Towards a Romanian phrasal academic lexicon

Madalina Chitez[1], **Ana-Maria Bucur**[2], **Andreea Dinca**[1], **Roxana Rogobete**[1]
[1]West University of Timisoara, Timisoara, Romania
[2]University of Bucharest, Bucharest, Romania
{madalina.chitez,andreea.dinca,roxana.rogobete}@e-uvt.ro
ana-maria.bucur@drd.unibuc.ro

Abstract

The lack of NLP based research studies on academic writing in Romania results in an unbalanced development of automatic support tools in Romanian compared to other languages, such as English. For this study, we use Romanian subsets of two bilingual academic writing corpora: the ROGER corpus, consisting of university student papers, and the EXPRES corpus, composed of expert research articles. Working with the Romanian Academic Word List / Ro-AWL, we present two phrase extraction phases: (i) use Ro-AWL words as node words to extract collocations according to the thresholds of statistical measures and (ii) classify extracted phrases into general versus domain-specific multi-word units. We show how manual rhetorical function annotation of resulting phrases can be combined with automatic function detection. The comparison between academic phrases in ROGER and EXPRES validates the final phrase list. The Romanian phrasal academic lexicon (ROPAL), similar to the Oxford Phrasal Academic Lexicon (OPAL), is a written academic phrase lexicon for Romanian language made available for academic use and further research or applications.

Keywords: Romanian academic writing corpora, Romanian phrasal academic lexicon, EXPRES corpus.

1 Introduction

We present the first approach to creating the Romanian phrasal academic lexicon (ROPAL)[1] for written Romanian language. Academic writing is a challenge for students and experienced writers alike. Studies (Saberi et al., 2020; Chitez and Dinca, 2023) have pointed out that linguistic support is effective at all levels of language use and acquisition. Among the multitude of support options, e.g. grammatical accuracy, lexical diversity, paraphrasing (Strobl et al., 2019), multi-word units with a rhetorical function, i.e. academic phrases, are the most productive in enhancing the coherence and persuasiveness of academic writing (AlHassan and Wood, 2015; Hinkel, 2003). Thus "opportunities to practice bundle usage so as to trigger acquisition of formulaic language" (Pérez-Llantada, 2014) are associated with higher academic writing proficiency. These phrases ensure that writers comply with specialized academic conventions and further contribute to the development of writers' language skills, facilitating their ability to articulate complex ideas and arguments with precision and refinement. From a computational linguistics perspective, phrasal academic lexicons can serve as invaluable resources for training large language models to improve their proficiency in language generation and paraphrasing. At the same time, academic phrase lists can contribute to training LLM models to automatically annotate rhetorical functions. As a result, AI tools can be improved, which can help with tasks such as genre classification, stylistic analysis, and discourse segmentation, thereby enhancing the automated processing and understanding of academic texts.

When compiling lists of academic phrases, specialized corpora such as academic writing corpora can be of great assistance in identifying prevalent language patterns and rhetorical strategies across different academic disciplines or cross-disciplinarily. Although most extraction and analysis methods have been performed for phraseology extraction and use in L2 (English) (Section 2), their replicability for academic writing in L1 has been demonstrated (Pérez-Llantada, 2014). In the present paper, we use the EXPRES corpus and the Romanian Academic World List / Ro-AWL (Bucur et al., 2023) (Section 3) to extract phrases

[1]https://github.com/chia-AR/ROPAL-Romanian-phrasal-academic-lexicon

in ROPAL. Highly frequent phrases (first 150 occurrences in ROPAL) are manually annotated for rhetorical functions (Section 5). We use an adaptation of two rhetoric function systems proposed by Morley (2018) and by OPAL (2019) for the English language. In the end, we verify whether ROPAL can detect learner academic writing phrase patterns by using the ROGER corpus (Chitez et al., 2022a) (Section 5).

