

**The Use of Corpus for Close Synonyms**

M. Naci KAYAOĞLU

naci@ktu.edu.tr

Suggested Citation:

Kayaoğlu, M. N. (2013). The use of corpus for close synonyms. *The Journal of Language and Linguistic Studies*, 9(1), 128-144. <http://www.jlls.org/vol9no1/128-144.pdf>

**Abstract**

*Problem Statement:* Using corpora is still in its infancy in foreign language classes in spite of its great benefits and potential to offer solutions to the various challenges in foreign language instruction both for teachers and learners. This partly stems from a lack of interest and practical knowledge about the pedagogic role that the corpora can play. There is a pressing need to convince teachers of the great benefits of corpora with empirical data.

*Purpose of Study:* This research aims to explore the feasibility of using a corpus to help students differentiate between close synonyms which have similar meanings but cannot be substituted one for the other.

*Method:* This is quasi-experimental research based on a pre and post-test (one shot design) design. To this end, participants were introduced to the Corpus of Contemporary American English (COCA) and asked to refer to the corpus when deciding the appropriate close synonym in the 50 sentences given. Participants were also interviewed upon completion of the task with the COCA about their use of corpus.

*Findings and Results:* The t-test showed that the use of corpus for deciding on close synonyms proved to be very effective as there was a statistically significant difference in participants' performance on the vocabulary portion of the pretest and post-test.

*Conclusion and Recommendations:* The present paper suggests that corpora can be a very rich and effective source of empirical data both for teachers and students to make foreign language learning more meaningful and enjoyable. Providing learners more exposure to authentic examples, corpora can be utilized for pedagogic purposes from syllabus design to

materials development. Yet, it needs to be integrated into language courses. Teachers should be made fully aware of what corpora offer for language teaching.

*Keywords:* corpora, close synonyms, corpus of contemporary American English

### **Introduction**

Corpus Linguistics as a sub-branch of General Linguistics is one of the few fields that enable us to study authentic language use for various purposes. Given the current status of corpora in foreign language, it is fair to say that the benefits, the potential and the likely contributions of the corpus have been underestimated by language teachers for many years and thus the perceptions towards corpus use changed very slowly. With the advent of high speed computer technology and enormous storage capacity to process hundreds of millions of words, corpus linguistics has revolutionized the studies of language analysis in its real context, providing authentic real-life data and rich insights into the use of language in a variety of contexts.

### **The Use of Corpus for Language Teaching and Learning**

A corpus refers to a systematic collection of naturally occurring written or spoken language samples in context, stored on a computer readily available for qualitative and quantitative analysis. A corpus is generated according to an explicit design principle (content/genre/typology/register, etc) for a variety of purposes. “A corpus...can allow users to get hundreds of examples in context in a matter of seconds in a readily accessible tool without training in computational linguistics. So corpus access software enables us to analyze these real-life language samples in a magic way which had not before been possible” (Kayaoğlu 2009, p. 139). Some tend to view the corpus as a methodological basis for linguistic analysis and a set of methods and procedures for the exploration of language (McEnery & Hardie 2011; Kennedy 1998; Biber et al., 1998; Conrad & Levelle, 2008). According to McEnery and Wilson (1996, p. 9) “corpus is the most reliable source of evidence for such features as frequency”.

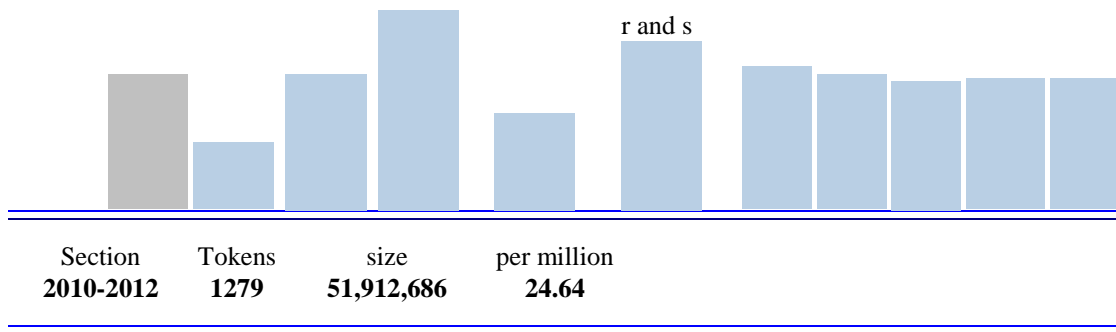
Generally corpus software packages are characterized by three ways in which data is processed. They display frequency, phraseology and collocation. It is possible to know how many times the words, for example, *publicity*, and *attraction* occur in a corpus of tourism,

