

a. The United Nations TV had ________ identified Mr. Smith as the ambassador. (*either*)

b. Because of this decision, some schools have been ________ punished. (*wrongly*)

c. It seems clear that whoever takes advantage of their temporary power over a child’s body to perform the operation must be abusing this power and acting _________. (*wrongly*)

d. These English learners pronounced some of the words _________. (*incorrectly*)

e. Some parents have been ________ charged with child abuse. (*wrongly*)

(2) **Exercise 2**: Decide whether *incorrectly*, *wrongly*, or either adverb may replace each underlined word while keeping the original meaning and tone of the sentence.

a. It has been found that some of the prisoners were *unjustly* convicted. (*wrongly*)

b. His mental disease prevented him from understanding that he acted *illegally* in not obeying the law. (*wrongly*)

c. I believe that Marissa Mayer was *inaccurately* identified as “Melissa.” (*either*)

d. As a result, a child may be *mistakenly* diagnosed with the disease. (*either*)

e. The newspaper *unfairly* blamed the lawyer for the problem. (*wrongly*)
Exercise 3: Some of the underlined uses of *incorrectly* and *wrongly* are inappropriate. Identify and correct them. Then translate the sentences into Korean.

a. I *incorrectly* identified Brit Hume as being located at the White House right now. *(correct)*

b. There is a complaint process, and we would encourage her to follow up on that if she thinks she been treated *incorrectly*. *(wrongly)*

c. Because of an editing error, David Hilliard's job title was *wrongly* described in a story yesterday on the Oakland mayoral campaign. *(correct although incorrectly preferable)*

d. By using tax money for personal trips, the governor certainly acted *incorrectly*. *(wrongly)*

e. It's reported that two men in California were *incorrectly* imprisoned for 17 years. *(wrongly)*

The reason for asking students to translate the sentences in Activity 3 is that research (Laufer & Girsi, 2008; Nation, 2001) has shown translation is very helpful for learning lexical items whose meaning and usage in L2 differ noticeably from those in learners' L1, because it helps raise learners' consciousness of the inter-lingual differences in the words being learned.

3.3. Teaching Synonymous Adjectives: *important* and *significant*

3.3.1. Corpus queries and analyses

Compared with *incorrectly/wrongly*, *important* and *significant* are more difficult to differentiate, as will be shown below. To effectively understand the difference between the two adjectives, we can begin by identifying the typical nouns they each modify, a practice that research has shown to be very effective for studies on synonymous adjectives (Justeson & Katz, 1995; Liu, 2010a). A query of “important [nn*]/significant [nn*]” where [nn*] stands for all nouns in both singular and plural forms will quickly generate the information we need. For lack of space, in lieu of screenshots, the results (concerning the top ten most common noun-collocates and their frequencies and MI scores) are summarized in Table 1. It is clear from the results that the nouns that the two adjectives each typically modify differ substantially, for, of the ten top nouns on each adjective’s list, only one (*role*) appears on both lists. An examination of the nouns on each list indicates that while most of the nouns modified by *important* refer to non-relational things, etc. (e.g., *thing/part/issue/step*), most of those modified by *significant* deal with either relationships (e.g., *difference/relationship* between two things and *effect/impact* of one upon another) or
numbers (*number/amount*). However, a close look at the entire lists of the nouns the adjectives each modify indicates that while some of nouns are modified much more frequently by one adjective, they also register a fairly substantial use with the other adjective. For example, while *difference* is modified by *significant* 6,024 times, it is also modified by *important* 584 times. Is there any difference in the meaning between *significant* and *important differences* and the other pairs like this? To answer this question would require us to look at additional collocation information and scrutinize concordance lines.

