

Bridging the Pragmatic Gap: Representing Georgian Discourse Markers in Lexicographic Resources

პრაგმატული ხარვეზის დაძლევა: ქართული დისკურსის მარკერების წარმოდგენა ლექსიკოგრაფიულ რესურსებში

ელენე ჯივალაგოვი

თბილისის სახელმწიფო უნივერსიტეტი

Elene Jivalagovi

Tbilisi State University

აბსტრაქტი

მოცემული კვლევა იკვლევს, თუ როგორ არის წარმოდგენილი ქართული დისკურსის მარკერები (დმ) ონლაინ ლექსიკონებში, კონკრეტულად კი ერთმონოლინგვალურ ქართულ განმარტებით ლექსიკონსა და ერთ ორენოვან ქართულ-რუსულ ლექსიკონში. ამ რესურსების შერჩევა განპირობებულია ქართულის მომხმარებლებსა და შემსწავლელებში მათი ფართოდ გამოყენებით და საუბრის ენის ლექსიკის მოცვით. მიზანია ქართული პრაგმატული მარკერების (ანუ დისკურსის მარკერების) ლექსიკოგრაფიული დამუშავების ხარისხისა და გამოყენებადობის შეფასება მეტალექსიკოგრაფიული პერსპექტივიდან. ნაშრომი ადარებს ერთენოვან და ორენოვან ჩანაწერებს შვიდი ფუნქციური კატეგორიის მიხედვით: დამატება, მიზეზი/შედეგი, თანმიმდევრობა, კონტრასტირება, შედარება, მაგალითით დადასტურება და დასკვნა. დისკურსის მარკერები (DM) შეიძლება გამოჩნდეს როგორც ცალკეული სიტყვები, ნაწილაკები ან მოკლე ფრაზები. გაანალიზებული მონაცემთა კატეგორიები მოიცავს გრამატიკულ ინფორმაციას (მეტყველების ნაწილი, ფორმა), სათარგმნელ ეკვივალენტებს, შესიტყვებს, გამოყენების მაგალითებსა და პრაგმატულ ფუნქციებს. ეს ელემენტები ინტეგრირებულია თითოეული მარკერის ერთიანი ლექსიკური პროფილის შესაქმნელად. ანალიზი აფასებს არა მხოლოდ ლექსიკურ შინაარსს, არამედ სამიზნე აუდიტორიისთვის - ქართველი მომხმარებლებისა და მეორე ენის შემსწავლელთათვის- ფუნქციურ შესაბამისობასაც. ლექსიკონების მიკროსტრუქტურისა და მაკროსტრუქტურის შესაფასებლად გამოიყენება მრავალთეორიული ჩარჩო, რომელიც აერთიანებს ჩარჩო სემანტიკას (ფილმორი), კორპუსის ლინგვისტიკას, კონტრასტულ ანალიზსა და თარგმანის თეორიას. მარკერების ემპირიული შერჩევა ეფუძნება ქართული ეროვნული

კორპუსის (ტანდაშვილი და გიპერტი, 2015) სიხშირული მონაცემებს, რაც ანალიზს უზრუნველყოფს გამოყენებაზე დაფუძნებულ საფუძველს.

საკვანძო სიტყვები: დისკურსის მარკერები, ლექსიკონები, ორენოვანი, ერთენოვანი, მეტალექსიკოგრაფია, ქართული

Keywords: discourse markers, dictionaries, bilingual, monolingual, metalexigraphy, Georgian

1. Introduction

Discourse markers (DMs) are words and expressions that facilitate natural, coherent, and pragmatically appropriate speech. In any language, these elements signal relationships between ideas, organise discourse, and convey speaker stance or attitude. They distinguish native-like fluency and are essential for everyday communication. Georgian is no exception. DMs connect utterances, structure narratives, and enhance textual or conversational flow, making speech more listener-friendly and natural.

