

A Model of a Georgian-English Learner's Online Dictionary of Collocations (based on the principles of collocations dictionaries and I. Mel'čuk's theory of Lexical Functions)

Salome Tchighladze
Ilia State University
salome.tchighladze.1@iliauni.edu.ge

Abstract

The present article aims to propose a model of a Georgian–English online learner's dictionary of collocations based on the principles of collocations dictionaries and Igor Mel'čuk's theory of lexical functions. It presents a model of dictionary entries in which collocations are organized according to parts of speech and enriched with relevant lexical functions and authentic usage examples. The proposed approach seeks to develop an active, text-synthesis-oriented dictionary that will contribute to Georgian learner lexicography and support the effective use of lexicographic resources in both teaching and research contexts.

In the process of foreign language acquisition, lexical competence extends beyond knowledge of the meanings of individual words. Effective communication largely depends on the ability to use correct and natural word combinations, enabling language learners not only to comprehend texts but also to produce well-formed sentences. From this perspective, collocations are of particular importance, as they constitute relatively stable and coherent combinations of lexemes. Knowledge of collocations contributes significantly to the naturalness and fluency of speech and distinguishes native-like language use from learners' language.

Although collocation dictionaries are well established in English lexicography, no comparable resource currently exists in Georgian-English learner lexicography. Most existing bilingual dictionaries focus primarily on the translation of individual words and devote little attention to their collocational potential. This creates a significant challenge for language learners, particularly in the context of text production and synthesis.

Keywords: dictionary of collocations; Mel'čuk's lexical functions; model of the dictionary entry; Georgian-English lexicography

1 Introduction

As early as the 1920s and 1930s, Harold Palmer and Albert Hornby emphasized the importance of word combinations in the process of language learning (Cowie 1999: 56). However,

this linguistic phenomenon as a term was first introduced into linguistics in the 1950s by the British linguist John Rupert Firth (1957). Following Firth, John Sinclair's research significantly expanded and deepened the understanding of the practical importance of collocations, which came to be regarded as one of the main principles of corpus-based research and lexical analysis.

In this regard, the COBUILD project (Collins Birmingham University International Language Database) deserves special mention, as it was the first electronic corpus. It brought together examples of the natural use of English, on the basis of which the first corpus-based dictionary – the *COBUILD English Dictionary* – was compiled. Through this project, Sinclair shifted the focus from grammatical rules to the issue of word combinability and collocational behavior (Sinclair 1991).

The further development of corpus linguistics provided an even stronger methodological foundation for the study of collocations. Particularly noteworthy in this context is the creation of the Sketch Engine platform, one of the major achievements of corpus linguistics. The program was developed in 2004 by Adam Kilgariff (Kilgariff et al. 2004) and his team and currently includes around 800 corpora covering more than 100 languages. It is equipped with advanced search tools, and it was this platform that introduced collocational search functionality into corpus analysis (Kilgariff & Rundell 2002). The Macmillan Dictionary was the first to make use of this functionality, adding the most frequently used collocations to learner dictionary entries (Rundell & Kilgariff 2011).

The issue of collocations is also central to Igor Mel'čuk's work, both in his theory of the Explanatory Combinatorial Dictionary (Mel'čuk 2012) and in his theory of lexical functions (Mel'čuk 1996). According to the theory of lexical functions, words are interconnected within different semantic contexts, and these relationships can be described as lexical functions. The theory identifies around 70 lexical functions, which Mel'čuk classifies into different categories, including syntagmatic and paradigmatic functions. Syntagmatic lexical functions primarily describe relationships between different parts of speech and specify the nouns, adjectives, adverbs, or verbs that typically co-occur with a given lexeme. Thus, these functions provide semantically expected and appropriate word combinations. Paradigmatic lexical functions (such as synonyms, antonyms, conversives, and others) describe words that belong to the same semantic category and can potentially replace a given lexeme. Syntagmatic lexical functions, by contrast, are used together with the given word in actual usage.

Within the framework of the present research, a new methodology was developed following a review of the theoretical literature on collocations, the principles of compiling collocations dictionaries, and Igor Mel'čuk's theory of lexical functions. According to this methodology, collocations in the entries of the Georgian-English learner's dictionary of collocations are classified according to parts of speech and are supplemented with the corresponding lexical functions. This approach enables users to gain a detailed understanding of the combinatorial potential of lexical units and, when necessary, to construct well-formed syntactic structures independently. From Mel'čuk's system of lexical functions, only those syntagmatic and paradigmatic functions were selected that are relevant to Georgian-English lexicographic goals within the context of a learner's dictionary of collocations. The article provides a detailed discussion of a noun-entry model compiled according to this methodology.