2 Related Work

Most research on academic multiword units was done in the context of English for Academic Purposes (EAP), where considerable attention was placed on finding a core academic phrasal lexicon that could be used across disciplines. These phraseological resources are informed by both EAP research and EAP practice, as academic writing is widely taught in university settings. Typically, EAP research uses academic writing corpora from which common multiword units are identified using computational methods. After extraction, the phrases are assigned rhetorical and functional categories (see e.g. Hyland (2008)). These findings are then integrated into various learning and teaching resources for academic writing. For instance, they may be presented into standalone phrase lists (e.g. Simpson-Vlach and Ellis (2010)) or phrasebooks (e.g. Morley (2018)) categorized by rhetorical and functional attributes. More recently, phrases have also been integrated into digital learners' dictionaries (see OPAL (2019)), or have become part of AI-powered academic writing platforms, such as Writefull[2].

In contrast to the abundance of resources available for English, research on academic phraseological units in other languages is scarce. To the best of our knowledge, no study to date has aimed at identifying cross-disciplinary academic phraseological units in a language other than English. Existing investigations predominantly adopt a comparative approach, such as Cortes (2008), which compares English and Spanish academic phraseological units. In the case of Romanian, research is rather characterized by heterogeneity, both in theoretical frameworks, terminologies and methodologies employed (Zafiu, 1989). While some researchers refer to 'pragmatic functions' (Rovenţa-Frumuşani, 2012; Stefănescu, 2007), others may use terms such as 'metalinguistic markers' (Bîtea, 1986), 'specialized syntagms' (Pricop, 2014). This lack of standardization in terminology makes it challenging to compile comprehensive online resources, such as lists of various types of phraseological units or academic vocabulary or conventions. Unlike languages with more robust digital resources, such as English, the availability of openly accessible linguistic data for Romanian is limited. Even though the last two decades have seen various initiatives launched (Tamba, 2017; Cristea et al., 2011), efforts have lacked sufficient coordination. While specific micro-studies have been conducted to analyze a restricted number of phraseological units across limited disciplines comparatively (Bocoş, 2018; Stoichițoiu-Ichim, 2001), only one study has focused on multiword units in Romanian academic writing (Muresan et al., 2022). It explored the difference between expert academic writing produced by Romanian scholars and novice academic writing produced by Romanian university students.

3 Method

Several language datasets are used in this study. The Corpus of Expert Academic Writing in Romanian and English / EXPRES (Chitez et al., 2022b) and the Romanian Academic Word List / Ro-AWL (Bucur et al., 2023) were used in the creation of ROPAL. Subsequently, The Corpus of Romanian Academic Genres / ROGER (Chitez et al., 2022a) is used to test ROPAL. The datasets and the extraction procedures are described in the following sections.

3.1 EXPRES and ROGER

EXPRES is a bilingual multidisciplinary corpus representative of expert academic writing in English and in Romanian. This investigation uses a subset of EXPRES containing academic writing in Romanian. As shown in Table 1, the dataset has a size of more than 3 million words, and more than 200 thousand unique words. The disciplines represented in the corpus are Linguistics (LG), Economics (EC), Information Technology (IT) and Political Science (PS). There are 800 research articles, 200 articles for each discipline.

ROGER is a bilingual comparable learner corpus containing academic writing by students enrolled in Romanian universities. The full corpus contains various text genres from five disciplines, written in either students' native language, Romanian, or in English as a Foreign Language. The

[2] https://www.writefull.com/

Domain	Tokens	Types
EC	1,092,846	48,807
LG	674,277	73,667
IT	750,236	40,494
PS	963,061	62,096
Total	3,480,420	225,064

Table 1: EXPRES Statistics.

Domain	Texts	Tokens	Types
EC	97	238,770	39,605
HUM	653	476,232	75,785
CS	42	118,507	21,131
PS	119	346,388	52,796
Total	911	1,179,897	139,283

Table 2: ROGER Statistics.

present study uses the Romanian subset of ROGER. As can be seen from Table 2, it contains text genres such as essays, or B.A. theses from the disciplines of Economics (EC), Humanities (HUM), Computer Science (CS) and Political Sciences (PS). The corpus amounts to more than one million words.

3.2 Ro-AWL

The Romanian Academic Word List[3] (Bucur et al., 2023) is derived from the EXPRES corpus and consists of a set of academic words that are frequently encountered in academic texts. It is made up of 673 lemmas, distributed among the main part-of-speech categories (nouns, verbs, adverbs, adjectives). The list is freely available, and a detailed description of its creation is available in Bucur et al. (2023).