which we can then compare with the occurrences in other corpora. This may give us quantitative data in a matter of seconds without training in computational linguistics. It is also possible to compare between synonyms and other semantically-related words. For example, it might be interesting to know the most frequent verbs or adjectives that appear with *publicity* and *attraction*. It is also possible to go beyond the frequency of the words by analyzing real-life language samples qualitatively. Corpora, depending on a diverse range of purposes, are classified into different categories such as spoken vs. written, monolingual vs. bi/multilingual, parallel vs. comparable corpora (translation corpora), general language purpose vs. specialised language purpose (academic writing), diachronic (historical) vs. synchronic, learner and pedagogic corpus. While a specialized corpus involves a particular type of text with a focus on a particular register, a comparable corpus includes more than one language or different varieties of a language. A particular interest for language teachers is obviously learner corpora which includes a collection of authentic texts generated by foreign/second language learners to be used to analyse learners' problems.

Regardless of the type, a corpus-based analysis enables us to discern the patterns of language use, language patterns ranging from morphological, discourse to lexicogrammatical ones. It allows us to observe current genre-based variation and recent diachronic shifts in the language. Corpus is of interest for not only lexicographers and researchers but also educators and language teachers as corpus linguistics offers a great deal of opportunities applicable to materials development, lexical diversity in production and receptive vocabulary, syllabus design and classroom activities. A query can provide a wide range of examples including synonyms, collocates, customized word lists, limits by certain period or genre including spoken, fiction, popular magazines, newspapers, and academic journals. One challenging issue for non-native speakers of English is to decide the right word for a particular register in which language often behaves differently. This is where mono or bilingual dictionaries fail to provide comprehensive and adequate information in relation to the use of words in different registers. Kettemann (1995) indicates the discrepancy between real language use and the prescriptive grammar. With the help of corpus analysis one can make a well-informed judgment about the right words in a specific register, for example, a search for *discover* gives the frequency in five different macro registers.

Table 1  
*Registers in corpus*

Section	All	Spoken	Fiction	Magazine	Newspaper	Academic	1990-1994	1995-1999	2000-2004	2005-2009	2010-2012
Freq	11745	1213	2318	3600	1693	2921	2816	2633	2505	2512	1279
Per mil	25.31	12.69	25.63	37.67	18.46	32.08	27.08	25.45	24.33	24.62	24.64



between 1990-94 and 2010-2012 (Corpus of Contemporary American English-COCA). This indicates that the word is most frequent in magazines, followed by fiction, newspaperspoken, and that over time there is a decreasing tendency in the frequency of the word *discover*. A further search to see the variation of the word in magazines results in the

Table 2  
*Search for variation in the register*

News/Opin	Financial	Sci/Tech	Soc/Arts	Religion	Sports	Entertain	Home/Health	Afric-Amer	Children	Women/Men
495	152	694	339	265	352	88	615	125	71	406
26.61	28.91	54.92	44.31	62.05	32.53	21.62	38.54	34.41	43.43	41.63



following information chart where the searched word occurred most frequently in Religion, followed by Sci/Tech. The following chart also indicates that the word appears most frequently in Humanities and Phil/Rel registers.

Table 3

*Detailed search in the related register*

History	Education	Geog/SocSc	Law/PolSci	Humanities	Phil/Rel	Sci/Tech	Medicine	Misc
301	333	340	173	621	345	494	57	255
24.58	35.26	21.01	20.12	52.07	51.18	35.10	8.51	59.91

A corpus can also be utilised for pedagogic purposes. For example, comparing *can* and *could* in a register or to find out what preposition most commonly follows *discover*, as displayed in Table 4 below, will provide us with many examples of the lexical item or pattern in its context of use. Obviously, a corpus does not prescribe a rule or tell the meaning of the words but with the help of concordance lines the corpus gives us a rich insight into how a lexico-grammatical query is used, how it collocates with other words and even idiomatic usage, from all of which we arrive at certain generalizations. When this is applied to the learner corpora by the teacher, it is quite possible to obtain very valuable information about interlanguage developmental stages of the learners. In addition to the frequency and distribution of the preposition of the verb, the corpus provides hundreds of real examples with expanded context as displayed in Table 4 through which we can arrive at well-informed interpretations and compensate for lack of native speaker intuition.