The chosen approach is considered to facilitate the creation of an active, text-synthesis-oriented dictionary, which significantly enhances the effective use of lexicographic resources in both teaching and research contexts.

Each dictionary entry includes not only collocations but also authentic usage examples accompanied by English translations, thereby providing language learners with a realistic understanding of how particular words function in everyday communication. The structure of the dictionary entry is based on the principles of collocations dictionaries employed in the *Oxford Collocations Dictionary* and the *Macmillan Collocations Dictionary for Learners of English*. In these dictionaries, collocations within each entry are grouped according to parts of speech, and in some cases extended phrases are also provided, which makes this type of dictionary particularly practical for language-learning purposes. A collocation-based dictionary model enriched with lexical functions represents a novel approach in the field of Georgian-English learner lexicography.

2 Collocation Dictionaries as a Lexicographic Genre

Corpus linguistics gave new impetus to collocation dictionaries as an independent lexicographic genre. Such dictionaries describe not only individual words and their definitions, but also collocational behavior – the ways in which words combine with one another in different contexts (Moon 1998). Collocation dictionaries are largely based on large-scale language corpora, which makes such dictionaries more reliable and empirically grounded (Sinclair 1991; Stubbs 2001).

The first collocations dictionary is generally considered to be *The BBI Combinatory Dictionary of English*, compiled by Morton Benson, Evelyn Benson, and Robert Ilson, first published in 1986 (Benson, Benson, & Ilson 1986). The BBI dictionary distinguishes between grammatical and lexical collocations. It does not include classical idioms, but rather focuses on word combinations in which the meanings of the components remain at least partially transparent. Grammatical collocations consist of a headword combined with prepositions or other grammatical elements, whereas lexical collocations mainly involve combinations of nouns, verbs, adjectives, and adverbs. It is noteworthy that the dictionary conforms to Firth's theoretical principles (Poulsen 1991).

In fact, the BBI dictionary was not the first of its kind. Its predecessors include Albrecht Reum's *Dictionary of English Style* (1961), which was intended to assist German learners of English in their language learning. Another work that can be classified within this genre is Joseph Isaac Rodale's *The Word Finder* (1947), which provides information useful for young writers.

The publication of the BBI dictionary marked an important milestone in collocations lexicography, as it established a standard for the presentation of collocations in dictionaries. From the 1980s onwards, other collocations dictionaries began to appear. Among the relatively newer generation of such dictionaries is the *Oxford Collocations Dictionary for Students of English*, first published in 2002 and based on the British National Corpus (BNC). Another example is the *Longman Collocations Dictionary and Thesaurus* (2013), which combines the principles of a collocations dictionary and a thesaurus. It provides users not only with infor-

mation on collocations, but also on synonym sets and semantic relations. The *Macmillan Collocations Dictionary* is another representative of this genre. These modern dictionaries have become important resources not only for language learners, but also for translators, lexicographers, and linguists in general.

2.1 The Oxford and Macmillan Collocations Dictionaries

In order to develop the research methodology, initially the principles underlying the compilation of existing collocation dictionaries and the methods used to present collocations in them were examined. For this purpose, two dictionaries were selected. The first is the *Oxford Collocations Dictionary* (online edition) (see Figure 1). In this dictionary, collocations are organized according to parts of speech: in the case of nouns, the dictionary lists the adjectives that typically modify them, as well as nouns that are used attributively with the given noun, and verbs that either precede or follow the noun. In some cases, collocations are accompanied by examples in the form of full sentences or extended phrases.

Below is an example from the aforementioned dictionary illustrating the noun entry for the word *book*. It's collocates are grouped according to parts of speech. For example, the entry presents adjectives that frequently occur in collocation with the noun *book*, such as *latest*, *new*, *recent*, *forthcoming*, *rare*, etc. It also provides an authentic example sentence, for example: *There's nothing like curling up with a mug of tea and a good book*. In some cases, the position of the headword within the collocation is also indicated. For example, the entry presents verbs that occur before the noun *book*, such as *look at* and *read*, as well as verbs that occur after it, including *appear* and *come out* (see Figure 1).