In recent years, discourse markers have attracted considerable attention in linguistic research (e.g., Fraser, 1999; Martín Zorraquino & Portolés, 1999; Pons Bordería, 2001; Fischer, 2006). Early interest in the topic first emerged within the frameworks of text grammar and discourse analysis (van Dijk, 1973, 1978). In these foundational studies, discourse markers were characterised as particles that facilitate the coherent interpretation of texts by serving as instructions for connecting different propositions and organising argumentation. Consequently, they are generally regarded as functional rather than lexical units, since they convey procedural rather than conceptual information. Nevertheless, DMs play a significant role in both written and oral communication. In addition to linking and structuring segments of discourse, they can signal subjectivity or speaker attitudes and may even regulate interaction between participants (Fox Tree, 2015). They thus constitute fundamental textual elements that occupy an intermediate position between grammar and the lexicon.

Discourse markers are notoriously difficult to identify and classify (Cartoni et al., 2013). They may consist of single words or multi-word expressions and can belong to a variety of grammatical categories, including conjunctions, adverbs, and prepositional phrases. The most widely adopted method for organising DMs is based on their functional similarity. Common functional categories include additive connectives (e.g., also, furthermore), contrastive connectives (e.g., however, nevertheless), causal connectives (e.g., consequently, for this reason), and numerous other groupings.

Although various approaches to categorising DMs have been proposed in discourse studies (Fuentes Rodríguez, 1987; Fraser, 1999; Martín Zorraquino & Portolés, 1999; Pons Bordería, 2001), they have rarely been described in dictionaries with sufficient detail and precision—most likely because of their inherent complexity and highly discursive nature. Comprehensive catalogues or specialised dictionaries of DMs remain relatively scarce. In

Spanish, notable examples include Briz et al. (2008), and Holgado Lage (2017). For other languages, taxonomies have been developed in English (Knott, 1996) among others.

In contrast to the present research, which adopts a lexicographic perspective, most publications in the field of computational linguistics that deal with DMs are concentrated in the area most closely related to discourse analysis.

This paper investigates how two prominent Georgian online dictionaries treat commonly used discourse markers: the monolingual Explanatory Dictionary of the Georgian Language (Chikobava et al., 1950–1964) and the Georgian–Russian online dictionary (Akhvlediani & Topuria, 1990). It addresses the following research questions:

1. How are Georgian DMs defined, grammatically tagged, and exemplified in each dictionary type?
2. To what extent do monolingual and bilingual entries differ in revealing pragmatic functions?
3. What are the strengths and limitations of each approach for users?

The study is organised as follows: theoretical framework, methodology, data collection and analysis procedures, results and findings, and conclusion.

2. Theoretical Framework

This study adopts an integrated theoretical framework that combines Frame Semantics, Corpus Linguistics, Contrastive Analysis, and Metalexicography (with particular attention to Interface Theory). The framework is designed to address the specific challenges of analysing discourse markers in lexicographic resources: their pragmatic, functional, and context-dependent nature, as well as the differences between monolingual and bilingual dictionary microstructures.

Frame Semantics (Fillmore, 2014) provides the primary lens for understanding how dictionary entries convey pragmatic meaning. According to Fillmore, lexical items are interpreted against background knowledge structures or “frames”—schematic representations of typical situations, participants, and relations. In the context of DMs, Frame Semantics explains how example sentences in dictionary entries function as mini-contexts that evoke cause–effect chains, temporal sequencing, contrast, or other discourse relations. A well-designed entry does not merely list synonyms or translations; it must make the pragmatic frame explicit so that users can reconstruct the marker’s function in real communicative situations. This approach is especially powerful for metalexicographic analysis because it reveals whether a dictionary entry successfully maintains the discourse-organising role of a marker or leaves it underspecified. For instance, the monolingual entry for *თუმცა* (although) evokes a clear frame of concession, whereas the bilingual equivalent-driven entry does not.

Corpus Linguistics supplies empirical grounding by privileging actual language use over intuition. Frequency data, collocational patterns, and authentic examples drawn from corpora allow the researcher to evaluate whether the dictionary’s descriptions and illustrations reflect contemporary spoken and written Georgian usage. This component ensures that the comparison

between monolingual and bilingual entries is not purely structural but also usage-based, highlighting cases where dictionaries diverge from real-world pragmatic behaviour (e.g., omission of spoken particles such as *ბო*). Within the domain of contrastive linguistics, we systematically compare Georgian DMs with their Russian counterparts, identifying both convergent and divergent pragmatic mappings (e.g., *თუმცა* – *хотя/впрочем*). Recent corpus-based lexicographic research on Georgian learners of English has also highlighted the importance of multi-word discourse units for pedagogic dictionaries (Makhatadze, 2024).