Table 4

*Expanded context (COCA)*

1	2011	FIC	Bk:NightStrangers	linked to the house. Sometimes Emily would find the sort of thing you might <b>discover in</b> the back of an antiques store, such as the broken
2	2011	FIC	AntiochRev	. Every time we went it was like entering a secret cave, expecting to <b>discover in</b> the damp, semi-dark nook special treasure or ordnance. We whispered amid green
3	2011	MAG	Esquire	of five seconds. You realize that anxiety really can be crippling. # You <b>discover in</b> successive waves some of what really happened. You
4	2011	NEWS	NYTimes	Qaddafi government might <b>in</b> fact have believed its

				own propaganda: that the journalists would <b>discover in</b> Zawiya radical Islamists or young people crazed by drugs supplied by Osama bin Laden
5	2011	NEWS	<u>Houston</u>	was at loose ends. But then he found his focus: helping young men <b>discover in</b> the wilderness that they too were capable of much more than they ever imagined
6	2011	NEWS	<u>Houston</u>	author # " I would say Bigfoot represents that there is a lot left to <b>discover in</b> our world. " # - Ken Gerhard, Gulf Coast Bigfoot
7	2011	NEWS	<u>CSMonitor</u>	in ourselves, or whether they are matters of temperament we can only try to <b>discover in</b> others. The sad part is to realize how infrequently
8	2010	SPOK	<u>NPR_Science</u>	, your PhD dissertation focused on the belief system of scientists. What did you <b>discover in</b> that? Dr- GRAFFIN: Well, <b>in</b> a sense, it did. I
9	2010	MAG	<u>GoodHousekeeping</u>	say on Christmas Eve, as though " house-proud " was a terrible thing to <b>discover in</b> a woman. She thought I was upstairs wrapping
10	2010	MAG	<u>ChristCentury</u>	, Christianity, Viking). FOOL ME TWICE: Paul Hoffman was editor o <b>Discover in</b> the 1990s when for five years the magazine published .
11	2010	NEWS	<u>CSMonitor</u>	year... making the sperm whale a carbon-neutral mammal. " What else might we <b>discover in</b> the 21st century about how whales contribute to keeping our seas healthy? #
12	2010	ACAD	<u>ChurchHistory</u>	, flatteringly like the States General of the United Provinces), one can not <b>discover in</b> the words of van der Donck any interest <b>in</b> religious
13	2009	SPOK	<u>NBC Today</u>	. Men said it was more like every day. " So what did you <b>discover in</b> talking to men and women all across the country?
14	2009	ACAD	<u>Poetry</u>	her way. But it's one of the inadvertent pleasures <b>in</b> reading Gunn to <b>discover in</b> his imagination a passion to propose new forms of
15	2009	ACAD	<u>SouthwestRev</u>	Mansoura Ez Eldin begins with a girl who is dreaming about death and wakes to <b>discover in</b> Kafkaesque horror that she no longer exists

It is also possible to search for the frequency of prepositions used with the word *discover* or for the most common nouns and adjectives. COCA indicates that “*in*” is the most common preposition used with “*discover*” followed by “*for*”, “*on*”, “*about*” “*through*” “*with*” and “*from*” as indicated in Table 5.

Table 5  
*Search for prepositions used with the selected verb*

1	in	127	
2	for	48	
3	on	36	
4	about	35	
5	through	17	
6	with	17	
7	from	15	
8	at	14	
9	by	14	
10	to	9	

In his discussion of Corpus of Contemporary American English (COCA), Davies (2009, p. 181) refers to the use of corpus as an advantageous over dictionaries as follows:





















While there are many different learner-oriented dictionaries of phrasal verbs in English (e.g. Longman Dictionary of Phrasal Verbs, Collins Cobuild Dictionary of Phrasal Verbs, NTC’s Dictionary of Phrasal Verbs and Other Idiomatic Verbal Phrases), almost none of them are based on the type of rich corpus data that one can obtain from a robust corpus like COCA.