Online OXFORD Collocation Dictionary

book *noun*

¹ **for reading**

ADJ. latest, new, recent | forthcoming | hardback, paperback | printed *one of the earliest printed books* | rare | second-hand | delightful, excellent, fascinating, fine, good, great, interesting, remarkable, useful *There's nothing like curling up with a mug of tea and a good book.* | famous, important, influential | controversial *a controversial book about the royal family* | favourite *a survey to find the nation's favourite children's book* | library | set *'Emma' was one of our set books for A level.* | children's, comic, cookery, guide, hymn, phrase, picture, prayer, reference, school, story, text (also textbook), travel | address, autograph, cheque, exercise, log, order, phone/telephone, sketch

QUANT. copy *How many copies of the book did you order?*

VERB + BOOK be deep/engrossed/immersed in, flick/skim through, look at, read | look up from *She looked up from her book and smiled at him.* | co-author, write | bring out, publish, put out | reprint | edit, proofread, revise | translate | illustrate | bind | ban, censor | dedicate, inscribe *The book is dedicated to his mother. Her name was inscribed in the book. The collector had many books inscribed to him by famous authors.* | review | borrow, have out, take out (= from a library) *How many books have you got out?* | return, take back (= to a library) | renew *Do you want to renew any of your library books?* | stock | plagiarize

BOOK + VERB appear, come out *His latest book will appear in December.* | be/go out of print

BOOK + NOUN title | shop (also bookshop) | review, reviewer | club

PREP. in a/the ~ *These issues are discussed in his latest book.* | ~ about/on *She's busy writing a book on astrology.* | ~ by a book by Robert Grout | ~ for a book for new parents | ~ from a new book from the publishing company, Bookworm | ~ of a book of walks in London 2 books company records

ADJ. account

VERB + BOOK audit, do, keep *She does the books for us.*

Figure 1: *Oxford Collocations Dictionary* (online), entry "book".

The second dictionary selected for the same purpose is the *Macmillan Collocations Dictionary for Learners of English* (see Figure 2). Like the *Oxford Collocations Dictionary*, the Macmillan dictionary organizes collocations according to parts of speech, which makes it easier for users to understand and remember related word combinations. However, the Macmillan dictionary not only groups collocations by parts of speech, but also separates collocates with similar meanings into distinct groups (see Figure 2).

The example from the *Macmillan Collocations Dictionary* illustrates the adjective entry for the word *impossible*. The collocations are organized according to the part of speech with which the headword combines.

In the section labeled *adv + ADJ*, the dictionary presents adverbs that frequently occur with the adjective *impossible*. For example, intensifying adverbs such as *absolutely*, *completely*, and *totally* are used in combinations like *absolutely impossible* and *totally impossible*. Some collocations are also illustrated through example sentences, such as: *I realized that it was absolutely impossible to get a job without a proper work permit*. The dictionary further groups collocations according to semantic similarity. For instance, adverbs such as *almost*, *nearly*, and *virtually* are grouped together because they express partial possibility, as in *almost impossible*. Similarly, adverbs such as *obviously* and *clearly* indicate certainty, producing collocations such as *obviously impossible* (see Figure 2).

impossible ADJ

unable to be done or to happen

- **adv+ADJ** completely **absolutely, completely, downright, quite, simply, totally, utterly** *I realized that it was absolutely impossible to get a job without a proper work permit.*
- ▶ **almost all but, almost, nearly, practically, virtually, well-nigh** *It's all but impossible to have a rational discussion about the issue. • The prison cells were so overcrowded that it was almost impossible for prisoners to sit down.*
- ▶ **obviously** **clearly, manifestly, obviously** *Bringing a life-sized tree onto the stage was obviously impossible.*
- ▶ **apparently** **apparently, probably, seemingly** *We should be grateful for his tireless efforts against seemingly impossible odds.*
- ▶ **in a particular way** **humanly, logically, logistically, mathematically, morally, physically, politically, technically** *It would have been physically impossible to read them all even if you did nothing else for a week.*
- **v+ADJ** **appear, be, become, be rendered, look, prove, seem, sound** *The game was rendered impossible by the wintry conditions. • It proved impossible to reach an agreement. • Telling your life story in 20 seconds sounds impossible.*

Figure 2: *Macmillan Collocations Dictionary for Learners of English*, entry “impossible”.

Another feature that distinguishes the *Macmillan Collocations Dictionary* from the *Oxford Collocations Dictionary* concerns the way semantic groups are labeled at the beginning of dictionary entries. In the *Oxford Learner's Dictionary*, the position of the headword within a collocation is indicated by patterns such as verb + headword, whereas in the *Macmillan Dictionary*, both elements of the collocation are explicitly identified by their parts of speech (e.g., adj + N).

2.2 Lexical Functions of Igor Mel'čuk

The theory of lexical functions is part of Igor Mel'čuk's Meaning-Text Theory (MTT) and describes the semantic and combinatorial relationships between lexical units. A lexical function (LF) is a key word associated with a specific lexeme that expresses its typical collocational behavior. It has no independent meaning and acquires its functional value only in relation to a given lexical item (Mel'čuk 1996).

The theory of LFs is presented as the principal method for compiling the Explanatory Combinatorial Dictionary (ECD). Unlike traditional dictionaries, the ECD is focused not only on explaining the meaning of a word, but also on describing its use in different contexts. This approach is particularly important for the creation of an active, text-synthesis-oriented dictionary, as it provides users with information about a word's typical collocational patterns (Mel'čuk 2012).

