

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის
ჰუმანიტარულ მეცნიერებათა ფაკულტეტი

სადოქტორო პროგრამა: ფილოლოგია
(ლექსიკოგრაფია)

მარინე მახათაძე

ინგლისური ენის სასწავლო კორპუსის შექმნა: ლექსიკური სიმდიდრე და
სიტყვათხმარების თავისებურებები ინგლისური ენის ქართულ შემსწავლელეში

ფილოლოგიის დოქტორის (Ph.D.) აკადემიური ხარისხის
მოსაპოვებლად წარმოდგენილი დისერტაცია

სამეცნიერო ხელმძღვანელი: მანანა რუსიშვილი
ფილოლოგიის მეცნიერებათა დოქტორი, პროფესორი
სამეცნიერო თანახელმძღვანელი: ნათია კენჭიაშვილი
ფილოლოგიის დოქტორი, ასოც. პროფ.

თბილისი

2024

შინაარსი

ანოტაცია	v
Abstract.....	viii
სქემები	x
სურათები	xii
დიაგრამები	xiii
შემოკლებები.....	xiv
შესავალი.....	1
თავი I. სასწავლო კორპუსი როგორც ინტერდისციპლინური კვლევის რესურსი (თეორიული მიმოხილვა)	10
1.1. სასწავლო კორპუსის გაგებისათვის	10
1.1.1. სასწავლო კორპუსის ტიპოლოგია.....	13
1.1.2. სასწავლო კორპუსის დიზაინი: გარემოს, დავალებისა და ენის შემსწავლელთა ცვლადები	17
1.1.3. სასწავლო კორპუსების მეთოდოლოგია.....	20
1.2. შესიტყვებებისა და მრავალსიტყვიანი ლექსიკური ერთეულების კვლევა სასწავლო კორპუსში	26
1.2.1. ფრაზეოლოგიზმების კვლევის სიხშირეზე დამყარებული მიდგომა.....	27
1.2.2. ფრაზეოლოგიზმების კვლევის მნიშვნელობაზე ორიენტირებული მიდგომა	30
1.2.3. შესიტყვებების სტრუქტურული ტაქსონომია.....	32
1.2.4. მრავალსიტყვიანი ლექსიკური ერთეულების (ლექსიკური კონების) ფუნქციური ტაქსონომია.....	35
1.3. ლექსიკური სიმდიდრის შეფასება კორპუსლინგვისტიკაში.....	38
1.4. სასწავლო კორპუსის მასალის ლექსიკოგრაფიული გამოყენების შესაძლებლობები	43
1.4.1 ლექსიკოგრაფიული ტექსტური გრაფები.....	43
1.4.2. მრავალკომპონენტიანი ლექსიკური ფრაზების სალექსიკონო მიკროსტრუქტურა..	48
1.5. შუალედური დასკვნა.....	51

თავი II. ინგლისური ენის ქართულ შემსწავლელთა კორპუსის (GLEAN) შექმნის პრინციპები	55
2. ზოგადი მიმოხილვა.....	55
2.1. ინგლისური ენის ქართულ შემსწავლელთა კორპუსის კრიტერიუმები.....	55
2.2. ჟანრული კლასიფიკაცია ინგლისური ენის ქართულ შემსწავლელთა კორპუსში	57
2.3. მონაცემთა შეგროვება ინგლისური ენის ქართულ შემსწავლელთა კორპუსისათვის...	60
2.4. ინგლისური ენის ქართულ შემსწავლელთა კორპუსის POS-ანოტაცია.....	65
2.5. შესიტყვებების კვლევის ასოციაციური საზომები	68
2.6. კონტრასტული ანალიზისთვის შერჩეული სასწავლო კორპუსები.....	74
2.6.1. ინგლისური ენის აზიელ შემსწავლელთა საერთაშორისო კორპუსი (ICNALE).....	75
2.6.2. მიჩიგანის უნივერსიტეტის ამერიკელი სტუდენტების ესეების სასწავლო კორპუსი (MICUSP)	76
2.7. შუალედური დასკვნა.....	78
თავი III. ინგლისური ენის ქართულ შემსწავლელთა კორპუსის (GLEAN) ანალიზი..	82
3. ზოგადი მიმოხილვა.....	82
3.1. ორკომპონენტური შესიტყვებების რაოდენობრივი ანალიზი.....	82
3.1.1. V+N სტრუქტურის მქონე ორკომპონენტური შესიტყვებები	84
3.1.2. N+N სტრუქტურის მქონე ორკომპონენტური შესიტყვებები	87
3.1.3. ADJ+N სტრუქტურის მქონე ორკომპონენტური შესიტყვებები	90
3.1.4. V+ADV სტრუქტურის მქონე ორკომპონენტური შესიტყვებები.....	94
3.1.5. ADV+V სტრუქტურის მქონე ორკომპონენტური შესიტყვებები.....	97
3.1.6. ADV + ADJ გამაძლიერებელი კოლოკაციები (<i>very, utterly, extremely, perfectly, absolutely, totally + ზედსართავი სახელი</i>)	99
3.2. სამ- და ოთხკომპონენტური ლექსიკური ფრაზების რაოდენობრივი ანალიზი	103
3.3. სამ- და ოთხკომპონენტური ლექსიკური ფრაზების ფუნქციურ-დისკურსული ანალიზი	106
3.3.1. კვლევაზე ორიენტირებული ლექსიკური ფრაზები	107
3.3.2. ტექსტზე ორიენტირებული ლექსიკური ფრაზები.....	111

3.3.3 მონაწილეზე ორიენტირებული ლექსიკური ფრაზები	114
3.4. ლექსიკური სიმდიდრის განსაზღვრა.....	116
3.4.1. „ფრაზეოლოგიური სათამაშო დათუნები“ GLEAN სასწავლო კორპუსში.....	116
3.4.2. Type/Token Ratio და სიხშირული სია ინგლისური ენის ქართულ შემსწავლელთა კორპუსში	118
3.5. შუალედური დასკვნა.....	120
IV თავი. ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსი (GLEAN), როგორც წყარო ინგლისურ-ქართულ სპეციალიზებული (EAP) ლექსიკონის სიტყვა-სტატიის შესადგენად.....	125
4. ზოგადი მიმოხილვა.....	125
4.1. ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები	126
4.2. შუალედური დასკვნა.....	144
4.3. სამომავლო პერსპექტივები.....	145
ნაშრომის ძირითადი დასკვნები და დებულებები	146
ბიბლიოგრაფია.....	155

ანოტაცია

ჩვენი ნაშრომის სათაურია „ინგლისური ენის სასწავლო კორპუსის შექმნა: ლექსიკური სიმდიდრე და სიტყვათხმარების თავისებურებები ინგლისური ენის ქართველ შემსწავლელებში“.

ენობრივი სიმდიდრის, სიტყვათხმარების, ენობრივი ქცევის ნორმების შესწავლის მიზნით, შევქმენით ცხრა მილიონამდე სიტყვისგან შემდგარი ინგლისური ენის ქართველი შემსწავლელების სასწავლო კორპუსი (GLEAN), რომელიც მოიცავს ინგლისური ენის ქართველი შემსწავლელების მიერ შესრულებულ სხვადასხვა ჟანრის მქონე ინგლისურენოვან წერილობით მონაცემს.

წინამდებარე სადისერტაციო ნაშრომი შეისწავლის ინგლისური ენის ქართველი შემსწავლელების ენის რეალიზების ასპექტებს, ლინგვისტურ მონაცემთა თანმიმდევრულ და მრავალმიზნობრივ (უმეტესად ლექსიკოგრაფიულ) შეგროვება-დამუშავებას. აქედან გამომდინარე, ნაშრომის კვლევით მიზანსა და ამოცანას წარმოადგენს:

- ფრაზეოლოგიზმების კვლევის სიხშირეზე დამყარებული და მნიშვნელობაზე ორიენტირებული მიდგომების საფუძველზე შესიტყვებების სტრუქტურული კლასიფიკაცია, სამ- და ოთხკომპონენტური ლექსიკური ფრაზების ფუნქციური ტაქსონომიის სისტემატიზაცია;
- კონტრასტული ანალიზის გზით დავადგინოთ: ა) რა სტრუქტურის მქონე შესიტყვებით მოდელებს ვხვდებით ჩვენ მიერ შექმნილ სასწავლო კორპუსსა და ინგლისური ენის აზიელი შემსწავლელების საერთაშორისო სასწავლო კორპუსში? ბ) რომელი მრავალკომპონენტური ლექსიკური ფრაზა გვხვდება ჭარბად ან იშვიათად ინგლისური ენის ქართველ შემსწავლელთა დისკურსში ამერიკელ სტუდენტთა სასწავლო კორპუსთან შედარებისას? გ) გამოვყოთ რომელია ყველაზე ხშირად გამოყენებული, სტატისტიკურად დაანგარიშებადი ლექსიკური ელემენტები კოლოკაციების ასოციაციური საზომების გამოყენებით და გამოვავლინოთ შესიტყვებების ვიზუალური გრაფების ანალიზის შედეგები;
- წარმოვადგინოთ ჩვენ მიერ შექმნილ სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტური ლექსიკური ფრაზის სალექსიკონო

სიტყვა-სტატია. აღნიშნული მასალა შესაძლებელია გახდეს სპეციალიზებული, აკადემიურ წერაზე ორიენტირებული (EAP) ლექსიკონის შემადგენელი ნაწილი. კვლევა წარმართულია კორპუსზე დამყარებული მეთოდოლოგიით. ნაშრომის ემპირიული ბაზა მოიცავს ინგლისური ენის 65 ქართველ შემსწავლელთა წერილობით მასალას, რომელიც რეპრეზენტატიული უნდა იყოს არა მხოლოდ რაოდენობრივად, არამედ თვისებრივადაც. იგი უნდა წარმოგვიდგენდეს უცხოური ენის განსხვავებულ ჟანრულ რეპერტუარს. კორპუსში შეტანილია 34,527 ტექსტი, ხოლო ტოკენების რაოდენობა 9,812,931.

I თავში - „სასწავლო კორპუსი როგორც ინტერდისციპლინური კვლევის რესურსი“ - ჩვენ ვეხებით სასწავლო კორპუსის კვლევისა და მეთოდოლოგიის თეორიულ წანამძღვრებს, მის როგორც თეორიულ, ისე პრაქტიკულ მხარეებს.

II თავში - „ინგლისური ენის ქართველ შემსწავლელთა კორპუსის შექმნის პრინციპები“ - განვიხილავთ კორპუსში შემავალი ტექსტების შერჩევის კრიტერიუმებს, ჟანრულ კლასიფიკაციასა და მონაცემთა მოგროვების დეტალებს. აგრეთვე, კონტრასტული ანალიზისათვის საერთაშორისო სასწავლო კორპუსების შერჩევის მიზეზებსა და კრიტერიუმებს. სასწავლო კორპუსში გამოვყავით ჟანრთა ოთხი კატეგორია (ესეისტიკა, აკადემიური, პუბლიცისტური, სასაუბრო) და მათი ქვეკატეგორიები. ესეისტიკის ჟანრში ერთიანდება არგუმენტირებული ესე, ნარატიული ესე, აღწერითი ესე, თავისუფალი კომპოზიციის ესე. აკადემიური ჟანრი მოიცავს კვლევით დარგობრივ სტატიებს ლინგვისტიკასა და ლიტერატურაში. პუბლიცისტის ჟანრში გაერთიანებულია საინფორმაციო ქრონიკა, რეპორტაჟი და საგაზეთო სტატიები როგორც პოლიტიკურ, ისე აპოლიტიკურ საკითხებზე, ხოლო ბლოგისა და კორესპონდენციის (არაოფიციალური მეილები) სასაუბრო ჟანრში გავაერთიანეთ.

III თავში - „ინგლისური ენის ქართველ შემსწავლელთა კორპუსის (GLEAN) ანალიზი“ - ჩვენი კვლევის საგანია ორკომპონენტური შესიტყვებების სტრუქტურულ-სემანტიკური ანალიზი, განსაკუთრებული ყურადღება გავამახვილეთ გამაძლიერებელი კოლოკაციების (very, utterly, extremely და ა.შ. + ზედსართავი სახელის) კვლევაზე. აგრეთვე წარმოდგენილია სამ- და ოთხკომპონენტური ლექსიკური ფრაზების ფუნქციურ-დისკურსული ანალიზი. **IV თავში - „ინგლისური**

ენის ქართველ შემსწავლელთა სასწავლო კორპუსი (GLEAN), როგორც წყარო ინგლისურ-ქართულ სპეციალიზებული (EAP) ლექსიკონის სიტყვა-სტატიის შესადგენად - ეხება სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტიანი ლექსიკური ფრაზის სალექსიკონო სიტყვა-სტატიის მაკრო- და მიკროსტრუქტურულ აღწერას. თითოეული სიტყვასტატიის ელემენტი მცირე მოდიფიკაციებით ეფუძნება დ. სიპმანის მიერ შემოთავაზებულ სამეთაურო ფრაზების ლექსიკონში ასახვის პრინციპებს.

საკვანძო სიტყვები: *სასწავლო კორპუსი, პრაგმატიკული მარკერები, მრავალსიტყვიანი ლექსიკური ფრაზები, შესიტყვებები, სპეციალიზებული ლექსიკოგრაფია.*

Abstract

The title of our work is "*Compiling English Learner Corpus: Lexical Richness and Peculiarities of Word Usage in Georgian Learner English.*"

To study lexical richness, word usage, and norms of linguistic behavior, we have created a corpus of Georgian learners of English (GLEAN) consisting of up to nine million words. This corpus includes English written data of different genres performed by Georgian learners of English.

The present thesis will study the aspects of language realization of Georgian learners of English, as well as the consistent and multi-purpose (mostly lexicographic) collection and processing of linguistic data. Therefore, the research goals of the work are:

- Structural classification of collocations based on the frequency-based approach and the significance-oriented approach, systematization of functional taxonomy of three- and four-word lexical bundles;
- Through a contrastive interlanguage analysis: a) to demonstrate the peculiarities of the use of phraseological units and collocations (adverb + adjective, verb+noun, etc.) in the ICNALE and GLEAN and MICUSP corpora; b) based on statistical measures, which phraseological units are overused or underused in the writings of Georgian students; c) to present some effective lexicographic ways through which the learner corpus data is applied into the English-Georgian dictionary microstructure, for example some usage notes are modified and included to the dictionary entries.
- To present a dictionary entries of 30 multi-word lexical bundles based on the learning corpus created by us. This material can become part of a specialized academic writing (EAP) dictionary.

The empirical base includes the written material of 65 Georgian learners of English. The data represents a different genre repertoire of a foreign language. The number of texts and tokens in the learner corpus is distributed as follows: the corpus contains 34,527 texts, and the number of tokens is 9,812,931.

In Chapter I - "Learner Corpus as a Resource for Interdisciplinary Research" - we touch upon the theoretical foundations of learning corpus research and methodology, both its theoretical and practical aspects.

In Chapter II - "Principles of Creating a Corpus of Georgian Learners of English" - we discuss the selection criteria of texts included in the corpus, genre classification, and details of data collection—also, reasons and criteria for selection of international study corpora for contrast analysis. In the training corpus, we have identified four categories of genres (essay, academic,

journalistic, conversational) and their subcategories. The genre of essay writing includes argumentative essays, narrative essays, descriptive essays, and free composition essays. The academic genre includes field research articles in linguistics and literature. In the publicistic genre, news chronicles, reports, and newspaper articles on political and non-political issues are combined, and in the conversational genre of blogs and correspondence (informal and semi-formal emails).

In chapter III - "Analysis of the corpus of Georgian learners of the English language (GLEAN)" - the subject of our research is the structural-semantic analysis of two-component expressions; we paid particular attention to the research of the booster and maximizer collocations (very, utterly, extremely, etc. + adjective). Functional discourse analysis of three- and four-component lexical phrases is also presented.

Chapter IV - "Learning Corpus of Georgian Learners of English Language (GLEAN) as a Source for Compiling Word-Articles of English-Georgian Specialized (EAP) Dictionary" - refers to the macro- and micro-structural description of lexical word articles of 30 multi-component lexical phrases based on the learner corpus. Each word article element is based on D. with minor modifications—the principles of reflection in the dictionary of head-phrases proposed by Sipman.

Keywords: learner corpus, pragmatic markers, multi-word lexical units, collocations, specialized lexicography.

სქემები

სქემა 1. სასწავლო კორპუსების კლასიფიკაცია (გილქინი, 2008a: 12).....	13
სქემა 2. სასწავლო კორპუსის ცვლადები (გრეინჯერი და სხვ., 2015: 16)	17
სქემა 3. საერთაშორისო სასწავლო კორპუსები (ინგლისური კორპუსლინგვისტიკის ცენტრი, 2024).....	22
სქემა 4. ლექსიკური და გრამატიკული შესიტყვებები (The BBI Combinatory Dictionary of English, 1997).	34
სქემა 5. ლექსიკური კონების ფუნქციური ტაქსონომია ჯ. ო'ფლინის (2022: 81) მიხედვით	36
სქემა 6. GLEAN სასწავლო კორპუსის მასალის რაოდენობრივი განაწილება.....	57
სქემა 7. GLEAN სასწავლო კორპუსის ჟანრული კლასიფიკაცია	58
სქემა 8. სასწავლო კორპუსში შემავალ სტუდენტურ ესეთა სათაურების ნიმუშები	63
სქემა 9. ასოციაციური საზომები სიხშირისა და ექსკლუზიურობის მიხედვით (ბრეზინა, 2018: 71).....	71
სქემა 10. ლექსიკური შესიტყვებების რაოდენობა მილიონ სიტყვაში (pmw).....	83
სქემა 11. შესიტყვებების საერთო აბსოლუტური რაოდენობა თითოეული ლექსიკური მოდელის მიხედვით	83
სქემა 12. GLEAN სასწავლო კორპუსში V+N შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები	85
სქემა 13. GLEAN სასწავლო კორპუსში არსებული <i>take</i> კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით	86
სქემა 14. N+N წყვილის ყველაზე ხშირად გამოყენებული 50 შესიტყვებები GLEAN კორპუსში.....	88
სქემა 15. GLEAN სასწავლო კორპუსში არსებული <i>take</i> კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით	89
სქემა 16. GLEAN სასწავლო კორპუსში ADJ+N შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები	91
სქემა 17. GLEAN სასწავლო კორპუსში არსებული <i>good</i> კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით	92

სქემა 18. GLEAN სასწავლო კორპუსში V+ADV (ლოკატივი) შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები	94
სქემა 19. GLEAN სასწავლო კორპუსში არსებული <i>forward</i> კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით	95
სქემა 20. GLEAN სასწავლო კორპუსში ADV+V შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები	97
სქემა 21. GLEAN სასწავლო კორპუსში არსებული <i>further</i> კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით	98
სქემა 22. GLEAN სასწავლო კორპუსში ADV+ADJ შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები	100
სქემა 23. GLEAN სასწავლო კორპუსში გამოვლენილი სამ- და ოთხკომპონენტური ლექსიკური ფრაზები	105
სქემა 24. ლექსიკური ფრაზების ჭარბი/იშვიათი მოხმარება GLEAN სასწავლო კორპუსში log-likelihood სანდოობის ტესტით.....	117

სურათები

სურათი 1. ლონგმენის აფრიკანსის ლექსიკონის მიკროსტრუქტურა (გოუესი და სხვ., 2010: 502).....	46
სურათი 2. სასწავლო ორენოვანი ლექსიკონის ტექსტური გრაფა <i>bored/boring</i> შესახებ (გოუესი და პრინსლო, 2010: 133).....	47
სურათი 3. სამკომპონენტო ლექსიკური ფრაზის ლექსიკონში წარმოდგენის ნიმუში (სიპმანი, 2004: 322).....	51
სურათი 4. POS-ანოტაციის შედეგი ინგლისური ენის ქართველ შემსწავლელთა კორპუსში.....	67
სურათი 5. *_JJ _JJ _NN1* თეგის კომბინაციის შედეგი ინგლისური ენის ქართველ შემსწავლელთა კორპუსში	67
სურათი 6. შესიტყვების გრაფა <i>love</i> (ბრეზინა, 2018: 76).....	72
სურათი 7. საკვანძო სიტყვების <i>time</i> და <i>money</i> საერთო შესიტყვებების (ბრეზინა, 2012: 78).....	73
სურათი 8. MICUSP სასწავლო კორპუსის საძიებო სისტემა (Michigan Corpus of Upper-level Student Papers, 2009. https://elicorpora.info/main).....	78
სურათი 9. GLEAN სასწავლო კორპუსში <i>take</i> და მასთან დაკავშირებული ლექსემების შესიტყვებითი გრაფა, საზომი z-score.....	87
სურათი 10. GLEAN სასწავლო კორპუსში <i>life</i> შესიტყვების გრაფა (Z-score).....	90
სურათი 11. GLEAN სასწავლო კორპუსში <i>good</i> შესიტყვების გრაფა (Z-score).....	93
სურათი 12. GLEAN სასწავლო კორპუსში <i>forward</i> შესიტყვების გრაფა (MI)	96
სურათი 13. სამ- და ოთხკომპონენტო ლექსიკური ფრაზების ძიება AntConc-ით (ენტონი, 2011).....	104
სურათი 14. ყველაზე ხშირად განმეორებადი „სიტყვათა ღრუბელი“ GLEAN კორპუსში (log-ratio)	119
სურათი 15. სიტყვა-სტატია <i>as for the</i> სპეციალიზებული (EAP) ინგლისურ-ქართული ლექსიკონისათვის	127

დიაგრამები

დიაგრამა 1. დარგების განაწილება GLEAN სასწავლო კორპუსში.....	59
დიაგრამა 2. სასწავლო კორპუსში შემავალი ტექსტების ავტორების ინგლისური ენის ცოდნის დონე (თვითშეფასების მიხედვით)	64
დიაგრამა 3. შესიტყვებების მოდელების პროცენტული განაწილება ICNALE და GLEAN სასწავლო კორპუსებში.....	84
დიაგრამა 4. GLEAN და MICUSP სასწავლო კორპუსებში გამაძლიერებელი ზმნიზედებისა+ADJ შესიტყვებების რაოდენობრივი გადანაწილება.....	101
დიაგრამა 5. GLEAN კორპუსში ყველაზე ხშირად ხმარებული სამ- ან ოთხკომპონენტური ლექსიკური კონები	105

შემოკლებები

ADJ	ზედსართავი სახელი
ADV	ზმნიზედა
BNC	ბრიტანული ეროვნული კორპუსი
CIA	კონტრასტული ინტერენობრივი ანალიზი
EAP	ინგლისური ენა აკადემიური მიზნებისათვის
GLEAN	ინგლისური ენის ქართველ შემსწავლელთა კორპუსი
GNC	ქართული ენის ეროვნული კორპუსი
ICLE	საერთაშორისო სასწავლო კორპუსი
ICNALE	ინგლისური ენის შემსწავლელ აზიელ სტუდენტთა სასწავლო კორპუსი
KWIC	საკვანძო სიტყვა კონტექსტში
LINDSEI	ლუვენის ზეპირმეტყველების ინგლისურის საერთაშორისო ბაზა
LTR	ლექსიკონების თანაფარდობა
MICUSP	მიჩიგანის უნივერსიტეტის სასწავლო კორპუსი
N	არსებითი სახელი
pmw	შემთხვევების რაოდენობა ერთ მილიონში
POS	მეტყველების ნაწილები
TTR	ტიპ/ტოკენების თანაფარდობა
V	ზმნა

“You shall know a word by the company it keeps.”

(ფირსი, 1957: 11)

შესავალი

ენის ინტერდისციპლინური კვლევის წარმართვის ერთ-ერთი განმაპირობებელი ფაქტორი ენის დოკუმენტირებაა. კორპუსლინგვისტიკა თავისი არსით სოციუმის შემოქმედებითი პოტენციალის რეალიზებისა და მასზე დაკვირვების შესაძლებლობას იძლევა. საკითხის ამ კუთხით დანახვამ წინა პლანზე წამოსწია არა მხოლოდ მშობლიური ენის, როგორც ნიშანთა სისტემის დოკუმენტირების, არამედ უცხოური ენის შემსწავლელთა ლინგვისტური თავისებურებების დისციპლინათაშორისი კვლევის აუცილებლობა.

უცხოური ენის შემსწავლელები ახალი ენის ათვისებისას გრძელ გზას გადიან, სანამ სპონტანური, ინტუიციური დამოკიდებულება არ იქცევა სტანდარტიზებულ, საგანგებოდ შემუშავებულ წესებზე დაფუძნებულ პროცესად. ენობრივი სიმდიდრის, სიტყვათხმარების, ენობრივი ქცევის ნორმების შესწავლის მიზნით, შევქმენით ცხრა მილიონამდე სიტყვისგან შემდგარი ინგლისური ენის ქართველი შემსწავლელების სასწავლო კორპუსი (GLEAN), რომელიც მოიცავს ინგლისური ენის ქართველი შემსწავლელების მიერ შესრულებულ სხვადასხვა ჟანრის მქონე ინგლისურენოვან წერილობით მონაცემს. ჩვენი კვლევის ობიექტი არის ენა არა როგორც მხოლოდ ნიშანთა აბსტრაქტული სისტემა, არამედ როგორც ენობრივი ნიშნების გამოვლენის შესაძლებლობების ჯამი.

ჩვენი საკვლევი თემის **აქტუალობა** განპირობებულია იმ ფაქტით, რომ ქართულ სამეცნიერო სინამდვილეში აქამდე არ შექმნილა ინგლისური ენის სასწავლო კორპუსი, რის გამოც შეუძლებელი იყო აღნიშნული ტიპის ტექსტური მასალის მასშტაბურად კვლევა. სირთულესთან იყო დაკავშირებული დაგვედგინა, როგორ ხასიათდება ინგლისური ენის შემსწავლელთა ენობრივი კომპეტენცია, მოგვეხდინა ენობრივი მიმოქცევის ნორმების პროგნოზირება და სწორად დაგვეგეგმა სპეციალიზებული ლექსიკოგრაფიული საქმიანობა. ამდენად, სასწავლო კორპუსის შექმნის საჭიროება და აუცილებლობა კვლევის თანამედროვე მოთხოვნილებებით არის განპირობებული.