3.3 Extraction and annotation of the phrases

The EXPRES corpus is organized into .txt files. We removed specific tags such as {FIG}, {REF_LIST}, {JOURNAL_TITLE}, {AUTHOR_NAME}, or tags used to indicate the title (<TITLE>, </TITLE>), abstract (<ABS_INT>, </ABS_INT>), keywords (<KW_INT>, </KW_INT>), etc. We used Stanza (Qi et al., 2020) for lemmatization, and all the lemmas from the texts were converted to lowercase for further analysis. We used Ro-AWL terms as node words and extracted collocations of 2 to 6 words from EXPRES that contain the academic words found in Ro-AWL.

[3] https://github.com/bucuram/Ro-AWL

We compiled the phrasal academic lexicon using the criteria of frequency and dispersion, following the works of Ädel and Erman (2012) and Ebeling and Hasselgård (2015). Given that we extracted collocations containing up to 6 words, we opted for a lower threshold for frequency – a collocation had to occur at least 10 times per million words. Dispersion was used to complement frequency measures to ensure that a collocation was not idiosyncratically confined to a limited number of texts. This is why the collocations that appeared in less than 5 different texts were excluded from the analysis. Manual filtering was further performed to remove collocations that were part of larger phrases. For example, the collocation "punct de vedere" (En: "point of view") was manually removed, as it was contained into "din punct de vedere" (En: "from the point of view"). Even if the purpose of the lexicon was to encompass 6-grams, all collocations with 6 words were excluded from the final list either because of their low frequency and dispersion, or because they were removed during manual filtering.

In this study, the frequency of the collocations is presented in two ways: raw frequency, which is the number of times a collocation appears in the dataset, and normalized frequency per 1,000 words (Lancaster, n.d.). Since the datasets differ in size, it is common practice to use normalized frequencies for comparing the results.

4 The Romanian phrasal academic lexicon

4.1 Statistics

Collocation	**English Translation**	**Frequency** Normalized per 1,000 words
de asemenea	also	0.67
de tip	of type	0.49
de exemplu	e.g.	0.42
din punct de vedere	from the point of view	0.38
cu privire la	regarding	0.36
în funcţie de	depending on	0.36
avea în vedere	considering	0.33
în timp ce	while	0.32
pe de altă parte	on the other hand	0.25
pe termen	on term	0.22

Table 3: Top 10 collocations in ROPAL.

ROPAL includes a total of 794 collocations, which are comprised of 477 2-grams, 252 3-grams, 52 4-grams and 13 5-grams. Table 3 displays the 10 most frequent collocations in ROPAL. The number

Category	Example translated from Romanian	No of collocations
Perspective and Viewpoint, Considering or Taking into Account a Specific Perspective	from the point of view; regarding; considering	28
Miscellaneous (e.g. Expressing quantity; Timeframe)	short term; real time	17
Compare and contrast (discourse)	while; on the other hand	14
Defining; Explaining terms	of type; consists of	14
Evaluation; Expressing degree	extremely; quality	13
Signaling transition; Introduce additional information	also; furthermore	9
Conclusion; Summary; Hypothesis	thus; this work	9
Literature review; Discussing related work; Appeal to Authority; Referring to other texts	in the specialized literature; in discussion	9
Explaining causality	as a result; having as a goal	7
Giving examples	e.g.; for example in	3

Table 4: Classification of academic language functions in ROPAL.

is influenced by the choice of the method extraction, i.e. use of Ro-AWL list (see Sections 3.2 and 3.3).

To verify ROPAL's reliability, we tested the coverage of the list in two academic writing corpora, EXPRES and ROGER. The coverage of ROPAL in the EXPRES corpus is 3.6%. When it comes to the coverage of ROPAL in writings by university students from the ROGER corpus, the coverage is lower, at 1.6%. ROPAL used EXPRES for validation since the list was based on Ro-AWL, which, in turn, used another reference corpus for list validation (Bucur et al., 2023). At the same time, ROPAL list was manually filtered, thus being quite different from the automatically generated list.