Collocation has been recognized as an area of major difficulty for language learners and as playing a fundamental role in a learner’s interlanguage continuum, accounting for a significant proportion of learner errors in language production. Following the growing interest in vocabulary research generated with the availability of large corpora, the focus has been shifted from theoretical perspective (descriptive corpus-based research) to the explicit learning of collocation in the language classroom because collocation runs through

the language. Combinations of words in a language are highly predictable to a native-speaker but may be quite challenging to a learner. For example, a learner learning Turkish may be faced with the choice of two verbs “rahmetli oldu ” and “öldü” given as equivalents for the verb *die* in a dictionary. A choice to make a sentence in relation to the cat s/he lost may results in a funny statement or a bemused smile. It takes a great degree of competence with Turkish to combine them correctly in productive use. As to the word “*discover*”, it may be quite valuable to know what words combinations are used with it as displayed in Table 6 through the use of COCA.

Table 6

*Collocations with corpus*

1. Things	78		11. People	21	
2. Truth	73		12. World	21	
3. Card	61		13. Meaning	20	
4. Ways	58		14. Secrets	20	
5. Magazine	36		15. Source	20	
6. Secret	31		16. Cure	19	
7. Life	29		17. Value	19	
8. Way	24		18. Differences	17	
9. Information	21		19. Cause	16	
10. Nature	21		20. Evidence	16	

Learners are expected to derive generalizations, rules and patterns from multiple examples. For this reason corpus-based learning is mostly associated with inductive rather than deductive approach. In a study by Sun and Wang (2003) learning grammatical collocations inductively through the use of online concordancers was found significantly more effective than the deductive approach.

To sum up, a corpus can:

- provide insights into language use where dictionaries and resource books simply fail or lead us to the wrong information.
- be a source of a variety of language teaching materials.



- provide students with hands-on opportunities for language learning.
- provide models of language use and language patterns
- allow for both quantitative and qualitative analytical techniques
- provide words frequency lists for particular area (chemistry, humanities)
- enable learners to improve their vocabulary (both fluency and accuracy)
- compensate for lack of the native speaker's intuition
- provide exploratory tasks and activities rather than ready-made mechanic drills and exercise
- cultural study e.g. differences found in two comparable corpora
- verify certain hypotheses about languages
- be more convincing than the teachers with supporting evidence
- be used to check grammar
- avoid human bias in data analysis
- allow the enrichment of data with metadata
- enable learners to discern subtle differences
- enable language learners to make well-informed decision similar to a native-speaker
- help learners to detect the nuances of language items (inhabitants, residents, dwellers)
- allow the students to arrive at their own generalizations from plenty of examples in the corpus
- be a good source for translation purposes
- enable learners to be familiar with the vocabulary of a particular section (medical, sports, science/technology)
- be a useful performance-enhancing tool in confirming intuitive decisions
- allow learners to quickly and easily see how native speakers use the language in a wide variety of naturally occurring texts
- be utilized for pedagogic purposes (finding out what preposition most commonly follows the verbs frequently used in the student's textbook)
- help learners develop their reasoning skills inductively and take an active role in their own learning

- provide learners more exposure and authentic examples instead of artificial ones not used very often daily
- enable learners to identify the key words in or more texts.
- be used for a range of comparative purposes (British English-American English)
- create direct interaction with learners as they get their hands authentic materials

For all these purposes, a list of corpus online is given here in Table 7.

