



**Վ. ԲՐՅՈՒՍՈՎԻ ԱՆՎԱՆ ՊԵՏԱԿԱՆ  
ՀԱՄԱԼՍԱՐԱՆ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ В. БРЮСОВА  
BRUSOV STATE UNIVERSITY**

**ԲԱՆԲԵՐ  
Վ. ԲՐՅՈՒՍՈՎԻ ԱՆՎԱՆ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆԻ  
ВЕСТНИК ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА ИМЕНИ  
В. БРЮСОВА  
BULLETIN OF BRUSOV STATE UNIVERSITY**

---

**ՄԱՆԿԱՎԱՐԺՈՒԹՅՈՒՆ ԵՎ ՀԱՍԱՐԱԿԱԿԱՆ ԳԻՏՈՒԹՅՈՒՆՆԵՐ**

**ПЕДАГОГИКА И СОЦИАЛЬНЫЕ НАУКИ**

**PEDAGOGY AND SOCIAL SCIENCES**

**2(55)**

**Վ. ԲՐՅՈՒՍՈՎԻ ԱՆՎԱՆ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆԻ  
«ԼԻՆԳՎԱ» ՀՐԱՏԱՐԱԿՉՈՒԹՅՈՒՆ**

**ԵՐԵՎԱՆ - 2021**

## CORPUS-BASED METHOD IN TEACHING VOCABULARY

LILIT MOVSESYAN

**Keywords:** *corpus, corpus linguistics, quantitative analysis, qualitative analysis, collocation, lexical nuances, register, semantic priming, conceptual metaphor*

Corpus linguistics is one of the fastest-growing methodologies in contemporary linguistics. It utilizes a large and principled collection of natural texts, known as a corpus, as the basis for analysis. At the same time, corpus linguistics is more than a methodological approach, because it has enabled researchers to ask fundamentally different kinds of research questions, sometimes resulting in radically different perspectives on language variation and use from those taken in previous research. And to make good use of corpus resources one needs a modest orientation to the routines involved in retrieving information from the corpus, and most importantly training and experience in how to evaluate that information. Hence, this paper presents corpus techniques in the field of teaching addressing the possible areas of its usage in vocabulary development through some activities requiring quantitative and qualitative analyses.

Corpus studies have boomed from 1980 onwards, as corpora, techniques and new arguments in favor of the use of corpora became more apparent. Currently this boom continues and corpus linguistics (henceforth CL) is maturing methodologically and the number of languages addressed by CL is growing.

Some linguists view CL as a research tool or methodology, others refer to it as a discipline or theory in its own right. German linguists K. Kuebler and H. Zinsmeister acknowledge that the answer to the question whether CL is a theory or a tool is that it can be both. It depends on how CL is applied (Kuebler, Zinsmeister 2015: 14).

With regard to this question, Hans Lindquist claims that CL is a methodology, comprising a large number of related methods which can be used by scholars of many different theoretical leanings. On the other hand, he points out that it cannot be denied that CL is also frequently associated

with a certain outlook on language. The center of this outlook is determined by the rules of language being usage-based and by the changes occurring when speakers use language to communicate with each other (Lindquist 2009: 105). The argument put forward by H. Lindquist is that if we are interested in the workings of a particular language, like English, it is a good idea to study language in use. One efficient way of doing this is to use corpus methodology.

Others' observations also indicate that CL can be regarded as primarily a methodological approach. At the same time, they state that CL is more than a methodological approach, on the grounds that these methodological innovations have enabled researchers to ask fundamentally different kinds of research questions, sometimes resulting in radically different perspectives on language variation and use from those taken in previous research. R. B. Kaplan contends that CL research offers strong support for the view that language variation is systematic and can be described using empirical, quantitative methods (Kaplan 2010: 548).

S. Hunston puts forward a concluding point that corpus is a test bed for hypotheses used to add a quantitative dimension to many linguistic studies (Hunston 2006: 235-247).

Pointing to the effectiveness of corpus in teaching and learning environment, Italian author and editor Elena Tognini-Bonelli believes that in the context of the classroom the methodology of corpus linguistics is congenial for students of all levels because it is a 'bottoms-up' study of the language requiring very little learned expertise to start with. She states that the students that come to linguistic enquiry without a theoretical apparatus learn very quickly to advance their hypotheses on the basis of their observations rather than received knowledge, and test them against the evidence provided by the corpus (Tognini-Bonelli 2001: 14).

Focusing on the good use of corpora resources in language teaching, Scottish professor John McHardy Sinclair marks the need for the teacher to have a modest orientation to the routines involved in retrieving information from the corpus, and most importantly training and experience in how to evaluate that information (Sinclair 2004:2).