Metalexicography and Interface Theory address the medium-specific aspects of online dictionaries. In the digital environment, lexicographic quality is no longer determined solely by content but also by accessibility, searchability, hyperlinking, and user-interface design (see, e.g., Tarp, 2008; Gouws, 2017). This perspective evaluates how the online platforms (lev.ge/ganmartebiti and the bilingual resource) present DM information—whether through layered senses, clickable examples, or cross-references—and how these features either facilitate or hinder users' ability to grasp pragmatic functions. For example, the complete omission of *ბო* from the bilingual resource constitutes not just a lexical gap but an interface failure, as users cannot retrieve information about a high-frequency spoken marker through search.

The study examines not only what information is provided about Georgian DMs but also how that information is structured, contextualised, and made usable for different audiences (native speakers, language learners, and translators).

3. Methodology

This study employs a qualitative metalexicographic design that combines comparative analysis of dictionary microstructure with systematic functional evaluation of discourse markers. The approach is grounded in the multi-theoretical framework outlined above and proceeds through three interrelated stages: (1) development of an analytical checklist, (2) systematic data collection from the selected dictionaries, and (3) comparative data analysis.

3.1 Analytical Procedure

A structured checklist was constructed to ensure consistent and replicable evaluation of each discourse marker entry. The checklist comprises four core dimensions:

Grammatical information: part of speech, morphological features, and inflectional patterns.

Semantic and pragmatic description: clarity of definitions, number of senses, synonyms/paraphrases, and explicit pragmatic functions (e.g., additive, contrastive, causal).

Usage evidence: quality and authenticity of example sentences, collocations, proverbs, and literary citations.

Lexicographic adequacy: presence and/or absence of procedural information, cross-references, and user-oriented features (especially relevant for online interfaces).

This checklist operationalises Frame Semantics by assessing whether entries successfully activate relevant discourse frames, while Corpus Linguistics informs judgements about the ecological validity of examples.

3.2 Instruments

Data were collected from two publicly accessible, high-use Georgian online dictionaries:

- **Monolingual resource:** Explanatory Dictionary of the Georgian Language (Chikobava et al., 1950–1964; <https://lev.ge/ganmartebiti>). Originally compiled as a foundational eight-volume print edition, its digital adaptation remains the most authoritative intralingual resource for Georgian. It was selected for its comprehensive semantic descriptions, inclusion of multiple layered senses, and rich illustrative material drawn from literature and proverbs.

Bilingual resource: the Georgian–Russian online dictionary, based on the second edition of the authoritative Georgian–Russian Dictionary by Akhvlediani and Topuria (1990). Unlike the monolingual resource, its microstructure is explicitly interlingual, focusing on translation equivalents rather than semantic definitions. The reliance on a print edition from 1990 is noted as a potential limitation, as it may not fully reflect contemporary spoken language or recent lexical innovations.

3.3 Data Collection Procedures

Seven high-frequency discourse markers were purposefully selected to represent the major functional categories identified in the literature (addition, cause/effect, sequencing, contrasting, comparison, exemplification, and conclusion). Selection was strictly empirically driven. Frequency queries were conducted using the Georgian National Corpus (Tandashvili & Gippert, 2015). The aim was to isolate and retrieve the highest-frequency discourse markers currently active in the language. Because discourse markers are a functional category rather than a strict morphological class, they are not assigned a unique part-of-speech (POS) tag within the GNC architecture. We filtered the lemmas tagged under the specific POS categories known to harbour pragmatic markers: conjunctions, particles, and adverbs. The retained discourse markers were ranked by their normalised frequency (occurrences per million words). The top seven markers, alongside the highest-frequency spoken particles, were selected for the lexicographic analysis. Table 2 (see Section 4) presents the frequency data for the selected markers.