ნაშრომის მეცნიერული **სიახლე** განპირობებულია იმ ფაქტით, რომ წინამდებარე ნაშრომში პირველად ვაანალიზებთ ინგლისური ენის ქართულ შემსწავლელთა წერილობით ნიმუშებს კორპუსლინგვისტური კვლევის თანამედროვე მეთოდების გამოყენებით, რომელიც ტექსტების პროგრამულ დამუშავებასთან არის დაკავშირებული. ბუნებრივია, რომ ენის სისტემურობის შეცნობის სურვილიდან გამომდინარე, ლექსიკონების უმრავლესობა სწორედ დესკრიფციული ხასიათისაა. დესკრიფციული კვლევა, როგორც კვლევის პარადიგმის შემადგენელი წევრი, სასწავლო კორპუსების მეშვეობით ახალ შინაარსს იძენს, როდესაც კვლევის ფოკუსში ექცევა ენის გამოყენებითი ასპექტი – ენა, როგორც სოციოლინგვისტიკის კვლევის ობიექტი.

კვლევა აღწუსხავს, იკვლევს და აანალიზებს იმ ენობრივ არსენალს, რომელიც ქართულ შემსწავლელებში ინგლისური ენის ათვისებისას არის დამახასიათებელი. ნაშრომის სიახლე განისაზღვრება იმითაც, რომ მიღებული კვლევის შედეგები სპეციალიზებული ლექსიკოგრაფიის განვითარებისთვისაც მნიშვნელოვანია, კერძოდ, ისეთი ტიპის ლექსიკონების შესაქმნელად/გასაუმჯობესებლად, რომლებიც ორიენტირებულია აკადემიური ინგლისურის შესწავლაზე. ამჟამად არსებულ ინგლისურ-ქართული ორენოვანი ბეჭდური თუ ელექტრონულ მატარებელზე გადასული ლექსიკონების მიკროსტრუქტურაში არ მოიძებნება ტექსტური გრაფები, ან სხვა სახის შიდატექსტობრივი მონაცემები, რომელიც ენის შემსწავლელს კომენტარის სახით სიტყვათხმარების კონკრეტული კომბინაციების გამოყენების თავისებურებებს ან მოსალოდნელ შეცდომათა გამაფრთხილებელ ნიშანს ამცნობს.

ენის შემსწავლელების მიერ უცხოურ ენაზე შესრულებულ ტექსტებს შეისწავლის სხვადასხვა დისციპლინა. მის სხვადასხვა ასპექტს დამოუკიდებელად იკვლევს ისეთი დარგები, როგორიცაა დიდაქტიკა, განათლების მეცნიერება, ენათმეცნიერება, ლექსიკოგრაფია, სტილისტიკა, ფსიქოლინგვისტიკა, ტექსტის ლინგვისტიკა და სხვ. ამ სფეროებში ენის შემსწავლელთა მიერ შესრულებული ტექსტების ანალიზისას იკვეთება კვლევის ინტერდისციპლინური ხასიათი და საკვლევი საზღვრების დადგენის პრობლემა. წინამდებარე კვლევაც ინტერდისციპლინური ხასიათისაა, რაც განაპირობებს ჩვენი კვლევის სამეცნიერო სიახლესაც.

წინამდებარე სადისერტაციო ნაშრომი შეისწავლის ინგლისური ენის ქართველი შემსწავლელების ენის რეალიზების ასპექტებს, ლინგვისტურ მონაცემთა თანმიმდევრულ და მრავალმიზნობრივ (უმეტესად ლექსიკოგრაფიულ) შეგროვება-დამუშავებას. აქედან გამომდინარე, ნაშრომის კვლევით მიზანსა და ამოცანას წარმოადგენს:

I. თეორიული საკითხების დასმა, განხილვა და გადაჭრა:

- სასწავლო კორპუსების შესწავლის თეორიული წინამძღვრების განხილვა, კორპუსლინგვისტიკის თეორიისა და პრაქტიკის შესწავლა ქართული და უცხოური სამეცნიერო ლიტერატურის ფონზე;
- სასწავლო კორპუსის შექმნისათვის ტექსტების შერჩევის კრიტერიუმების ჩამოყალიბება: ტექსტის ტიპი, წყარო, ჟანრი, ენის შემსწავლელთა ცვლადები;
- დასავლური თანამედროვე საერთაშორისო გამოცდილების გათვალისწინება (ლუვენის ინგლისური ენის საერთაშორისო მონაცემთა ბაზის ან უფსალის უნივერსიტეტის სტუდენტური სასწავლო კორპუსის მაგალითზე) და მოგვეცეს არსებული სამეცნიერო ღირებულებებისა და ცოდნის საქართველოში დანერგვის კონკრეტული შესაძლებლობა;
- ფრაზეოლოგიზმების კვლევის სიხშირეზე დამყარებული და მნიშვნელობაზე ორიენტირებული მიდგომების საფუძველზე შესიტყვებების სტრუქტურული კლასიფიკაცია, სამ- და ოთხკომპონენტიანი ლექსიკური ფრაზების ფუნქციური ტაქსონომიის სისტემატიზაცია;
- სასწავლო კორპუსის მასალის ლექსიკოგრაფიული გამოყენების შესაძლებლობების გამოვლენა.

II. პრაქტიკული ასპექტით კვლევის გაშლა:

ამ მხრივ, საჭიროა:

- თავი მოვუყაროთ ინგლისური ენის ქართველი შემსწავლელების მიერ შესრულებულ ინგლისურენოვან ტექსტებს, ხოლო ჟანრულად და თემატურად მრავალფეროვანი მასალის მოპოვების მიზნით, რესურსი შევავსოთ პოლიტიკური სტატიებით, სტუდენტური ბლოგებით, სხვადასხვა ტიპის ესეებითა და კორესპონდენციით.

- მოვახდინოთ ჩვენ მიერ შექმნილ სასწავლო კორპუსში (*GLEAN Georgian Learner English and Newspaper Corpus*) შემავალი სიტყვაფორმების გრამატიკული დახასიათება, მორფოლოგიური POS-ანოტაცია, რომლის დროსაც ყველა ლემას ენიჭება მეტყველების ნაწილის აღმნიშვნელი მარკერი;
- გამოვყოთ შეპირისპირებითი ანალიზისათვის საერთაშორისო სასწავლო კორპუსების შერჩევის კრიტერიუმები;
- გავაანალიზოთ შესიტყვებათა კვლევის ასოციაციური საზომების უპირატესობა და კონტრასტული ანალიზისათვის საჭირო სანდოობის (სიგნიფიკანტურობის) ტესტის log-likelihood მნიშვნელობა.

III. ანალიზი და დასკვნების გამოტანა საკვლევ ტექსტებში გამოვლენილი მაგალითების საფუძველზე (კვლევის ინდუქციური მეთოდი):

- კონტრასტული ანალიზის გზით დავადგინოთ: ა) რა სტრუქტურის მქონე შესიტყვებით მოდელებს ვხვდებით ჩვენ მიერ შექმნილ სასწავლო კორპუსსა და ინგლისური ენის აზიელი შემსწავლელების საერთაშორისო სასწავლო კორპუსში? ბ) რომელი მრავალკომპონენტური ლექსიკური ფრაზა გვხვდება ჭარბად ან იშვიათად ინგლისური ენის ქართველ შემსწავლელთა დისკურსში ამერიკელ სტუდენტთა სასწავლო კორპუსთან შედარებისას? გ) გამოვყოთ რომელია ყველაზე ხშირად გამოყენებული, სტატისტიკურად დაანგარიშებადი ლექსიკური ელემენტები კოლოკაციების ასოციაციური საზომების გამოყენებით და გამოვავლინოთ შესიტყვებების ვიზუალური გრაფების ანალიზის შედეგები;
- გავაანალიზოთ სამ- და ოთხკომპონენტური ლექსიკური ფრაზების ფუნქციურ-დისკურსული მახასიათებლები;
- ტიპ/ტოკენებისა და ლემა/ტოკენების შეფარდების სიდიდეზე დაყრდნობით დავადგინოთ ინგლისური ენის ქართველ შემსწავლელთა ენობრივი სიმდიდრე, როგორც მრავალგანზომილებიანი ფენომენი;
- წარმოვადგინოთ ჩვენ მიერ შექმნილ სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტური ლექსიკური ფრაზის სალექსიკონო სიტყვა-სტატია. აღნიშნული მასალა შესაძლებელია გახდეს

სპეციალიზებული, აკადემიურ წერაზე ორიენტირებული (EAP) ლექსიკონის შემადგენელი ნაწილი.

ნაშრომის ემპირიული ბაზა მოიცავს ინგლისური ენის 65 ქართველ შემსწავლელთა წერილობით მასალას, რომელიც რეპრეზენტატიული უნდა იყოს არა მხოლოდ რაოდენობრივად, არამედ თვისებრივადაც. იგი უნდა წარმოგვიდგენდეს უცხოური ენის განსხვავებულ ჟანრულ რეპერტუარს: იქნება ეს თავისუფალი კომპოზიციის ესეები, სამეცნიერო ტექსტები, პრესის მასალები და ა.შ. სასწავლო კორპუსში ტექსტებისა და ტოკენების რაოდენობა შემდეგნაირად ნაწილდება: მასში შეტანილია 34,527 ტექსტი, ხოლო ტოკენების რაოდენობა 9,812,931. ინგლისური ენის ქართველ შემსწავლელთა კორპუსის მონაცემების შეგროვება სამი წლის განმავლობაში მიმდინარეობდა. ჟანრული მრავალფეროვნებისა და თემატური დაბალანსებულობის პრინციპიდან გამომდინარე შევაგროვეთ, ერთი მხრივ, ახალგაზრდა ქართველი ჟურნალისტების მიერ შესრულებული ინგლისურენოვანი პრესა: პოლიტიკური და აპოლიტიკური სტატიები. მეორე მხრივ, სასწავლო კორპუსის რესურსი შეივსო ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის ჰუმანიტარულ მეცნიერებათა ფაკულტეტის, ბაკალავრიატისა და მაგისტრატურის საფეხურის, ინგლისური ფილოლოგიის მიმართულების სტუდენტების ინგლისურენოვანი ესეებით, კორესპონდენციით (ელექტრონული მეილებით) და ინგლისურენოვანი ბლოგებით.

დისერტაციის ძირითადი თეორიული ბაზა ეფუძნება საკითხისადმი ინტერდისციპლინურ მიდგომას და გამოიხატება შემდეგი მიმართულებებით: კორპუსლინგვისტიკა (ტექსტკორპუსების ტიპოლოგია და კლასიფიკაცია (სინკლერი, 1995), სასწავლო კორპუსების კვლევა (გრეინჯერი და სხვ., 2015), შესიტყვებების კვლევა სასწავლო კორპუსში (ნესელჰაუფი, 2005); მეტალექსიკოგრაფია (ელექტრონული ლექსიკოგრაფიის უახლესი მიღწევები, ლექსიკოგრაფიის პრინციპები და პრაქტიკა (გოუესი და პრინსლო, 2010); დისკურსის ანალიზი (დისკურსის მარკერები სხვადასხვა ენაში (სიპმანი, 2004), დისკურსის მარკერები აკადემიური წერაში (ო'ფლინი, 2022), სემანტიკა და დისკურსი: კოლოკატები, საკვანძო სიტყვები (ბრეზინა, 2018); განათლების მეცნიერებები (სასწავლო კორპუსის კვლევა და მეორე/უცხოური ენის სწავლების მეთოდოლოგია (ლუ, 2016).

კვლევის მეთოდოლოგია. კვლევა წარმართულია კორპუსზე დამყარებული მეთოდოლოგიით, რომლის დროსაც მეცნიერული პრობლემის დასმის წყაროს წარმოადგენს იმთავითვე კორპუსში წარმოებული კვლევის შედეგების ანალიზი. კვლევა წარმართა იმ პროგრამული ინფრასტრუქტურით, რომელიც სხვადასხვა სახის მონაცემების ავტომატური ანოტირების საშუალებას იძლევა. სასწავლო კორპუსის მასალის შეგროვებისას, კვლევაში მონაწილეებმა შეავსეს ენობრივი და ავტობიოგრაფიული კითხვარი.

ჩვენი კვლევისას უპირატესობა ენიჭება ლინგვისტიკის კვლევის ინდუქციური მეთოდის გამოყენებას, რაც გვამღებს ემპირიული მასალის განზოგადებიდან ძირითადი დასკვნების გამოტანისა და მათი შემდგომი გადამოწმების საშუალებას. პირველ რიგში, დადგენილია სიხშირის, სტრუქტურისა და ფუნქციის თვალსაზრისით რომელი კოლოკაციური წყვილები, სინტაგმები, „ლექსიკური სათამაშო დათუნიები“ (ფართოდ გავრცელებული, მაღალი სიხშირის მქონე ფრაზეოლოგიზმები) გვხვდება ქართველი შემსწავლელების ინგლისურენოვან ტექსტებში. ენაში რეალიზებული სიტყვათხმარების ტიპოლოგიური დახასიათებისას დისერტაციაში გამოყენებულია კ. ჰაილენდის, ჯ. ო'ფლინის, ისევე როგორც დ. სიპმანის მიერ მოწოდებული მეტადისკურსის ფუნქციების კლასიფიკაცია აკადემიურ ტექსტებში (ჰაილენდი, 1998, 2008; სიპმანი, 2004, ო'ფლინი, 2022).

კვლევის რაოდენობრივი მეთოდი დაგვეხმარა სასწავლო კორპუსში წარმოდგენილი წერილობითი მასალისა და სხვა მეტამონაცემების სხვადასხვა კრიტერიუმით შესწავლაში, ასევე საკვლევ ტექსტებში მოძიებული ინფორმაციის რაოდენობრივ შედარებასა და პროცენტული მაჩვენებლების გამოთვლაში. ამ პროცესში დისერტაციაში გამოყენებულია ვ. ბრეზინას შემოთავაზებული ასოციაციური საზომები: კოლოკაციათა გავრცელებულობა, ექსკლუზიურობა და მიმართულებითობა, რომლებიც ენის შემსწავლელებში შესიტყვებების ცოდნისა და გამოყენების სიხშირეს გამოავლენს (ბრეზინა, 2018). მონაცემების სიხშირის გამოთვლის შემდეგ, გამოკვლეულია და კონტრასტული გზით შედარებულია ინგლისური ენის ქართველი შემსწავლელების სიტყვათხმარების ასპექტები ამერიკელ სტუდენტთა ესეების სასწავლო კორპუსთან მიმართებით. ხოლო,

ლექსიკური სიმდიდრე, როგორც მრავალგანზომილებიანი კონსტრუქტი, შეპირისპირებულია აზიელი სტუდენტების სასწავლო კორპუსთან.

მონაცემების სიხშირის გამოთვლის შემდეგ, მეტალექსიკოგრაფიული კვლევის კუთხით, გამოკვლეულია ის შესაძლებლობები, რომელიც ჩვენი სასწავლო კორპუსიდან EAP ლექსიკონებში ჯეროვანი მასალის ექსტრაპირების საშუალებას იძლევა. სასწავლო კორპუსის ანალიზის საფუძველზე ნიმუშის სახით წარმოვადგენთ 30 სიტყვა-სტატიას და მასში გაერთიანებულ ტექსტობრივ გრაფებს, რომელიც სპეციალიზებული ტიპის (EAP) ინგლისურ-ქართული ლექსიკონის მიკროსტრუქტურის ნაწილი შეიძლება გახდეს.

გარდა ამისა, დისერტაციაში განხილული თეორიული და პრაქტიკული პრობლემატიკა, მისი დასკვნები შეიძლება გამოყენებულ იქნეს განათლების მეცნიერებების, სასწავლო ლექსიკოგრაფიის, მეტალექსიკოგრაფიის, ფსიქოლინგვისტიკის, ბუნებრივი ენის დამუშავებისა და ინგლისურის, როგორც მეორე ენის სწავლების კურსებში. დისერტაციის პრობლემატიკა საინტერესოა, ასევე, ზოგადად, უცხოური ენის შესწავლითა და ლექსიკური ფონდით დაინტერესებული მკითხველისთვის. სასწავლო კორპუსი ერთ-ერთი წინაპირობაა საქართველოში ემპირიული ცნობიერების განვითარებისათვის, როგორც საზოგადოების ისე სამეცნიერო საზოგადოების ფარგლებში, განსაკუთრებით ეფექტური მეცნიერული კვლევის წარმართვის მიზნით. კერძოდ: მეთოდოლოგიური თვალსაზრისით, იგი სერიოზულ ემპირიულ საფუძველს უქმნის ენის ინტერნალიზაციისას შეცდომების გამოვლენაზე ორიენტირებულ კვლევებს. რაც შეეხება პრაქტიკულ შედეგებს, ინგლისური ენის ქართველი შემსწავლელების კორპუსი საშუალებას მოგვცემს ენის სტრუქტურისა და განვითარების ტენდენციები თანამედროვე ტექნოლოგიური საშუალებებით ვიკვლიოთ.

დისერტაცია შედგება თეორიული და პრაქტიკული ნაწილებისაგან. თეორიული მხარე მოიცავს სასწავლო კორპუსების არსს, მისი კვლევის მეთოდოლოგიას, სასწავლო კორპუსებში შესიტყვებებისა და მრავალსიტყვიანი ლექსიკური ფრაზების შესწავლის ასპექტებს უცხოური ლიტერატურის გათვალისწინებით. ნაშრომის პრაქტიკული მხარე კი შეიცავს ლინგვისტურ, ლექსიკოგრაფიულ კვლევას, რომელიც წარიმართა ინგლისური ენის ქართველ

შემსწავლელთა სასწავლო კორპუსის ტექსტობრივ მასალაზე დაყრდნობით. დისერტაციის ძირითადი ნაწილებია: შესავალი, ოთხი თავი და ძირითადი დასკვნები და დებულებები, ბიბლიოგრაფიის სია. ნაშრომში წარმოდგენილია სქემები, სურათები, დიაგრამები და შემოკლებათა სია.

I თავში - *„სასწავლო კორპუსი როგორც ინტერდისციპლინური კვლევის რესურსი“* - ჩვენ ვეხებით სასწავლო კორპუსის კვლევისა და მეთოდოლოგიის თეორიულ წანამძღვრებს, მის როგორც თეორიულ, ისე პრაქტიკულ მხარეებს: 1. სასწავლო კორპუსის ტიპოლოგიის, დიზაინის (გარემოს, დავალებისა და ენის შემსწავლელთა ცვლადების) მიმოხილვა ს. გრეინჯერის თეორიის მიხედვით; 2. შესიტყვებისა და მრავალსიტყვიანი ლექსიკური ერთეულების კვლევის ასპექტები სასწავლო კორპუსში. განხილულია ფრაზოლოგიზმების კვლევის მნიშვნელობაზე ორიენტირებული და სიხშირეზე დამყარებული მიდგომა და მათი უპირატესობები ან ნაკლოვანებები; 3. მოცემულია შესიტყვებათა ლექსიკური და გრამატიკული პარადიგმები, ლექსიკური ფრაზების კ. ჰაილენდის, დ. სიპმანისა და ჯ. ო'ფლინის მიერ შემოთავაზებული ფუნქციური ტაქსონომიები; 4. სასწავლო კორპუსის მასალის ლექსიკოგრაფიული გამოყენების შესაძლებლობები, სადაც განხილულია ლექსიკოგრაფიული ტექსტური გრაფების გამოყენების მნიშვნელობა და მრავალკომპონენტიანი ლექსიკური ფრაზების სალექსიკონო მიკროსტრუქტურის საკითხები.

II თავში - *„ინგლისური ენის ქართველ შემსწავლელთა კორპუსის შექმნის პრინციპები“* - ჩვენ განვიხილავთ კორპუსში შემავალი ტექსტების შერჩევის კრიტერიუმებს, ჟანრულ კლასიფიკაციასა და მონაცემთა მოგროვების დეტალებს. აგრეთვე, კონტრასტული ანალიზისათვის საერთაშორისო სასწავლო კორპუსების შერჩევის მიზეზებსა და კრიტერიუმებს.

III თავში - *„ინგლისური ენის ქართველ შემსწავლელთა კორპუსის (GLEAN) ანალიზი“* - ჩვენი კვლევის საგანია ორკომპონენტიანი შესიტყვებების სტრუქტურულ-სემანტიკური ანალიზი, განსაკუთრებული ყურადღება გავამახვილეთ გამაძლიერებელი კოლოკაციების (very, utterly, extremely და ა.შ. + ზედსართავი სახელის) კვლევაზე. აგრეთვე წარმოდგენილია სამ- და ოთხკომპონენტიანი ლექსიკური ფრაზების ფუნქციურ-დისკურსული ანალიზი.

IV თავში - „*ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსი (GLEAN), როგორც წყარო ინგლისურ-ქართულ სპეციალიზებული (EAP) ლექსიკონის სიტყვა-სტატიის შესადგენად*“ - ეხება სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტიანი ლექსიკური ფრაზის სალექსიკონო სიტყვა-სტატიის მაკრო- და მიკროსტრუქტურულ აღწერას. თითოეული სიტყვა-სტატიის ელემენტი მცირე მოდიფიკაციებით ეფუძნება დ. სიპმანის მიერ შემოთავაზებულ სამეთაურო ფრაზების ლექსიკონში ასახვის პრინციპებს. ლექსიკონის მაკროსტრუქტურის ჩამოყალიბებისას ვიყენებთ კონცეპტუალურ, თემატურ მეთოდს, რომელიც ემყარება როგორც ჯ. ო'ფლინის ფუნქციურ ტაქსონომიას, ისე დ. სიპმანის დისკურსის მარკერების თემატურ-ფუნქციურ მწკრივებსაც. სიტყვა-სტატიების ტექსტური გრაფებში ასახული ინფორმაცია, საილუსტრაციო წინადადებები, ფუნქციური მარკერები ეფუძნება GLEAN სასწავლო კორპუსის მასალას. ბოლოს კი განვსაზღვრავთ სამომავლო პერსპექტივებს, მაგალითად, შეცდომათა ტეგერის შექმნასა და სასწავლო კორპუსის მასალით უფრო გამდიდრებას.

თავი I. სასწავლო კორპუსი როგორც ინტერდისციპლინური კვლევის რესურსი (თეორიული მიმოხილვა)

1.1. სასწავლო კორპუსის გაგებისათვის

ციფრული რესურსების ავტომატური გადამუშავების ტექნოლოგიების განვითარებამ, სრულიად შეცვალა ენობრივი რესურსების მოპოვების, სტრუქტურირებისა და ანალიზის შესაძლებლობები. ტექსტური კორპუსები ენობრივ მონაცემთა თანმიმდევრულ და მრავალმიზნობრივ, ინტერდისციპლინარული კვლევებისათვის (მაგალითად, ლექსიკოგრაფიული, ფსიქოლინგვისტური თუ ბუნებრივი ენის დამუშავების) გათვლილ რესურსად იქცნენ. შეგვიძლია სკრუპულოზულად შევისწავლოთ ენაში მიმდინარე ცვლილებები და მათი გამომწვევი მიზეზები, გამოვიკვლიოთ ენობრივი ქცევის ნორმები, მოვახდინოთ ენობრივი სიტუაციიდან გამომდინარე შეცდომების პროგნოზირება, სწორად დავგეგმოთ და განვახორციელოთ ენობრივი პოლიტიკა. კორპუსლინგვისტიკის მეთოდოლოგიის მასშტაბურობის საჩვენებლად, ჯ. ლიჩი ჰაბლის ტელესკოპის ანალოგიას მოიხმობს, რომელიც კოსმოსის ყოვლისმომცველი, დეტალური და ზუსტი კვლევის საშუალებას იძლევა, ზუსტად ისე, როგორც კორპუსლინგვისტური მეთოდოლოგია, რომელიც ენობრივი მოდელების დანახვასა და აღმოჩენას ამარტივებს (ლიჩი, 2011).

კორპუსლინგვისტიკის განვითარების საწყის ეტაპზე უცხოური ენის შესწავლის ასპექტები დღის წესრიგში არ იდგა. ვითარება ბოლო 40-50 წლის განმავლობაში შეიცვალა, როდესაც კორპუსზე დამყარებულმა კვლევებმა განაპირობა დეტალურად აღწერილიყო წერისა და მეტყველების ისეთი განსხვავებული რეგისტრები, როგორიცაა არაოფიციალური საუბარი, აკადემიური წერა, ოფიციალური მეტყველება, პუბლიცისტური სტილი და ა.შ. საკითხის ამ კუთხით დანახვამ წინა პლანზე წამოსწია სასწავლო კორპუსის შექმნის საჭიროება და კვლევის ფართო სპექტრს დაუდო სათავე. მან მნიშვნელოვანი კავშირი დაამყარა ერთი შეხედვით თითქოს არსებითად განსხვავებულ დარგებს – კორპუსლინგვისტიკას, უცხოური/მეორე ენის ცოდნის დონის კვლევასა და ლექსიკოგრაფიის თეორიას შორის.

სასწავლო კორპუსი წარმოადგენს წერილობითი და ზეპირი ტექსტების ერთობლიობას, რომლებიც უცხოური ენის შემსწავლელთა ენობრივ რეპერტუარს ასახავს. სასწავლო კორპუსების ერთ-ერთი მთავარი მახასიათებელია ლექსიკურ თავისებურებებსა და შეცდომების გამოვლენაზე ორიენტირება: მათი კატეგორიზაცია, ანალიზი და აღბეჭდვა. იგი ერთგვარი „შავი ყუთია“, რომელშიც რეგისტრირდება, გენერირდება და, საბოლოო ეტაპზე, ნათელი ეფინება ენის შემსწავლელთა ლექსიკისთვის დამახასიათებელ ნიშან-თვისებებს.

ასეთი ტიპის კორპუსი მნიშვნელოვანია იმ თვალსაზრისითაც, რომ სასწავლო კორპუსი უცხოური ან მეორე ენის (L2) შემსწავლელთა ენას წარმოაჩენს მთელი სისავსით. ეს არსობრივი ელემენტი სამეცნიერო ლიტერატურაში სასწავლო კორპუსის ცნების განსაზღვრისას მკაფიოდაა ხაზგასმული. მაგალითად, ნ. ნესელჰაუფი ასეთ დეფინიციას გვთავაზობს: *„სასწავლო კორპუსი არის სისტემატური, ელექტრონულ მატარებელზე გადატანილი ტექსტების ნაკრები, რომელიც ენის შემსწავლელთა მიერაა შესრულებული“* (ნესელჰაუფი, 2005: 125). ცნებაში „სისტემატური“ იგულისხმება კორპუსში მოცემული ტექსტების რეპრეზენტაციულობა და დაბალანსებულობა.