Two main approaches to corpus study are outlined: *quantitative* and *qualitative*. Quantitative techniques are essential for corpus-based studies. For example, if we want to compare the language use of patterns for the words *tourist* and *traveler*, we would need to know how many times each word occurs in the corpus, how many different words co-occur with each of these nouns (i.e. via the collocations), and how common each of those

collocations is. These are all quantitative measurements. Apparently, CL has been criticized for allowing only the observation of relative quantity and for failing to expand the explanatory power of linguistic theory. It is crucial for the corpus-based approach to go beyond the quantitative patterns to propose deep interpretations explaining why the patterns exist, i.e. implying qualitative approach. This suggests that a large amount of effort in corpus-based studies should be devoted to explaining and exemplifying quantitative patterns.

So in CL quantitative and qualitative analyses are extensively used in combination. Geoffrey Leech marks the characteristics of CL, proposing to begin with quantitative findings, then work toward qualitative ones. He points out that the procedure may have cyclic elements. Generally, quantitative results are subjected to qualitative scrutiny attempting to explain why a particular frequency pattern occurs, for example. But on the other hand, qualitative analysis may be the means for classifying examples in a particular corpus by their meanings and this qualitative analysis may then be the input to a further quantitative analysis, one based on meaning (Leech, Hundt, Mair, & Smith 2009: 32).

Taken together, it is crucial to note that the quantitative and qualitative importance of corpus findings greatly depends on the well-defined research question, i.e. corpus is best used to answer a research question which is well composed to address. This suggests that the corpus data we select to explore a research question must be well matched to that research question.

Evidently, to make good use of corpus resources one needs a modest orientation to the routines involved in retrieving information from the corpus, and most importantly training and experience in how to evaluate that information. So this paper presents corpus techniques in the field of teaching addressing the possible areas of its usage in vocabulary development through activities requiring quantitative and qualitative analyses.

#### **Activity A: Semantic Preference**

**Objective: Description of meaning of lexical items based on their contextual distribution in naturally occurring language.**

Today one of the most extensively used methods in corpus linguistics is the analysis of collocations with its fundamental place in the research on contextual semantics. J. M. Sinclair indicates that the meaning of a word is heavily context-dependent and it is to be analyzed not in terms of word-inherent characteristics but rather through the lexical and grammatical elements with which the word co-occurs (Sinclair 1991: 108).

Corpus generates *frequency lists*, sets out relevant *concordance lines* and presents *collocations* to thoroughly depict the distributional behavior of the lexical item. Most corpora allow the user to conduct collocational searches per different criteria, importantly through the specification of collocation window, i.e. *what our collocations are depends on exactly how we calculate them*.

To put it simply, corpus provides the option to specify the span and have the words that occur within the selected span. A collocation window span (CWS) of 4:4 is often used (4 words before and after the node). The preferred CWS may differ between studies. Studies show that the simplest collocate list tends to include words that are not particularly significant for a given word as a node but that are very frequent in the language as a whole.

Importantly, working with corpus huge data *one must avoid any selective approach*. As there is too much evidence, one must abstain from conscious selection of corpus data. This suggests not to have any motivated selection of examples to favor those examples that fit the hypothesis.

**Task A: Study the collocation behavior of the noun TOURISM in different collocation patterns.**

Table 1.

Corpus: BNC  
 Query pattern 1:  
**[TOURISM+N]**  
 CWS: R<sup>1</sup> (the highest frequency list)

N	Collocates	Frequency
1	INDUSTRY	4353
2	SECTOR	1470
3	DEVELOPMENT	1310
4	BOARD	841
5	MINISTER	601
6	DESTINATION	461
7	PRODUCT	394
8	MARKET	320
9	MANAGEMENT	317
10	PROMOTION	314

Table 2.

Corpus: BNC  
 Query pattern 2:  
**[ADJ]+TOURISM]**  
 CWS: L<sup>1</sup> (the highest frequency list)

N	Collocates	Frequency
1	SUSTAINABLE	1065
2	MEDICAL	710
3	CULTURAL	610
4	RESPONSIBLE	605
5	INTERNATIONAL	461
6	LOCAL	344
7	NATIONAL	308
8	MASS	291
9	DOMESTIC	280
10	GLOBAL	198

Table 3.

Corpus: BNC  
 Query pattern 3:  
**[V+TOURISM]**  
 CWS: L<sup>1</sup> (the highest frequency list)

N	Collocates	Frequency
1	[PROMOTE]	550
2	[BOOST]	213
3	[SAY]	107
4	[MAKE]	86
5	[DEVELOP]	85
6	[BASE]	84
7	[AFFECT]	73
8	[ENCOURAGE]	71
9	[INCLUDE]	67
10	[INCREASE]	56

Typically, three standards are suggested to classify collocations. These are: (i) distance, (ii) frequency, and (iii) exclusivity. The *distance* defines the period around the node where we are searching for collocates. This span is called the ‘*collocation window*’. The distance of the collocate from the node can be as small as one word, for example, in the adjectives that directly precede a noun in English or a span of four to five words on either side of the node. The second criterion, *frequency of use*, is an important indicator of the typicality of a word association. The third criterion is the *exclusivity* of the collocates, that is the extent to which the two words appear purely or exclusively in the company of each other, usually expressed in terms of the relationship between the amount of times they are seen together and the amount of times they are seen separately in the corpus (Sinclair 2004: 42-48).