4.2 Classification

To classify the most frequent collocations in ROPAL, we used a mixed approach, by adapting established models such as the OPAL framework (OPAL, 2019) and the Manchester Academic Phrasebank (Morley, 2018). The decision to draw from these models was motivated by the lack of a standardized classification system in Romanian (Stefanescu, 2017). These models provided valuable insights into the categorization of academic language functions and served as guiding templates for the refinement process for Romanian phrases in ROPAL. The final 10 categories were developed for the first 129 units/collocations, which occurred in the corpus with a frequency of at least 0.05. We simplified overlapping concepts and reorganized them into broader groups. The expertise of the team, in both Romanian linguistics and didactics, contributed to the creation of a unique classification model inspired by previous international models. For example, categories like "Being critical" and "Describing trends" were integrated into broader categories such as "Perspective and Viewpoint; Considering or Taking into Account a Specific Perspective". We also merged categories such as "Making contrast" and "Comparing". Finally, we developed a more versatile category - "Literature review; Discussing related work; Appeal to Authority/ Referring to other texts", which encompasses classes such as "Hedging", "Writing about the past" and "Describing trends". The final functional areas developed for this study can be seen in Table 4.

Results show that phrases falling under the category of "Perspective and Viewpoint/Considering or Taking into Account a Specific Perspective" are the most common in research articles from EXPRES corpus. This is because academic writing requires authors to discuss or evaluate various viewpoints or theoretical frameworks (particularly within sections like the literature review) to present their own arguments.

Furthermore, the "Compare and contrast" section is well represented, since scholars often choose a comparative analysis or contextualize their research within the existing literature and academic community.

4.3 Evaluation

To describe how the phrasal academic units from ROPAL are distributed among disciplines, we conducted a comparative linguistics analysis looking at how these units are used in four disciplines.

The EXPRES platform (https://expres-corpus.org/) where the corpus is freely available for use, was used to extract examples. The top three collocations in each of the four disciplines, are presented in Table 5. Most of these collocations serve as linguistic tools for authors to introduce, contextualize, and evaluate different viewpoints, while contributing to the overall coherence and logical flow of the text. By looking at discipline-based phrases in ROPAL, common patterns, phrase overlaps and discipline specific units be highlighted. Research articles in Economics contain specialized phrases such as "at a national level", "on term", "on long-term", "at the global level". The field of Economics appears to be the most specialized among the disciplines analyzed, given the presence of N-grams that are entirely absent in the other disciplines examined, like "from a statistical point of view", "had a positive impact on", "at an average pace", "annual average of", "growth trend", "influencing factors". When shifting our attention to Information Technology, structures such as "database analysis", "real time", "model of", "in the database", "of classification" are specific to data management and analysis. An N-gram which occurs only in IT is "allow access", emphasizing the central role of security, databases, and information processing for the domain. The field of linguistics employs most of the rhetorical strategies and connective phrases, such as "e.g.", "also", "by point of view", "of type", "depending on", "considering", "while". A structure like "in the paradigm" occurs only in this domain, which appears to have a predilection for expansive explanations and exemplification. The last analyzed discipline, political sciences, is characterized by inserting multiple perspectives, since the most frequent structures are "regarding", "also", "by ... point of view", "on the other hand", "while". This field appears to be the most nonspecific in using academic phrases, since there is none used in political sciences and absent in other disciplines (even an N-gram like "in the public space" appears also in other corpora).

Overall, it can be observed that political sciences appear less prone to specific linguistic patterns, perhaps due to its interdisciplinary nature, while academic writing in linguistics tends to overuse explanations. Although information technology seems to have a more technical focus, it often employs a more complex discourse. The field that exhibits the most distinctive academic phrase patterns is Economics.

Collocation	Examples from the EXPRES
Economics	
regarding	"member states will have reports regarding financial aspects"
considering	"an analysis considering a causal relationship between macroeconomic variables and [...]"
from the point of view	"the most developed regions from the economical point of view"
Information Technology	
depending on	"depending on the GPU memory"
from the point of view	"from the point of view of data management functionalities"
by analysis	"The functionality and effectiveness of MOOC projects will be highlighted by traffic and event log analysis"
Linguistics	
from the point of view	"from a semantic point of view"
depending on	"depending on conjugation"
considering	"Considering semantic equivalences/ analogies [...]"
Political Sciences	
regarding	"we have identified several approaches regarding foreign affairs"
from the point of view	"from the point of view of the political route"
believes that	"a large part of the population believes that the state should intervene"

Table 5: Top 3 collocations for each discipline in ROPAL.