Mel'čuk's theory describes approximately 70 LFs, which are divided into syntagmatic and paradigmatic functions. Syntagmatic LFs describe typical word combinations and indicate the adjectives, verbs, or adverbs that commonly occur with a given lexeme. For example, the LF *Magn* denotes an intensifying function (*to condemn* - **strongly** *condemn*, *patience* - **infinite** *patience*), the LF *Son* refers to the characteristic sound associated with a lexeme (a dog **barks**, wind **howls**), and the LF *Figur* expresses figurative or metaphorical usage (**wall** of fog, **flames** of passion). Paradigmatic lexical functions, by contrast, describe words that belong to the same semantic category, such as synonyms, antonyms, collective nouns and some others. For example, *Syn* expresses a synonymic relationship (*to telephone* - *to phone*), *Anti* denotes a semantic opposite (*victory* - *defeat*), and *Mult* refers to a collective noun (*ship* - *fleet*).

3 The Structure of a Dictionary Entry

After reviewing the models of collocations dictionaries and the principles underlying their compilation, a new approach to the presentation of collocations in the dictionary was developed. Specifically, collocations in the entries of the Georgian-English learner's dictionary are presented according to a mixed principle: each entry follows the structure of collocations dictionaries by grouping the collocations of a headword according to parts of speech, while also being enriched with the syntagmatic and paradigmatic LFs proposed by Igor Mel'čuk.

Using this method, entry models were developed for nouns, adjectives, and verbs. However, the present article focuses exclusively on the noun-entry model. Below is the dictionary entry for the noun წვიმა/ts'vima 'rain' (see Figure 3).

3.1. Fields of a dictionary entry

The dictionary entry for *rain* consists of three fields. The first field includes collocations grouped according to parts of speech: adjective + rain, rain + noun, and rain + verb (e.g. constant rain, beating of rain, rain started, etc.). Syntagmatic lexical functions in the entry are marked in red (see Figure 3). *Magn* is one of the syntagmatic lexical functions identified by Igor Mel'čuk (as mentioned above), meaning an “intensifier” (e.g. pelting rain). When the cursor is placed over the term, the meaning of the lexical function appears on the screen (see Figure 4).

Son denotes the syntagmatic lexical function referring to the “sound produced by rain” (e.g. beating of rain, drumming of rain), while *Figur* refers to the lexical function “figurative” (e.g. rain curtain).

წვიმა rain

შესიტყვებები:

(ზედსართავი + წვიმა) გადაუღებელი წვიმა constant rain, incessant rain; გაუთავებელი წვიმა endless rain; ხანმოკლე წვიმა brief shower; **Magn** კოკისპირული წვიმა pelting rain; შხაპუნა წვიმა heavy rain / downpour; (წვიმა + არსებითი) **Son** წვიმის შხაპუნი beating of rain; წვიმის რაკარუკი drumming of rain; **Figur** წვიმის ფარდა rain curtain; (წვიმა + ზმნა) წვიმა დაიწყო rain started; წვიმამ დაუშვა the rain fell / came down; წვიმამ გადაიღო rain stopped

პარადიგმატული ლექსიკური ფუნქციები >

იდიომ(ებ)ი:

სცენაზე წვიმად წამოვიდა ყვავილები flowers showered down upon the stage

Figure 3: The entry for *rain* in the Georgian-English learner's dictionary of collocations.

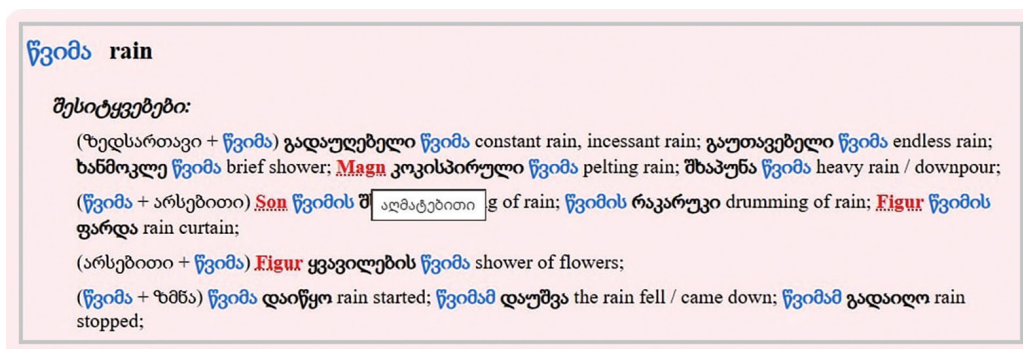


Figure 4: A fragment of the entry for *rain* in the Georgian-English dictionary of collocations.