Moreover, corpus allows us to understand one idea or concept in terms of another through the study of the collocation behavior of the node consisting of more than one token. For example, if we are interested in the concept of “happiness” with the noun *happiness* explicitly given, then we can run the following query:

Corpus: GloWbE  
 Query pattern:  
**[(Happiness is)+N]**  
 CWS: R<sup>1</sup> (the highest frequency list)

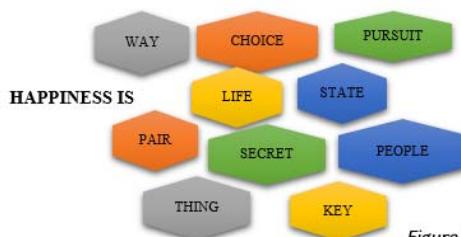


Figure 1

### Activity B: Lexical Nuances of Style and Meaning

**Objective: Context-based display of semantic subtle differences between synonyms through their areas of collocation and usage frequency.**

A word can express a myriad of implications, connotations, and attitudes in addition to its basic denotational meaning. And a word often has near-synonyms that differ from it solely in these nuances of meaning. So, in

order to find the right word to use in any particular situation - the one that precisely conveys the desired meaning and yet avoids unwanted implications - one must carefully consider the differences between all of the options.

**Task B: Compare the semantic peculiarities of nouns TOURIST and TRAVELER observing their collocation behavior in [N+N] patterns**

Table 4.

The corpus study shows that the two instances of collocates shared by TOURIST and TRAVELER are the nouns **site** and **coach** (not concurring in the top 10 frequency lists, thus not given in Table 4). In all other cases they display differences.

Areas of Collocation		
N	TOURIST	TRAVELER
1	Tourist board	Traveler families
2	Tourist information	Traveler children
3	Tourist industry	Traveler return
4	Tourist office	Traveler boys
5	Tourist attraction	Traveler battery
6	Tourist trade	Traveler things
7	Tourist authority	Traveler site
8	Tourist season	
9	Tourist center	
10	Tourist map	

Usage Frequency in [N+N] pattern	
TOURIST	TRAVELER
1342	27

**Activity C: Register Preference**

**Objective: Description of lexical items based on their genre and sub-genre (corpus section) distribution in naturally occurring language.**

Register is defined as situation of language use. Different language is used with different audience at different times and for different purposes. Register can be broadly defined- spoken versus written- or more narrowly defined- conversation versus news or even separate parts of a research paper.

**Task 3: Define the semantic differences between the nouns TOURIST and TRAVELER based on their corpus section distribution.**

Table 5.

Keyword: TOURIST

Section	All	Spoken	Fiction	Magazine	Newspaper	Non-acad.	Acad.	Misc
Freq.	1958	82	189	233	481	210	69	694
Words (M)	100	10.0	15.9	7.3	10.5	16.5	15.3	20.8
Per Mil	19.58	8.23	11.88	32.08	45.96	12.73	4.50	33.31

Table 6.

Keyword: TRAVELER

Section	All	Spoken	Fiction	Magazine	Newspaper	Non-acad.	Acad.	Misc
Freq.	843	24	144	63	89	142	71	310
Words (M)	100	10.0	15.9	7.3	10.5	16.5	15.3	20.8
Per Mil	8.43	2.41	9.05	8.68	8.50	8.61	4.63	14.88

Evidently, the top three sections for the nouns under comparison are the following:

Table 7.

	TOURIST			TRAVELER		
Section	Non-acad.	Misc.	Fiction	Non-acad.	Misc.	
Freq.	210	694	144	142	310	

As seen in Table 6 and Table 7, TOURIST displays much higher frequency than the noun TRAVELER in the corpus sections except for the academic one. Here there is not a great difference yet the frequency number is comparatively higher with TRAVELER.