უცხოური ენის შემსწავლელთა ენობრივი მასალის ელექტრონულ კრებულზე წვდომას ორი საგულისხმო უპირატესობა აქვს: პირველ რიგში, ვინაიდან ენობრივი მასალა გროვდება საკმაო რაოდენობის ენის შემსწავლელებისგან, შეგვიძლია დავასკვნათ, რომ ლინგვისტური მონაცემები რეპრეზენტაციულია; მეორე მხრივ, ელექტრონულ ფორმატში, მასალის დამუშავება ანალიზის პროცესს მნიშვნელოვნად აჩქარებს და არ საჭიროებს შრომატევად მანუალურ დამუშავებას. კერძოდ, მეტყველების ნაწილების ანოტაციის მეშვეობით სასწავლო კორპუსის თითოეულ სიტყვაფორმას ენიჭება შესაბამისი გრამატიკული კატეგორიის თეგი, რომელიც ამარტივებს ენის შემსწავლელთა მიერ გრამატიკული კატეგორიების, მაგალითად, წინდებულებისა და დამხმარე ზმნების გამოყენების თავისებურებებზე დაკვირვებას. კონკორდანსი კი ლექსიკისა და ფრაზეოლოგიის კვლევა-ძიებას შესაძლებლად ხდის სასწავლო კორპუსში ცალკეული სიტყვებისა და ფრაზების განმეორებადობის კვლევას; წარმოაჩენს ენობრივი ერთეულების კონტექსტში გამოყენების შემთხვევებს და გვიჩვენებს სიტყვათა კოლოკაციებს და სიტყვათა კლასტერებს. მიუხედავად იმისა,

რომ სასწავლო კორპუსების კვლევა შედარებით ახალი მიმართულებაა, უკვე განვითარების ანგარიშგასაწევ საფეხურებს მიაღწია.

აქტუალურია საკითხი იმის თაობაზე, თუ რა ხარისხითაა სასწავლო კორპუსის ლინგვისტური მასალა ბუნებრივი და ცოცხალი. მიუხედავად იმისა, რომ ნებისმიერი ლინგვისტური კორპუსი მანქანურად დამუშავებად ავთენტურ ენობრივ რესურსებს უყრის თავს (შეიძლება შეიცავდეს როგორც წერილობით მონაცემებს, ისე ზეპირმეტყველების ტრანსკრიბირებულ მასალას), უტყუარობის ცნება, გარკვეული თვალსაზრისით სადავო და პრობლემატურია ენის შემსწავლელთა შემთხვევაში. ს. გრეინჯერი სასწავლო კორპუსის დეფინიციაში დამატებით ხაზს უსვამს იმ გარემოებას, რომ სასწავლო კორპუსი არის *თითქმის* ბუნებრივი ენობრივი რესურსების კრებული. ბუნებრიობასთან მიახლოებულობა იმ მიზეზით აიხსნება, რომ უცხოური ენის შემსწავლელთათვის სამიზნე ენის გამოყენება რიგ შემთხვევაში შეზღუდულია და მოქცეულია საკლასო ოთახის კონტექსტის საზღვრებში (გრეინჯერი, 2008a: 338).

სასწავლო კორპუსების ავთენტურობის დონე, შესაძლოა, ვარიირებდეს და ეს რამდენიმე ფაქტორით არის გამოწვეული. კერძოდ, ენობრივი მასალის შეგროვებისას სამიზნე ჯგუფებისა და რესპოდენტებისათვის განკუთვნილი დავალების ტიპი ან დროის ლიმიტი სპონტანურობის ხარისხს რაღაც დოზით ამცირებს. მაგალითად, გერმანელი და ამერიკელი სტუდენტების დისკუსიის კორპუსი (Telekorp) გაცილებით დაუგეგმავი და იმპულსურია, ვიდრე წინასწარგანზრახული და შედარებით „ხელოვნური“ ჩარლი ჩაპლინის მუნჯი კინოს თხრობის კორპუსი (Giessen-Long Beach Chaplin Corpus), რომლის შექმნის ფარგლებში მონაწილეების ერთ ნაწილს სთხოვეს ჩარლი ჩაპლინის მუნჯი ფილმის, „ემიგრანტის“, დასაწყისი ენახათ, ხოლო მეორე ნაწილს - დასასრული. კვლევის შემდეგ ეტაპზე კი მათ დაევაღათ, მოეთხროთ ნანახი და ფილმის რამდენიმე საინტერესო ასპექტიც განეხილათ (ჯაკერი, 2003). მიუხედავად იმისა, რომ აღნიშნულ კორპუსში ავთენტურობის ხარისხი დარღვეულია, ავტორების საკვლევ ობიექტს - თუ რამდენად სცდება უცხოური ენის ცოდნა გრამატიკის საზღვრებს და ენის პრაგმატიკა რამდენად ამახვილებს ყურადღებას მეტყველების აქტზე, რელევანტურობა არ დაუკარგავს. მიღებული კვლევის შედეგებზე დაყრდნობით შესწავლილ იქნა ფუნქციური კატეგორიები და დისკურსის მარკერები,

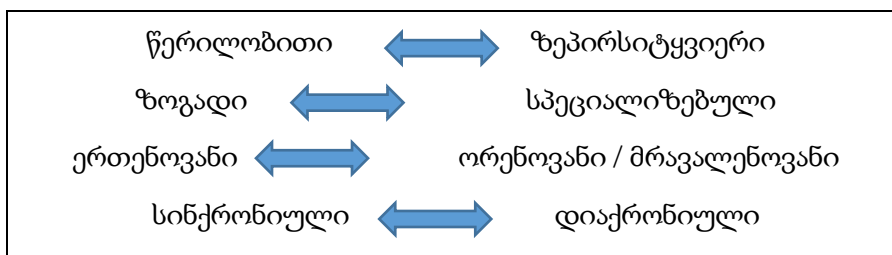
როგორიცაა *well, so, like, you know*, შედარდა ორივე მონაწილე ჯგუფის ლინგვისტური თავისებურებები და მათი ფარდობითი სიხშირე სათანადო გარემო პირობებში (მიულერი, 2005: 260).

თუკი შევაჯამებთ სასწავლო კორპუსის ზემოთხსენებულ განმარტებებს, მისი განსაზღვრებები ერთგვაროვანი არ არის. ნ. ნესელჰაუფისა და ს. გრეინჯერის მიერ მოცემული დეფინიციებისა და სასწავლო კორპუსებისათვის დამახასიათებელი მნიშვნელოვანი ნიშან-თვისებების გათვალისწინებით, ვასკვნით, რომ სასწავლო კორპუსის კვლევა გასცდა ერთი კონკრეტული დისციპლინის, კერძოდ, კორპუსლინგვისტიკის შესწავლის ფარგლებს და გადაიქცა მულტიდისციპლინურ მიმართულებად. სასწავლო კორპუსი წარმოადგენს სისტემატურ, ელექტრონულ მატარებელზე გადატანილ ტექსტების ნაკრებს, რომელიც უცხოური ენის შემსწავლელთა მიერაა შესრულებული. ვინაიდან ავთენტურობის დონე, გარკვეული თვალსაზრისით, დარღვეულია ენის შემსწავლელთა ლინგვისტური მასალის შეგროვების შემთხვევაში, სამართლიანია დავასკვნათ, რომ სასწავლო კორპუსები ნაწილობრივ ბუნებრივ, ავთენტურ მასალას მოიცავენ. თუმცა ბუნებრიობის ეს ხარისხი საკვლევი საკითხების გადაჭრის პროცესში დაბრკოლებას არ წარმოადგენს.

1.1.1. სასწავლო კორპუსის ტიპოლოგია

კორპუსის სტრუქტურისა და გამოყენების მიზნებიდან გამომდინარე, განასხვავებენ სასწავლო კორპუსის რამდენიმე სახეს. ისინი კლასიფიცირდება განზომილებებისა და დიქტომიის მიხედვით. გრაფიკულად წარმოვადგენთ სასწავლო კორპუსების დიქტომიურ ტიპოლოგიას (იხ. სქემა 1).

სქემა 1. სასწავლო კორპუსების კლასიფიკაცია (გილქინი, 2008a: 12).



პირველი განზომილება განსაზღვრავს თუ როგორ, რა სახით გროვდება და გარდაიქმნება ლინგვისტური მონაცემები კორპუსად. სასწავლო კორპუსი შედგება წერილობითი ტექსტური მასალის ან/და დისკურსის ტრანსკრიფციებისგან. ლუვენის უნივერსიტეტის მიერ შედგენილი სასწავლო კორპუსების სიის მიხედვით (“Learner Corpora around the World, LCW”), დროის, ფინანსებისა და ადამიანური რესურსების გათვალისწინებით, დღესდღეობით წერილობითი სასწავლო კორპუსები უფრო ფართოდაა გავრცელებული, ვიდრე ზეპირი ტექსტების ნაკრებები. ზეპირმეტყველების სასწავლო კორპუსებში განირჩევა, ერთი მხრივ, ტრანსკრიბირებული, ანუ ტექსტად გარდაქმნილი მეტყველება და, მეორე მხრივ, აუდიო-ვიზუალური ფაილებით (ხმოვანი და ვიდეოჩანაწერებით) შექმნილი მულტიმედიური კორპუსი, მაგალითისათვის MAELC (the Multimedia Adult ESL Learner Corpus), რომლის მეშვეობითაც მკვლევარი იკვლევს არა მარტო ენის შემსწავლელთა მეტყველების პროსოდიულ და ლექსიკურ თავისებურებებს, არამედ მათ ჟესტიკულაციასა და მზერასაც კი. აღნიშნულ კორპუსზე დამყარებულმა ერთ-ერთმა უახლესმა კვლევამ ნათელყო სოციო-კულტურული ქცევასა და ენის შემსწავლელთა წიგნიერების დონეს შორის არსებული კორელაცია (რამირეს-ესპარზა და სხვ., 2012).

ვინაიდან განირჩევა წერილობითი და წერილობით დოკუმენტირებული ზეპირმეტყველების მასალების სასწავლო კორპუსები, მასალის შეგროვების გზებიც შესაბამისად განსხვავდება. წერილობითი სასწავლო კორპუსის მასალა ხელნაწერია ან ელექტრონულად აკრეფილია. მკვლევრის მიერ ენის შემსწავლელთა ხელნაწერების გაციფრულება საფრთხილოა და ტექსტების რეპროდუცირება ზუსტად ისე უნდა მოხდეს როგორც არის, ყოველგვარი ჩარევისა და მანიპულაციის გარეშე დაშვებული შეცდომების ჩათვლით. ხელნაწერების დიგიტალიზაციის შედარებით უფრო მარტივი და სანდო მეთოდი ხორციელდება ნაწერების დასკანერებით სპეციალურად შექმნილი პლატფორმის გამოყენებით.

ზეპირმეტყველების კორპუსის მასალის შეგროვება ორთოგრაფიული ტრანსკრიპტის გაკეთებით ან ძალიან დეტალური ფონოლოგიური და ფონეტიკური ტრანსკრიბირებით ხორციელდება. საინტერესო მაგალითია The LeaP (Learning Prosody in a Foreign Language) კორპუსი, რომელიც მოიცავს არამხოლოდ ორთოგრაფიულ

ტრანსკრიფციას, არამედ მარცვლების, სეგმენტების, ტონისა და ტონალობის შემცველ ინფორმაციასაც. ზოგადად, თუ კორპუსი იქმნება სამეტყველო ენის ლექსიკური ანალიზის გამოსაკვლევად, მაშინ ორთოგრაფიული ტრანსკრიფცია სრულებით საკმარისია. ხოლო, თუ სასწავლო კორპუსი მიზნად ისახავს ენის შემსწავლელთა წარმოთქმისა და პროსოდიის გამოკვლევას, უმჯობესი იქნება მოხდეს შეგროვებული მასალის ფონეტიკური ტრანსკრიბირება.

მეორე განზომილება, რომლის მიხედვითაც ვახდენთ სასწავლო კორპუსების კატეგორიზებას, არის ჟანრული კლასიფიკაცია. ბოლო ათწლეულის განმავლობაში სასწავლო კორპუსის სპეციალიზებული, დარგობრივი მიზნებისთვის შექმნა და დანერგვა უფრო მოთხოვნადი გახდა. იგი შედგება სხვადასხვა დისციპლინისა და ჟანრული მრავალფეროვნების ტექსტებისაგან, მაგალითად, ესეების, მოხსენებების, ოფიციალური და არაოფიციალური წერილების, შემაჯამებელი ანგარიშებისა და ბიზნესწერილებისაგან (გრეინჯერი და ფაქო, 2013: 3142). სპეციალიზებული სასწავლო კორპუსის ერთ-ერთი მნიშვნელოვანი ნიმუში საბუნებისმეტყველო მეცნიერებათა ფაკულტეტზე მყოფი იაპონელი სტუდენტების კვლევითი ნაშრომების ერთობლიობაა, Active Learning of English for Science Students (ALESS), რომელშიც ერთიანდება საბუნებისმეტყველო მეცნიერებათა სპეციალობების შემსწავლელი იაპონელი სტუდენტების კვლევითი ნაშრომები. მისი შექმნის იდეა უკავშირდებოდა იმის დადგენას, თუ რამდენად ეფექტიანი იყო ინგლისური ენის შესწავლის კურიკულუმები საბუნებისმეტყველო ფაკულტეტის სტუდენტებისათვის (ალენი, 2010).

სასწავლო კორპუსები აგრეთვე განირჩევა სამიზნე ენების შერჩევის მიხედვით. სიაში ინგლისური ენა დომინირებს, თუმცა ბოლო წლებში ახალ პროექტებს ჩაეყარა საფუძველი, რომელთა მიზანიც სხვა ფართოდ გავრცელებული ენების შემსწავლელების მასალის შეგროვებაა, როგორიცაა ფრანგულის (FLLOC, French Learner Language Oral Corpora), გერმანულისა (Falko, Fehlerannotiertes Lernerkorpus) და ესპანური ენის (CEDEL2, Corpus Escrito del Español L2) მასალა. ძირითადად სასწავლო კორპუსები მონოლინგუურია, შეიცავს მხოლოდ ერთი სამიზნე ენის ლინგვისტურ რესურსებს, თუმცა არსებობს მრავალენოვანი სასწავლო კორპუსებიც, მაგალითად MiLC Corpus, რომელშიც გაერთიანებულია კატალანური, ინგლისური, ფრანგული და

ესპანური ენების შემსწავლელთა წერილობითი ნამუშევრები და ორიენტირებულია შეცდომათა ანოტირებასა და აღმოჩენაზე (ტანინი, 2006).

სასწავლო კორპუსი, ერთი მხრივ, შესაძლოა იყოს სამიზნე ენის შემსწავლელთა ცოდნის მომენტალურად ამსახველი, რასაც სინქრონიული სასწავლო კორპუსი მოიცავს, ან, მეორე მხრივ, იყოს დიაქრონიული და ასახოს მოსწავლეთა ენობრივი ცოდნის ევოლუცია რაღაც დროის განმავლობაში. ჯ. ბელცი და ნ. ვიატკინა დიაქრონიული სასწავლო კორპუსების სინონიმად იყენებენ „ევოლუციურ სასწავლო კორპუსს“ (ბელცი და ვიატკინა, 2008: 33). პირველი ტიპის სასწავლო კორპუსები უფრო გავრცელებულია იმ ლოგიკური მიზეზით, რომ დიაქრონიული სასწავლო კორპუსის შედგენა დროის ხანგრძლივი პერიოდის განმავლობაში მიმდინარეობს და დამატებით რესურსებს მოითხოვს. დიაქრონიული სასწავლო კორპუსის ერთ-ერთი ნიმუშია, LONGDALE მონაცემთა ბაზა (The Longitudinal Database of Learner English), რომელშიც ენაზე დაკვირვების, პროგრესის ასახვისა და კორპუსის გამდიდრების მიზნით ერთსა და იმავე შემსწავლელთა მასალა გროვდებოდა რამდენიმე წლის განმავლობაში.

დღეს არსებული სასწავლო კორპუსების უმრავლესობა განივი კვეთის (cross-sectional) ტიპისაა, რომელშიც მონაცემები შეგროვებულია დროის ერთ შუალედში ინფორმანტების მრავალრიცხოვანი რაოდენობის ჩართულობით. მეორე მხრივ, სასწავლო კორპუსს აქვს პოტენციალი აჩვენოს უცხოური/მეორე ენის შესწავლის განვითარების ეტაპები და ეს შესაძლებლობა რეალიზებულია გრძივი ტიპის (*longitudinal*) სასწავლო კორპუსებში. ასეთი ტიპის კორპუსში მოგროვებულია მცირე რაოდენობის რესპოდენტთა ენობრივი მონაცემები პერიოდულად, ხანგრძლივი დროის განმავლობაში. ამგვარი გზით გამოაშკარავდება რამდენად იცვლება შემსწავლელთა ენობრივი ცოდნა და ლინგვისტური თავისებურებები გარკვეული დროის შემდეგ. ამას გარდა, არსებობს ფსევდო-გრძივი (ნაწილობრივ-გრძივი) ტიპის კორპუსი (*quasi-/pseudo-longitudinal corpus*), რომელშიც ენის შემსწავლელთა რამდენიმე ჯგუფის, კატეგორიის მასალა ერთ კონკრეტულ დროსაა შეგროვებული (გრეინჯერი და სხვ., 2015).

როგორც ვხედავთ, სასწავლო კორპუსები საკვლევი პრობლემის შესაბამისად არაერთგვაროვანი შეიძლება იყოს. ისინი იყოფა რამდენიმე განზომილების მიხედვით:

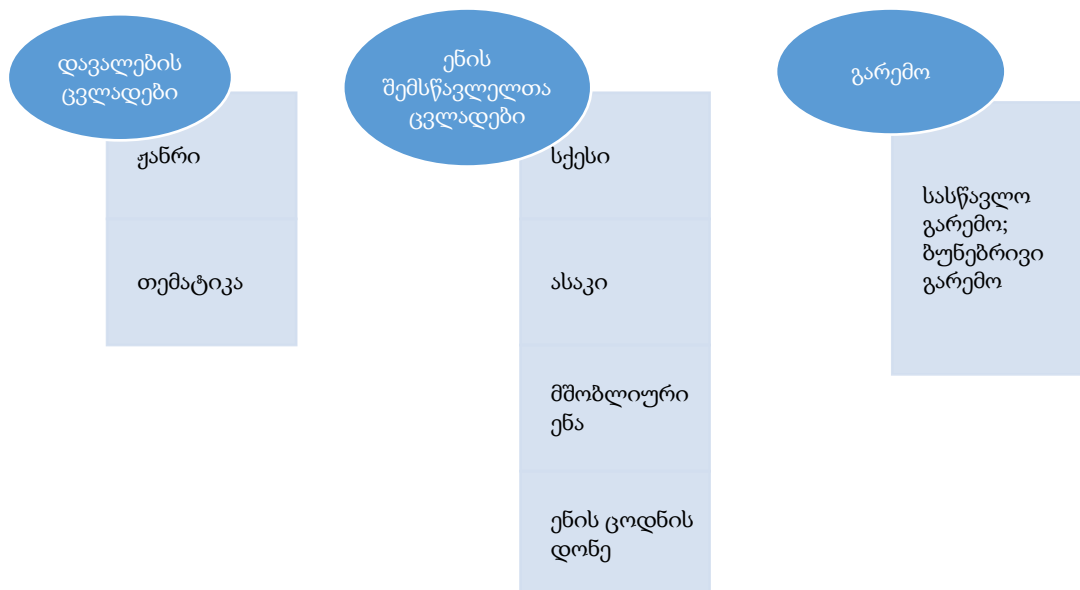
ა) რა ტიპის ლინგვისტური მასალა (ზეპირმეტყველების თუ წერილობითი)

გარდაიქმნება კორპუსად; ბ) ჟანრული კლასიფიკაციის თვალსაზრისით, სპეციალიზებულ, რომელიმე დარგზე ორიენტირებულ მასალას ეფუძნება სასწავლო კორპუსი თუ უფრო ზოგადი ხასიათისაა; გ) სამიზნე ენების შერჩევის მიხედვით; დ) დროითი კატეგორიის მიხედვით, კერძოდ, ცოდნის მომენტალური აღმზეჭდავი არიან თუ უცხოური ენის შემსწავლელთა ენას დიაქრონიულად შეისწავლიან.

1.1.2. სასწავლო კორპუსის დიზაინი: გარემოს, დავალებისა და ენის შემსწავლელთა ცვლადები

კორპუსის შექმნისას წინასწარ განსაზღვრული კრიტერიუმების შემუშავება და გათვალისწინება სამეცნიერო ლიტერატურაში მნიშვნელოვან ადგილს იკავებს. სასწავლო კორპუსში თავმოყრილი ენობრივი მასალა ჰეტეროგენული ხასიათისაა და გარემო პირობების, დავალებებისა თუ ენის შემსწავლელთა ცვლადების მიმართ მგრძნობიარეა (ატკინსი და სხვ., 1992). სასწავლო კორპუსების დაგეგმვისას გასათვალისწინებელია ისეთი წინაპირობები, როგორიცაა გარემო, შესასრულებელი დავალებები და საკუთრივ ენის შემსწავლელების შესახებ არსებული ინფორმაცია (იხ. სქემა 2).

სქემა 2. სასწავლო კორპუსის ცვლადები (გრეინჯერი და სხვ., 2015: 16)



სანამ სასწავლო კორპუსის შეგროვების პროცესი დაიწყება, მკვლევარმა საკვლევი პრობლემის დასმის შემდეგ, უნდა გადაწყვიტოს, რა სახის მონაცემები შევა კორპუსში, რადგან არაერთგვაროვანი და ერთმანეთთან შეუთავსებელ მონაცემთა სიმრავლე კვლევის წარმოებისთვის უსარგებლოა (გრეინჯერი, 2013a: 3235). მაგალითად, წინა ქვეთავში განხილული განზომილებების გათვალისწინებით, თუ მკვლევარს სურს ენის შემსწავლელთა ზეპირი ან წერილობითი მონაცემების შეგროვება, კორპუსის შექმნაზე, მის ანალიზსა და ინტერპრეტაციაზე გარემო პირობების, შესასრულებელი დავალებებისა და მონაწილეთა მეტამონაცემები გავლენას ცალსახად მოახდენს.

წინასწარ მხედველობაში უნდა იქნეს მიღებული ის გარემო, რომელშიც მასალა გროვდება, ბუნებრივ სასწავლო გარემოში თუ საგანმანათლებლო გარემოში. ამ ფაქტორების ერთმანეთისაგან გამოჯვანა განსაკუთრებით მნიშვნელოვანია მეორე ენის სასწავლო კორპუსის შექმნისას და შედეგების ინტერპრეტაციისას, რადგან მეორე ენა სოციალური კონტექსტის უფრო მრავალფეროვან სპექტრში გამოიყენება და ბუნებრივია, უცხოური, მესამე ენის შემსწავლელთა ენობრივი სურათისგან განსხვავებულ მაჩვენებელს მივიღებთ.

კვლევაში მონაწილეთა მიერ შესასრულებელი დავალებების ცვლადები მჭიდრო კავშირშია ჟანრისა და მედიუმის ცნებებთან. მაგალითად, არგუმენტირებული ესეს წერის ან ზეპირსიტყვიერად ფოტოს აღწერის შემთხვევაში, სავსებით განსხვავებული კვლევითი მექანიზმი ამოქმედდება. წერილი დავალებების შესრულების დროს გასათვალისწინებელია დროის ფაქტორი (აქვთ თუ არა კვლევაში მონაწილეებს დროის ლიმიტი), დამხმარე საშუალებების ხელმისაწვდომობა (ლექსიკონები ან გრამატიკის სახელმძღვანელოები), კომპიუტერიზაცია (ენის შემსწავლელებმა დავალება ხელით, წერილობით შეასრულეს თუ კომპიუტერის მეშვეობით). ზეპირი მეტყველების სასწავლო კორპუსის მასალის შეგროვების დროს კი ანგარიშგასაწევია მოსამზადებელი დრო (ენის შემსწავლელებს ჰქონდათ თუ არა ამა თუ იმ დავალების შესრულებისას წინასწარ მოსამზადებელი დრო), წერილი დამხმარე საშუალებები (ბარათები ჩანაწერისათვის) და ა.შ.

ენის შემსწავლელთა ასაკის, სქესის, ეროვნების, მშობლიური ენის, უცხოური ენის ცოდნის და სხვა ტიპის მეტაინფორმაცია გროვდება კითხვარებით, რომელთაც

თან ერთვის ინფორმირებული თანხმობის ფორმა მონაწილეთა მონაცემების გამოყენების ნებართვისათვის სამეცნიერო-კვლევითი მიზნებისათვის. როგორც კი სასწავლო კორპუსის დაგეგმვის პროცესი სრულდება, მასალის შეგროვების დაწყებით პირველი ქმედითი ნაბიჯი იდგმება. კვლევაში მონაწილეთა შერჩევისას, ხშირად, კორპუსის შემქმნელი დამოკიდებულია რესპოდენტების/მონაწილეების კეთილ ნებაზე. უფრო სარწმუნოა თუ თავდაჯერებული და მოტივირებული ენის შემსწავლელები კვლევაში მონაწილეობის სურვილს გამოთქვამენ. ამ შემთხვევაში, მკვლევარს უწევს თავადვე შეარჩიოს საკვლევი პოპულაცია, რამაც შესაძლოა სასწავლო კორპუსის თვისობრივი მახასიათებლები (რეპრეზენტაციულობა და ბალანსირება) საფრთხის წინაშე დააყენოს არსებული მიკერძოების გამო. თუმცა, ტ. მაკენერის პოზიციით, „სასწავლო კორპუსის ბალანსირება და რეპრეზენტაციულობა შეიძლება გააზრებულ იქნეს შედარებით პერსპექტივაში, ანუ კორპუსი უნდა იყოს იმ მასშტაბით რეპრეზენტაციული, როგორც ამის შესაძლებლობა არსებობს. კორპუსის შექმნა სრულყოფილებისა და პრაქტიკულობის მჭიდრო კავშირში მდგომარეობს“ (მაკენერი და სხვ., 2006: 73). ენის შემსწავლელთა ცვლადების სიუხვით ხასიათდება უცხოური ენის ფრანგი შემსწავლელების ზეპირმეტყველების კორპუსი (PARallèle Oral en Langue Etrangère), რომელშიც თითოეული რესპოდენტის მოსმენის უნარ-ჩვევები, გრამატიკული და ლექსიკური კომპეტენციისა და მოტივაციის ცვლადები ინდივიდუალურადაა გამოკვლეული (ჰილტონი, 2008).

დასკვნის სახით შეგვიძლია ვთქვათ, რომ რასაკვირველია ყველა ცვლადის კონტროლი შეუძლებელია, მაგრამ სასწავლო კორპუსის დიზაინის შექმნისას, მკვლევარმა საერთო, მუდმივი მახასიათებლები უნდა გამოჰყოს, გაითვალისწინოს გარემოს, დავალებისა და ენის შემსწავლელთა ცვლადები, რადგან ამ ფაქტორების გაუთვალისწინებლობის შემთხვევაში, როგორც ლ. ბარნარდი აღნიშნავს, მეტამონაცემების გარეშე, მკვლევარს ხელთ სხვა არაფერი აღმოაჩნდება, გარდა გაურკვეველი წარმომავლობის და შეუცნობელი სიტყვებისა (ბარნარდი, 2005: 31).