The second field of the entry is devoted to paradigmatic lexical functions. This section can be toggled on or off (see Figure 5). It includes synonyms of *rain* (e.g. fine drizzle, heavy downpour), a hypernym (meteorological phenomenon), co-hyponyms (snow, hail, wind, storm), and derivatives (rain-washed, rainy, etc.) (see Figure 5). Although paradigmatic functions are not collocations, their presentation in this format was considered important in a learner's dictionary.

The third field is the field of idioms.

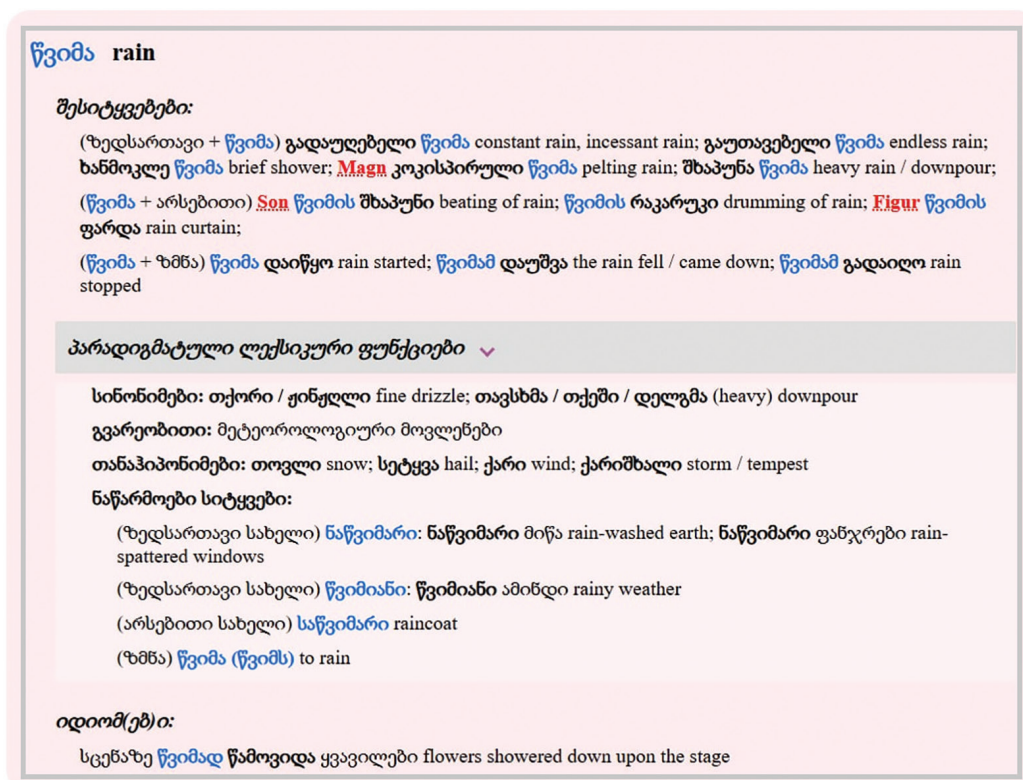


Figure 5: The full entry for *rain* with toggled on field of paradigmatic lexical functions.

3.2 Related words

In addition to the fields discussed above, some entries also include a field of related words associated with particular syntagmatic lexical functions. These fields can be toggled on or off. For example, consider the entry for მგელი / *mgeli* ‘wolf’, which contains two lexical functions: *Mult*, meaning “collective noun,” and *Son*, denoting “the sound produced by a wolf” (see Figure 6). This entry therefore includes two corresponding fields of related words linked to the lexical functions *Mult* and *Son*. When activated, the first field displays words associated with the lexical function *Son*, i.e. sounds produced by a wolf. These include expressions referring to sounds produced by other animals and birds, such as the grunting of a pig, cackling of a hen, whinnying of a horse, and twittering of birds, etc (see Figure 7).

The second field of related words is associated with the lexical function *Mult*, meaning “a collective noun,” or “a pack of wolves” in this case. When toggled on, it displays collective nouns for different animals and birds, such as a flock of sheep, a herd/drove of horses, a herd of swine/pigs, a flock of deer, and a flock of birds (see Figure 7).