Table 7

*The list of available corpus online*

Survey of Corpora		How to find out more
Corpus of Contemporary American English (COCA)	450 million words	<a href="http://corpus.byu.edu/coca/">http://corpus.byu.edu/coca/</a>
American National Corpus (ANC)	22 million words	<a href="http://americannationalcorpus.org">http://americannationalcorpus.org</a>
Bergen Corpus of London Teenage Language – (COLT)	500,000 words	<a href="http://torvald.aksis.uib.no/colt/">http://torvald.aksis.uib.no/colt/</a> <a href="http://icame/newcd.htm">http://icame/newcd.htm</a>
British National Corpus (BNC)	100 million words	<a href="http://www.natcorp.ox.ac.uk/what/index.html">http://www.natcorp.ox.ac.uk/what/index.html</a>
Brown Corpus	1 million words	<a href="http://khnt.hit.uib.no/icame/manuals/brown/INDEX.HTM">http://khnt.hit.uib.no/icame/manuals/brown/INDEX.HTM</a>
Collins Wordbanks Online English corpus	56 million words	<a href="http://www.collins.co.uk/Corpus/CorpusSearch.aspx">http://www.collins.co.uk/Corpus/CorpusSearch.aspx</a>
Cambridge and Nottingham Corpus of discourse in English (CANCODE)	5 million words	<a href="http://www.cambridge.org/elt/corpus/cancode.htm">http://www.cambridge.org/elt/corpus/cancode.htm</a>
Cambridge International Corpus	1 billion words	<a href="http://www.cambridge.org/elt/">http://www.cambridge.org/elt/</a>

(CIC)

---

Corpus of English as a Lingua Franca in Academic Settings (ELFA)	500,000 words	<a href="http://www.uta.fi/laitokset/kielet/engf/research/elfa/project.htm">http://www.uta.fi/laitokset/kielet/engf/research/elfa/project.htm</a>
Frown Corpus	1 million words	<a href="http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM">http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM</a>
International Corpus of Learner English – ICLE	2 million words	<a href="http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/icle.htm">http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/icle.htm</a>
International Corpus of English – (ICE)	in progress	<a href="http://www.ucl.ac.uk/english-usage/ice/">http://www.ucl.ac.uk/english-usage/ice/</a>
Lancaster-Oslo/Bergen corpus (LOB)	1 million	<a href="http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM">http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM</a>
Longman Written American Corpus	100 million words	<a href="http://www.longman.com/dictionaries/corpus/lcawritt.html">http://www.longman.com/dictionaries/corpus/lcawritt.html</a>
The Longman Learners' Corpus	10 million word	<a href="http://www.longman.com/dictionaries/corpus/lclearn.html">http://www.longman.com/dictionaries/corpus/lclearn.html</a>
London-Lund Corpus	500,000 words	<a href="http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM">http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM</a>

---

It appears that corpora have many applications in the area of language teaching as well as learning from materials design to communicative activities. Perhaps, one of the most concrete uses of language corpora is the analysis of lexical items. This is a revolution for EFL settings given the fact that language learners have suffered long when are compelled to discern subtle differences between words or differences between close synonyms especially when the mother tongue doesn't have a corresponding one. They have to rely on prescriptive books or dictionaries which can be at odds with actual use. Together with register and style this issue poses a serious handicap for language learners as stressed by Kayaoğlu (2009: 90) 'dictionaries, grammar and similar reference books provide

explanation about particular Word usages or grammatical forms but do not explain how or in what context a particular Word is used'. With this in mind, this study was designed to see whether the use of a corpus can be a viable option for foreign language learners when they are challenged to make a choice between close synonyms or commonly confused words.

## **Method**

### **Research**

The aim of this research was to explore the feasibility of using a corpus for the solution of the challenging word-choice problems in which close synonyms which have similar meanings but may not substituted one for the other in the given test. A total of 23 intermediate level students majoring in English (in the English Prep program) participated in the study. Students were given a vocabulary test of 30 English sentences in seven set of questions with a blank each to fill from the group of close synonyms as shown in Table 8. Students were advised to mark the most challenging questions whose answers (the right words) they were not very sure about. Students were not told the results of the test. On the same day students were made familiar with the COCA following a 40-minute training. Students were then asked to do the same test with the same task of choosing the most appropriate English words for the given blanks by referring to the COCA in the computer lab. The students were encouraged to act as language detectives; to observe nuances of meaning and patterns of usage in problematic words, using COCA. The corpus obviously does not prescribe anything other than a collection of real language samples. For this reason the students were expected to make their own judgment about the challenging on the basis of the corpus.