**TABLE 1**

<table>
<thead>
<tr>
<th>Nouns modified</th>
<th>Frequency</th>
<th>MI</th>
<th>Nouns modified</th>
<th>Frequency</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>thing(s)</td>
<td>6,241</td>
<td>6.57</td>
<td>difference(s)</td>
<td>6,024</td>
<td>9.34</td>
</tr>
<tr>
<td>role(s)*</td>
<td>3,191</td>
<td>7.80</td>
<td>change(s)</td>
<td>1,365</td>
<td>5.67</td>
</tr>
<tr>
<td>part(s)</td>
<td>2,954</td>
<td>6.28</td>
<td>number(s)</td>
<td>1,269</td>
<td>5.90</td>
</tr>
<tr>
<td>issue(s)</td>
<td>2,266</td>
<td>6.31</td>
<td>effect(s)</td>
<td>1,160</td>
<td>6.70</td>
</tr>
<tr>
<td>factor(s)</td>
<td>1,709</td>
<td>7.46</td>
<td>amount(s)</td>
<td>1,053</td>
<td>7.45</td>
</tr>
<tr>
<td>question(s)</td>
<td>1,490</td>
<td>5.45</td>
<td>role(s)</td>
<td>1,005</td>
<td>6.66</td>
</tr>
<tr>
<td>point(s)</td>
<td>1,442</td>
<td>5.20</td>
<td>relationship(s)</td>
<td>819</td>
<td>6.41</td>
</tr>
<tr>
<td>aspect(s)</td>
<td>1,136</td>
<td>7.92</td>
<td>impact(s)</td>
<td>801</td>
<td>7.22</td>
</tr>
<tr>
<td>step(s)</td>
<td>834</td>
<td>5.59</td>
<td>increase(s)</td>
<td>782</td>
<td>6.24</td>
</tr>
<tr>
<td>element(s)</td>
<td>763</td>
<td>7.21</td>
<td>predictor(s)</td>
<td>789</td>
<td>10.67</td>
</tr>
</tbody>
</table>

*Words in italics are on the lists of both adjectives.*

One useful additional type of collocate to examine is the typical adverbs that are used to modify the adjective-noun pairs or simply before the two adjectives. A query of “[r*] important”/“[r*] significant” generates the results summarized in Table 2. On the surface, the results do not appear to be very informative because the top ten adverbs on both lists look fairly similar, e.g., six of them (e.g., *most/more/very*) overlap. Also, most of the adverbs including all of the six overlapping ones are degree or intensifying adverbs. These adverbs can modify any adjectives and do not have other unique meanings. Hence, they cannot help reveal the meanings of the adjectives they modify. However, a closer look indicates that there are two non-degree/intensifying adverbs on the list of *significant* (*statistically/clinically*) that can help differentiate it from *important*. It is particularly worth noting that *statistically* is by far the most frequent modifier of *significant* and also that, of all of the *adverb-important/significant* collocations in the list, the *statistically-significant*
and *clinically-significant* collocations boast the highest MI scores, indicating their strong bond. Furthermore, these two adverbs are uniquely meaningful. When we say something is *statistically/clinically significant*, we mean that the referent is meaningful/important in statistical/clinical sense. So *significant* typically means meaningful/important in a particular sense/context or from a particular perspective, often based on some measurement (i.e., statistics). This explains why *number* and *amount* are among the most frequently modified nouns of *significant*. It also explains why *marginally* is a typical modifier of *significant* (a collocation registering the third highest MI score). This is because something may be statistically *significant* but the statistical significance may be *marginal* (e.g., with a p value of .049 in a situation where a p of < .05 is considered significant). An examination of some of the concordance lines of the query results can help further explain this important difference between the two adjectives.

**TABLE 2**

<table>
<thead>
<tr>
<th>Modifying adverbs</th>
<th>Frequencies</th>
<th>MI</th>
<th>Modifying adverbs</th>
<th>Frequencies</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>most</em></td>
<td>21,257</td>
<td>8.09</td>
<td>statistically</td>
<td>3,622</td>
<td>14.48</td>
</tr>
<tr>
<td><em>more</em></td>
<td>12,775</td>
<td>6.25</td>
<td><em>most</em></td>
<td>2,702</td>
<td>7.56</td>
</tr>
<tr>
<td><em>very</em></td>
<td>11,722</td>
<td>7.30</td>
<td><em>more</em></td>
<td>1,200</td>
<td>5.28</td>
</tr>
<tr>
<td><em>so</em></td>
<td>3,289</td>
<td>4.25</td>
<td><em>very</em></td>
<td>976</td>
<td>6.15</td>
</tr>
<tr>
<td><em>really</em></td>
<td>1,973</td>
<td>5.36</td>
<td><em>highly</em></td>
<td>361</td>
<td>8.55</td>
</tr>
<tr>
<td><em>less</em></td>
<td>1,326</td>
<td>5.68</td>
<td><em>less</em></td>
<td>201</td>
<td>5.40</td>
</tr>
<tr>
<td><em>extremely</em></td>
<td>1,221</td>
<td>8.43</td>
<td><em>clinically</em></td>
<td>170</td>
<td>12.08</td>
</tr>
<tr>
<td><em>equally</em></td>
<td>1,216</td>
<td>8.92</td>
<td><em>particularly</em></td>
<td>166</td>
<td>6.07</td>
</tr>
<tr>
<td><em>especially</em></td>
<td>1,110</td>
<td>6.35</td>
<td><em>marginally</em></td>
<td>126</td>
<td>11.93</td>
</tr>
<tr>
<td><em>particularly</em></td>
<td>1,032</td>
<td>6.86</td>
<td><em>especially</em></td>
<td>118</td>
<td>5.56</td>
</tr>
</tbody>
</table>