1.1.3. სასწავლო კორპუსების მეთოდოლოგია

სასწავლო კორპუსები წარმოაჩენენ სისტემის მიხედვით ორგანიზებულ ბუნებრივ და კონტექსტუალიზებულ ენას, რომელსაც უცხოური ან მეორე ენის შემსწავლელები იყენებენ. ისევე, როგორც მონაცემთა შეგროვებისათვის საჭირო იყო განსაზღვრული ინსტრუმენტების შერჩევა, მეთოდოლოგიის არჩევაც დამოკიდებულია მკვლევრის მიერ დასმულ საკვლევ საგანსა და კითხვებზე. კორპუსის მეთოდოლოგიურად გამოყენების რამდენიმე ძირითადი გზა არსებობს. მონაცემთა ამორჩევის, ანალიზისა და ინტერპრეტაციის მიხედვით, განარჩევენ კორპუსით ინფორმირებულ (*corpus-informed*), კორპუსზე დამყარებულ (*corpus-based*) და კორპუსით განპირობებულ (*corpus-driven*) მიდგომებს. გასათვალისწინებელია, რომ თითოეული მათგანის პრინციპები ნაწილობრივ ერთმანეთს ემთხვევა და მათი იზოლირებულად აღქმა მართებული არაა.

კორპუსით ინფორმირებული მიდგომის მიხედვით, სასწავლო კორპუსი გამოიყენება მკვლევრის ინტუიციის შესამოწმებლად, ტექსტებში რაიმე კონკრეტული ლინგვისტური თავისებურების (მაგ., სიტყვის სიხშირის, ფრაზისა ან კონსტრუქციის) ან რომელიმე ტიპის შეცდომის გამოსავლენად. კორპუსზე დამყარებული მიდგომა გულისხმობს ისეთ მეთოდოლოგიას, რომელიც არსებულ თეორიებსა და პრინციპებს ერთგვარად განმარტავს, ამოწმებს და მათ ილუსტრირებას ახდენს (ტონინი-ბონელი, 2001: 65). ასეთი ტიპის ლინგვისტური კვლევით, მკვლევრები ორი სასწავლო კორპუსის კონტრასტული ანალიზის მეშვეობით წარმოაჩენენ ენის შემსწავლელთა ლინგვისტური ბუნებას და, ამ გზით, არსებულ ჰიპოთეზებს ამყარებენ ან აბათილებენ. კორპუსით განპირობებული მიდგომა მკვლევრის ყველაზე ნაკლებ ჩართულობას გულისხმობს, რადგან კვლევის პროცესი მკაცრად ეფუძნება მონაცემთა ამოღებისა და შეფასების კომპიუტერულ მეთოდებს. მკვლევარს გამოაქვს დასკვნები იმის მიხედვით, რასაც კორპუსის მონაცემების სტატისტიკური ანალიზი გამოააშკარავებს. ლინგვისტური ჰიპოთეზის ჩამოყალიბება იმთავითვე უკავშირდება სასწავლო კორპუსს და მეცნიერული პრობლემის დასმის წყაროს წარმოადგენს სასწავლო კორპუსის მასალის ანალიზი (თანდაშვილი და ფურცხვანიძე, 2014).

კიდევ ერთი მთავარი განმასხვავებელი ნიშანი სასწავლო კორპუსის გამოკვლევის პროცესში არის რაოდენობრივი და თვისებრივი ანალიზი. კვლევის მეთოდოლოგია, რომელიც რაოდენობრივ ანალიზზეა დაფუძნებული ძირითადად დედუქციურია, შედეგზე ორიენტირებული და კონკრეტული ჰიპოთეზის შემოწმებისათვის გამოდგება. რაოდენობრივი მონაცემები უტყუარია და სრულყოფილი სტატისტიკური ანალიზის (მონაცემთა იდენტიფიცირების, დათვლისა და კლასიფიცირების) შესაძლებლობას იძლევა. ხარისხობრივი ანალიზი, მეორე მხრივ, ევრისტიკულია, პროცესსა და აღმოჩენაზე ორიენტირებული, კონცენტრირდება ლინგვისტური ფენომენის საფუძვლიან შესწავლაზე და ეფუძნება ავთენტურ და საკომუნიკაციო ენას. უცხოური ენის შემსწავლელელებზე დაკვირვება და მათი ლინგვისტური თავისებურებების აღწერა-ახსნა ბუნებრივ სოციალურ და კულტურულ გარემოში ფუნდამენტურად მიიჩნევა (ჰასკო, 2013a).

შეიძლება ითქვას, რომ დღეს ყველაზე ფართოდ გავრცელებულია რაოდენობრივი ანალიზის პრინციპებზე დაფუძნებული, კორპუსზე დამყარებული მეთოდოლოგია. ნაშრომში დეტალურად იქნება განხილული მსგავსი ტიპის სასწავლო კორპუსების ნიმუშები, ხოლო მათი შექმნის პრინციპები და მეთოდოლოგია დანერგილი იქნება ჩვენს მიერ შედგენილ ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსში.

კორპუსზე დამყარებული მიდგომები უცხოური ენის შემსწავლელთა ლექსიკური მახასიათებლების დეტალურად ასახვისა და ენობრივი ვარიაციების ნათლად დანახვის შესაძლებლობას იძლევა. დღეს სასწავლო კორპუსები ისეთი აქტუალურია როგორც არასდროს. ამიტომ უჩვეულო არაა, რომ დასავლეთევროპულ სამეცნიერო წრეებში რამდენიმე ისეთი ფასეული სასწავლო კორპუსი შეიქმნა, რომლებიც თავიანთი მეთოდოლოგიითა და საკვლევი საკითხების სიღრმით ნიმუშად მისაჩნევად გვევლინება. წარმოვადგენთ შეგროვებული მასალის ტიპის, დავალების ცვლადისა და ენის განზომილების მიხედვით კლასიფიცირებულ დასავლეთევროპულ და ამერიკულ სასწავლო კორპუსების ნუსხას თავიანთი მახასიათებლებით (იხ. სქემა 3).

სქემა 3. საერთაშორისო სასწავლო კორპუსები (ინგლისური კორპუსლინგვისტიკის ცენტრი, 2024)

როგორც სქემა 3 ასახავს, სასწავლო კორპუსების სამიზნე ენა, ძირითადად

N	სასწავლო კორპუსი	მასალის ტიპი	კვლევის სუბიექტი/დავალება	სამიზნე ენა
1.	LINDSEI Louvain International Database of Spoken English Interlanguage)	ზეპირმეტყველება	ლუვენის უნივერსიტეტის III-IV კურსის სტუდენტების ინტერვიუები	ინგლისური
2.	Leiden Learner Corpus	ზეპირმეტყველება და წერილობითი	ლაიდენის უნივერსიტეტის სტუდენტები / მზის თხრობა; სურათების აღწერა, ესეები	ჰოლანდიური, ფრანგული, იტალიური, პორტუგალიუ რი
3.	Spanish Learner Language Oral Corpus (SPLLOC)	ზეპირმეტყველება	ინტერვიუები, სურათების აღწერა	ესპანური
4.	The Longman Learners' Corpus	წერილობითი	საგამოცდო ესეები	ინგლისური
5.	The Icelandic L2 Error Corpus (IceL2EC)	წერილობითი	არგუმენტირებული ესეები	ისლანდიური
6.	The Cambridge Learner Corpus (CLC)	წერილობითი	ინგლისური ენის კომპეტენციის დამადასტურებელი საგამოცდო ნაწერები	ინგლისური
7.	French Interlanguage Database (FRIDA)	წერილობითი	თავისუფალი კომპოზიციის ესეები, პრესა	ფრანგული

შემთხვევაში, ინგლისური ენაა, რაც განპირობებულია ინგლისური ენის, როგორც საერთაშორისო ენის მნიშვნელობითა და გავრცელებულობით. მიუხედავად იმისა, რომ შესაძლოა რომელიმე ქვემოთ გაანალიზებული კორპუსის შექმნის მიზანი ჩვენი კვლევის მიზანს არ ემთხვეოდეს, მაინც ღირებულად მიგვაჩნია არსებული გამოცდილების გათვალისწინება და მათი ერთგვარ საყრდენად გამოყენება ლექსიკოგრაფიულ საქმიანობაში.

უპირველეს ყოვლისა, ლუვენის ინგლისური ენის ზეპირმეტყველების საერთაშორისო მონაცემთა ბაზასთან დაკავშირებული საკითხების გაანალიზება დავალების ცვლადებისა და რესპოდენტთა მეტამონაცემების კონტექსტში, ჩვენ მიერ შექმნილ სასწავლო კორპუსს არაერთი მნიშვნელოვანი გამოცდილებით გაამდიდრებს. იგი ფართო აუდიტორიისთვის (მკვლევრებისა და პედაგოგებისათვის) განკუთვნილი რესურსია. აღნიშნულ კორპუსზე მუშაობა 1995 წელს დაიწყო და დღემდე მიმდინარეობს. ამჟამად სასწავლო კორპუსში გაერთიანებულია მილიონზე მეტი სიტყვა, რომელიც მოიცავს ინგლისური ენის შემსწავლელთა მეტყველების ამსახველ 130 საათიან ინტერვიუებს. საშუალოდ ინტერვიუების ხანგრძლივობა თოთხმეტი წუთია და თითოეული მათგანი ორთოგრაფიულად ტრანსკრიბირებულია. მინიმალური პროსოდიული ინფორმაცია მოცემულია პაუზების (მოკლე, საშუალო და გრძელი), მარცვლების გახანგრძლივების და ხმოვანთა სახით. მოცემულ კორპუსის მეტამონაცემები გვაწვდის ცნობას ენის შემსწავლელთა ასაკზე, ეროვნებასა და მათ ენობრივ კომპეტენციაზე. კვლევაში მონაწილეთა საშუალო ასაკი 22 წელია, ხოლო უმეტესობა, კონკრეტულად, 79% მდედრობითი სქესისაა. რესპოდენტთა უმეტეს ნაწილს არ უცხოვრია ინგლისურ ენოვან ქვეყანაში. დავალების ცვლადების ანალიზისას იკვეთება, რომ LINDSEI-ში თითოეული ტექსტი შედგება სამი ნაწილისაგან, რომელთაგან ორი დიალოგური ხასიათისაა და ერთი - მონოლოგური: (1) დაძაბულობის განმუხტვისათვის შერჩეული შეკითხვებია, მაგალითად, ქვეყანა, რომელიც მოინახულეს; ფილმი/სპექტაკლი, რომელმაც მათზე შთაბეჭდილება მოახდინა და ა.შ.; (2) დისკუსია თავად მონაწილეთა მიერ არჩეული საკითხების შესახებ; და (3) სურათის აღწერის დავალება (სტეიპლსი და სხვ., 2013: 261).

მონაცემთა შეგროვებისა და ტექსტოლოგიური თვალსაზრისით კიდევ ერთი საინტერესო სასწავლო კორპუსის ნიმუშია ლაიდენის უნივერსიტეტში წარმოებული

პროექტი, რომლის მიზანიც კროს-ლინგვისტური ანალიზით ჰოლანდიელი სტუდენტების რომანული ენების (ფრანგული, იტალიური, ესპანური და პორტუგალიური) ათვისების პრობლემის კვლევაა. სასწავლო კორპუსში თავმოყრილია როგორც წერილობითი, ისე ზეპირმეტყველების მასალაც სხვადასხვა ენობრივი ცოდნის მქონე სტუდენტებისაგან. მისი მიზანია, შედარებითი ანალიზის გზით წარმოადგინოს, სიტყვათა ხმარებისას არსებული სირთულეები და ენობრივი ინტერნალიზაციის სტრატეგიები ცხადად გამოაშკარავოს. მიუხედავად იმისა, რომ ლაიდენის სასწავლო კორპუსი საპილოტე რეჟიმშია, მისი მასალა საჯაროდაა ხელმისაწვდომი, რათა ხელი შეუწყოს მეორე ენის შესწავლის კვლევის წინსვლას.

ლაიდენის სასწავლო კორპუსს აინტერესებს არა მხოლოდ ლექსიკურ სირთულეებზე დაკვირვება, არამედ უცხოური/მეორე ენის ათვისების პროგრესის წარმოჩენაც. ავტორები სამომავლოდ გეგმავენ რესპოდენტების განმეორებით გამოკითხვას და მათ ენობრივ განვითარებაზე დაკვირვებას. კვლევაში მონაწილეობდა 182 რესპოდენტი: საუბრის საერთო დრო 11 საათია და წერილობითი მასალა აღემატება 100,000 სიტყვას. ზეპირმეტყველების მასალა სპონტანური ხასიათის იყო და შეეხებოდა ერთ-ერთი ილუსტრირებული, უტექსტო ზღაპრის („ბაყაყის ამბავი“) ვერბალიზაციას. შედეგმა აჩვენა, რომ მონაწილეთათვის ამბის თხრობის, მონაცემთა შეგროვების ეს ხრიკი ბევრად მასტიმულირებელი და ეფექტური აღმოჩნდა, ვიდრე კონსტრუირებული ინტერვიუებია, რომლის დროსაც მონაწილეები შესამჩნევად ნერვიულობდნენ. წერითი დავალებები მოიცავდა ისეთ საკითხებს, როგორიცაა „დაახასიათეთ თქვენი საუკეთესო მეგობარი“, „აღწერეთ თქვენი იდეალური დღე“, „რას გააკეთებდით ლატარეაში მოგებული ერთი მილიონი ევროთი?“. რაც შეეხება დამატებით კითხვარს, დამოუკიდებელი ცვლადების შესახებ ინფორმაცია შეგროვდა მათი ასაკის, სქესის, საქმიანობის, უცხოურ ენასთან შემხებლობისა და ოჯახის წევრების განათლების შესახებ.

მნიშვნელოვანი კვლევითი სამუშაოს ჩატარების კიდევ ერთ შესაძლებლობას იძლევა ისლანდიური ენის სასწავლო კორპუსი, რომელიც ისლანდიური ენის შემსწავლელთა შეცდომათა გამოვლენაზეა ორიენტირებული. კორპუსლინგვისტიკაში ამ მიზნობრიობის მქონე სასწავლო კორპუსები გამოსადეგია როგორც სასწავლო რესურსებისა და მასალის შედგენაში, ისე მართლწერისა და

გრამატიკის შემსწორებელი პროგრამების დასახვეწად. ისლანდიური ნეიროლინგვისტიკის მდიდარ ტრადიციას ავსებს სპეციალიზებული სახის სასწავლო კორპუსები, მაგალითად, ისლანდიური დისლექსიის შეცდომათა კორპუსი და ბავშვთა მეტყველების შეცდომათა კორპუსი. ორივე მათგანში ყველა ენობრივი ლაფსუსი მანუალურადაა ანოტირებული (არნარდოთირი და სხვ., 2022: 245).

უშუალოდ ლექსიკოგრაფიული მიზნებისათვის, სალექსიკონო მიკროსტრუქტურული ველების გასაუმჯობესებლად არსებული მნიშვნელოვანი რესურსია შეცდომების გამოვლენაზე ორიენტირებული ლონგმენის სასწავლო კორპუსი, რომელიც აერთიანებს ესეებსა და საგამოცდო წერილობით ნამუშევრებს მთელი მსოფლიოდან. სასწავლო კორპუსის მასალაზე დაყრდნობით თითოეული პრობლემური და ბუნდოვანი სიტყვა, რომელიც ძირითადად ენის შემსწავლელების შეცდომის წყაროა, ლექსიკონში აისახება.

თუკი შევაჯამებთ ზემოთ მოხმობილი დასავლეთევროპული და ამერიკული სასწავლო კორპუსების შექმნის მეთოდოლოგიას, დიზაინსა და კონცეფციას, ჩვენი კვლევისათვის რამდენიმე გამოსადეგი პრინციპის დანერგვა შეგვიძლია:

- **მონაცემთა შედარებითი ანალიზისათვის ლაიდენის სასწავლო კორპუსის მოდელის** გამოყენება ხელსაყრელია, ვინაიდან, ამ კორპუსში მშობლიურ ენაზე მოსაუბრეთა ტექსტობრივი მონაცემები შეპირისპირებულია უცხოური ენის შემსწავლელთა მასალას. ამ გადაწყვეტილებამ შესაძლებელი გახადა სხვადასხვა კონტექსტში არსებული სირთულეები და ენობრივი ინტერნალიზაციის სტრატეგიები ცხადად გამოამჟღავნებელიყო;
- **მიღებული შედეგების ლექსიკოგრაფიული დამუშავების თვალსაზრისით, ლონგმენის სასწავლო კორპუსის მოდელის** გამოყენება გონივრული იქნება იმ მხრივ, რომ ამ სასწავლო კორპუსში ასახული ენობრივი შეცდომები პრაქტიკულ ფუნქციას ჰპოვებენ სალექსიკონო სტატიის მიკროსტრუქტურულ ველებში. ამჟამად არსებულ ინგლისურ-ქართულ სასწავლო ლექსიკონების (ბეჭდური თუ ელექტრონულ მატარებელზე გადასული) მიკროსტრუქტურაში არ მოიძებნება ტექსტური გრაფები, ან სხვა სახის შიდატექსტობრივი მონაცემები, რომელიც ენის შემსწავლელს კომენტარის სახით სიტყვათხმარების კონკრეტული

კომბინაციების გამოყენების თავისებურებებს ან მოსალოდნელ შეცდომათა გამაფრთხილებელ ნიშანს ამცნობს.

1.2. შესიტყვებებისა და მრავალსიტყვიანი ლექსიკური ერთეულების კვლევა სასწავლო კორპუსში

სასწავლო კორპუსზე დაფუძნებული და ინფორმირებული კვლევებით ხორციელდება არა მხოლოდ შეცდომების გამოვლენაზე ორიენტირება, მათი აღბეჭდვა, კლასიფიკაცია და ანალიზი, არამედ რაოდენობრივი ანალიზის გზით კონკრეტული სიტყვების, ფრაზებისა თუ სტრუქტურების ჭარბად ან არასაკმარისად გამოყენების შესწავლაც.

სასწავლო კორპუსების კვლევის მრავალპლანიანობა განაპირობებს უცხოური ენის შემსწავლელთა ფორმულარული ერთეულების, მენტალური ლექსიკონის ერთეულებზე წვდომისა და კოგნიტური პროცესების სიღრმისეულად დამუშავებას. ვინაიდან, უცხოური ენის ლექსიკის შესწავლა კუმულაციური პროცესია და იწყება კონტექსტში მნიშვნელობის ცნობით, სრულდება მისი შესაფერისი რეალიზაციით, უაღრესად მნიშვნელოვანია კოლოკაციების, გავრცელებულ სიტყვათშეხამებათა, სიტყვათა კომბინაციების გამოყენება უცხოური ენის სრულყოფილად ასათვისებლად. ფრაზეოლოგიური ლექსიკის სირთულეს ენის შემსწავლელთა სწავლის პირველივე, დაწყებით საფეხურზე ეჩეხებიან და როგორც ფ. ჰაუსმანი აღნიშნავს: „*კომპლექსური ლექსემების დაუფლება მხოლოდ სიდიდით, მოთმინებითა და ფუტკრის მსგავსი განუწყვეტელი შრომით ხდება. განა არსებობს რაიმე უკეთესი, ვიდრე ყვავილიდან ყვავილზე ზუზუნით ფრენა, ნექტრის მოგროვება და შიგადაშიგ მათი ჩხვლეტა, ვინც შენს შეჩერებას ცდილობს?*“ (ჰაუსმანი, 1993: 17).

ინგლისურ ენაში სიტყვების ერთმანეთთან შეთავსებისა და გვერდიგვერდ თანაარსებობის კვლევის ლინგვისტური ინტერესი მეოცე საუკუნის დასაწყისიდან იღებს სათავეს, როდესაც ჯ. ფირსმა (1951), ო. იესპერსენმა (1924) და ჰ. პალმერმა (1981) შესიტყვებებსა და მყარ გამონათქვამებზე თეორიული ნაშრომები გამოაქვეყნეს.

მოგვიანებით, მეოცე საუკუნის მიწურულს, ფრაზეოლოგიზმების კვლევას ფორმულარული გამონათქვამების თვისებრივი ანალიზიც შეემატა. თუმცა, ამ პერიოდში არსებული ანალიტიკური ინსტრუმენტების სიმწირის გამო, ადრეული ფრაზეოლოგიური კვლევები ძირითადად ფოკუსირდებოდა მცირე რაოდენობის გამონათქვამების კლასებზე, იდიომებსა და მყარ შესიტყვებებზე და უგულებელყოფდა უფრო რთულ ენობრივ კონსტრუქციებს.

ცნება „შესიტყვება“ დღეისთვის მეტნაკლებად დიფერენცირებულად გამოიყენება ენის სხვადასხვა დონეზე. კოლოკაცია აღნიშნავს ლექსმათა თანაარსებობას და განპირობებულია რამდენიმე ფაქტორით. ესენია: სემანტიკური ასოციაცია, არსობრივი ურთიერთგანპირობებულობა დენოტატებს შორის, ლექსიკური სოლიდარობა, კონცეპტუალური სტერეოტიპები. კორპუსლინგვისტიკაში კოლოკაცია აღნიშნავს ყველაზე ხშირ, სტატისტიკურად დაანგარიშებად ლექსიკური ელემენტების (ტოკენების) თანაარსებობას. რაც უფრო მაღალია შესიტყვებების სიხშირე, მით მეტია იმის ალბათობა, რომ საქმე გვაქვს მყარ სინტაგმებთან (თანდაშვილი და ფურცხვანიძე, 2014: 54).

მიუხედავად ამ ცნების გაგების მრავალსახოვნებისა, შესიტყვების განსაზღვრებათა ჯაჭვში სულ მცირე ერთი საერთო მნიშვნელის პოვნა შეგვიძლია და რაზეც ძირითადად სამეცნიერო წრეები თანხმდებიან არის ის, რომ სიტყვებს შორის არსებული სინტაგმატური კავშირები შესიტყვებების მთავარი განმსაზღვრელი ფაქტორია. ამ ცნების უკეთესად გასაანალიზებლად საჭიროა მისი კვლევის ორი მთავარი თვალსაზრისის: სიხშირეზე (სტატისტიკაზე) დამყარებული მიდგომისა და ფრაზეოლოგიური (მნიშვნელობაზე) ორიენტირებული მიდგომის განსაზღვრა. ამ კონტექსტში ს. გრეინჯერი გვთავაზობს მოცემული შესაძლებლობიდან მაქსიმუმის გამოყენებას და ორივე მიდგომის ერთმანეთთან შეჯერებას (გრეინჯერი და სხვ., 2008: 45).

1.2.1. ფრაზეოლოგიზმების კვლევის სიხშირეზე დამყარებული მიდგომა

ერთ-ერთი შეხედულების მიხედვით, შესიტყვება გულისხმობს ენობრივ ერთეულთა უშუალო კონტექსტუალურ, გვერდიგვერდ თანაარსებობას. ამ კონცეპტის

კვლევა ჯ. სინკლერის აღმოჩენებმა განაპირობა და ფრაზეოლოგიის ახალი, ინდუქციური მიდგომით, **დისტრიბუციული** (ევერტი, 2005), სხვაგვარად, **სიხშირეზე დამყარებული მიდგომით** შესწავლის შესაძლებლობა გააჩინა (ნესელჰაუფი, 2004: 11).

ცვლილება, რაც კორპუსლინგვისტიკამ მოიტანა დაკავშირებულია ლექსიკისა და სინტაქსის ტრადიციულ დიქტომიის გადააზრებასთან. ჯ. სინკლერის ორმა რევოლუციურმა კონცეპტმა - იდიომური პრიციპისა და ლექსიკური გრამატიკის პრინციპმა, კორპუსზე ორიენტირებულ ლექსიკო-გრამატიკულ კვლევას გზა გაუკვალა. იდიომის პრინციპი უკავშირებდა ენის ფრაზეოლოგიურ მიდრეკილებას, რომლის დროსაც სიტყვები გვერდიგვერდ ჩნდება არა ერთმანეთისაგან იზოლირებულ ერთეულებად, არამედ ისინი ერთმანეთს ერწყმიან რაიმე მნიშვნელობის მისაღებად. აღნიშნული პრინციპი განსხვავდება ღია არჩევანის პრინციპისგან, რომლის დროსაც სიტყვები ინდივიდუალურად არის არჩეული წინადადებაში გარკვეული ადგილის შესავსებად. ენის მომხმარებელს აქვს მისთვის ხელმისაწვდომი ნახევრად, წინასწარკონსტრუირებული ფრაზების დიდი ოდენობა, რომლებიც ქმნიან თითო არჩევანს (სინკლერი, 1991: 109-115). ჯ. სინკლერმა დიდი მოცულობის კორპუსული მტკიცებულების საფუძველზე ლექსიკისა და გრამატიკის ურთიერთგადაჯაჭვულობა და განუყოფლობა დაინახა, რის გამოც მან ლექსიკური გრამატიკის ცნება შემოგვთავაზა, ანუ მნიშვნელობა და სტრუქტურა ავტორისეული ხედვით ერთ მთლიანობად განიხილება.

ფრაზეოლოგიური ერთეულების იდენტიფიკაციის დროს სიხშირეზე დაფუძნებული მეთოდი სცდება ტრადიციულ ჩარჩოებსა და მიდგომებს, რომლებიც აქცენტს ისეთ ლექსიკურ ფენომენებზე აკეთებენ, როგორიცაა იდიომები, ანდაზები და მსგავსი ფიქსირებული გამონათქვამები. დისტრიბუციული, ანუ სიხშირეზე დამყარებული მიდგომა არ ეფუძნება სემანტიკურ კრიტერიუმებს იმის დასადგენად, არის თუ არა ესა თუ ის ფრაზეოლოგიური ერთეული ნამდვილად ფრაზეოლოგიზმი. მაშასადამე, სინკლერისეული კონტექსტუალისტური ტრადიციის მიხედვით, ის შეისწავლის სიტყვას არა იზოლირებულად, არამედ კონტექსტთან თანაარსებულ ლექსიკურ კავშირს. ხშირად ლექსიკური წყვილები დამოუკიდებელ ელემენტებად აღიქმება, რომელთა შესწავლაც ენის სწავლების ძირითად ქვაკუთხედად ითვლება; მათ გარეშე ენობრივი ქმედება შეუძლებელია, თუმცა, თუ დავუკვირდებით, ვნახავთ,

რომ სიტყვები დამოუკიდებელ ერთეულებად კი არ გვხვდება ენაში, არამედ ყოველთვის ტექსტის ჩარჩოშია მოქცეული და ხშირ შემთხვევაში მათი მნიშვნელობის დადგენაც მხოლოდ კონტექსტით გამომდინარე შეიძლება.