The *Corpus of Contemporary American English (COCA)*, which was released online in 2008, is the first large, diverse and publicly-available corpus of American English that contains texts from a wide range of genres and texts from the past fifteen years. It contains more than 385 million words from 1990–2008 (20 million words each year), balanced between spoken, fiction, popular magazines, newspapers, and academic journals. In the COCA, texts are evenly divided between spoken (20%), fiction (20%), popular magazines (20%), newspapers (20%) and academic journals (20%). This composition holds for the corpus overall, as well as for each year in the corpus. This means that researchers

can compare data diachronically across the corpus, and be reasonably sure that the equivalent text composition from year to year will accurately show changes in the language. For each year, the texts within each of the five genres are balanced between the sub-genres or domains just mentioned. For example, each year the newspapers are evenly divided between the ten newspapers (approximately 400,000 words each); approximately 10% of the fiction texts (400,000–500,000 words each year) come from movie scripts; the popular magazines maintain roughly the same composition from year to year (African-American, current events, sports, science, religion, health, etc.), and the same is true for the academic journals (science, history, religion and philosophy, technology, education, etc.) (Davies, 2009).

Table 8

*List of close synonyms*

1	refuse	deny	decline	reject	rebuff	
2	career	job	work	employment	profession	
3	find out	detect	discover	search	seek	examine
4	memorise	remember	remind	recall	recognise	recollect
5	common	ordinary	usual	normal	regular	
6	tour	travel	voyage	journey	trip	cruise
7	precious	valuable	invaluable	priceless	worthless	economic economical

**Results and Discussion**

The research is based on pre and post-test design. In the pre-test students were asked to choose the most appropriate English words for the given blanks. In the second part of the experiment, students were asked to do the same test by using COCA. A t-test was used to see whether there was a significant difference between pre-test and post-test results in terms of appropriate word choice.

Table 9

*Paired samples t-test statistics*

	Mean	N	Std. Deviation	Std. Error Mean	t	df	Sig (2-tailed)
test1	57,6087	23	7,66775	1,59884	-11.33	22	000
test2	73,2609	23	5,95603	1,24192			

The t-test analysis in Table 9 indicates that the use of corpus for deciding on close synonyms proved to be very effective as there was a statistically significant difference between the pretest and post-test results. The study is very much limited to quantitative data. However, interviews with five randomly selected students provided evidence for the fact that using the corpus (COCA) was useful and at the same time demanding. While some students were found having derived great pleasure from discovering regularities of patterning in the corpus, some were overwhelmed by the existence of huge number of variations and examples from which they were expected to arrive at some conclusions. For example, one participant stated “*when you click the button, the corpus gives you hundreds of examples but it is not that much easy to come to a conclusion in hundreds of examples in short time*”(s5), referring that it is difficult to process large amount of information.

It is obvious that the corpus only presents information but leaves it to the learner to analyse and interpret. This may also be partly to do with the learners’ readiness to adopt an autonomous approach to language learning because some students enjoy having the inherent culture of being in a teacher-led classroom. Yet, to characterize commonly accepted experience “*it is really amazing to see how much convincing is the corpus with so many examples from real-life in context and in certain area(register) eventhough it can be some time-consuming*”(s3). Another participant was quite pleased with what the corpus offered in terms of vocabulary-related problems “*I am an English majoring student and I read English papers almost every day but I do this for fun. Verb constructions and verb with adjectives and collocates have been always serious problem for me. Although I refer to dictionary of phrasal verbs, thesaurus, I am not always convinced of the information given in these sources. There is sometimes mismatch or conflict between what is stated and what we have, for example, in a newspaper. I feel the corpus may of great help to us at least*

*when we are desperate*” (s5). This is what a corpus-linguistic software can offer for individual learners, that’s learning how to learn. In this respect, corpus data can be of great help for autonomous language learners, if not to create autonomous learners.

### **Conclusion**

In view of the fact that the students were able to significantly improve their performance on vocabulary use and the quality of their intuitive decisions for words choice in the test, the use of corpora proved to be effective and has great potential for EFL classes. However, corpus seems not to have gained its due value and recognition in foreign language teaching and learning in spite of its wide use. It would be wrong to think of the use of corpus as a magic touch to solve lexical, grammatical or discourse related problems. Yet, it can be a very rich source of data to be utilized in activities from designing syllabi to instructional materials for classroom use. In order for the corpora to be of pedagogic value and an effective tool, teachers need to be made familiar with the potential uses of the corpus programs in the teaching practice through in-service teacher training sessions. For example, it can be quite interesting for teachers to know that learner corpora provide empirical data for the identification of the mistakes most commonly made by their students. An important point to note that information generated by the corpora may be too much to digest for beginners. Therefore, tasks should be adapted to their level and needs. Activities created from real-life examples of authentic language use would be more interesting than the exercises made up by material developers.