მგელი wolf

შესიტყვებები:
 (ზედსართავი + **მგელი**) რუხი მგელი a gray wolf; ჟღალი მგელი a red wolf; ზებერი მგელი an old wolf, an elder wolf; ვეება მგელი a huge wolf; კავკასიური მგელი a Caucasian wolf;
 (არსებითი + **მგელი**) მამა მგელი a wolf-father; ძე მგელი a she-wolf / bitch wolf;
 (მგელი + არსებითი) მგლის ლეკვები wolf cubs; მგლის ჯილაგი wolf-stock; მგლის კვალი, მგლის ნაკვალავი a trail of a wolf; **ჰიი** მგლის ყმუილი howl of a wolf; **Mult** მგლის ხროვა a wolf-pack / pack of wolves;

Son დაკავშირებული სიტყვები >

Mult დაკავშირებული სიტყვები >

პარადიგმატული ლექსიკური ფუნქციები >

იდიომი:
 მგლის მუხლი გამოაბა it gave him a wolf's running powers; მგელი ცხვრის ტყავში a wolf in sheep's clothing.

Figure 6: The entry for *wolf* in the Georgian–English dictionary of collocations.

მგელი wolf

შესიტყვებები:

(ზედსართავი + მგელი) რუხი მგელი a gray wolf; ჟღალი მგელი a red wolf; ბებერი მგელი an old wolf, an elder wolf; ვეება მგელი a huge wolf; კავკასიური მგელი a Caucasian wolf;

(არსებითი + მგელი) მამა მგელი a wolf-father; ძე მგელი a she-wolf / bitch wolf;

(მგელი + არსებითი) მგლის ლეკვები wolf cubs; მგლის ჯილაგი wolf-stock; მგლის კვალი, მგლის ნაკვალევი a trail of a wolf; **Son** მგლის ყმული howl of a wolf; **Mult** მგლის ხროვა a wolf-pack / pack of wolves;

Son დაკავშირებული სიტყვები ▼

ღორის ღრუტუნი grunting of a pig;
 ქათმის კაკანი cackling of a hen;
 ცხენის ჭიხვინი whinnying of a horse;
 ჩიტების ჭიკჭიკი twittering of birds;
 ძაღლის ყევა barking of a dog;
 ძაღლის წკაფწკავი yelping of a dog.

Mult დაკავშირებული სიტყვები ▼

ცხვრის ფარა flock of sheep;
 ცხენების რემა / ჯოგი herd / drove of horses;
 ღორების კოლტი herd of swine / pigs;
 ირმების ჯოგი flock of deer;
 ფრინველების გუნდი flock of birds;

Figure 7: The entry for *wolf* with toggled on fields of words related to the lexical functions *Son* and *Mult*.

The entry for *wolf* also includes a field of paradigmatic lexical functions, comprising a generic term, co-hyponyms, derivatives and so on (see Figure 8).

მგელი wolf

შესიტყვებები:

(ზედსართავი + მგელი) რუხი მგელი a gray wolf; ჟღალი მგელი a red wolf; ზებერი მგელი an old wolf, an elder wolf; ვეება მგელი a huge wolf; კავკასიური მგელი a Caucasian wolf;

(არსებითი + მგელი) მამა მგელი a wolf-father; ძე მგელი a she-wolf / bitch wolf;

(მგელი + არსებითი) მგლის ლეკვები wolf cubs; მგლის ჯილაგი wolf-stock; მგლის კვალი, მგლის ნაკვალევი a trail of a wolf; **სი** მგლის ყმული howl of a wolf; **მულა** მგლის ხროვა a wolf-pack / pack of wolves;

Son დაკავშირებული სიტყვები >

Mult დაკავშირებული სიტყვები >

პარადიგმატული ლექსიკური ფუნქციები v

გვარუბითი: ნადირი beast; ცხოველი animal

თანაჰიპონიმები: მელია fox; ყარსალი polar fox; ტურა jackal; კოიოტი coyote; აფთრისებრი ძაღლი African wild dog

ნაწარმოები სიტყვები:

(ზედსართავი სახელი) მგლისფერი wolf-coloured;

(ზმნიზედა) მგლურად wolfishly;

მგელივით in wolf-fashion; like a wolf; მგელივით დაბუმბულვდა prowled / trotted like a wolf;

მგელივით შიოდა he was ravenously hungry;

Figure 8: The entry for *wolf* with the paradigmatic lexical functions toggled on.

4 Sources of the *Georgian-English Online Learner's Dictionary of Collocations*

The main sources of the *Georgian-English Online Learner's Dictionary of Collocations* are the following:

The English-Georgian Parallel Corpus.¹ The corpus consists of scientific texts and works of fiction, comprising texts available in both Georgian and English (Margalitadze, Meladze, & Furtskhvanidze 2022). At present, the corpus contains up to 158 text groups, 6,062 text sets, 668,048 manually aligned sentence pairs and 17 million tokens. It also incorporates material from *The Comprehensive English–Georgian Dictionary*. Since 2024, the corpus has been available on a new platform that includes a collocational search function. This platform proved particularly useful, as it allows collocations to be retrieved together with their English translations (see Figure 9), thereby serving as the primary source for the present study.