5 Utilization

In this section, we will test the ROPAL list on the Romanian section of the ROGER corpus (Chitez et al., 2022a), also available online (https://roger-corpus.org/). We first selected the most frequent 10 N-grams from ROPAL, and searched them in the whole ROGER corpus, then in similar disciplinary datasets in EXPRES (economics, humanities, political sciences, computer science). The numbers are listed in Table 6 (normalized frequency in the entire ROGER corpus vs normalized frequency in the analyzed disciplines).

The distribution of the ROPAL academic phrases related to discourse cohesion follows, in ROGER,

Collocation English Translation	ROGER Total	EC	HUM	CS	PS
also	0.32	0.26	0.45	0.22	0.21
of type	0.16	0.15	0.20	0.25	0.09
e.g.	0.20	0.20	0.25	0.19	0.12
from the point of view	0.26	0.20	0.42	0.11	0.15
regarding	0.13	0.12	0.15	0.11	0.12
depending on	0.11	0.10	0.15	0.11	0.07
considering	0.14	0.12	0.19	0.12	0.10
while	0.12	0.17	0.11	0.04	0.11
on the other hand	0.10	0.07	0.12	0.03	0.09
on term	0.05	0.16	0.01	0.03	0.05

Table 6: Frequencies of the top 10 ROPAL discourse cohesion collocations in the ROGER corpus.

an expected pattern, since we identified common features across disciplines. In fact, the novice writing samples from ROGER do not present sufficient discipline-specific characteristics. For instance, the domain-specific units identified in EXPRES are poorly represented in ROGER. Students tend to focus more on elaborating their discourse than on the development of a discipline-specific language. Therefore, greater exposure to specialized articles would enable students to familiarize themselves with the conventions and expectations of their micro-academic community.

6 Conclusions

The findings presented in the study highlight several key points regarding the development and application of the Romanian Phrasal Academic Lexicon (ROPAL) in computational linguistics and language education. Firstly, since the main objective of the current study was to verify the capacity of our generated academic phrase lexicon to support the automatic assessment of academic writing proficiency in the native language Romanian (which was demonstrated by the fact that coverage percentages are lower in novice versus expert writing), IAA was not performed.

Secondly, the extraction and annotation of academic phrases from the EXPRES corpus demonstrate the didactic applicability of using corpora to compile resources adapted for the needs of Romanian academic writers. The phrase extraction and classification approaches provide insights into prevalent language patterns and rhetorical strategies across different academic disciplines, but also across disciplines. Thus, a pilot ROPAL list (i.e. ROPAL for teaching), similar to the OPAL list (OPAL, 2019), to be used for teaching and education purposes, was compiled[4]. It contains the prominent rhetorical categories in Romanian academic phrases, based on the complete OPAL list generated for computational purposes. Such a list is particularly novel because, until now, no equivalent resource has existed for the Romanian academic setting.

Most items in ROPAL for teaching represent general academic writing multi word units (e.g. [translation from Romanian into English] “also”, “from the point of view”, “regarding”) whereas others are rather discipline-specific or have a higher probability of being used more frequently in a particular disciplinary field. This list will be adapted for various disciplines and validated through interrater agreement (IAA) methods. Further studies using ROPAL and the rhetoric function annotation for each phrase may result in the creation of AI tools for academic text correction, text suggestion and text assessment in the Romanian language. ROPAL enhances students’ ability to engage with and produce academic work that adheres to the rhetorical standards of their field.

Acknowledgments

This study was supported by the UVT 1000 Develop Fund of the West University of Timisoara.

References

Annelie Ädel and Britt Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of english: A lexical bundles approach. *English for specific purposes*, 31(2):81–92.

Lina AlHassan and David Wood. 2015. The effectiveness of focused instruction of formulaic sequences in augmenting l2 learners’ academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, 17:51–62.

Ioan N Bîtea. 1986. Conectorii metalingvistici: adverbe ori conjuncții? *Editura Academiei Republicii Socialiste România*.