დისტრიბუციული მიდგომა უკავშირდება ისეთ ცნებას, როგორიცაა, ერთი მხრივ, მ. ჰოუის ლექსიკური პრაიმინგი, როდესაც ერთ ლექსიკურ ერთეულს შეუძლია ფასილიტირება (პრაიმინგი), გაუწიოს, ანუ გაუადვილოს მომდევნო სიტყვის გადამუშავებისა და ამოცნობის პროცესი. სიტყვების ცოდნა განპირობებულია რა კონტექსტში გვხვდება ის, როგორ ვიყენებთ, რა ენობრივ სტრუქტურებსა და ტექსტურ ტიპებში ვხვდებით მათ (ჰოუი, 2005); მეორე მხრივ, დისტრიბუციული მიდგომა აღსანიშნავია ა. სტეფანოვიჩისა და ს. გრიესის კოლოსტრუქტურული ანალიზისას, რომლის უმთავრესი ამოცანაა სიტყვათა და გრამატიკულ სტრუქტურათა შორის არსებული ასოციაციების დადგენა, ან რა სახის განზიდულობა ახასიათებთ მათ და რამდენად ხშირად ან, პირიქით, ნაკლებად გამოიყენება კონკრეტული ენობრივი კონსტრუქციები (სტეფანოვიჩი და გრიესი, 2003: 210).

ჯ. სინკლერი შესიტყვებას განმარტავს, როგორც ტექსტში ერთმანეთთან ახლო მანძილზე მყოფ ორ ან ორზე მეტ ლექსიკურ ერთეულს. შესიტყვებების კვლევისათვის ტექსტკორპუსში გასათვალისწინებელია საძიებო, ანუ საკვანძო სიტყვასა და მის მოსაზღვრე სიტყვებს შორის არსებული მანძილის განსაზღვრა (სინკლერი, 1991: 170). მანძილი, ძირითადად, შეადგენს ოთხ სიტყვას საკვანძო სიტყვამდე და მის შემდეგ. მაგალითად, მოცემულ ტექსტში სიტყვა *red* გაანალიზებისას He drove a car. When he stopped at the red light, traffic grew heavier, რომელშიც სიტყვები the, at, stopped, he, light, traffic, grew, heavier მიიჩნევა *red*-ის კოლოკატებად.

იმ ავტორების მოსაზრებები, რომლებიც შესიტყვებების კვლევისათვის სიხშირეზე დამყარებულ მიდგომას იყენებენ, გვიჩვენებს აზრთა სხვადასხვაობას რამდენიმე საკითხთან მიმართებით. პირველ რიგში, რა რაოდენობისა და თანმიმდევრობის ლექსიკური ერთეულები ქმნიან კოლოკაციას და შესიტყვებებში შემავალ ელემენტებს შორის როგორი სინტაქსური მიმართებებია საჭირო. გ. კიელმერი აღნიშნავს, რომ მაგალითისათვის, წინადადებაში, როგორიცაა At night, he had left her free to follow her own inclinations, to be whatever she wanted to be, ლექსიკური წყვილი night he არ უნდა ჩაითვალოს შესიტყვებად, მიუხედავად იმისა, რომ სიხშირის

კრიტერიუმი ამ ელემენტებს კოლოკაციად აღიქვამს (კიელმერი, 2014). როგორც ვხედავთ, სიხშირეზე დამყარებული მიდგომის მიხედვით, ელემენტებს შორის არსებული სინტაქსური კავშირები ზოგჯერ არ ქმნიან შესიტყვებებს. ისინი კოოკურენტებად უნდა მოვიაზროთ, რადგან ისინი უბრალოდ ასახავენ ლექსემათა შორის არსებულ ლოგიკურ-კოჰერენტულ ურთიერთმიმართებას.

მხოლოდ დისტრიბუციული, ანუ სტატისტიკაზე დამყარებული მიდგომის გამოყენება კოლოკაციათა კვლევის უფრო ნათელ და ზუსტ შედეგებს ვერ მოგვცემს, ამიტომ არსებული დანაკლისი მნიშვნელობაზე ორიენტირებული მიდგომითაც უნდა შეივსოს.

1.2.2. ფრაზეოლოგიზმების კვლევის მნიშვნელობაზე ორიენტირებული მიდგომა

შესიტყვებების ცნების კიდევ ერთი განსაზღვრების მიხედვით, რომელიც რუსული ფრაზეოლოგიური სკოლის ძლიერი გავლენის შედეგია, კოლოკაციები სიტყვის კომბინაციის ტიპებადაა დანახული. ამგვარ შეხედულებას მნიშვნელობაზე ორიენტირებულ მიდგომას უწოდებენ (ჰერბსტი, 1996: 380). შესიტყვებების კვლევის ფრაზეოლოგიური მიდგომის მიხედვით, ა. ქოუი შესიტყვებების კვლევისას, სიტყვათა კომბინაციებს ორ კატეგორიად, კომპოზიტებად და ფორმულა-გამონათქვამად ჰყოფს. ეს უკანასკნელი, მაგალითისათვის, How are you? ან Good night პრაგმატული ფუნქციის მქონეა, ხოლო შესიტყვება კომპოზიტების ჯგუფს განეკუთვნება და ხასიათდება, როგორც სინტაქსური ფუნქციის მატარებელი (ქოუი, 1994: 3169).

ავტორი კომპოზიტების ჯგუფს გამჭვირვალობისა და შენაცვლებადობის კრიტერიუმების მიხედვით ჰყოფს. გამჭვირვალობა გულისხმობს, რომ კომბინაციაში შემავალ ელემენტებს (ლექსემებს) ერთმანეთთან პირდაპირი თუ გადატანითი სემანტიკური კავშირები აქვთ. ხოლო შენაცვლებადობა მიემართება იმას, თუ რა ხარისხით არის შესაძლებელი კომბინაციაში შემავალი ელემენტების ჩანაცვლება სხვა ენობრივი ერთეულით. ამის საფუძველზე, ავტორი განარჩევს ოთხი ტიპის სინთეზს და ხაზს უსვამს იმ ფაქტს, რომ მოცემულ ტიპებს საზღვრებში ვერ მოვაქცევთ და უმჯობესია აღქმულ იქნეს უწყვეტ კონტინუუმად:

ა) თავისუფალი კომბინაციები (მაგ., drink coffee): სადაც სიტყვათა კომბინაციის ყველა ელემენტი პირდაპირი მნიშვნელობით გამოიყენება.

ბ) შეზღუდული შესიტყვებები (მაგ., perform a task): ელემენტთა ჩანაცვლება შესაძლებელია, მაგრამ ჩანაცვლებისას პირობითი შეზღუდვები მაინც არსებობს; სულ მცირე ერთ ელემენტს აქვს პირდაპირი მნიშვნელობა, სულ მცირე ერთი ელემენტი გამოიყენება ირიბი მნიშვნელობით; მთლიანი კომბინაცია გამჭვირვალეა.

გ) ფიგურალური (გადატანითი) იდიომები (მაგ., do a U-turn): ელემენტების ჩანაცვლება იშვიათადაა შესაძლებელი; კომბინაციას აქვს გადატანითი მნიშვნელობა.

დ) სრული იდიომები (მაგ., blow the gaff): ელემენტების ჩანაცვლება შეუძლებელია და კომბინაცია გადატანითი მნიშვნელობისაა.

მსგავსი ტიპის კლასიფიკაციას ვხვდებით ლ. იუანგენის კომბინაციებშიც. ავტორი წარმოადგენს იმ მთავარ კრიტერიუმებს, რომელთაც კოლოკაციები აკმაყოფილებენ: ა) სადაც ერთი კომპონენტი თავისუფალია, ხოლო მეორეს ჩანაცვლება შესაძლებელია, მაგალითად, „collect information“, „collect stamps“; ბ) სადაც შესაძლებელია ორივე კომპონენტის ჩანაცვლება („draw attention“, „draw a conclusion“, „pay attention“, „reach a conclusion“); გ) სადაც ერთი კომპონენტის ჩანაცვლება შეგვიძლია და მეორე მათგანის – არა („biological parents“, „adoptive parent“); დ) სადაც ორივე კომპონენტი შეზღუდულია, ერთგვარ საზღვრებშია მოქცეული („electric currents“, „supreme court“). ორივე ავტორის შესიტყვებების კვლევის ფრაზეოლოგიური მიდგომის გათვალისწინების შემთხვევაში, შესიტყვებების შერჩევისას მკვლევარი თავს ვერ აარიდებს ადამიანურ განსჯას, თუ რა ჩათვალოს კოლოკაციად და რა უგულებელყოს (ლუ, 2017: 47).

სიხშირეზე დამყარებული მიდგომისგან განსხვავებით, ფრაზეოლოგიური მიდგომა მუდმივად საჭიროებს შესიტყვებების ელემენტებს შორის სინტაქსური ურთიერთკავშირის დადგენას. ფ. ჰაუსმანი უფრო შორსაც კი მიდის კოლოკაციების დეფინირებისას და მხოლოდ ისეთ შესიტყვებებს მიიჩნევს ვალიდურად, რომელიც მის მიერვე წინასწარ განსაზღვრულ სინტაქსური კავშირების მოდელებსა და კომბინაციებს ერგება (ჰაუსმანი, 1989: 1010).

სიხშირეზე დაფუძნებულ მიდგომასა და ფრაზეოლოგიურ მიდგომას შორის ერთ-ერთი საერთო ნიშან-თვისება არის ის, რომ კოლოკაციებში შემავალი ერთეულების რაოდენობა მერყეობს ორს ან ორზე მეტ ერთეულს შორის.

მას შემდეგ, რაც გავაანალიზეთ შესიტყვებების რაობა, მისი კვლევის მიდგომების საერთო და განმასხვავებელი თვისებები, შეგვიძლია დავასკვნათ, რომ შესიტყვებების კვლევის სტატისტიკაზე ორიენტირებული მიდგომა ეფუძნება სიტყვებს შორის არსებულ სინტაგმატურ კავშირებს, ანუ ტექსტკორპუსში საკვანძო სიტყვასა და მის მოსაზღვრე სიტყვებს შორის არსებულ ბმებს, რაც მკვლევარს სუბიექტური შეხედულებისა და მიკერძოებული განსჯისგან ათავისუფლებს. ამის საპირისპიროდ, შესიტყვებების კვლევის ფრაზეოლოგიური მიდგომის მიხედვით, კოლოკაცია მერყეობს თავისუფალი კომბინაციების კატეგორიიდან იდიომებამდე და, ვინაიდან, რთულია ზუსტი საზღვრების დაწესება, მათ შორის, შესიტყვებების განსაზღვრა მეტწილად დამოკიდებულია ადამიანურ განსჯაზე.

სასწავლო კორპუსებში შესიტყვებების შესწავლასთან არსებული შეზღუდვების გათვალისწინებით, საჭიროდ მივიჩნევთ რამდენიმე არსებითი მაღიმიტირებელი ფაქტორის დასაბალანსებლად და კვლევის ძირითადი მიზნის გათვალისწინებით, ჩვენ მიერ შექმნილ სასწავლო კორპუსში არსებულ შესიტყვებებს გამოვიკვლევთ როგორც სიხშირეზე დაფუძნებული მიდგომის გათვალისწინებით, ისე მნიშვნელობაზე ორიენტირებული მიდგომით.

1.2.3. შესიტყვებების სტრუქტურული ტაქსონომია

მას შემდეგ, რაც გავაანალიზეთ შესიტყვებების რაობა და ძირითადი მახასიათებლები, ჩვენი კვლევის ძირითადი მიზნის გათვალისწინებით, ჩვენ მიერ შექმნილ სასწავლო კორპუსში არსებულ შესიტყვებებს გამოვიკვლევთ სიხშირეზე დამყარებული მიდგომის გათვალისწინებით. ვინაიდან კორპუსში შესიტყვებების დადგენა ფრაზეოლოგიური მიდგომის გამოყენებით მანუალურად არის შესაძლებელი, რაც ცალსახად მეტ ადამიანურ და მატერიალურ რესურსს მოითხოვს,

ჩვენი საკვლევი პრობლემების გადაჭრისათვის ლოგიკურია იმ სტატისტიკური საზომების გამოყენება, რომლებიც სანდოობის მაღალი ხარისხით გამოირჩევა.

უპირველესად, შესიტყვებების დასადგენად ვიყენებთ სიხშირეზე დაფუძნებულ მიდგომას და ამისათვის მოვიშველიებთ რამდენიმე გზას შესიტყვებების შესწავლისათვის. ესენია: n-gram (დანაწევრებული ტექსტობრივი ფრაგმენტები: კლასტერები, „ლექსიკური დასტები“, კოლგრამები და ა.შ.), შესიტყვების ფანჯარა და შესიტყვების ქსელები. N-gram ამოიცნობს ტექსტკორპუსში სიტყვის მომიჯნავე კომბინაციებს. შესიტყვების ფანჯარა ეძებს შესიტყვებებს კონკრეტული ფანჯრის მონაკვეთში, მაგალითად, 5L, 5R გულისხმობს სამიეზო სიტყვიდან მარცხნივ განლაგებულ 5 სიტყვას და 5 სიტყვას მარჯვნივ. დაბოლოს, შესიტყვების ქსელები აერთიანებს ენასა და დისკურსში სიტყვებს შორის არსებულ ურთიერთკავშირებს (ბრეზინა და სხვ., 2015).

თუმცა, მეორე მხრივ, როგორც უკვე აღვნიშნეთ, სტატისტიკაზე ორიენტირებულ მიდგომას თავისი შეზღუდვებიც გააჩნია და საჭიროდ მივიჩნევთ რამდენიმე არსებითი მალიმიტირებელი ფაქტორის დასაბალანსებლად, ჩვენს კვლევაში გავითვალისწინოთ სტრუქტურული (ლექსიკო-გრამატიკული) პარადიგმებიც. სტატისტიკაზე ორიენტირებული მიდგომის გამოყენების დროს გაჩენილ სიცარიელეს შევავსებთ სასწავლო კორპუსში შესიტყვებების მოდელების მანუალური მოძიებით. შესიტყვებათა სტრუქტურული ძიებისათვის ვიყენებთ კოლოკაციურ მოდელებს, რომელშიც წარმოდგენილია ლექსიკური შესიტყვებების 9, ხოლო გრამატიკული შესიტყვებების 10 პარადიგმა (ბენსონი და სხვ., 1997: 9). იგი ეფუძნება BBI კომბინატორულ ლექსიკონს (The BBI Combinatory Dictionary of English, 1997).

ლექსიკური შესიტყვებები შედგება არსებითი სახელის, ზმნის, ზედსართავი სახელის და ზმნიზედისაგან, არსებითი სახელების წინ განსაზღვრული და განუსაზღვრელი არტიკლის ჩათვლით. გრამატიკული შესიტყვებები მოიცავს მთავარ სიტყვას, წინდებულს ან ისეთ გრამატიკულ სტრუქტურას, როგორიცაა ინფინიტივი ან რთული წინადადების ნაწილი (იხ. სქემა 4).

სქემა 4. ლექსიკური და გრამატიკული შესიტყვებები (The BBI Combinatory Dictionary of English, 1997).

მოდელები	
ლექსიკური კოლოკაციები	verb + noun
	verb + determiner + noun
	verb + modifier + noun
	noun + noun
	adjective + noun
	adverb + adjective
	verb + adverb
	adverb + verb
გრამატიკული კოლოკაციები	verb + . . . + preposition
	verb + adverb
	preposition + noun
	noun + preposition
	adjective + preposition
	noun + to-infinitive
	noun + that-Clause
	adjective + to-infinitive
	adjective + that-Clause

კორპუსლინგვისტიკის თეორიაში შესიტყვებების შესწავლის არასტატისტიკური გზის გამოყენების ერთ-ერთი მეთოდი აღნიშნული კოლოკაციური პარადიგმების კონკორდანსის გზით კვლევაა. ამ ხერხის ძირითადი პრინციპი ეყრდნობა არა ალგორითმს ან ავტომატურ სტატისტიკურ პროცედურას, არამედ

ლინგვისტის ინტუიციას, როდესაც იგი კონკორდანსში სიტყვა-ფორმების კონტექსტებს იკვლევს, ეძებს მაგალითებსა და შესიტყვებების მოდელებს.

1.2.4. მრავალსიტყვიანი ლექსიკური ერთეულების (ლექსიკური კონების) ფუნქციური ტაქსონომია

ენის ფუნქციებისა და ტექსტური ურთიერთმიმართებების კლასიფიკაციისას სისტემურ ხასიათს იღებს მრავალსიტყვიანი ლექსიკური ერთეულების მახასიათებლების კვლევა. გამართული ენობრივი ფუნქციონირებისათვის მნიშვნელოვანია ფრაზეოლოგიზმების (კოლოკაციების, მრავალსიტყვიანი ლექსიკური ერთეულების, ფიქსირებული ფრაზების, იდიომების და ა.შ.) თანმიმდევრული გამოყენება. მათი სრულფასოვანი ხმარება ენობრივი რეალიზაციისას განაპირობებს ენის შემსწავლელთა ცოდნის დონეს, ხოლო ლექსიკური კლასტერების გამოყენების დეფიციტი მეტყველებასა თუ წერილობით დისკურსში მიუთითებს ენობრივ „გამოუცდელობაზე“. ამიტომ, სასწავლო კორპუსის ანალიზისას განსაკუთრებულ ყურადღებას ვუთმობთ ფრაზეოლოგიზმის კვლევას.

მრავალსიტყვიანი ლექსიკური ერთეულების (ლექსიკური კონების) ფუნქციური ანალიზის ჩარჩო დ. ბიბერმა (ბიბერი და სხვ. 2004) შემოგვთავაზა. აღნიშნული ტაქსონომიის წყარო ზეპირმეტყველებისა და წერილობითი სახის კორპუსული მონაცემები გახდა, რომელიც მოიცავდა ყოველდღიურ, საყოფაცხოვრებო დიალოგებს, სახელმძღვანელო და ინსტიტუციონალურ ტექსტებს. შესაბამისად, ლექსიკური კლასტერების ფუნქციური ტაქსონომია შეზღუდულია და, ძირითადად, პერსონალური ტიპის ჟანრს მოიცავს. მეორე მხრივ, ჯ. ო'ფლინის (2022: 81) კლასიფიკაციაში, რომელიც დ. ბიბერის (2006) სისტემას ეფუძნება, მეტადისკურსი სამი კატეგორიისაგან შედგება (იხ. სქემა 5).

3-4 კომპონენტიანი ლექსიკური კონების ფუნქციური კატეგორია	ფუნქცია
კვლევაზე ორიენტირებული	ეხმარება მწერალს, წარმოადგინოს ან განიხილოს ესა თუ ის საკითხი.
ადგილმდებარეობა	მიემართება დროსა და ადგილს.
პროცედურა	მიემართება როგორ და რატომ ხდება ესა თუ ის.
რაოდენობა	მიემართება რისამე რაოდენობას ან ხარისხს.
აღწერა	მიემართება რისამე ფიზიკურ ან აბსტრაქტულ აღწერას.
ტექსტზე ორიენტირებული	ეხმარება მწერალს, მოახდინოს ტექსტის ორგანიზება.
სტრუქტურირებისთვის დამხმარე კავშირები	მარკერები, რომლებიც თხრობის ორგანიზებას ახდენენ.
მიზეზ-შედეგობრივი კავშირები	აღნიშნავენ ელემენტებს შორის არსებულ დასკვნით ან შედეგის გამომწვევ მიმართებებს.
ამხსნელ-დამაზუსტებელი კავშირები	ამტკიცებს არგუმენტებს შემზღუდავი პირობების დაზუსტებით.
მონაწილეზე ორიენტირებული	ეხმარება მწერალს, გამოხატოს თავისი დამოკიდებულება.
მოსაზრების გამომხატველი კავშირები	გადმოსცემს მწერლის დამოკიდებულებებსა და შეფასებებს.
ჩართულობის გამომხატველი კავშირები	პირდაპირ მკითხველს მიემართება.

აღნიშნული კლასიფიკაცია ოთხი დისციპლინის (ბიოლოგიის, განათლების მეცნიერებების, ბიზნესისა და ინჟინერიის) აკადემიური ტექსტების დისკურის ანალიზის შედეგია. თითოეული დარგის ტექსტში მრავალსიტყვიანი ლექსიკური ერთეულების სხვადასხვაგვარი ფუნქციური ხმარების ტენდენცია იკვეთება.

მრავალკომპონენტიან დისკურსულ მარკერებს შორის საზღვრები იმდენად არამდგრადია, რომ მათ მკაცრად განკუთვნილ ფუნქციურ ჩარჩოში ვერ მოვაქცევთ. ფუნქციონალურად და სემანტიკურად ისინი ახლოს დგანან პირველი დონის დისკურსის მარკერებთან. დ. სიპმანი სამ- ან ოთხკომპონენტიან ლექსიკურ ფრაზებს

(მისი ინტერპრეტაციით, „შორ მანძილზე მყოფ კოლოკაციებს“ (long-distance collocations) სხვადასხვა ფუნქციურ კატეგორიაში მოაქცევს და თემატურ-ფუნქციურ მწკრივებსაც გამოყოფს (სიპმანი, 2004: 82). ეს დისკურსის მარკერებია: ა) კონტრასტული: it is one thing...It is another... Compare this...with/to, analogously, the same can be said for; ბ) აღიარებითი: it would be a mistake (to +inf.), also worth remembering that..., it could be argued; გ) მაგალითით ილუსტრირების: for the sake of illustration, as an example; დ) ამხსნელი: this is because, let us now see why; ე) განმმარტებელი: an X is a Y such that...; ვ) მთვლელი: beyond this, it should be added that, one (point/reason/ factor) is that; ზ) შემაჯამებელი: to round off this picture, in conclusion, I would say that; თ) მიზეზ-შედეგობრივი: number of factors account for this, this is because; ი) მიზნობრივი: I will now briefly describe, my purpose is, კ) (საკითხის) მაინიცირებელი: it is often said that, turning to; ლ) თემიდან გადახვევის (დიგრესიული): it should be mentioned in passing that; მ) კითხვა-პასუხის: the next obvious question is, the answer may lie in... ნ) აზრის აქცენტირებისთვის: it must be noted that, on inspection (my emphases); ო) ინფორმაციის მიმწოდებელი: a first point is that, it should be recognized that; პ) დამაზუსტებელი: the (key) point is that, but that is not the point; ჟ) მიმანიშნებელი: it is widely accepted that, it will be readily seen that; რ) ჰიპოთეზის: let us imagine that, I hypothesized that, assume that; ს) შემზღუდავი: to further confound the picture, to make matters worse, a further problem is that; ტ) ვისამე მიმაკუთვნებელი: [name] argues, as [name] writes, it has been proposed that; უ) ხელახლა ფორმულირების: put another way, in other words, as discussed, to recapitulate.

ვინაიდან, ჩვენ მიერ შექმნილი სასწავლო კორპუსი ჟანრული სხვადასხვაგვარობით ხასიათდება და თითოეულ ტექსტში მრავალსიტყვიანი ლექსიკური ერთეულების განსხვავებული ფუნქციური ხმარების ტენდენცია იკვეთება, ჯ. ო'ფლინის ტაქსონომიის დანერგვა მიზანშეწონილად გვეჩვენება. მეორე მხრივ, ლექსიკოგრაფიული სამუშაოს წარმოებისას სპეციალიზებული (EAP) ლექსიკონის მაკროსტრუქტურის შერჩევისას დ. სიპმანის სამ- ან ოთხკომპონენტური ლექსიკური ფრაზების ფუნქციურ ტაქსონომიას გამოვიყენებთ.

1.3. ლექსიკური სიმდიდრის შეფასება კორპუსლინგვისტიკაში

ლექსიკური სიმდიდრე განისაზღვრება ენის შემსწავლელთა ლექსიკური ფონდის მოცულობითა და განვითარების დონით. კორპუსლინგვისტიკის მეთოდოლოგიაში ტექსტობრივი მონაცემების სიმდიდრის შეფასების რამდენიმე გზა არსებობს, რომელთა დახმარებით ენობრივი სიმდიდრის განმსაზღვრელი სიდიდე დგინდება. იგი მიიჩნევა უცხოური ენის შემსწავლელთა წერილობითი ოსტატობისა და ენობრივი კომპეტენციის შეფასების ერთ-ერთ ყველაზე ეფექტურ საშუალებად. ლექსიკური სიმდიდრე, როგორც მრავალგანზომილებიანი კონსტრუქტი, შედგება ლექსიკური სიმჭიდროვის (lexical density), ლექსიკური მრავალფეროვნებისა (lexical diversity) და ლექსიკური სირთულისაგან (lexical sophistication). ამ სისტემის სარგებლიანობა ენის შემსწავლელთა საერთო ლინგვისტური პროგრესის შეფასებაშიც მდგომარეობს და კვლევის პროცესში სამივე საზომი ტოლფასოვანი და თანაბარღირებულია (ქანგი და სხვ., 2021).

ლექსიკური სიმჭიდროვე განსაზღვრავს ტექსტში ლექსიკური სიტყვების პროპორციულ რაოდენობას. დაანგარიშებისას აღნიშნული საზომი არ ითვალისწინებს გრამატიკულ სიტყვებს და ამის შედეგად, ინფორმაციული თვალსაზრისით ლექსიკური სიმჭიდროვე იმის ინდიკატორია, თუ რამდენად კომპლექსურია ტექსტი. დაბალი ლექსიკური სიმჭიდროვის სიდიდის შემთხვევაში, ტექსტი აღსაქმელად და გასაგებად უფრო მარტივია, ხოლო მაღალი ლექსიკური სიმჭიდროვის ხარისხი მიანიშნებს დისკურსის უფრო რთულ, სპეციალიზებულ ბუნებაზე.

ლექსიკური სიმჭიდროვის კონცეპტზე მსჯელობისას, განვასხვავოთ ლექსიკური თვისებების მქონე სიტყვები და ისეთი ერთეულები, რომელთაც არ გააჩნიათ ლექსიკური მახასიათებლები. ეს უკანასკნელნი გრამატიკულ-სინტაქსური ფუნქციის მატარებლები არიან. მოცემულ სიდიდეზე გავლენას უფრო მეტად ახდენს არსებითი სახელების, ზმნების, ზედსართავი სახელებისა და ზმნიზედების რაოდენობა, ვიდრე პირის ნაცვალსახელების, წინდებულების და კავშირებისა. გასათვალისწინებელია ენობრივი ფენომენი hapax legomenon, ისეთი სიტყვები, რომლებიც მხოლოდ ერთხელ გვხვდება ტექსტკორპუსში, რომელთაც ცალსახად შეუძლიათ ლექსიკური სიმჭიდროვის სიდიდის გაზრდა (იოჰანსონი, 2008: 61).