1 <https://enkacorporus.iliauni.edu.ge/en/>



ინგლისურ ქართული
პარალელური კორპუსი

EN

დეტალური ძიება

კოლოკაციები

გამოყენების ინსტრუქცია

პროექტის შესახებ

კონტაქტი

წვიმა

მოთხოვნილი სიტყვის სიხშირე: 181

კოლოკაციები

ნაჩვენებია 1 - 105 ჩანაწერი 105 ჩანაწერიდან

1

მარცხნივ					მარჯვნივ				
კოლოკანტი	სიხშირე		LOG DI	MI	კოლოკანტი	სიხშირე		LOG DI	MI
	კოლოკაცია	კოლოკანტი სიტყვა				კოლოკაცია	კოლოკანტი სიტყვა		
კოკისპირული	11	13	10.86	16.29	წამოვიდა	10	149	9.96	12.63
სანამ	8	4348	5.86	7.44	არ	7	72324	1.66	3.19
თავსხმა	7	37	10.04	14.12	გადაიღებს	7	12	10.21	15.75
და	5	280790	-0.78	0.75	დაიწყო	5	2470	5.95	7.58
რომ	3	86512	0.18	1.71	და	5	280790	-0.78	0.75
ლაპარაკობდა	3	424	7.34	9.38	შეწყდა	5	200	8.75	11.2
ეტყობა	3	505	7.16	9.13	მოღის	4	791	7.08	8.9
მხოლოდ	2	14592	2.15	3.69	გადაღებას	4	19	9.36	14.28
ჟუჟუნა	2	6	8.45	14.94	მოდიოდა	3	406	7.39	9.45
ფანჯრებიდან	2	61	8.08	11.6	ცრიდა	2	10	8.42	14.2
როცა	2	10740	2.59	4.14	კი	2	27067	1.27	2.8

Figure 9: English-Georgian Parallel Corpus - Collocates of the Word *rain*.



ინგლისურ ქართული
პარალელური კორპუსი

EN

დეტალური ძიება

კოლოკაციები

გამოყენების ინსტრუქცია

პროექტის შესახებ

კონტაქტი

<

Figure 10: English-Georgian Parallel Corpus - Collocates of the Word *rain* - detailed search function.

*The Corpus of Georgian Neologisms*² (Laluashvili & Margalitadze 2025) - a resource developed in 2024 at Ilia State University. It contains textual material from Georgian online newspapers and magazines, television and other media websites, as well as the websites of governmental agencies and non-governmental organizations. Like the *English-Georgian Parallel Corpus*, this corpus also includes a collocational search function.

2 <https://neologism.iliauni.edu.ge/>

*The Georgian National Corpus*³ (Gippert & Tandashvili 2015). The National Corpus of the Georgian Language includes several different corpora, among them the corpora of Old and Middle Georgian, the Dialect Corpus of the Georgian Language, the Corpus of Political and legal texts, also the Megrelian and Svan corpora. Within the framework of the present study, the Reference Corpus of the Georgian Language (217,130,870 tokens) was used to search for collocations. The aforementioned corpus and the Georgian Neologism Corpus were used as supplementary sources, while the primary source for identifying collocations was the *English-Georgian Parallel Corpus*.

The research showed that, these sources are sufficient for obtaining the relevant data necessary for the compilation of a *Georgian-English Online Learner's Dictionary of Collocations*.

5 Conclusion

The article examines the microstructure of a *Georgian-English Online Learner's Dictionary of Collocations*, which is based on the compilation principles of the Oxford and Macmillan collocations dictionaries, as well as on Igor Mel'čuk's theory of lexical functions.

A collocation is a fixed or semi-fixed combination of words that occurs naturally and recurrently in a language. The study of collocations is important because they reflect the lexical and semantic characteristics of a language, as well as the natural relationships between words. In modern lexicography, dictionaries of collocations have developed into an independent genre, as they play a significant role in language teaching, translation, and linguistic research. From this perspective, the Georgian language remains insufficiently described, therefore, the systematic study of collocations and the compilation of a corresponding dictionary are of particular importance for revealing the lexical potential and usage patterns of the Georgian language for the study process.

The proposed approach involves grouping collocations within dictionary entries according to the principles used in the Oxford and Macmillan collocations dictionaries and describing them through the application of lexical functions.

The research results demonstrate both the effectiveness of the proposed mixed approach in the context of learner lexicography, and the role and reliability of existing lexicographic resources used in the study, the *English-Georgian Parallel Corpus*, the *Corpus of Georgian Neologisms*, and the *Georgian National Corpus*. Moreover, collocations were retrieved semi-automatically, which indicates the potential for further automation of the process; work in this direction would significantly accelerate the compilation of the dictionary.