Although the use of computerized corpora is well established in linguistic research, the corpus is still in its maturation in language teaching profession. Once its potential is discovered, it is more like to be used in language teaching.

### References

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Conrad, S. M. & Levelle, K. R. (2008). Corpus linguistics and second language instruction. In B. Spolsky and F. M. Hult (Eds.), *The handbook of educational linguistics*. Oxford: Blackwell Publishing.
- Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008+) Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(7), 159-190.
- Kayaoğlu, M. N. (2009). Corpora to compensate for lack of sixth sense in mastering a foreign language. Paper presented at the *1st İstanbul Arel University International ELT Conference: Managing Innovative Changes in TEFL: New Insights Beyond Methods*, İstanbul.
- Kayaoğlu, M. N. (2009). Using a corpus for well-on formed intuitive decisions on appropriate word choice in translation process. Paper presented at *3rd International IDEA Conference Studies in English*. Ege University Publication of the Faculty of Letters. 137-146.
- Kennedy, C. & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77-90.
- Kettemann, B. (1995). Concordancing in English language teaching. *TELL and CALL*, 4, 4-15.
- McEnery, A. M. & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. Cambridge, MA: MIT Press.

**Dr. M. Naci Kayaoglu**, having received his MA from Bilkent University, holds his PhD from Bristol University, the U.K. He is the author of *Language Learning Strategies: Theory, Practice and Issues* (2011) and one of the authors of *Music, Language and Second Language Acquisition*. He is the coordinator of Applied Linguistics Graduate Programme (MA-PhD) at Karadeniz Technical University.



## Yakın Eşanlamlı Kelimelerde Korpus Kullanımı

### Öz

*Problem tanımı:* Yabancı dil eğitimine olan katkısına ve yabancı dil eğitiminde karşılaşılan sorunlara çözümler sağlamasına rağmen, korpora kullanımı yabancı dil sınıflarında hala çok yaygın değildir. Bu durum, korporanın pedagojik rolü hakkında yeteri kadar uygulanabilir bilginin olmadığından kaynaklanmaktadır. Bu nedenle, deneysel veriler sayesinde korporanın önemi öğretmenlere kavratılmalıdır.

*Çalışmanın amacı:* Bu çalışmada, öğrencilerin benzer anlamları olan fakat birbirleri yerine kullanılamayan yakın eş anlamlı kelimeler arasında ayırım yapabilmesine yardımcı olması amacıyla korpora kullanımının uygulanabilirliği incelenmektedir.

*Çalışmanın yöntemi:* Bu yarı deneysel çalışma, ön test ve son test desenine (tek seferlik desen) göre hazırlanmıştır. Bu amaçla, katılımcılara Çağdaş Amerikan İngilizcesi Korpusu (COCA) gösterilmiştir ve katılımcılardan, verilen 50 cümledeki en yakın eş anlamlı kelimeye karar verirken korpusa başvurmaları istenmiştir. Ayrıca, katılımcılarla COCA ile çalışmalarını tamamlarken korpus kullanımı hakkında görüşleri sorulmuştur.

*Bulgular ve Tartışma:* Katılımcıların verilen ön test ve son testteki kelime bölümündeki performanslarında istatistiksel olarak önemli ölçüde fark olduğundan yakın eş anlamlı kelimelere karar verirken korpus kullanımının çok etkili olduğu T-testi sonucunda belirlenmiştir.

*Sonuç ve Öneriler:* Bu çalışma, hem öğretmenler hem de öğrenciler açısından yabancı dil eğitimini daha anlamlı ve eğlenceli hale getirmek için korporanın çok zengin ve etkili bir deneysel veri kaynağı olabileceğini göstermektedir. Öğrencileri daha fazla özgün kaynağa yönelterek içerik oluşturmadan materyal geliştirmeye kadar tüm pedagojik amaçlar için korpora kullanılabilir. Bu bağlamda (yabancı dil derslerine de entegre edilmelidir. Öğretmenler, dil öğretiminde korporanın neler sunabileceğinin farkına varmalıdırlar.

*Anahtar Kelimeler:* bütüncü, yakın eş anlamlı sözcükler, çağdaş Amerikan İngilizcesi bütüncüsü