The markers examined are:

აგრეთვე / ასევე (addition); რადგან / რადგანაც (cause/effect); შემდეგ (sequencing); თუმცა (contrasting); ვიდრე (comparison); მაგალითად (exemplification); საბოლოოდ (conclusion); ხო (confirmation / spoken interaction).

For each marker, the full dictionary entry was extracted, including headword, grammatical labels, definitions/senses, translation equivalents (in the bilingual resource), collocations, and all illustrative examples.

3.4 Data Analysis Procedures

A side-by-side comparative analysis was conducted for every marker. We evaluated monolingual and bilingual entries against the analytical checklist. To evaluate the bilingual entries, a strict criterion for functional equivalency was applied. Because discourse markers rarely have one-to-one literal translations, a Russian equivalent was deemed “adequate” only if it successfully activated the same pragmatic frame (e.g., temporal sequencing or logical conclusion) as the original Georgian marker in that specific context. Particular attention was paid to: congruence between the two resources in part-of-speech assignment and pragmatic framing; presence or absence of key functional information (e.g., spoken vs. written usage); degree of alignment with actual usage patterns as understood through Frame Semantics and corpus-informed expectations; and user-oriented strengths and limitations of each dictionary type.

The analysis remains qualitative, as the aim is metalexicographic insight rather than statistical generalisation.

4. Results and Findings

The monolingual and bilingual dictionaries adopt distinct yet complementary approaches. The monolingual dictionary prioritises intralingual explanation, providing rich Georgian-internal definitions, multiple senses, and authentic literary or proverbial examples. It excels at revealing the language’s internal logic and pragmatic nuances. The bilingual dictionary, by contrast, focuses on interlingual equivalence, offering Russian translations and occasional collocations, but frequently exhibits inconsistencies in part-of-speech labelling and pragmatic precision.

Table 1. Comparative Analysis of Selected Discourse Markers

Marker	Category	Monolingual entry	Bilingual entry
რადგან	Cause/effect	Conjunction (კავშ.); synonyms ვინაიდან, იმიტომ რომ; proverb-based causal frame.	Multiple Russian equivalents (так как, ибо, раз).
ასევე	Addition	Synonyms (ამგვარადვე); literary examples.	Single equivalent (также).

შემდეგ	Sequencing	Multiple senses (temporal, logical); diverse literary examples.	Long list of translations (после, потом, затем) plus phrases (და ასე შემდეგ).
თუმცა	Contrasting	Clear pragmatic contrastive function with literary support.	Equivalents (хотя, впрочем).
მაგალითად	Exemplification	Adverb (გმგნ.); “for example”; literary example.	Introductory word (вводное слово); standard equivalent (например).
საბოლოოდ	Conclusion	Adverb; “finally, in conclusion”; summarises and closes discourse.	Range of equivalents (окончательно, в конечном счете) plus adjectival forms → part-of-speech confusion.
ხო	Confirmation (spoken)	Colloquial particle marking agreement in spoken interaction.	Absent

The comparative analysis revealed significant divergence in how the dictionaries handle pragmatic framing, particularly concerning contrastive markers. In the monolingual Explanatory Dictionary, the marker *თუმცა* (although / however) is not merely defined syntactically but is embedded within a clear semantic frame of concession. The entry provides authentic literary citations that show the marker's role in introducing a proposition that contrasts with the preceding discourse expectation. In the bilingual dictionary (Akhvlediani & Topuria, 1990), *თუმცა* has Russian equivalents (*хотя, впрочем*). While these translations are semantically accurate at the word level, the entry strips away the pragmatic frame. It fails to provide usage examples that would instruct a learner on whether these equivalents are interchangeable or how they affect the syntactic structure of the Georgian clause.

The limitations of the equivalence-driven approach are most starkly revealed in the treatment of highly frequent, spoken discourse markers. The empirical data extracted from the Georgian National Corpus (Tandashvili & Gippert, 2015) confirm that the colloquial particle *ხო* (yeah / right / isn't it?) is one of the most pervasive markers of confirmation and interactional alignment in everyday Georgian speech. Table 2 presents the normalised

frequency data (occurrences per million words) for all selected markers, confirming their empirical relevance.