The study also addresses an important theoretical issue, namely the extent to which Igor Mel'čuk's theory of lexical functions can be adapted and applied in the context of the Georgian language. The results confirm that this theoretical framework can be successfully integrated into Georgian learner lexicography.

3 <http://gnc.gov.ge/gnc/page>

References

- Benson, M., E. Benson, & R. Ilson. (1997). *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.bbi> [7/01/2025].
- Corpus of Neologisms*. Accessed at: <https://neologism.iliauni.edu.ge/> [30/11/2025].
- Cowie, A. P. (1999). *English Dictionaries for foreign learners: A History*. Oxford University Press.
- English–Georgian Parallel Corpus* (new platform). Accessed at: <https://enkacorpor.iliauni.edu.ge/> [30/11/2025].
- Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930-55. In *Studies in Linguistic Analysis*, by John R. Firth, pp. 1-31.
- Firth, J. R. (1957). *Papers in linguistics, 1934 – 1951*. London: Oxford University Press.
- FreeCollocation.com. (n.d.). *Free English collocations dictionary*. Accessed at: <https://www.freecollocation.com/> [7/01/2025].
- Georgian National Corpus*. Accessed at: <http://gnc.gov.ge/gnc/page?page-id=gnc-main-page> [9/01/2023].
- Gippert, J., & M. Tandashvili. 2015. Structuring a diachronic corpus: The Georgian National Corpus project. In *Historical Corpora: Challenges and Perspectives. Corpus Linguistics and Interdisciplinary Perspectives on Language*, Vol. 5, edited by Jost Gippert and Ralf Gehrke, 305–22. Tübingen: Narr.
- Kilgarrieff, A., & M. Rundell. (2002). Lexical Profiling Software and its Lexicographic Applications: A Case Study. In *Proceedings of the Euralex 2002 Conference, ELX02-090, 13-17 August 2002*. Denmark: Center for Sprogteknologi, pp. 807-818. Available at: <https://euralex.org/publications/lexical-profiling-software-and-its-lexicographic-applications-a-case-study/> [7/01/2025].
- Kilgarrieff, A., P. Rychlý, P. Smrž, & D. Tugwell. (2004). The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress, 6-10 July 2004*. France: Université de Bretagne-Sud, Faculté des lettres et des sciences humaines, pp. 105-115. Available at: <https://euralex.org/publications/the-sketch-engine/> [7/01/2025].
- Laluashvili, T. & T. Margalitzadze. (2025). Semi-Automatic Detection of New Words in Modern Georgian. *Lexikos* 35 (1), pp. 399-413. <https://doi.org/10.5788/35-1-2044>.
- Longman Collocations Dictionary and Thesaurus. (2013). Pearson Education Ltd., ISBN 9781408252253.
- Macmillan Collocations Dictionary. (2010). *Macmillan Collocations Dictionary for Learners of English*. Macmillan Education.
- Margalitzadze, T., G. Meladze, & Z. Purtskhvanidze. (2022). English–Georgian Parallel Corpus and Its Use in Georgian Lexicography. In *Lexikos*, 32(2), pp. 158-176. <https://doi.org/10.5788/32-2-1701> [7/01/2025].

- Mel'čuk, I. (1996). Lexical functions in Lexicography and Natural Language Processing. In *Lexical Functions: A Tool for the description Lexical Relations in the Lexicon*, pp. 37-102.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Year-book of Phraseology* 3(1): 31–56. <https://doi.org/10.1515/phras-2012-0003>.
- Poulsen, S. (1991). The BBI Combinatory Dictionary of English: A Guide to Word Combinations. Compiled by Morton Benson, Evelyn Benson and Robert Ilson. John Benjamins Publishing Company/Munksgaards Ordbøger, Amsterdam/Philadelphia 1986. In *HERMES - Journal of Language and Communication in Business*, 4(7), pp. 129–137. <https://doi.org/10.7146/hjlc.v4i7.21483> [7/01/2025].
- Reum, A. (1961). *Dictionary of English Style*. Philosophical Library.
- Rodale, J. I. (1947). *The Word Finder*. Funk & Wagnalls.
- Rundel, Michael, & Adam Kilgariff. 2011. Automating the creation of dictionaries: where will it all end? In *A taste for Corpora: In honor of Sylviane Granger*, edited by Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin, and Magali Paquot, pp. 257–82. <https://doi.org/10.1075/scl.45.15run> [7/01/2025].
- Rundell M., & G. Fox. (2002). *Macmillan English Dictionary : For Advanced Learners*. London: Macmillan.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell Publishers.