*Words in italics are on the lists of both adjectives.*

For example, COCA has quite a few sentences related to 911 or some other phone number as being an *important* number.” These numbers are indeed *important*, but not *significant*, because the importance of these numbers is not based on any sense of measure. In contrast, the significance of a number or a *significant number* has to be based on measurement, especially measurement of proportion, as is shown in the following sentence: “37 percent black . . . live in extreme poverty areas. . . that's a fairly *significant* number.” Therefore, a *significant something* is generally not the same as an *important something* or vice versa. More importantly, there are some things that are *important* in their own right.
while some other things are generally significant in particular terms: “Freedom/food/energy are each important for humans” but “The correlation/difference/effect are often statistically significant.” After explaining the differences between the two adjectives with these corpus analyses/results, we can then develop additional useful exercises like the following to help students reinforce their understanding.

3.3.2. Sample learning activities

(1) **Exercise 1**: Decide whether important or significant fills in each blank better semantically; write either if you believe either adjective works equally well.
   a. What is the most ______ decision you’ve made this year? (important)
   b. His anxiety is at a clinically ______ level. (significant)
   c. A ______ number of Americans still haven’t decided who to vote for. (significant)
   d. Liberty is the most ______ thing for many Americans. (important)
   e. He made ______ contributions to the profession. (either)

(2) **Exercise 2**: Decide whether important, significant, or either adjective may replace each underlined word while keeping the original meaning and tone of the sentence.
   a. Four is a very meaningful number in this culture. (important)
   b. A large number of students failed the test. (significant)
   c. The two methods differ in a very substantial manner. (important)
   d. Government has a key role to play in shared prosperity. (either)
   e. The patient showed clinically meaningful improvements. (significant)

(3) **Exercise 3**: Some of the underlined uses of important and significant are inappropriate.
   Identify and correct them. Then translate the sentences into Korean.
   a. I will ask for your opinion on all significant questions. (important)
   b. The test results are clinically important. (significant)
   c. Immigration is a very important issue in this country. (correct)
   d. At the moment, we’ve more significant things to discuss than scheduling. (important)
   e. The difference is statistically important. (significant)
3.4. Teaching Synonymous Verbs: demand, request, and require

3.4.1. Corpus queries and analyses

To distinguish synonymous verbs often requires us to examine their typical object and subject collocates as well as their common modifying adverbs. Because demand and request can each also be used as a noun, a simple “[nn*] [demand]”/“[nn*] [request]” query for find what subject nouns are used before each verb would generate many irrelevant tokens where demand/request are each used as nouns, such as market/oil demand and budget/information request. So we have to go through the results and eliminate the irrelevant tokens. This is another example showing the complexity involved in corpus queries. Despite the complexity challenge, a careful query and scrutiny of the typical subjects, objects, and modifying adverbs of the three verbs will reveal that while they all can mean “ask for,” the verbs differ quite noticeably in agent (subject), theme (object), and manner (shown by their adverbials), as well as in the structural type of their objects.

Concerning the most typical (top ten) subject nouns of each verb in COCA (reported in Table 3), those of demand are mostly humans/human institutions (e.g., students/nations/Congress/protesters) although the verb also takes non-human subjects (e.g., situation). The fact that human nouns like students/citizens/protesters often serve as the subjects of demand suggests that the verb is frequently used for situations where individuals are making demands to authorities. In comparison, the subjects of request are exclusively humans/human institutions (e.g., students/administration/stores; store used as the subject of request refers to the owners/employees of the store, not the physical facility). In contrast, the subjects of require are mostly nonhuman nouns (laws/rules/regulations/tasks), although a few (e.g., states) are human institutions. These findings suggest that demand is similar to request in the sense that their subjects are mostly human, but demand is sometimes used like require when taking nonhuman subjects. However, we need more information to more clearly define their differences.
TABLE 3