Table 2. Normalised Frequency of Selected Discourse Markers in the Georgian National Corpus

Marker	Frequency (per million words)
რადგან	312.4
ასევე	287.1
შემდეგ	245.8
თუმცა	198.3
მაგალითად	176.5
საბოლოოდ	89.2
ხო	682.7

As shown in Table 2, *ხო* is by far the most frequent marker in the spoken sub-corpora, with a frequency of 682.7 occurrences per million words. The monolingual Explanatory Dictionary adequately captures this communicative reality and dedicates a specific entry to its function as a colloquial particle marking agreement in spoken interaction. In contrast, the bilingual Georgian–Russian dictionary exhibits a critical lexicographic gap: the complete omission of *ხო*. From the perspective of Interface Theory, this omission is highly detrimental to the user, as it prevents learners from accessing information about one of the most common pragmatic particles in spoken Georgian.

The comparative analysis further revealed inconsistencies in how discourse markers are categorised, particularly within the bilingual resource. An example is the concluding marker *საბოლოოდ* (finally / in conclusion). The monolingual Explanatory Dictionary accurately identifies this unit as an adverb (*გმნიშვლად*) and explicitly outlines its pragmatic function as a procedural, discourse-closing element used to summarise preceding arguments. On the other hand, the bilingual Georgian–Russian dictionary (Akhvlediani & Topuria, 1990) shows significant morphosyntactic ambiguity, because the entry lists a conflated range of Russian equivalents and mixes adverbial discourse phrases (*в конечном счете*) with forms that suggest standard adjectival or propositional usage (*окончательно*). The blurring of lines between adjectival modifiers and discourse-organising adverbs fundamentally misrepresents the marker's grammatical behaviour and may lead to user confusion.

We can argue that the monolingual dictionary proves superior for understanding Georgian-internal pragmatic structure, while the bilingual dictionary serves as a practical bridge for translation. Both resources are most effective when used together.

5. Conclusions

Drawing on Frame Semantics, Corpus Linguistics, Contrastive Analysis, and Translation Theory, this metalexicographic study demonstrates that monolingual and bilingual dictionaries employ complementary strategies. Monolingual entries adopt a descriptive, meaning-centred approach that foregrounds native pragmatic functions. Bilingual entries adopt a user-oriented, equivalence-driven approach that prioritises cross-linguistic accessibility. However, the bilingual resource (Akhvlediani & Topuria, 1990) exhibits critical shortcomings: the complete omission of high-frequency spoken markers such as *ხო*, inconsistencies in part-of-speech assignment for markers like *საბოლოოდ*, and a general lack of pragmatic framing that would help users understand the functional context of each marker.

Neither format is inherently superior; rather, their strengths are maximised through complementary consultation. The monolingual dictionary teaches users to think in Georgian, while the bilingual dictionary facilitates intercultural communication. Future lexicographic work on Georgian DMs should incorporate corpus-derived frequency data, more consistent pragmatic labelling, and improved interface design to enhance usability for both native speakers and learners. Additionally, the bilingual resource would benefit from explicit attribution of authorship and a systematic revision of entries for spoken-language markers.

AI Use Disclosure Statement

Portions of this manuscript were refined with the assistance of a large language model (Grok, developed by xAI) for purposes of language editing and formatting of APA 7th-edition citations and references. All intellectual content, analytical claims, data interpretation, and final decisions regarding the text remain the sole responsibility of the author.

References

1. Akhvlediani, G., & Topuria, V. (1990). ქართულ-რუსული ლექსიკონი [Georgian–Russian dictionary] (2nd ed.). Ganatleba https://opac.sciencelib.ge/cgi-bin/koha/opac-detail.pl?biblionumber=6390&shelfbrowse_itemnumber=9348
2. Briz, A., Pons Bordería, S., Borreguero Zuloaga, M., & Portolés, J. (Coords.). (2008). Diccionario de partículas discursivas del español (DPDE). <http://www.dpde.es>
3. Cartoni, B., Zufferey, S., & Meyer, T. (2013). Annotating the Meaning of Discourse Connectives by Looking at their Translation: The Translation Spotting Technique. *Dialogue & Discourse*, 4(2), 65–86. <https://doi.org/10.5087/dad.2013.203>