Cristina Bocoş. 2018. Unităţile frazeologice cu termeni religioşi în română şi spaniolă. o analiză comparativ-contrastivă. *Philologica Jassyensia*, 14(2):337–340.

Ana-Maria Bucur, Andreea Dincă, Mădălina Chitez, and Roxana Rogobete. 2023. Automatic extraction of the romanian academic word list: Data and methods. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 234–241.

[4]ROPAL for teaching is freely accessible online at: https://codhus.projects.uvt.ro/corpus-tools/.

Mădălina Chitez, Loredana Bercuci, Andreea Dincă, Roxana Rogobete, and Karla Csürös. 2022a. Corpus of romanian academic genres (ROGER). West University of Timisoara.

Madalina Chitez and Andreea Dinca. 2023. On corpora and writing. In *Digital Writing Technologies in Higher Education: Theory, Research, and Practice*, pages 385–403. Springer.

Mădălina Chitez, Valentina Muresan, Roxana Rogobete, and Andreea Dincă. 2022b. Corpus of expert writing in romanian and english (EXPRES). West University of Timisoara.

Viviana Cortes. 2008. A comparative analysis of lexical bundles in academic history writing in english and spanish. *Corpora*, 3(1):43–57.

Dan Cristea, Gabriela Haja, Alex Moruz, Marius Răschip, and Mădălin Pătraşcu. 2011. Statistici parţiale la încheierea proiectului edtlr–dicţionarul tezaur al limbii române în format electronic. *În volumul Lucrările conferinţei naţionale Limba română: ipostaze ale variaţiei lingvistice*, pages 3–4.

Signe Oksefjell Ebeling and Hilde Hasselgård. 2015. Learners' and native speakers' use of recurrent word-combinations across disciplines. *Bergen Language and Linguistics Studies*, 6.

Eli Hinkel. 2003. *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Routledge.

Ken Hyland. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1):4–21.

University of Lancaster. n.d. Comparing frequencies for corpora of different sizes.

John Morley. 2018. *Academic phrasebank: A compendium of commonly used phrasal elements in academic English in PDF format*. The University of Manchester.

Valentina Muresan, Roxana Rogobete, Ana-Maria Bucur, Madalina Chitez, and Andreea Dinca. 2022. Phraseology in romanian academic writing: Corpus based explorations into field-specific multiword units. *Recent Advances in Digital Humanities. Romance Language Applications, Peter Lang*, pages 29–48.

OPAL. 2019. Oxford phrasal academic lexicon.

Carmen Pérez-Llantada. 2014. Formulaic language in l1 and l2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14:84–94.

Alina-Mihaela Pricop. 2014. Tendinţe în tratarea sintagmelor specializate. *Lexicografia academică românească*, page 139.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108.

Daniela Rovenţa-Frumuşani. 2012. *Analiza discursului: ipoteze şi ipostaze*. Tritonic.

Dariush Saberi, John Lee, and Jonathan Webster. 2020. Automatic assistance for academic word usage. In *The 28th International Conference on Computational Linguistics (COLING'2020)*, pages 2163–2168. International Committee on Computational Linguistics.

Rita Simpson-Vlach and Nick C Ellis. 2010. An academic formulas list: New methods in phraseology research. *Applied linguistics*, 31(4):487–512.

Ariadna Stefănescu. 2007. Conectori pragmatici.

Maria Stefanescu. 2017. Some remarks on phraseological units in the thesaurus dictionary of the romanian language. *Dacoromania*, 22(1):51–62.

Adriana Stoichițoiu-Ichim. 2001. *Semiotica discursului juridic*. Editura Universității din București.

Carola Strobl, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Proske, and Christian Rapp. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & education*, 131:33–48.

Elena Isabelle Tamba. 2017. Clre. corpus lexicografic românesc esenţial. 100 de dicţionare din bibliografia dlr aliniate la nivel de intrare şi la nivel de sens. haja, gabriela (éd.). 2017. *Lexicografia academică românească. Studii. Proiecte*, pages 221–234.

Rodica Zafiu. 1989. Câteva observaţii asupra conectorilor pragmatici din limba română. *Studii şi cercetări lingvistice, Vol. 40, nr. 3,(1989); p. 315-319.*