სამეცნიერო ლიტერატურაში აღნიშნული სიდიდის კვლევის რამდენიმე მოდელი არის შემოთავაზებული. ყველაზე პოპულარული მეთოდი არსებითი სახელების სიმჭიდროვის რაოდენობრივი ანალიზია, რომლის დროსაც, ტექსტკორპუსში არსებული არსებითი სახელების რაოდენობა შეეფარდება ტოკენების მთლიან ოდენობას. კიდევ ერთი ვარიანტი ზმნის, ან ზედსართავი სახელის, ან ზმნიზედის რაოდენობების შეფარდებაა კორპუსის მთლიან ლექსიკურ ფონდთან (კობი და ჰორსტი, 2015: 188-189).

ლექსიკური სიმჭიდროვის ცნების დახვეწა მ. ჰალიდის სახელთანაცაა დაკავშირებული. თავის ნაშრომში „სამეტყველო და წერილობითი ენა“ (ჰალიდი, 1985) მსჯელობს ლექსიკურ და გრამატიკულ ერთეულთა შორის არსებულ განსხვავებებსა და კავშირებზე. საკითხის სიღრმისეულად განხილვისას მაგალითად მოჰყავს ერთზე მეტ სიტყვიანი ენობრივი ერთეული turns up, რომელსაც ის დანაწევრებულად კი არ მიიჩნევს, არამედ აღიქვამს, როგორც ერთ მთლიანობად. ავტორის მიხედვით, ბავშვის ენობრივი განვითარების საწყისი ეტაპები ცხადყოფს ამ ორი კლასის (ლექსიკური და გრამატიკული ერთეულების) არსებობას. ბავშვები ისე აგებენ წინადადებებს, რომ უგულებელყოფენ გრამატიკულ ერთეულებს. ჰალიდის აზრით, ლექსიკიდან გრამატიკამდე კონტინუუმი არსებობს და, აგრეთვე, ზმნიზედების ზოგიერთი კლასი ლექსიკურ და გრამატიკულ ერთეულთა გზაგასაყარზეა (ჰალიდი, 1985).

ჩვენს კვლევაში ლექსიკური სიმჭიდროვის დადგენისას უპირატესობას ვანიჭებთ მოცემული კონცეპტის ჰალიდისეულ გააზრებას. ჩვენს კორპუსულ მონაცემებში ლექსიკური სიმჭიდროვე დაანგარიშებული იქნება ლექსიკური ერთეულების შეფარდებით ტექსტკორპუსის სიტყვათა მთლიან რაოდენობასთან.

ლექსიკური მრავალფეროვნება სამეცნიერო ლიტერატურაში ხშირად გამოიყენება ლექსიკური სიმდიდრის ეკვივალენტად, თუმცა ხშირად იგი განიხილება როგორც ლექსიკური სიმდიდრის მრავალგანზომილებიანი ბუნების მხოლოდ ერთი ნაწილი (მალვერნი და სხვ., 2004). ჩვენს კვლევაში ვიზიარებთ მალვერნის პოზიციას და მოცემულ საზომებს სასწავლო კორპუსში შემავალ ტექსტების ხარისხობრივი ანალიზისათვის გამოვიყენებთ, ლექსიკური მრავალფეროვნების დასადგენად, ტექსტკორპუსში სიტყვათა ტიპებს ვითვლით. რაც უფრო მეტად მრავალგვარ და

მრავალფეროვან ლექსიკას ვხვდებით ტექსტში, მით მეტია ლექსიკური მრავალფეროვნება. იმისათვის, რომ ტექსტი ლექსიკურად სხვადასხვაგვაროვანი იყოს, უცხოური ენის შემსწავლელმა ბევრი განსხვავებული სიტყვა, მცირედი განმეორებითობის ხარისხით უნდა გამოიყენოს.

ლექსიკური მრავალფეროვნების ტრადიციული საზომი სხვადასხვა სიტყვების (ტიპების) შეფარდებაა სიტყვათა (ტოკენების) საერთო რაოდენობასთან, ანუ ტიპ-ტოკენის მიმართებით (Type/Token Ratio), სხვა სიტყვებით რომ ვთქვათ, კორპუსული მონაცემების ტიპებისა და ტოკენების რაოდენობების ურთიერთშესაბამისობის ინდექსით განისაზღვრება. რაც უფრო ახლოსაა ინდექსის სიდიდე 1-თან, მით უფრო მრავალფეროვანია სასწავლო კორპუსის ბაზა ფორმობრივი თვალსაზრისით (კობი და ჰორსტი, 2015: 192).

ს. გრეინჯერი (2000) ლექსიკური სიმდიდრის ამსახველი საზომი ერთეულების კვლევაში ამტკიცებს არსებული სიდიდის ოპტიმიზაციის მნიშვნელობას. ტიპ-ტოკენის შესაბამისობების სიდიდის გამოყენებისას გასათვალისწინებელია ერთი მნიშვნელოვანი დეტალი. როდესაც სასწავლო კორპუსი შეიცავს დიდი რაოდენობით მართლწერის შეცდომებს, ამ ფაქტორმა შესაძლოა გავლენა მოახდინოს TTR სიდიდის მართებულობაზე (გრეინჯერი და ვინი, 2000: 253). მაგალითად, სიტყვა *exactly*, ტექსტში მოხვედრისას ჩაითვლება როგორც ერთი ტოკენი, მაგრამ თუკი კორპუსის მონაცემებში გვაქვს ისეთი მართლწერის შეცდომები, როგორიცაა *exectly* ან *exactlly*, ავტომატური კალკულაციით ისინიც ჩაითვლება განსხვავებულ ტოკენებად.

აქედან გამომდინარე, საჭიროდ მიგვაჩნია ლექსიკური სიმდიდრის ინდექსის დადგენისას, გათვალისწინებულ იქნეს ლექსიკური მარაგის შეფასების კიდევ ერთი საზომი *Lemma/Token Ratio*, ანუ სიტყვათფორმების საწყისის, ლექსიკური ერთეულების ძირითადი ფორმის მიხედვით სიდიდის განსაზღვრა. როგორც ფ. მაკკარტი აღნიშნავს, პედაგოგიური და ლექსიკოგრაფიული მიზნებისათვის სასწავლო კორპუსის ანალიზისას სიტყვების ფლექსიური ფორმები პროგრამულად ერთ ტიპად უნდა იქნეს აღქმული (მაკკარტი, 1990: 73). ენის შემსწავლელს, რომელიც იყენებს ზმნის *begin* ხუთ განსხვავებულ ფორმას (*begin, begins, began, begun, beginning*) ერთსა და იმავე ტექსტში, უფრო ნაკლებ მრავალფეროვანი ლექსიკური მარაგი აქვს,

ვიდრე მას, ვინც ხმარობს ხუთ განსხვავებულ ლემას, (როგორცაა commence, start, launch, lead, initiate).

ლექსიკური სიმდიდრის განსაზღვრისათვის ვიკვლევთ კიდევ ერთ სტატისტიკურ საზომს – ლექსიკურ სირთულეს, რომელსაც სხვაგვარად უწოდებენ ლექსიკურ იმჟიათობასაც. ტექსტკორპუსში არსებული იმჟიათი ან რთული სიტყვების სტატისტიკური ანალიზის შედეგად ვიღებთ მოცემულ სიდიდეს. ენის შემსწავლელთა ნაწერებში ლექსიკური სირთულის დასადგენად და კონკრეტული სიტყვების იდენტიფიცირებისა თუ ლექსიკური სიხშირის აღწერის მიზნით გამოიყენება VocabProfile პროგრამა (კობი, 2003), რომელიც სასწავლო კორპუსის ტექსტს ამუშავებს სიხშირული პრინციპით რამდენიმე კატეგორიის მიხედვით. პლატფორმა ეფუძნება ინგლისური ენის ყველაზე ხშირად გამოყენებულ 2000 სიტყვას, ბრიტანული ეროვნული კორპუსის „სიტყვათა ოჯახების“ მრავალფეროვან სიასა და ინგლისური ენის აკადემიური ტექსტების ლექსიკას.

ლექსიკური მრავალფეროვნების კვლევისათვის შესაძლებელია, განხორციელდეს ორი სასწავლო კორპუსის კონტრასტული ანალიზი ორ სხვადასხვა შემთხვევაში: ა) როდესაც ორივე სასწავლო კორპუსი მოიცავს სხვადასხვა მშობლიურ ენაზე მოსაუბრეთა (მაგალითად, ფრანგულად, გერმანულად, არაბულად და ა.შ.); წერილობით ან ზეპირსიტყვიერ პროდუქციას, ხოლო სამიზნე ენა (მაგალითად, ინგლისური) ერთი და იგივეა. ასეთი ტიპის კვლევა მიზნად ისახავს უცხოური ენის შემსწავლელთა ენობრივი მახასიათებლების, ლინგვისტური და მეტალინგვისტური ბუნების დადგენას. კორპუსების ერთმანეთთან შეფარდების გზით ფასდება მეტამონაცემებიც, მაგალითად, ენის შემსწავლელთა ასაკი, მათი გაწაფულობის დონე, დავალების ტიპი და სასწავლო გარემო; ბ) როდესაც ერთი სასწავლო კორპუსი მოიცავს მშობლიური ენის წერილობით ან ზეპირსიტყვიერ მასალას, ხოლო მეორე კორპუსი – იმავე ენის შემსწავლელთა ენობრივ მასალას.

სასწავლო კორპუსების შედარებითმა ანალიზმა და ზემოხსენებული პროგრამების გამოყენებამ არსებულ სასწავლო კორპუსებში აჩვენა, რომ უცხოური ენის შემსწავლელები კონკრეტულ ლექსიკურ ერთეულებს, როგორც ჰ. ჰასელგრენი „ლექსიკურ სათამაშო დათუნიებს“ მსგავს ძალზე ნაცნობ, ფართო გამოყენებისთვის განკუთვნილ სიტყვებს უწოდებს, სისტემატურად იყენებენ ესეებსა თუ სხვა ტიპის

ნაწერებში. კვლევამ განსაზღვრა რამდენიმე ლექსიკური ერთეულის ჭარბად ან შედარებით ნაკლებად გამოყენების ტენდენცია ნაწერებში. მაგალითისათვის, ენის შემსწავლელები იშვიათად ხმარობენ არსებით სახელებს (issue, debate, suggestion), მაკავშირებელ სიტყვებს (however და therefore), რომლებსაც, მეორე მხრივ, მშობლიურ ენაზე მოლაპარაკეები არგუმენტების ორგანიზებისას ხშირად იყენებენ (ჰასელგრენი, 1994: 237).

თუკი შევაჯამებთ ლექსიკური სიმდიდრის განმარტებებს და მისი დადგენის გზებს კორპუსლინგვისტიკაში, გამოვყოფთ ლექსიკური სიმდიდრისათვის დამახასიათებელ ნიშანდობლივ თვისებებს:

- სასწავლო კორპუსების კვლევის ერთ-ერთი ასპექტი გულისხმობს უცხოური ენის შემსწავლელთა წერილობითი ოსტატობისა და ენობრივი კომპეტენციის შეფასებას. ამ მიზნის მისაღწევად კორპუსლინგვისტიკის მეთოდოლოგიაში ტექსტობრივი მონაცემების სიმდიდრის შეფასების რამდენიმე გზა არსებობს, რომელთა დახმარებით ენობრივი სიმდიდრის განმსაზღვრელი სიდიდე დგინდება.
- ჰ. ჟანგის და სხვ. (2021) მიერ განსაზღვრულია ლექსიკური სიმდიდრე, როგორც მრავალგანზომილებიანი კონსტრუქტი და შედგება ლექსიკური სიმჭიდროვის, ლექსიკური მრავალფეროვნებისა და ლექსიკური სირთულისაგან. ამ სისტემის სარგებლიანობა ენის შემსწავლელთა საერთო ლინგვისტური პროგრესის შეფასებაშიც მდგომარეობს და კვლევის პროცესში სამივე საზომი ტოლფასოვანი და თანაბარდირებულია.
- ჩვენს კვლევაში ლექსიკური სიმდიდრის ობიექტური და სანდო სიდიდის მისაღებად ოპტიმალურია ლექსიკური ვარიაციის საზომების (type/token ratio და lemma/token ratio) კომბინირებული გამოყენებით.

1.4. სასწავლო კორპუსის მასალის ლექსიკოგრაფიული გამოყენების შესაძლებლობები

მშობლიური ენის ამსახველმა ერთენოვანმა კორპუსებმა, თამამად შეიძლება ითქვას, რომ რევოლუცია მოახდინა ლექსიკოგრაფიაში, განსაკუთრებით, ფრაზეოლოგიის ლექსიკოგრაფიულად დამუშავების პროცესში, რომლის ფარგლებშიც, სულ უფრო დიდი და დივერსიფიცირებული ერთეულები მოგვევლინა ლექსიკონების სიტყვა-სტატიებად. მოპოვებულმა რაოდენობრივმა და თვისობრივმა ინფორმაციამ ლექსიკოგრაფებს საშუალება მისცა, უკეთესად დაენახათ განსამარტავი სიტყვებისა თუ ფრაზეოლოგიზმების პოლისემიური მნიშვნელობები და ავთენტურ კონტექსტში რეალიზებული საილუსტრაციო მასალით შეევსოთ სიტყვა-სტატიები.

1.4.1 ლექსიკოგრაფიული ტექსტური გრაფები

სასწავლო კორპუსის კვლევის შედეგების მსგავსი მასშტაბის გავლენა ლექსიკოგრაფიაზე ნაკლებ სარწმუნოა, მაგრამ უდაოდ პერსპექტიულია. სასწავლო კორპუსებიდან მიღებული ლექსიკოგრაფიული სარგებელი უპირველესად პრევენციული ხასიათისაა. ასეთი მასალა ლექსიკონში აისახება იმ მოტივით, რომ ენის შემსწავლელებს, რომლებიც ხშირად უშვებენ ენობრივ შეცდომებს, თავიდან აარიდოს სიტყვების, შესიტყვებების, ფრაზების არასწორ კონტექსტში გამოყენების მიდრეკილება. ლექსიკოგრაფებს ხშირად ლინგვისტური ალლო კარნახობს იმასთან დაკავშირებით, თუ რომელ ასპექტში აწყდებიან ენის შემსწავლელები სირთულეებს, მაგრამ საჭიროა, ეს ინტუიცია დაეფუძნოს სასწავლო კორპუსების ემპირიულ მასალას.

აღნიშნული პრაქტიკული იდეა თავდაპირველად ს. მეინგეისა და მ. რანდელის მიერ იქნა განხორციელებული ლონგმენის სასწავლო ლექსიკონმა თანამედროვე ინგლისური ენის ლექსიკონის მესამე გამოცემაში (1995), რომელიც „ლონგმენის სასწავლო კორპუსის“ მონაცემებს ეფუძნებოდა: გაჩნდა ჩანაწერები, “Get it right” ტექსტობრივი გრაფები, რომლებიც ენის შემსწავლელებს გავრცელებულ ენობრივ შეცდომებზე მიაწვდიდა ინფორმაციას (რანდელი და გრეინჯერი, 2007). რომელშიც დოკუმენტირებულია ინგლისური ენის შემსწავლელთა წერილობითი მასალები მათი

ენობრივი საჭიროებებისა და სიძნელეების დასადგენად. მიუხედავად იმისა, რომ მოცემული კორპუსი დაუმუშავებელი და არაანოტირებული იყო, ლონგმენის სასწავლო კორპუსმა სალექსიკონო დეფინიციების დახვეწაში დიდი როლი ითამაშა. კორპუსული მეთოდოლოგიით შეიქმნა შემსწავლელთათვის განსაკუთრებული სირთულის მქონე სიტყვათა სია და თავი მოეყარა ე. წ. „დამხმარე ყუთებში“.

მოგვიანებით მსგავსი მიდგომა სხვა სასწავლო ტიპის ერთენოვანმა ლექსიკონებმაც დანერგეს. მაგალითად, კემბრიჯის სასწავლო ლექსიკონის პირველი გამოცემა დაეფუძნა ნაწილობრივ-ანოტირებულ კემბრიჯის სასწავლო კორპუსს. ლექსიკოგრაფებმა შეძლეს შეცდომების ანოტაციის (error-tagging) საფუძველზე შეერჩიათ შეცდომათა ის კატეგორიები, რომელთა ლექსიკონში ასახვა ენის შემსწავლელთათვის ხელშემწყობი იქნებოდა. დღეს მომხმარებელს შეუძლია კემბრიჯის ონლაინლექსიკონის გვერდზე სპეციალურად განკუთვნილი ველის ქვეშ იპოვოს (“easily-confused words”) სიტყვები, რომლებიც ყველაზე ხშირად იწვევს მოსწავლეთა დაბნევას (დე კოკი და გრეინჯერი, 2005).

ლექსიკოგრაფიის თეორია (მეტალექსიკოგრაფია), რომელიც ორიენტირებულია ლექსიკოგრაფიის ისტორიისა და ზოგადი თეორიის კვლევაზე, ლექსიკონების კრიტიკასა და გამოყენების ასპექტებსაც შეისწავლის. თანამედროვე მეტალექსიკოგრაფიის ერთ-ერთი ყველაზე მნიშვნელოვანი პარადიგმული ცვლილება მდგომარეობს იმის გააზრებაში, რომ აუცილებელი არაა, ლექსიკონებში ყველა სიტყვა-სტატიას ერთი და იგივე სტრუქტურა ჰქონდეს. სავალდებულო და მუდმივად გამოყენებული სტატიის ერთგვაროვანი სტრუქტურის ნაცვლად, ლექსიკოგრაფს უფლება აქვს აირჩიოს ნაკლებად ხისტი ჰეტეროგენული სტატიის სტრუქტურა. სხვადასხვა ტიპის მიკროსტრუქტურები ლექსიკოგრაფებს აძლევს შესაძლებლობას, წარმოადგინონ ლექსიკური ელემენტები ისე, რომ საუკეთესოდ მოერგონ მათი სამიზნე მომხმარებლების საჭიროებებსა და მითითების უნარებს. მომხმარებლის პერსპექტივა თამაშობს დომინანტურ როლს ლექსიკონების დაგეგმვისა და საბოლოო შედგენის პროცესში. ეს ასევე ეხება მონაცემთა განაწილებას და მონაცემთა პრეზენტაციას ნებისმიერი ტიპის ლექსიკონში.

ლექსიკონების შეფასების დროს სიტყვა-სტატიების მიკროსტრუქტურის ერთ-ერთ მნიშვნელოვან ელემენტს – ტექსტურ გრაფას ექცევა განსაკუთრებული

ყურადღება. მონაცემთა გრაფები ლექსიკონის ფუნქციური ჩანაწერებია, რადგან ისინი ხელს უწყობენ ლექსიკონის ფუნდამენტური, უპირველესი ფუნქციის შესრულებას. ამ დანიშნულების განხორციელებისთვის ლექსიკოგრაფებმა ცხადად უნდა განარჩიონ ერთმანეთისაგან ტექსტური გრაფებისა და სხვა ლექსიკოგრაფიული ინსტრუმენტების, მაგალითად, კვალიფიკაციების გამოყენება. მიუხედავად იმისა, რომ ამ ყუთების ღირებულება შეიძლება არასდროს შეილახოს მისი ჭარბი და არასათანადო გამოყენებით სალექსიკონო სტატიის მიკროსტრუქტურაში, მაინც მნიშვნელოვანია, ლექსიკოგრაფმა ადეკვატურად შეაფასოს ტექსტური გრაფის გამოყენების საჭიროება.

გრაფები რიგ შემთხვევაში ამა თუ იმ სიტყვის მნიშვნელობის გაანალიზებისათვის მეგზურად გვევლინებიან, ზოგჯერ რომელიმე სიტყვის გამოყენების დიაპაზონის შეზღუდვებზე, წარმოთქმასა და რეგისტრზე გვაწვდიან ცნობებს. ლექსიკონის სიტყვის სიაში წარმოდგენილი მონაცემების გარდა, ლექსიკოგრაფს შეუძლია გამოიყენოს მონაცემთა განაწილების ის ხერხი, რომელიც გარკვეულ მონაცემებს განათავსებს ლექსიკონის წინა და/ან უკანა განყოფილებებში. როდესაც კონკრეტული ლექსიკოგრაფიული ჩანაწერის წარმოდგენა სტატიის გასაუმჯობესებლად აუცილებელია და ლექსიკონის მაკროსტრუქტურა არ გვთავაზობს შესაფერის პოზიციას ასეთი მონაცემების ჩანაწერებისთვის, ლექსიკოგრაფს შეუძლია მონაცემთა ყუთების გამოყენება – ან სტატიის შიდა- ან გარეთექსტების სახით. თუმცა, აქაც ლექსიკოგრაფიული ინფორმაციის მოწოდებისას საზღვრის დაწესებაა საჭირო, რ. გოუვსისა და დ. პრინსლოს (2010) აზრით, ლექსიკონების სტრუქტურული გადატვირთვა მეორე უკიდურესობაა. მიკროსტრუქტურულ ველში დეფინიციები უნდა შეიცავდეს მონაცემებს ზომიერად, გადაუჭარბებლად, რათა არ გამოიწვიოს სამიზნე აუდიტორიის (მომხმარებლის) დაბნევა და ლექსიკოგრაფიული საჭიროებების უგულებელყოფა (გოუვსი და სხვ., 2010: 501).

ლექსიკოგრაფიული მონაცემების გრაფების გამოყენების სიხშირემ ბეჭდურ და ელექტრონულ ლექსიკონებში, განსაკუთრებით სასწავლო ტიპის ლექსიკონებში ბოლო ათწლეულების განმავლობაში საგრძნობლად მოიმატა. ეს მოვლენა იმ ფაქტორით აიხსნება, რომ სასწავლო ლექსიკოგრაფიის თეორია მუდმივად იხვეწება,

ხოლო კონკურენტულმა ბაზარმა და გამომცემლობებს შორის პირველობისათვის ბრძოლამ ასევე იქონია გავლენა ლექსიკონების სტრუქტურისა და შინაარსის ცვლილებაზე. ამ ცვლილებებმა ლექსიკონებს თავისი პირვანდელი ღირებულება დაუბრუნა (გოუვსი და სხვ., 2010: 238). მაგალითად, ოქსფორდის სასკოლო აფრიკაანს-ინგლისური/ინგლისურ-აფრიკაანსი ლექსიკონი იყენებს ე.წ. შეფერილ გრაფებს, ანუ სიტყვათხმარების გრაფებს, მომხმარებლების დასაკვალიანებლად „შესაძლო სირთულის მქონე სიტყვებისა და ხშირად დაშვებული შეცდომების თავიდან ასაცილებლად“.

ტექსტური ველები ასევე შეიძლება გამოყენებულ იქნეს მოცემული კონკრეტული ლექსიკოგრაფიული ფუნქციის მისაღწევად ლექსიკონში. როგორც წესი, განმარტებითი ლექსიკონების სიტყვა-სტატიების ენციკლოპედიური ინფორმაციით გადატვირთვა მიღებული არაა, მაგრამ, მცირე დოზით, ტექსტური გრაფები შეიძლება გამოყენებულ იქნეს ენციკლოპედიური ხასიათის მქონე მონაცემების მოსათავსებლადაც. მსგავსი მიდგომა ორენოვანი ლექსიკონებისთვისაც გამოსადეგია, მაგალითად, კულტურული ინფორმაციის გრაფებში მოსაქცევად, სიტყვათხმარების კონკრეტული კომბინაციების გამოყენების თავისებურებებისა ან კოლოკაციების, როგორც გამაფრთხილებელი ფუნქციის მქონე მიკროსტრუქტურული ელემენტი (გოუვსი და სხვ., 2010: 502). წარმოვადგენთ ლონგმენის აფრიკაანსის ლექსიკონიდან (Longman-HAT Afrikaans Dictionary) აღებულ სიტყვა-სტატიას, რომელშიც გამოკვეთილია ტექსტური კომენტარების გრაფა (იხ. სურათი 1).

სურათი 1. ლონგმენის აფრიკაანსის ლექსიკონის მიკროსტრუქტურა (გოუვსი და სხვ., 2010: 502).

eg·ter (*ietwat deftig*)

voegwoordelike bw. om 'n teëstelling uit te druk; ewenwel, nietemin, tog, nogtans: *Ons het hom geroep; hy het egter nie geantwoord nie.*

(a) in Afrikaans begin 'n sin nie met *egter* nie (in Engels kan *however* wel 'n sin inlei). Ook die volgende is foutief: *Piet, egter, was nie teenwoordig nie.*

(b) aangesien *maar* en *egter* dieselfde teenstellende funksie verrig, word hulle nie in dieselfde sin gebruik nie, bv. **maar die ouers het egter nie saamgestem nie.*

აფრიკანსის ლექსიკოგრაფიული სკოლისათვის დამახასიათებელია სასწავლო ტიპის ლექსიკონის მიკროსტრუქტურული გადააზრება და ტექსტური კომენტარების გრაფებით მისი შევსება. მაგალითისათვის, მონოლინგვური ლექსიკონის „Nuwe Woordeboek sonder Grense“ ერთ-ერთ ტექსტობრივ ყუთში ფუტკრის ილუსტრაციით ახსნილია ორი ლექსემის boring და bored შორის არსებული მორფოლოგიურ-სემანტიკური განსხვავება და მისი კონტექსტში რეალიზების მაგალითები (გოუვსი და პრინსლო, 2010: 133) (იხ. სურათი 2).

სურათი 2. სასწავლო ორენოვანი ლექსიკონის ტექსტური გრაფა bored/boring შესახებ (გოუვსი და პრინსლო, 2010: 133).



ამჟამად არსებულ ინგლისურ-ქართულ სასწავლო ლექსიკონების მიკროსტრუქტურაში არ მოიძებნება ტექსტური გრაფები, ან სხვა სახის შიდატექსტობრივი მონაცემები, რომელიც ენის შემსწავლელს კომენტარის სახით დამატებით ლექსიკოგრაფიულ ცნობას მიაწვდის განმარტებული სიტყვის თაობაზე. ამიტომ, ზემოთ განხილული კორპუსული კვლევის შედეგებიდან გამომდინარე, წარმოგიდგენთ სიტყვა-სტატიებში მოცემულ ტექსტობრივ გრაფებს, რომელიც სასწავლო ან სპეციალიზებული ტიპის ინგლისურ-ქართული ლექსიკონის კომპლექსური მიკროსტრუქტურის ნაწილი შეიძლება გახდეს.