<table>
<thead>
<tr>
<th>Subject Noun</th>
<th>Demand F*</th>
<th>MI</th>
<th>Request F*</th>
<th>MI</th>
<th>Require F*</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>student(s)</td>
<td>138</td>
<td>7.25</td>
<td>students(s)</td>
<td>39</td>
<td>3.57</td>
<td>law(s)</td>
</tr>
<tr>
<td>government(s)</td>
<td>113</td>
<td>6.37</td>
<td>administration</td>
<td>38</td>
<td>5.81</td>
<td>state(s)</td>
</tr>
<tr>
<td>society</td>
<td>102</td>
<td>6.52</td>
<td>parent(s)</td>
<td>35</td>
<td>4.57</td>
<td>rule(s)</td>
</tr>
<tr>
<td>situation(s)</td>
<td>102</td>
<td>5.47</td>
<td>President</td>
<td>28</td>
<td>3.17</td>
<td>program(s)</td>
</tr>
<tr>
<td>nation(s)</td>
<td>93</td>
<td>4.81</td>
<td>store(s)</td>
<td>25</td>
<td>4.92</td>
<td>regulation(s)</td>
</tr>
<tr>
<td>official(s)</td>
<td>75</td>
<td>6.82</td>
<td>Bush (president)</td>
<td>24</td>
<td>4.34</td>
<td>system(s)</td>
</tr>
<tr>
<td>investor(s)</td>
<td>65</td>
<td>8.12</td>
<td>client(s)</td>
<td>23</td>
<td>5.75</td>
<td>job(s)</td>
</tr>
<tr>
<td>citizen(s)</td>
<td>64</td>
<td>6.83</td>
<td>teacher(s)</td>
<td>23</td>
<td>3.95</td>
<td>progress</td>
</tr>
<tr>
<td>Congress</td>
<td>55</td>
<td>9.46</td>
<td>department(s)</td>
<td>21</td>
<td>4.35</td>
<td>task(s)</td>
</tr>
<tr>
<td>protester(s)</td>
<td>40</td>
<td>6.02</td>
<td>company(ies)</td>
<td>20</td>
<td>3.03</td>
<td>legislation</td>
</tr>
</tbody>
</table>

*Frequency

In terms of the most frequent object nouns of the verbs (reported in Table 4), there are a few important points worth noticing. Except for three nouns that appear on the lists of two of the verbs (i.e., attention and change(s) appearing on the lists of both demand/require and money on the lists of both demand/request), the most common objects of each verbs are unique. First, while quite a few nouns of demand relate to legally/ethically right things to do (justice/respect/accountability), most of the nouns of request concern assistance from others (permission/assistance/help). On the other hand, the most common nouns of require are humans (students/employers/teachers, which are each followed by an infinitive, e.g., requires students to write a paper). The results also suggest some similarity between demand and request, for both may take concrete nouns (such as money/payments) as their objects. There is also some similarity between demand and require, as both share some abstract object nouns, such as attention/changes. When having such abstract nouns as their objects, demand/require are used in the sense of need, e.g., “These issues demand/require attention” = “These issues need attention.” Given that what is often demanded is something rightful (at least from the perspective of the person/persons making the demands), the verb demand appears to be the strongest of the three in terms of the intensity involved. Request is the weakest because what is typically requested is assistance from others. Require is unique in the sense that its typical subject is a law/rule/regulation or an agency in charge of implementing the law/rule/regulation; as a result, what is required is often something official that one has to do or provide with no room for negotiation or exception.
Table 4
Ten Most Frequent Object Nouns of Demand/Request/Require in COCA

<table>
<thead>
<tr>
<th>Object Noun</th>
<th>Demand</th>
<th>Request</th>
<th>Require</th>
</tr>
</thead>
<tbody>
<tr>
<td>attention**</td>
<td>138</td>
<td>7.25</td>
<td>222</td>
</tr>
<tr>
<td>answer(s)</td>
<td>113</td>
<td>6.37</td>
<td>142</td>
</tr>
<tr>
<td>action(s)</td>
<td>102</td>
<td>6.52</td>
<td>140</td>
</tr>
<tr>
<td>money***</td>
<td>102</td>
<td>5.47</td>
<td>77</td>
</tr>
<tr>
<td>change(s)**</td>
<td>93</td>
<td>4.81</td>
<td>52</td>
</tr>
<tr>
<td>justice</td>
<td>75</td>
<td>6.82</td>
<td>21</td>
</tr>
<tr>
<td>payment(s)</td>
<td>65</td>
<td>8.12</td>
<td>18</td>
</tr>
<tr>
<td>respect</td>
<td>64</td>
<td>6.83</td>
<td>18</td>
</tr>
<tr>
<td>accountability</td>
<td>55</td>
<td>9.46</td>
<td>18</td>
</tr>
<tr>
<td>access</td>
<td>40</td>
<td>6.02</td>
<td>17</td>
</tr>
</tbody>
</table>

* Frequency
** An italicized word with one ** also appears on the noun list of "require."
*** An italicized word with two *** also appears on the noun list of "request."