Considerable differences in numbers are observed in the following sections:

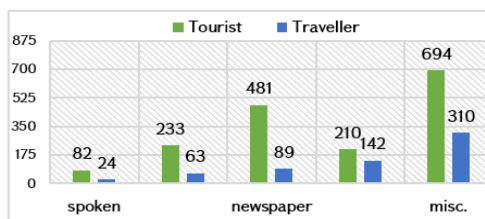


Figure 2. Major differences in frequency: a contrastive display

#### Activity 4: Semantic Priming

**Objective: Every word is mentally primed for collocational use.**

M. Hoey combines corpus linguistics methods with cognitive psychological theory. He claims that collocation is primarily a psychological association between two word forms which is evidenced by their statistically



significant tendency to co-occur in corpus data. Collocation, as a psychological phenomenon, is formed by the cognitive process of priming. A speaker's mental representation of language is affected by his or her implicit memory of all previously experienced instances of language use. Since every speaker has a partially different experience of language use, words can be differently primed for each person. The notion of semantic priming is used to discuss the way a "priming" word may provoke a particular "target" word (Hoey 2005:2).

**Task 4: Reveal the semantic associations the lexical unit TO TAKE is likely to be primed for.**

Four general patterns of collocations with TO TAKE may be observed: [TAKE+N], [TAKE+ADV], [TAKE+PRP], [TAKE+PRN]. The pattern [TAKE+N] stands out as the most applicable one in corpus among other patterns. And this is the pattern used to sketch those seven semantic associations the lexical unit TAKE is likely to be primed for. This classification of associations may be extended to more detailed interpretation revealing some possible semantic association subtypes, like semantic frames<sup>1</sup>:

**[MIND]:** {[analyzing], [reasoning], [learning], [obtaining], [performing], [leading]}

**[ATTITUDE]:** {[negative feeling], [positive feeling], [spiritual feeling]}

**[BEING]:** {[human being], [proper name], [common noun]}

**[ACTION]:** {[with tools], [with body parts], [status change], [position change]}

**[ENTITY]:** {[object], [transportation], [finance], [nourishment], [image], [attire], [interest], [form], [measure]}

**[SPACE]:** {[spatial unit], [geographic name]}

**[TIME]:** {[time unit], [month], [day], [year]}



Figure 3.

To conclude, corpus is the basis of thousands of academic articles, theses, and dissertations, and they form the backbone of courses on language and

<sup>1</sup> The theory of **Frame Semantics** relates linguistic semantics to encyclopaedic knowledge developed by Charles J. Fillmore. The basic idea is that the word activates or evokes a semantic frame of encyclopaedic meaning relating to the specific concept it refers to (or highlights, in frame semantic terminology). Words do not only highlight individual concepts, but also specify a certain perspective in which the frame is viewed (Fillmore 1982:37).

linguistics throughout the world, at all levels of instruction. The activities discussed in this paper are just a few instances of myriads of activities to exercise with the good use of corpus focusing different aspects of language skills. Using corpora in language teaching and learning is challenging. It may have a big potential especially in foreign language classrooms when having teachers trained on how to design suitable corpus-based tasks so as to help their students get exposed to a broader framework of language system.

#### REFERENCES

1. Fillmore Ch. Frame Semantics. The Linguistic Society of Korea (ed.), Linguistics in the Morning Calm, Seoul, Hanshin, 1982, pp. 11-137.
2. Hoey M. Lexical Priming: A New Theory of Words and Language. London: Routledge, 2005, pp. 2-13.
3. Hunston S. Corpus Linguistics. University of Birmingham, UK: Elsevier Ltd. 2006, pp. 235-247.
4. Kaplan R. The Oxford Handbook of Applied Linguistics. Publisher: Oxford University Press; 2<sup>nd</sup> edition, 2010, p. 548.
5. Kübler S., Zinsmeister H. Corpus Linguistics and Linguistically Annota. Bloomsbury, London, 2015, p. 14.
6. Leech G., Hundt M., Mair C., Smith N. Change in contemporary English: A grammatical study (Studies in English language). Cambridge: Cambridge University Press, 2009, p. 32.
7. Lindquist H. Corpus Linguistics and the Description of English. Edinburgh University Press, 2009, p. 105.
8. Sinclair J. Trust the Text: Language, Corpus and Discourse. London: Routledge. 2004, p. 2.
9. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991, p. 108.
10. Tognini-Bonelli E. Corpus Linguistics at Work. Amsterdam: John Benjamins, Publishing Company, Amsterdam and Philadelphia, 2001, p. 14.

#### ԼԻԼԻԹ ՄՈՎԱԵՍՅԱՆ - ԲԱՌԱՊԱՇԱՐԻ ՈՒՍՈՒՑՄԱՆ ԿՈՐՊՈՒՍԱՅԻՆ ՄԵԹՈԴԱԲԱՆՈՒԹՅՈՒՆ

*Հիմնաբառեր՝ կորպուս, կորպուսային լեզվաբանություն, քանակական վերլուծություն, որակական վերլուծություն, բառակապակցություն, իմաստային նրբերանգներ, ռեգիստր, ներակա հիշողության բառային հետք, հասկացության փոխաբերություն*

1980-ական թվականներին կորպուսային հեղափոխության արդյունքում կորպուսը դարձավ լեզվի ուսուցման արդյունավետ միջոցներից