4. Chikobava, A. (Ed.). (1950–1964). *ქართული ენის განმარტებითი ლექსიკონი* [Explanatory dictionary of the Georgian language] (Vols. 1–8). Georgian Academy of Sciences. <https://lev.ge/ganmartebiti>
5. Fillmore, C. J. (2014). Frame semantics. In *Linguistics in the morning calm* (pp. 111–137). Hanshin Publishing Co. (Original work published 1982)
6. Fischer, K. (Ed.). (2006). *Approaches to discourse particles* (Vol. 1). Elsevier.
7. Fox Tree, J. E. (2015). *Discourse markers in speech*. In *The handbook of discourse analysis* (pp. 800–818). John Wiley & Sons. <https://doi.org/10.1002/9781118584194.ch37>
8. Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31(7), 931–952. [https://doi.org/10.1016/S0378-2166\(98\)00101-5](https://doi.org/10.1016/S0378-2166(98)00101-5)
9. Fuentes Rodríguez, C. (1987). *Enlaces extraoracionales*. Alfar.
10. Gouws, R. H. (2017). Dictionaries and access. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 43–58). Routledge. <https://doi.org/10.4324/9781315104942>
11. Holgado Lage, A. (2017). *Diccionario de marcadores discursivos para estudiantes de español como segunda lengua*. Peter Lang. <https://doi.org/10.3726/b11878>
12. Knott, A. (1996). *A data-driven methodology for motivating a set of coherence relations* [Doctoral dissertation, University of Edinburgh].
13. Makhatadze, M. (2024). *Toward lexical bundles in Georgian Learner Corpus of English with implications for pedagogic lexicography*. *Online Journal of Humanities ETAGTSU*, (9), 22. <https://doi.org/10.52340/PUTK.2024.2346-8149.09.06>
14. Martín Zorraquino, M. A., & Portolés, J. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 4051–4213). Espasa Calpe.
15. Pons Bordería, S. (2001). *Conectores y marcadores discursivos*. Ediciones de la Universidad de Navarra.
16. Tandashvili, M., & Gippert, J. (2015). Structuring a diachronic corpus: The Georgian National Corpus project. In *Historical corpora: Challenges and perspectives* (pp. 305–322). Narr.
17. Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-knowledge*. Niemeyer. <https://doi.org/10.1515/9783484970434>
18. van Dijk, T. A. (1973). *Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics*. Mouton.
19. van Dijk, T. A. (1978). *Text and context: Explorations in the Semantics and Pragmatics of Discourse*. Longman.

ავტორის შესახებ:

ჭივალაგოვი ელენე ინგლისური ფილოლოგიისა და ინგლისური ენის თეორიის მაგისტრი, ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის კურსდამთავრებული. მისი სამაგისტრო ნაშრომი ეყრდნობა სოციოლინგვისტურ

კვლევას თემაზე „სოციალური ქსელები, როგორც ციფრული იდენტობის კონსტრუირების ინსტრუმენტი“, რომლის ფარგლებშიც გაანალიზდა ვირტუალური იდენტობის ფორმირების პროცესები და სტრატეგიები.

About the author:

Elena Jivalagova is holder of Master's degrees in English Philology and English Language Theory from Ivane Javakhishvili Tbilisi State University. Her Master's thesis is based on a sociolinguistic study on social media as a tool for constructing digital identity, in which the processes and strategies of virtual identity formation were analyzed.

E-mail: elene.jivalagova035@hum.tsu.edu.ge

სამადლობელი:

მადლობას ვუხდით ქართული ეროვნული კორპუსის შემქმნელებს მონაცემების მიწოდებისთვის, აგრეთვე რეცენზენტებსა და რედაქტორებს სტატიის გაუმჯობესებაში შეტანილი წვლილისთვის.

Acknowledgments:

I would like to thank the creators of the Georgian National Corpus for providing the data, as well as the reviewers and editors for their valuable contributions to improving the article.