A query of the typical modifying adverbs of the three nouns also confirms these findings about the semantic/usage patterns of the three verbs. Due to space limit, I will just report the main findings related to the modifying adverbs. While all three verbs share some additive/ frequency/time adverbs (e.g., also/often/never), they each have their own unique adverbs that help divulge their meanings: publicly/angrily/loudly demand; formally/politely/respectfully request; and constitutionally/legally require. Furthermore, a close reading of the concordance lines of the verbs also indicates that while request/require are sometimes used in the Verb + Noun + Infinitive structure (e.g., request/require students to write an essay), demand is never used (i.e., cannot be used) with this structure (i.e., we generally do not say *demand someone to do something). Instead, demand is often used in the Verb + That-clause structure (e.g., demand that the city government keep the bus service). The above analysis and findings should help learners attain a good understanding of the semantic/usage patterns of each of the three synonymous verbs. However, to help learners reinforce their learning, teachers can develop useful exercises based on the corpus findings, such as the following.

3.4.2. Sample learning activities

(1) Exercise 1: Decide whether demand, request, or require fills in each blank the best semantically; if you believe more than one words may work equally well, write
them down.
a. The law _______ that at least one parent/guardian be notified before an abortion. 
   (requires)
b. Many demonstrators held a rally to _______ free elections in this country. 
   (demanded)
c. Most knee problems don't _______ surgery. (require)
d. I'd like to respectfully _______ that you support my candidacy for the office. 
   (request)
e. This crisis _______ immediate attention. (either demand or require work)

(2) Exercise 2: Decide whether demand, request, or require may best replace each 
underlined word while keeping the original meaning and tone of the sentence.
a. Teachers should help students learn how to _______ help. (request)
b. To the best of my knowledge, the Saudi government invited, indeed, asked 
   American troops to come into Saudi Arabia. (requested)
c. In the lawsuit, they _______ the payment of $2.5 billion as compensation for the 
   use of water from the Alto Lerma Canal from 1970 up to that moment. (demanded)
d. This task needs time and effort. (demands or requires)
e. An 1891 federal law _______ that goods be marked with their country of origin. 
   (required)

(3) Exercise 3: Some of the underlined uses of demand, request, and require are 
incorrect. Identify and correct them. Then translate the sentences into Korean.
a. Beijing _______ that the Sudanese government implement the resolution. (correct)
b. The airplane, low on fuel, has _______ permission to land in Caracas. (requested)
c. The procedure requests patience and self-control. (demands or requires)
d. The angry demonstrators _______ that the government release their leader. 
   (demanded)
e. All three tests _______ students to use more than content knowledge. (require)

3.5. Teaching Related Verbs: doubt and suspect

3.5.1. Corpus queries and analyses

Unlike all of the above examples, doubt and suspect are not synonymous. In fact, they 
are actually antonyms (i.e., having opposite meanings) in English. They are listed as 
related verbs here simply because they are translated into the same word in Korean and 
some other Asian languages. Because of the nature of the two words in English, examining
their typical collocates will not help reveal their semantic difference, for they can be placed in the exactly the same linguistic context: “I doubt/suspect he is a spy.” Yet, which of the two verbs is used completely changes the meaning of the sentence: the use of doubt will mean “I tend not to believe he is a spy,” while the use of suspect will mean I tend to believe he is a spy. To identify this semantic difference will require the examination of the context beyond the utterance, often the discourse immediately before the utterance. Here are two examples adapted from COCA:

a. He has been absent from the Senate lately and his appearance has changed so much. I doubt that many of his colleagues would recognize him unless he wore a name tag.

b. He has had short breaths for quite sometime. Doctors suspect he has a congenital heart block.

In example A, the sentence before the “I doubt” utterance makes clear that the “he” senator has been absent for a while and his appearance has changed a lot. So it follows that the speaker “I” tends not to believe (hence the verb doubt) that many of the senator’s colleagues would recognize him. Suspect would not work in the context, as it would mean the speaker believes the senator would be recognized by his colleagues. In contrast, in example B, the first sentence states that the “he” person has had short breaths for quite sometime; hence it is logical for doctors to believe (suspect) the he has a heart blocking problem from birth. Doubt would not work here because the use of it would mean that the doctors tend not to believe the person has a heart problem. Teachers can have students read and analyze such corpus examples in class so as to help students better understand the difference between the two verbs. Teachers can also use corpus data to develop exercises for students to practice and learn the use of the two verbs. The following are some examples.

3.5.2. Sample learning activities

(1) Exercise 1: Fill in each blank with either doubt or suspect based on its discourse context.

a. She said that she didn’t know anything but I _______ she actually knew. (suspect)

b. The meeting was supposed to be an hour, but I _______ it was actually that short. (doubt)

c. If you know or simply _______ any illegal practice is going on, you should report it to police. (suspect)
d. As a matter of fact, I'm feeling scared. I ________ we are at a very dangerous point. (suspect)
e. Some say it's a good time to invest in stocks. I ________ it because the economy is still in a deep recession. (doubt)

(2) Exercise 2: Decide whether doubt or suspect can replace each underlined phrase while keeping the original meaning of the sentence.
a. These are the things in her life that I would like to think she doesn't want others to know about. (suspect)
b. He lives in a dream world now, all alone, and I have trouble believing that he can distinguish between what is real and what is not. (doubt)
c. She murdered her first husband. When her second husband died, the police immediately considered the possibility that she killed him. (suspected)
d. He didn't believe her story because he was inclined to think she was not telling him the truth. (suspected)
e. He said he missed the class because he was sick, but his teacher was not included to believe his story. (doubted)

(3) Exercise 3: Some of the underlined uses of doubt and suspect are incorrect. Identify and correct them. Then translate the sentences into Korean.
a. You've helped edit my paper so well that my teacher may doubt that someone else wrote it for me. (suspect)
b. He says it took him just an hour to complete the essay. Yet she suspects it took him at least twice as much time. (corrects)
c. Tom has lied many times before. So I suspect he's telling the truth this time. (doubt)
d. Russian President Putin and his wife haven't appeared together for quite sometime. Some doubt they've been divorced. (suspect)
e. He often tells the teacher what we say and do, so we all doubt that he told the teacher about our plan to pretend sick this time. (suspect)

4. CONCLUSION

Drawing on existing research, this paper has shown that the semantic/usage differences among synonyms are best manifested by their typical distributional patterns (especially their typical collocational patterns) and that corpus analysis of such distribution information can often effectively differentiate synonyms and help L2 learners better grasp their semantic/usage patterns. Corpus-based/driven learning also provides learners with
excellent discovery learning opportunities and ample authentic language input. Using specific examples, this paper has also illustrated how corpus analysis and its results can be used to help learners grasp difficult-to-distinguish L2 words that are undifferentiated in learners’ L1. Of course, the learning activities and exercises given are meant only as examples. Many other types of activities and exercises can be developed.

It is also important to reiterate that, because corpus analysis of the semantic/usage patterns of synonyms is often complex and labor-intensive, using the approach in teaching can be very challenging. Hence, generally, teachers should not involve students (except for those with an upper-intermediate or up language proficiency) in direct corpus queries. When engaging upper-level students in direct corpus queries and analyses, the teacher should, however, provide them with adequate training and guidance on query and analysis procedures, using as many specific examples as possible (Liu, 2010b, 2011; Liu & Jiang, 2009). Without such training, students can easily become frustrated. To help students succeed in their corpus queries and analyses, the teacher can have them work in groups, as group work has been found to be effective in corpus-based learning (Liu, 2010b, 2011; Liu & Jiang, 2009). For low- and intermediate-level students, the teacher can do the query before class and give students screened results of the query for them to analyze. Or, if the corpus query for the usage patterns of a given set of synonyms is fairly simple and straightforward, the teacher can have students do the query together in class and provide whatever guidance and assistance that is necessary. Finally, while it is important to be fully aware of the challenges, it is equally important not to let the challenges stop us from trying this corpus analysis-based approach because the potential benefits of this teaching practice appear to be too great for us not to use it.

REFERENCES


Language Journal, 68, 130-137.


Tsui, A. (2004). What teachers have always wanted to know—and how corpora can help. In J. M. Sinclair (Ed.), How to use corpora in language teaching (pp. 39-61). Amsterdam: Benjamins.


Applicable levels: All

Dilin Liu
Applied Linguistics/TESOL
English Department
University of Alabama
United States of America

Received in June 2013
Reviewed in July 2013
Revised version received in August 2013