თუკი შევაჯამებთ ნარატივის ლექსიკოგრაფიული ტექსტური გრაფების/ველების მნიშვნელობასა და გამოყენების მოტივებს სასწავლო ან სპეციალიზებულ ლექსიკონებში, გამოვყოფთ მისთვის დამახასიათებელ ფუნდამენტურ ნიშან-თვისებებს:

- ძირითადი აქცენტი გაკეთებულია სხვადასხვა ტიპის მიკროსტრუქტურების მნიშვნელობაზე, რომლის დახმარებით ლექსიკოგრაფებს საშუალება აქვთ წარმოადგინონ ლექსიკური ელემენტები ისე, რომ საუკეთესოდ მოერგონ მათი სამიზნე მომხმარებლების საჭიროებებსა და მითითების უნარებს. მომხმარებლის პერსპექტივა დომინანტურ როლს თამაშობს ლექსიკონების დაგეგმვისა და საბოლოო შედგენის პროცესში.
- ამჟამად არსებულ ინგლისურ-ქართულ სასწავლო ლექსიკონების (ბეჭდური თუ ელექტრონულ მატარებელზე გადასული) მიკროსტრუქტურაში არ მოიძებნება ტექსტური გრაფები, ან სხვა სახის შიდატექსტობრივი მონაცემები, რომელიც ენის შემსწავლელს კომენტარის სახით სიტყვათხმარების კონკრეტული კომბინაციების გამოყენების თავისებურებებს ან კოლოკაციებს მიუთითებს.
- რ. გოუესისა და დ. პრინსლუს მიხედვით, სასწავლო კორპუსის შედეგების გათვალისწინება იმ მიზნით, რომ სალექსიკონო მიკროსტრუქტურულ ველებში საჭირო კულტურული, ენციკლოპედიური, კოლოკაციური და სხვა სახის ინფორმაცია შევიტანოთ, ნაყოფიერია ენის შემსწავლელთა ცოდნის გასადრმაველად.

1.4.2. მრავალკომპონენტური ლექსიკური ფრაზების სალექსიკონო მიკროსტრუქტურა

ნებისმიერი ლექსიკონი, იქნება ეს სასწავლო ტიპის თუ სპეციალიზებული, წარმოადგენს მონაცემთა ბაზას, რომლის უმთავრესი დანიშნულებაა ლინგვისტური ინფორმაციის შენახვა კონკრეტული ენობრივი ერთეულების შესახებ. ნაშრომის საკვლევი საკითხიდან გამომდინარე შესიტყვებებისა და სამ- ან ოთხკომპონენტური ლექსიკური ფრაზების ლექსიკოგრაფიული ასახვის პრობლემატიკა ჩვენთვის მნიშვნელოვანია.

განსაკუთრებით ნიშანდობლივია ის სირთულეები, რაც პრაგმატიკული მრავალსიტყვიანი მარკერების განმარტებით ან ორენოვან ლექსიკონებში ასახვას სდევს თან. პირველ რიგში, ერთ-ერთი გადასაჭრელი საკითხი მაკროსტრუქტურული ხასიათისაა, ანუ, მაგალითად, სიტყვა-სტატიის რომელ კატეგორიაში უნდა იქნეს

შეტანილი ლექსიკური ფრაზა *the same goes for?* ჯ. სინკლერი ასეთი ტიპის მრავალსიტყვიან ლექსიკურ ერთეულებს ლექსიკონში დამოუკიდებელ წევრებად მიიჩნევს და მათ „მეთაური სიტყვის“ ნაცვლად „მეთაურ ფრაზებს“ უწოდებს (სინკლერი, 2010: 37).

ფრაზეოლოგიის და მათ შორის, ლექსიკური ფრაზების ლექსიკონებში მოთავსების გადაწყვეტილება დამოკიდებულია ლექსიკონების მიზნებსა და სამიზნე აუდიტორიის ლინგვისტურ მოთხოვნებზე. როგორც წესი, განმარტებით ლექსიკონებში მრავალსიტყვიან კომბინაციათა ჩართვის ყველაზე ტიპური გზაა, სიტყვა-სტატიის ქვეკატეგორიაში მოქცევა, როგორც რომელიმე ლემის მიკროსტრუქტურის ერთი ნაწილი. ასეთი გადაწყვეტილება ელექტრონულ ან ონლაინლექსიკონებში განმარტებული სიტყვების ძიების პროცესს ართულებს. მეტიც, დისკურსის მარკერების, ან ზოგადად, მრავალკომპონენტური ლექსიკური ფრაზების პრაგმატიკული ფუნქციების ილუსტრირება ლექსიკონის შეზღუდულ კომპასში ადვილად ვერ მოხერხდება. ამის გათვალისწინებით, ნაკლებად არის გასაკვირი, რომ ბეჭდური ლექსიკონები ძირითადად რუტინულ ფრაზეოლოგიზმებს (*could you please...?; nice to meet you* და ა.შ.) განმარტავენ, როგორც შეზღუდული რაოდენობით, ისე ლიმიტირებული თითო საილუსტრაციო წინადადებით.

ერთმანეთს შევადაროთ, თუ როგორ გადაიჭრება დასავლეთევროპულ ლექსიკონებში ფრაზეოლოგიზმების ჩართვის პრობლემა. ჰოლანდიური ენის განმარტებით ლექსიკონში (Van Dale Groot Woordenboek Der Nederlandse Taal), მრავალსიტყვიანი ლექსიკური კომბინაციები დანაწევრებულია რამდენიმე თემატურ ველად: *uitdrukkingen* (გამოთქმები), *zegswijzen* (ფრაზები), *spreekwoorden* (ანდაზები), *vergelijkingen* (შედარებები) and *formules* (ფორმულარული გამონათქვამები) (ბერგენჰოლცი და გოუვსი, 2014). ფრანგულ განმარტებით ლექსიკონებშიც მსგავსი ტენდენცია შეინიშნება. საუკეთესო შემთხვევაში, სიტყვა-სტატიაში მოცემულია ყველაზე გავრცელებული დისკურსის მარკერები, მაგალითად, *il convient de* და მისი ეკვივალენტური, სინონიმური ვარიაციები. ინგლისურენოვანი განმარტებითი ლექსიკონების შემთხვევაში, ლექსიკური ფრაზები საგანგებოდ გამოყოფილი, მაკროსტრუქტურულად დამოუკიდებელ სიტყვა-სტატიებში თავსდება. რაც შეეხება, ინგლისურ-ქართულ სასწავლო ან დიდ ლექსიკონებს, ლექსიკურ

კლასტერებს ინდივიდუალურად მინიჭებული მაკროსტრუქტურა არ აქვთ და ერწყმიან ერთსიტყვიანი სიტყვა-სტატიის მიკროსტრუქტურას. თუმცა, არც ერთ ზემოთ ჩამოთვლილ ლექსიკონში დისკურსის მარკერების ფუნქციური აღმნიშვნელები არ გვხვდება და არც სასწავლო კორპუსიდან საილუსტრაციო მასალად გადმოტანის მცდელობა არსებობს.

ჩვენს კვლევაში ვიზიარებთ ჯ. სინკლერის (2010) მოსაზრებას, რომ რამდენიმეკომპონენტიან ლექსიკურ ერთეულებს დამოუკიდებელი სამეთაურო ფრაზის სტატუსი მიენიჭოთ (EAP) ტიპის სპეციალიზებულ ლექსიკონებში, რომლებიც აკადემიური წერის უნარ-ჩვევების გაუმჯობესების მიზნითაა შექმნილი. მეთაური ფრაზის სტატუსი ყველა ტიპის ფრაზეოლოგიურ ერთეულთან მიმართებით არარეალისტური იქნება, თუმცა სამ- ან ოთხკომპონენტიანი ლექსიკურ კონების შემთხვევაში ეს გადაწყვეტილება ლოგიკური მიზეზებითაა განპირობებული.

მრავალკომპონენტიანი ლექსიკური ფრაზების სალექსიკონო მიკრო- და მაკროსტრუქტურაში მოთავსების საკითხის დანახვის ალტერნატიულ გზად გვევლინება დ. სიპმანის მოსაზრება, რომლის მიხედვით ლექსიკური ფრაზები შეიძლება დალაგდეს კონცეპტუალურად, ანუ თემატურად, ლექსიკური ფრაზების დისკურსული ფუნქციების მიხედვით, ან ანბანთრიგზე: ა) ანბანთრიგზე პოზიციურად პირველი კომპონენტის მიხედვით; ბ) ანბანთრიგზე სემანტიკურად ყველაზე მნიშვნელოვანი კომპონენტის მიხედვით; გ) ანბანთრიგზე გრამატიკულად ყველაზე მნიშვნელოვანი კომპონენტის მიხედვით (სიპმანი, 2004: 297).

ავტორი ორენოვანი ლექსიკონის მრავალკომპონენტიანი ლექსემების მიკროსტრუქტურაში მოთავსების პრინციპების განხილვისას, გამოყოფს შემდეგ ელემენტებს: ფუნქციურ ინფორმაციას, სემანტიკურ ინფორმაციას, დარგობრივ კვალიფიკაციას (მაგალითად, რომელ ჟანრსა და დისციპლინაში გამოიყენება ესა თუ ის დისკურსის მარკერი უმეტესწილად) და ბოლოს კონტექსტებს. მაგალითისათვის, მეთაურ ფრაზად აღებული აქვს *es ist anzumerken* და ავტორის მიერ შემოთავაზებული სიტყვა-სტატია შემდეგნაირად გამოიყურება (იხ. სურათი 3).

FUNKTION	es ist anzumerken
KONTEXT	einen erklärenden, häufig kritischen Zusatz zu einem vorhergehenden Teiltext machen meist in wissenschaftlichen Texten, selten in Zeitungstexten
STRUKTUR UND KOLLOKATOREN	zu etw./bei etw.//hierzu/dazu/hierbei ist/bleibt (kritisch/negativ//positiv// generell/ überhaupt...//noch/ferner/außerdem) anzumerken, daß... Es ist (kritisch/noch/ferner/negativ...) (zu etw.) anzumerken, daß... (Kritisch/Negativ//Positiv//Generell//Überhaupt...//allerdings/jedoch...//ferner/...) ist (zu etw.) (jedoch/indes//noch/ferner/außerdem/...) anzumerken, daß...(kritisch/negativ//positiv...) anzumerken ist (jedoch/indes//ferner/außerdem/...), daß... (hier[zu]/dazu/an dieser Stelle...) ist/wäre (einiges/folgendes/etliches...) anzumerken
1 BEISPIEL	In der hier gewählten Perspektive erwiesen sich Demarkations- und Differenzierungsprozesse als untrennbar aufeinander bezogen; ihr Wechselspiel zeigte sich als wesentlich für die Ausgestaltung schizophrener Phänomene, wobei ein erweiterter Kognitionsbegriff sich als sinnvoll erwiesen hat, um neurophysiologische und psychische Prozesse gemeinsam zu erfassen, ohne zu reduktionistischen Lösungen zu gelangen. Allerdings ist hierzu kritisch anzumerken, daß der weit gefaßte Kognitionsbegriff MATURANAs nicht nur Vorteile bietet. (psychologischer Fachtext) [...]

მეთაური ფრაზის *es ist anzumerken* („უნდა აღინიშნოს“) შედგება შემდეგი მიკროსტრუქტურული ელემენტებისგან: ა) ფუნქცია: ამხსნელ-განმმარტებელი, ხშირად გამოიყენება წინარე მსჯელობის კრიტიკულ გაგრძელებად; ბ) კონტექსტი: უმეტესწილად სამეცნიერო ტექსტებში, იშვიათად - საგაზეთო ჟანრში; გ) სტრუქტურა და კოლოკაცია: zu etw./bei etw.//hierzu/dazu/hierbei ist/bleibt; დ) მაგალითი: Allerdings ist hierzu kritisch anzumerken, daß der weit gefaßte Kognitionsbegriff MATURANAs nicht nur Vorteile bietet. (psychologischer Fachtext).

1.5. შუალედური დასკვნა

წინამდებარე თავის მიზანი იყო წარმოგვედგინა სასწავლო კორპუსის კვლევისა და მეთოდოლოგიის თეორიული წანამდგრები, მისი როგორც თეორიული, ისე პრაქტიკული მხარეები.

როგორც აღმოჩნდა, სასწავლო კორპუსის განსაზღვრებები არაერთგვაროვანია. ნ. ნესელჰაუფისა და ს. გრეინჟერის მიერ მოცემული დეფინიციებისა და სასწავლო კორპუსებისათვის დამახასიათებელი მნიშვნელოვანი ნიშან-თვისებების გათვალისწინებით, ვასკვნით, რომ სასწავლო კორპუსის კვლევა გაცდა ერთი კონკრეტული დისციპლინის, კერძოდ, კორპუსლინგვისტიკის შესწავლის ფარგლებს და გადაიქცა მულტიდისციპლინურ მიმართულებად. სასწავლო კორპუსი წარმოადგენს სისტემატურ, ელექტრონულ მატარებელზე გადატანილი ტექსტების

ნაკრები, რომელიც უცხოური ენის შემსწავლელთა მიერაა შესრულებული. ვინაიდან ენის შემსწავლელთა ლინგვისტური მასალის შეგროვების შემთხვევაში, ავთენტურობის დონე, გარკვეული თვალსაზრისით, დარღვეულია, სამართლიანია დავასკვნათ, რომ სასწავლო კორპუსები ნაწილობრივ ბუნებრივ, ავთენტურ მასალას მოიცავენ. მეორე მხრივ, ბუნებრიობის ეს ხარისხი საკვლევი საკითხების გადაჭრის პროცესში დაბრკოლებას არ წარმოადგენს. ისინი კლასიფიცირდება განზომილებებისა და დიქტომიის მიხედვით: ა) რა ტიპის ლინგვისტური მასალა (ზეპირმეტყველებისა თუ წერილობითი) გარდაიქმნება კორპუსად; ბ) ჟანრული კლასიფიკაციის თვალსაზრისით, სპეციალიზებულ, რომელიმე დარგზე ორიენტირებულ მასალას ეფუძნება სასწავლო კორპუსი თუ უფრო ზოგადი ხასიათისა; გ) სამიზნე ენების შერჩევის მიხედვით; დ) დროითი კატეგორიის მიხედვით, კერძოდ, ცოდნის მომენტალური აღმბეჭდავი არიან თუ უცხოური ენის შემსწავლელთა ენას დიაქრონიულად შეისწავლიან.

მეორე საკითხი, არის სასწავლო კორპუსის შექმნისას გასათვალისწინებელი ცვლადები (გარემოს, დავალებისა და საკუთრივ ენის შემსწავლელთა ინფორმაცია). სანამ სასწავლო კორპუსის შეგროვების პროცესი დაიწყება, მკვლევარმა საკვლევი პრობლემის დასმის შემდეგ უნდა გადაწყვიტოს, თუ რა სახის მონაცემები შევა კორპუსში, რადგან არაერთგვაროვანი და ერთმანეთთან შეუთავსებელ მონაცემთა სიმრავლე კვლევის წარმოებისთვის უსარგებლოა. ამ კუთხით, ღირებულად მიგვაჩნია არსებული დასავლური გამოცდილების გათვალისწინება, მაგალითად, ლუვენის ინგლისური ენის ზეპირმეტყველების საერთაშორისო მონაცემთა ბაზისა და ლეიდენის სასწავლო კორპუსის შექმნის პრინციპების გაზიარება.

რაც შეეხება, შესიტყვებათა და ლექსიკურ ფრაზათა კვლევას სასწავლო კორპუსში, კვლევისას მონაცემთა ამორჩევის, ანალიზისა და ინტერპრეტაციის მიხედვით, განარჩევენ კორპუსით ინფორმირებულ (*corpus-informed*), კორპუსზე დამყარებულ (*corpus-based*) და კორპუსით განპირობებულ (*corpus-driven*) მიდგომებს. გასათვალისწინებელია, რომ თითოეული მათგანის პრინციპები ნაწილობრივ ერთმანეთს ემთხვევა და მათი იზოლირებულად აღქმა მართებული არაა. თავად ცნება „შესიტყვება“ დღეისთვის მეტ-ნაკლებად დიფერენცირებულად გამოიყენება ენის სხვადასხვა დონეზე. კოლოკაცია აღნიშნავს ლექსემათა თანაარსებობას და

განპირობებულია რამდენიმე ფაქტორით. მიუხედავად ამ ცნების გაგების მრავალსახოვნებისა, შესიტყვების განსაზღვრებათა ჯაჭვში, სულ მცირე, ერთი საერთო მნიშვნელის პოვნა შეგვიძლია და რაზეც ძირითადად სამეცნიერო წრეები თანხმდებიან, არის ის, რომ სიტყვებს შორის არსებული სინტაგმატური კავშირები შესიტყვებების მთავარი განმსაზღვრელი ფაქტორია.

ერთ-ერთი შეხედულების მიხედვით, შესიტყვება გულისხმობს ენობრივ ერთეულთა უშუალო კონტექსტუალურ, გვერდიგვერდ თანაარსებობას. ამ კონცეპტის კვლევა ჯ. სინკლერის აღმოჩენებმა განაპირობა და ფრაზეოლოგიის ახალი, ინდუქციური მიდგომით, დისტრიბუციული (ევერტი, 2005), სხვაგვარად, სიხშირეზე დამყარებული მიდგომით შესწავლის შესაძლებლობა გააჩინა. მხოლოდ დისტრიბუციული, ანუ სტატისტიკაზე დამყარებული მიდგომის გამოყენება კოლოკაციათა კვლევის უფრო ნათელ და ზუსტ შედეგებს ვერ მოგვცემს, ამიტომ არსებული დანაკლისი მნიშვნელობაზე ორიენტირებული მიდგომითაც უნდა შეივსოს. სტატისტიკაზე ორიენტირებული მიდგომის გამოყენების დროს გაჩენილ სიცარიელეს ამოვასებთ სასწავლო კორპუსში შესიტყვებების მოდელების მანუალური მოძიებით. შესიტყვებათა სტრუქტურული ძიებისათვის ვიყენებთ კოლოკაციურ მოდელებს, რომელშიც წარმოდგენილია ლექსიკური შესიტყვებების 9, ხოლო გრამატიკული შესიტყვებების 10 პარადიგმა (ბენსონი და სხვ., 1997: 9). იგი ეფუძნება BBI კომბინატორულ ლექსიკონს (The BBI Combinatory Dictionary of English, 1997).

ვინაიდან, ჩვენ მიერ შექმნილი სასწავლო კორპუსი ჟანრული სხვადასხვაგვარობით ხასიათდება და თითოეულ ტექსტში მრავალსიტყვიანი ლექსიკური ერთეულების განსხვავებული ფუნქციური ხმარების ტენდენცია იკვეთება, წარმოვადგინეთ ლექსიკური ფრაზების რამდენიმე ფუნქციური ტაქსონომია: დ. სიპმანის, კ. ჰაილენდისა და ჯ. ო'ფლინის კლასიფიკაციები. მრავალკომპონენტიან დისკურსულ მარკერებს შორის საზღვრები იმდენად არამდგრადია, რომ მათ მკაცრად განკუთვნილ ფუნქციურ ჩარჩოში ვერ მოვაქცევთ.

რაც შეეხება, სასწავლო კორპუსებში ლექსიკური სიმდიდრის ცნებას, ის მრავალგანზომილებიანი კონსტრუქტიაა და შედგება ლექსიკური სიმჭიდროვის, ლექსიკური მრავალფეროვნებისა და ლექსიკური სირთულისაგან. ამ სისტემის

სარგებლიანობა ენის შემსწავლელთა საერთო ლინგვისტური პროგრესის შეფასებაშიც მდგომარეობს და კვლევის პროცესში სამივე საზომი ტოლფასოვანი და თანაბარდირებულია. ჩვენს კვლევაში ლექსიკური სიმდიდრის ობიექტური და სანდო სიდიდის მისაღებად ოპტიმალურია ლექსიკური ვარიაციის საზომების – type/token ratio და lemma/token ratio – კომბინირებული გამოყენებით.

პირველი თავის ერთ-ერთი ამოცანა იყო სასწავლო კორპუსის მასალის ლექსიკოგრაფიული გამოყენების შესაძლებლობების დანახვა. როგორც გამოჩნდა, თანამედროვე მეტალექსიკოგრაფიაში სიტყვა-სტატიების სტრუქტურის განუყოფელი ნაწილია ტექსტური გრაფები, რომლებიც საჭირო კულტურულ, ენციკლოპედიურ, კოლოკაციურ და სხვა სახის ინფორმაციას ინახავენ და ნაყოფიერია ენის შემსწავლელთა ცოდნის გასაღრმავებლად. ფრაზეოლოგიის და მათ შორის, ლექსიკური ფრაზების ლექსიკონებში მოთავსების გადაწყვეტილება დამოკიდებულია ლექსიკონების მიზნებსა და სამიზნე აუდიტორიის ლინგვისტურ მოთხოვნებზე და მათი სალექსიკონო ერთეულად ჩართვა განსხვავებული სტრატეგიით ხორციელდება.

თავი II. ინგლისური ენის ქართველ შემსწავლელთა კორპუსის (GLEAN) შექმნის პრინციპები

2. ზოგადი მიმოხილვა

წინამდებარე თავში განვიხილავთ ინგლისური ენის ქართველი შემსწავლელების სასწავლო კორპუსს, მის შექმნასთან დაკავშირებულ ისეთ საკითხებს, როგორიცაა ტექსტობრივი მასალის შეგროვება, ლინგვისტური ანოტაცია, კონტრასტული ანალიზისთვის საერთაშორისო სასწავლო კორპუსების შერჩევის კრიტერიუმები და ა.შ. ჩვენ მიერ შექმნილი სასწავლო კორპუსი ამჟამად განთავსებულია პერსონალური წვდომის ლანკასტერის უნივერსიტეტის CQPweb სერვერზე, ტრიირის უნივერსიტეტის პროფესორ სებასტიან ჰოფმანის უშუალო მონაწილეობით. სასწავლო კორპუსზე წვდომა შესაძლებელია მომხმარებლის ანგარიშით ([CQPweb Main Page \(lancs.ac.uk\)](http://CQPweb.Main.Page(lancs.ac.uk)))¹. იგი მოიცავს 10 მილიონამდე სიტყვას.

2.1. ინგლისური ენის ქართველ შემსწავლელთა კორპუსის კრიტერიუმები

ინგლისური ენის ქართველი შემსწავლელების წერილობითი სასწავლო კორპუსის (Georgian Learner English and Newspaper Corpus) შემოკლებული დასახელება „GLEAN“ სიმბოლურ ხასიათს ატარებს. ზმნა to glean მკის დროს ყანაში დაბნეული თავთავების მარცვალ-მარცვალ აკრეფას აღნიშნავს, რაც ერთგვარად ჩვენი კვლევის პროცესს, ენაში რეალიზებული სიტყვათხმარების თავისებურებების ძიებასა და ამოკრებას ასახავს მეტაფორულად.

GLEAN სასწავლო კორპუსი განზომილებების მიხედვით, ხასიათდება როგორც სინქრონიული, სპეციალიზებული ტიპის სასწავლო კორპუსი, რომელიც ერთენოვან (მხოლოდ ინგლისურ ენაზე) შესრულებულ წერილობითი ხასიათის მასალას მოიცავს. მასალის შეგროვება მიმდინარეობდა სამი წლის განმავლობაში. კორპუსი ასახავს ენის შემსწავლელთა არა მხოლოდ ენობრივი ცოდნის ევოლუციას დროთა განმავლობაში,

¹ შენიშვნა: სასწავლო კორპუსის მოხმარებისათვის საჭიროა მითითებული მომხმარებლის სახელისა: Marine1995 და პაროლის: Lali30042880 გამოყენება.

არამედ მისი შესწავლის საფუძველზე, გამოვიკვლიოთ ენა და ენობრივი ქცევის ნორმები სინქრონიულ ასპექტში.

ქანრული განაწილების თვალსაზრისით, ინგლისური ენის ქართულ შემსწავლელთა კორპუსი სპეციალიზებული ხასიათისაა და მოიცავს დარგობრივი და ქანრული მრავალფეროვნების მქონე ტექსტებს. კორპუსში დარგობრივი წერილობითი მასალა შედგება როგორც ლინგვისტიკის, ისე ლიტერატურის მიმართულებით შესრულებული სტუდენტური აკადემიურ ნაშრომების, სხვადასხვა ტიპის ესეს, საგაზეთო სტილის პოლიტიკურ-სოციალური თემატიკის სტატიების, ბლოგების, დღიურებისა და კორესპონდენციისაგან. წერილობითი რეგისტრის შემთხვევაში, მასში მრავალფეროვნება ამგვარი სახით მიიღწევა. მეორე მხრივ, ზეპირმეტყველების კორპუსის სპეციალიზებულ ბუნებას განაპირობებს ასაკობრივი ჯგუფების სიჭრელე, საქმიანი მოლაპარაკების, სპეციფიკური სოციოლექტური დიალოგი ან თუნდაც ალკოჰოლის ზემოქმედების ქვეშ მყოფი ადამიანის მეტყველება.

GLEAN შექმნის თავდაპირველი მოტივაცია უკავშირდებოდა მხოლოდ სტუდენტთა ესეების შეგროვებას, თუმცა, რეპრეზენტაციულობისა და დაბალანსებულობის პრინციპებიდან გამომდინარე, მივიღეთ გადაწყვეტილება, ტექსტობრივი მასალა ინგლისური ენის შემსწავლელთა სხვადასხვა ქანრის წერილობითი ნამუშევრებითაც შეგვევსო. გადაწყვეტილება კორპუსის მოცულობითაც იყო განპირობებული, რადგან მხოლოდ სტუდენტთა ესეებისაგან შემდგარი, რამდენიმე ათეულათასიანი სიტყვის შემცველი სასწავლო კორპუსის ანალიზი მრავალმნიშვნელოვან, განმაზოგადებელ დასკვნებამდე ვერ მიგვიყვანდა. ჯ. სინკლერის მიერ შემოთავაზებული პრინციპის გათვალისწინებით, „კორპუსი მაქსიმალურად მოცულობითი უნდა იყოს და მუდმივად იზრდებოდეს“ (სინკლერი, 1991: 18). საბოლოოდ, მივიღეთ 10 მილიონამდე ტოკენის რაოდენობის სასწავლო კორპუსი. უფრო კონკრეტულად, ტექსტებისა და ტოკენების კორპუსში გადანაწილება გამოიყურება შემდეგნაირად (იხ. სქემა 6).

ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში:	რაოდენობა
ტოკენის რაოდენობა	9,776,931
ტიპის რაოდენობა	118,947
ტექსტების რაოდენობა	34,427
ენის შემსწავლელთა რაოდენობა	65

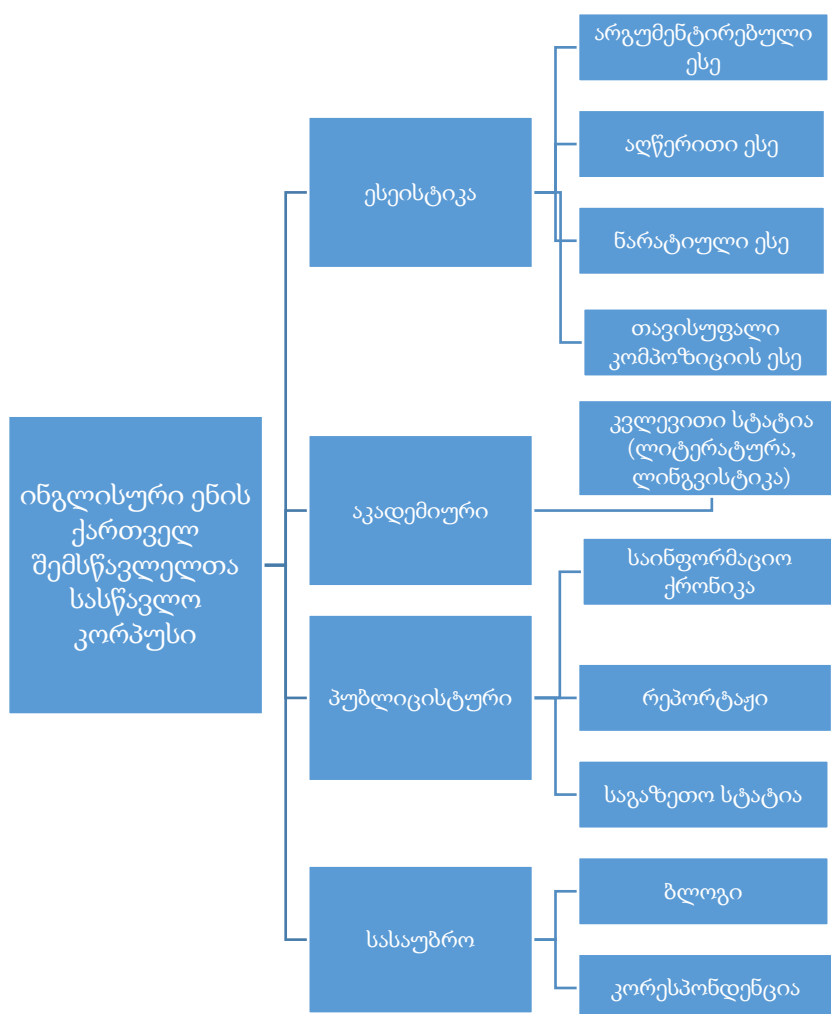
„ტოკენში“ იგულისხმება კორპუსის უმცირესი შემადგენელი ელემენტები და შეესაბამება ცნებას „სიტყვაფორმა“, ხოლო „ტიპი“ – ცალკეული სიტყვაფორმის ხმარების შემთხვევას, ანუ განსხვავებული/უნიკალური სიტყვაფორმების რაოდენობას განსაზღვრავს.

2.2. ჟანრული კლასიფიკაცია ინგლისური ენის ქართველ შემსწავლელთა კორპუსში

კორპუსის შექმნის ერთ-ერთი ბაზისური მახასიათებელი მისი დაბალანსებულობაა, რომელიც გულისხმობს კორპუსში ჩართული რესურსების ჟანრული მრავალფეროვნების გაწონასწორებას. რეპრეზენტაციულობის მსგავსად, დაბალანსებულობის ცნება სასწავლო კორპუსისთვის მეტ-ნაკლებად ბუნდოვან საკითხად რჩება და მისი სრულად მიღწევა თითქმის შეუძლებელია, თუმცა ჯ. სინკლერი მოგვიწოდებს, მაინც ვისწრაფოდეთ დაბალანსებულობის პრინციპისაკენ, იქნება ეს წერილობითი თუ ზეპირმეტყველების სასწავლო კორპუსი (სინკლერი, 1991).

ჩვენ მიერ შექმნილი სასწავლო კორპუსი (GLEAN) სინქრონულად, შეძლებისდაგვარად, დაბალანსებულია და მოიცავს ენის რეალიზების სხვადასხვა რეგისტრს. შეგროვებულ იქნა არამხატვრული ტიპის სხვადასხვა დარგის, ჟანრისა და მათივე ქვეკატეგორიაში შემავალი ტექსტები და აღნიშნული მასალის კლასიფიკაცია ამგვარად გამოიყურება (იხ. სქემა 7).

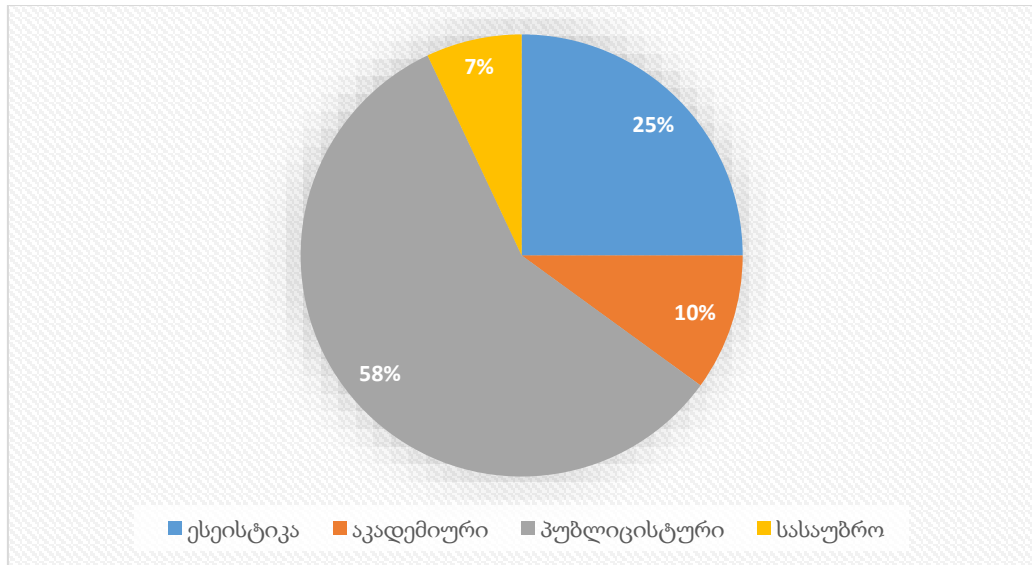
სქემა 7. GLEAN სასწავლო კორპუსის ჟანრული კლასიფიკაცია



სასწავლო კორპუსში გამოვყავით ჟანრთა ოთხი კატეგორია (ესეისტიკა, აკადემიური, პუბლიცისტური, სასაუბრო) და მათი ქვეკატეგორიები. ესეისტიკის ჟანრში ერთიანდება არგუმენტირებული ესე, ნარატიული ესე, აღწერითი ესე, თავისუფალი კომპოზიციის ესე. აკადემიური ჟანრი მოიცავს კვლევით დარგობრივ სტატიებს ლინგვისტიკასა და ლიტერატურაში. პუბლიცისტის ჟანრში გაერთიანებულია საინფორმაციო ქრონიკა, რეპორტაჟი და საგაზეთო სტატიები როგორც პოლიტიკურ, ისე აპოლიტიკურ საკითხებზე, ხოლო ბლოგისა და კორესპონდენციის (არაოფიციალური მეილები) სასაუბრო ჟანრში გავაერთიანეთ.

კვლევის კიდეც ერთი მხარეა სასწავლო კორპუსში დარგებისა და ჟანრების მოცულობის შესწავლა და ხვედრითი წილის განაწილება მთლიან კორპუსთან მიმართებით (იხ. დიაგრამა 1).

დიაგრამა 1. დარგების განაწილება GLEAN სასწავლო კორპუსში



როგორც ვხედავთ სასწავლო კორპუსის 58% პუბლიცისტური ჟანრის ტექსტებს უკავია, 25% – ესეისტიკას, აკადემიური ჟანრის მასალა GLEAN კორპუსის 10% წარმოადგენს, ხოლო ყველაზე მცირე 7% კი – სასაუბროს. დაბალანსებულობის მისაღწევად აუცილებელი არაა ტექსტების იდენტური რაოდენობის თანაარსებობა კორპუსში. სასწავლო და სხვა ტიპის კორპუსებს შორის არსებული ერთ-ერთი თვალსაჩინო განსხვავება მათ ჟანრულ გადანაწილებაშია. სასწავლო კორპუსებში არამხატვრული ტექსტები ჭარბობს, რაც იმით აიხსნება, რომ ენის შემსწავლელებისათვის მხატვრული ტექსტების (პოეზიის, ნოველების, ჩანახატების) წერა სირთულეებთანაა დაკავშირებული. შედარებისათვის, მაგალითად, თანამედროვე ამერიკული ინგლისურის კორპუსი (COCA), რომელიც 450 მილიონ სიტყვას მოიცავს, ჟანრული მრავალფეროვნებით გამოირჩევა და მოიცავს ზეპირმეტყველების, მხატვრულ ტექსტებს, პოპულარულ ჟურნალებს, პოლიტიკურ სტატიებსა და აკადემიურ-სამეცნიერო ტექსტებს. თითოეულ მათგანში ქვეკატეგორიები გამოიყოფა, კერძოდ, პოპულარულ ჟურნალებში განირჩევა ფინანსური, სპორტული, ფემინისტური და სხვა ტიპის ქვეჟანრები.

ჟანრული მრავალფეროვნების თვალსაზრისით, ღირებულია ქართული ენის ეროვნული კორპუსიც (GNC), რომელშიც თავმოყრილია მხატვრული და არამხატვრული ტექსტები და ჟანრები: რელიგია, პროზა, ფილოსოფია, ავტობიოგრაფია, იურისპრუდენცია, პრესა და სხვ.

რაც შეეხება ადრეული ახალი ინგლისურის კორპუსს, რომელიც 755 მილიონ სიტყვას ითვლის, შ. მერფი (2019: 66) ხუთი დარგის 20-ზე მეტ ჟანრულ კატეგორიას და ქვეჟანრს გვთავაზობს, მაგალითად, საცნობარო დარგში ერთიანდება ბიოგრაფია, წერილი, პამფლეტი; რელიგიურ დარგში შედის ბიბლიის, კათოლიციზმის, პროტესტანტიზმის, თეოლოგიის ჟანრები.

ამგვარად, ინგლისური ენის ქართველი შემსწავლელების კორპუსის (GLEAN) ჟანრული კლასიფიკაცია თვისებრივად ახლოს დგას სხვა ტიპის კორპუსებთან, თუმცა მთავარი განმასხვავებელი ნიშანი არამხატვრული ჟანრის მრავალფეროვნებაში მდგომარეობს.

2.3. მონაცემთა შეგროვება ინგლისური ენის ქართველ შემსწავლელთა კორპუსისათვის

კორპუსის შექმნის ერთ-ერთი ფუნდამენტური პრინციპი არის ტექსტობრივ ბაზაში შემავალი ყველა ინფორმაციის, კორპუსის სტრუქტურის, ენის შემსწავლელთა ინფორმაციის დეტალური ასახვა და ანალიზი. სასწავლო კორპუსის ანალიზისას მკვლევარმა შესაძლოა მიიღოს მისთვის მოულოდნელი შედეგები, რადგან, როგორც წესი, კორპუსი მოულოდნელობებითაა სავსე და მასალის ინტერპრეტაციისას ასეთი გაუთვალისწინებელი შედეგების სათავე არასწორად დოკუმენტირებულ მასალაში უნდა ვეძიოთ (სინკლერი, 2005: 7). ამ რისკის გათვალისწინებით, ინგლისური ენის ქართველ შემსწავლელთა კორპუსისათვის მონაცემთა შეგროვება რამდენიმე საფეხურისაგან შედგებოდა.

პირველ ეტაპზე, ერთი მხრივ, შეირჩა პოპულაცია, სტუდენტები, ქართველების მიერ შესრულებული ინგლისურენოვანი პრესის წყაროები და კორესპონდენციის მოპოვების გზები. ესეებისა და ელექტრონული მეილების მასალის შეგროვებისათვის კვლევაში მონაწილეობა მიიღეს ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის, ინგლისური ფილოლოგიის მიმართულების ბაკალავრიატისა და მაგისტრატურის საფეხურის სტუდენტებმა. კორესპონდენციის მასალაში ერთიანდება ინგლისურენოვანი ელექტრონული მეილები (საქმიანი,

არაოფიციალური და ა.შ.), რომლებიც სტუდენტებმა უკანასკნელი ერთი წლის განმავლობაში გაგზავნეს. მეილებიდან გაფილტრულ იქნა ადრესატისა და გამგზავნის შესახებ არსებული მეტაინფორმაცია. პრესის მასალის წყაროებად გამოყენებულია: ა) ახალი ამბების ონლაინპლატფორმა agenda.ge, რომელშიც თავს იყრის როგორც პოლიტიკური, ისე ტურისტული, ჯანდაცვის, საგანმანათლებლო, სპორტული და სასოფლო-სამეურნეო თემატიკის სტატიები; ბ) ქართველ სტუდენტთა ინგლისურენოვანი ბლოგები სამოგზაურო თავგადასავლებისა და პირადი გამოცდილებების შესახებ, აგრეთვე თავმოყრილია მიმოხილვები.

მონაცემთა შეგროვების წინ მათ გავაცანით ინფორმაცია კვლევის მიზნებისა და მონაწილეობის კონფიდენციალურობის შესახებ. თანხმობის მიღების შემდეგ, მათ დაევალათ ენობრივი და ავტობიოგრაფიული ველების შევსება, როგორიცაა: მშობლიური ენა; მშობლების მშობლიური ენა; რომელ ენას იყენებენ ყველაზე ხშირად ყოველდღიურად (შინ თუ გარეთ); ერთიან ეროვნულ გამოცდაზე ინგლისური ენის ქულა (სურვილისამებრ); ინგლისური ენის თვითშეფასება თითოეულ უნართან მიმართებით (მეტყველება, მოსმენა, წერა და კითხვა); ინგლისური ენის საერთაშორისო გამოცდის შედეგი (ასეთის არსებობის შემთხვევაში); სქესი; ასაკი; უცხო ქვეყანაში ცხოვრების ხანგრძლივობა; განათლება (უნივერსიტეტი/ფაკულტეტი/კურსი/კურსდამთავრებული); ხარისხი (ასეთის არსებობის შემთხვევაში). მეტამონაცემების შევსების შედეგად დადგინდა, რომ კვლევაში მონაწილეთა შორის სხვადასხვა ეროვნების სტუდენტები (მაგალითად, ეთნიკური უმცირესობის წარმომადგენლები) მოხვდნენ, თუმცა კვლევის მიზნებიდან და სადისერტაციო ნაშრომის სათაურიდან გამომდინარე, მათი მონაცემები სასწავლო კორპუსში შეტანილ არ იქნა.

შეგროვებული მასალა კომუნიკაციური ფუნქციის მატარებელია და იცავს მიუკერძოებლობის პრინციპს. ჯ. სინკლერის (2005: 15) მოსაზრებით, კორპუსის შემქნელებს ხშირად იზიდავთ შესაგროვებელი ტექსტების ლექსიკის, ისე სხვადასხვა ლინგვისტური სტრუქტურების კონტროლი და მათში ჩარევა. აღნიშნული მიდგომის გათვალისწინებით, GLEAN სასწავლო კორპუსის სტუდენტთა მიერ შესრულებული ესეები არის მხოლოდ სააუდიტორო მეცადინეობისას შესრულებული, ელექტრონული რესურსების, ხელოვნური ინტელექტის წვდომისგან თავისუფალი.

სამუშაო დროდ განსაზღვრული იყო მაქსიმუმ 50 წუთი და სტუდენტები სიტყვათა რაოდენობაში შეზღუდულნი არ იყვნენ. სასწავლო კორპუსის ტექსტების მოცულობის ვარირება კვლევის შედეგებს არ ცვლის და მის ვალიდურობას ეჭვქვეშ არ აყენებს. კვლევაში მონაწილე ენის შემსწავლელებს აქვთ ავტონომია, შეასრულონ სასურველი ზომის ტექსტი. პრესის, ბლოგებისა და კორესპონდენციის მასალის შეგროვება ვინაიდან ასინქრონულ რეჟიმში მოხდა, გარემოს ცვლადიც განსხვავებული და ნაკლებად კონტროლირებადია. კვლევაში მონაწილეები წერის პროცესში იყენებდნენ ლექსიკონს, როგორც დამხმარე საშუალებას და ეს ფაქტორი შედეგების ინტერპრეტაციისას გათვალისწინებულია. კონტრასტული ანალიზისათვის შერჩეული საერთაშორისო სასწავლო კორპუსის მასალაში ისეთი გარედაქტირებული სტუდენტური ესეებიცაა თავმოყრილი, რომელთა შესრულებისას ინგლისური ენის შემსწავლელებს ლექსიკონზე წვდომა ჰქონდათ.

ჰომოგენურობის პრინციპის მიხედვით კორპუსში არ უნდა იქნეს ისეთი უჩვეულო და საეჭვო ტექსტები შეტანილი, რომლებიც თავიანთ კატეგორიაში სხვებისგან რადიკალურად განსხვავდებიან. ამ პრინციპის მიხედვით, კორპუსლინგვისტებმა თავიდან უნდა აირიდონ ისეთ მასალას, რომელიც შესაბამის კრიტერიუმებს არ აკმაყოფილებს (სინკლერი, 2005: 20). ჩვენს სასწავლო კორპუსში უგულებელყოფილია (არ შევიტანეთ) რამდენიმე ტიპის ესე: 1. გარედაქტირებული ესეები, რომლებიც ენის ბუნებრიობისა და ავთენტურობის პრინციპს დაარღვევდა. სტუდენტების შესრულებული ესეები სასწავლო კორპუსში შევიტანეთ პირვანდელი სახით, ლექტორების ჩასწორებების გარეშე; 2. ესეები, რომლის ავტორებისთვისაც ქართული მშობლიური ენა არაა; 3. ძალიან მოკლე ზომის ესეები (ვგულისხმობთ 50-მდე სიტყვისგან შემდგარ ესეებს, რომელშიც აზრის ლოგიკური ბმა დარღვეული იყო და დაუსრულებელი შინაარსის შთაბეჭდილებას ტოვებდა). ეს უკანასკნელი არ ეხება ელექტრონულ წერილებს, რადგან კორესპონდენციის ჟანრის ერთ-ერთი მახასიათებელი ლაკონურობაა.

დავალების ცვლადის, თემატიკის ანალიზისას გამოიკვეთა კორესპონდენციის ჟანრში სტუდენტთა მიერ ამა თუ იმ ინფორმაციის მოთხოვნის ან დაზუსტების შინაარსის ტექსტები, უცხოური სასწავლებლებისა და დასაქმების პროგრამების მისამართით გაგზავნილი ელექტრონული წერილები.

ესეების თემატიკის შერჩევისას გამოვყავით მათთვის საინტერესო რამდენიმე თემატური კატეგორია, მათ შორის ინტერპერსონალური ურთიერთობები, მოგონებები, პროცესის ამსახველი ესეები, ორი კონცეპტის შედარების ესეები, ნახატების ან ფოტოგრაფიული ნამუშევრების აღწერითი ესეები და ა.შ. იხილეთ შესრულებული ესეების სათაურების ნიმუშები (იხ. სქემა 8).

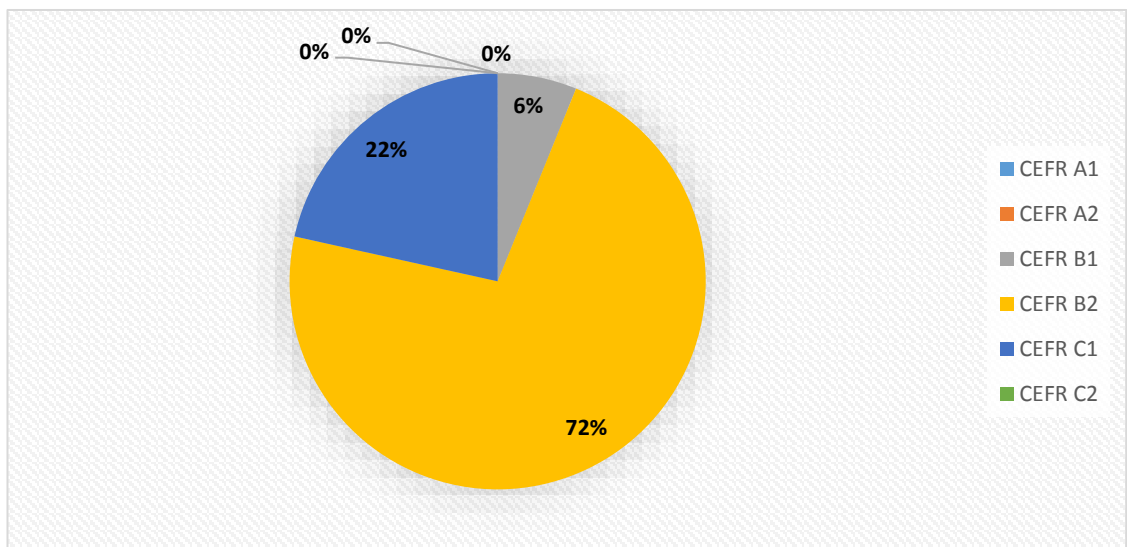
სქემა 8. სასწავლო კორპუსში შემავალ სტუდენტურ ესეთა სათაურების ნიმუშები

№	ესეების სათაურები
1.	Perfect sunset I have ever seen...
2.	How to enjoy the weekend for less than 50 GEL.
3.	Look at the photograph "Skier sunbathing in front of the Matterhorn Zermatt Switzerland" and not only describe what you can see, but discuss how it makes you feel. Use five senses.
4.	What does home mean to you?
5.	Express and reflect on your life experience which has impacted you deeply (a trip, break-up, university life etc.).
6.	How important is family for you? Is the family relationship the most enduring of all?
7.	Importance of make-believe games for children development.
8.	Recall one memorable day from your childhood.
9.	Happiness is there in front of our eyes, but we don't see it. Miracles do happen.
10.	There are plans to demolish an old and unused building in the town where you are a student. You feel that the building should be saved. You decide to write a proposal for the town council explaining why you think the building should be preserved, suggesting what could be done to modernise it and saying how the building could benefit the local people.
11.	"Marriages, like chemical unions, release upon dissolution packets of the energy locked up in their bonding" - John Updike "An orphaned swimming pool".
12.	What objects tell the story of your past, present and future? Why did you select these items? Describe them.

დავალების ცვლადის ასპექტში ესეების, საგაზეთო სტატიებისა და ბლოგების, ელექტრონული წერილების მოცულობის განსაზღვრაც მნიშვნელოვანია. GLEAN-ში ესეების მოცულობის მედიანა ≈ 510 -ია (სიტყვათა მაქსიმალური რაოდენობა ≈ 900 , მინიმალური ≈ 120); საგაზეთო სტატიებისა და ბლოგების მასალის მედიანა ≈ 350 (სიტყვათა მაქსიმალური რაოდენობა ≈ 600 , მინიმალური ≈ 100); კორესპონდენციის მედიანა ≈ 80 (სიტყვათა მაქსიმალური რაოდენობა ≈ 140 , მინიმალური ≈ 20).

ენის შემსწავლელთა ცვლადების – სქესის, მშობლიური ენის, ასაკისა და ინგლისური ენის ცოდნის დონის განსაზღვრამ გვიჩვენა, რომ ენის შემსწავლელთა საშუალო ასაკი 23 წელია, მათი 86% არის მდედრობითი სქესის, 14% კი – მამრობითი სქესის. კვლევაში მონაწილე ყველა პირისათვის ქართული ენა არის მშობლიური, ხოლო სტუდენტთა ინგლისური ენის სწავლის საშუალო წლიური მაჩვენებელი 13 წელია. კვლევაში მონაწილეთა (65 პირი) ინგლისური ენის ცოდნის დონე თვითშეფასების მიხედვით, გამოიყურება შემდეგნაირად (იხ. სქემა 2):

დიაგრამა 2. სასწავლო კორპუსში შემავალი ტექსტების ავტორების ინგლისური ენის ცოდნის დონე (თვითშეფასების მიხედვით)



ენობრივი კომპეტენციის ევროპული ჩარჩოს მიხედვით, GLEAN-ში შემავალი ტექსტების ავტორების 72% თვითშეფასების საფუძველზე თვლის, რომ მათი ინგლისური ენის ფლობის ზღვრული განვითარების დონეზეა, 22% მიიჩნევს, რომ

ინგლისურ ენას ფლობს პროფესიულ დონეზე, ენის შემსწავლელთა 6%-ის აზრით, ზღვრულ დონეზე შეუძლია ინგლისური ენის გამოყენება, ხოლო არც ერთი მათგანი არ მიიჩნევს ენის გასაღებისა (A1) და ენის საფუძველის (A2) დონეზე ყოფნას.

წინამდებარე კვლევისას ვითვალისწინებთ, ერთი მხრივ, ტექსტობრივ ბაზაში შემავალი ყველა ინფორმაციის, ენის შემსწავლელთა, დავალებისა და გარემოს ცვლადების მნიშვნელობას, ხოლო, მეორე მხრივ, კორპუსის შექმნის დაბალანსებულობის, რეპრეზენტაციულობის, ჰომოგენურობისა და მიუკერძოებლობის პრინციპებს (სინკლერი, 2005: 20).

2.4. ინგლისური ენის ქართველ შემსწავლელთა კორპუსის POS-ანოტაცია

კორპუსლინგვისტიკაში ლინგვისტური ანოტაცია სხვაგვარად ხაზთაშორისი გლოსირების პრინციპადაც მოიაზრება. ანოტაცია არის წერილობითი ან ზეპირმეტყველების მონაცემებისათვის ლინგვისტური ინფორმაციის დართვა, რომელიც განმარტებითი ხასიათისაა და მას დამოუკიდებელ ბჭვარებზე (ანოტაციის სხვადასხვა დონეზე) ასახავს. ანოტაციის დონეები, იმავდროულად, წარმოადგენენ გრამატიკულად რელევანტურ დესკრიფციულ მონაცემებს (ლიჩი, 1997a: 2). ისინი ძირითადად ფორმალური ნიშნების მიხედვით ახასიათებენ ენობრივ მონაცემებს, სიტყვაფორმებს. ლინგვისტური ანოტაციის გარეშე კორპუსის ტექსტი უბრალოდ დაუმუშავებელი ტექსტია, ამის საპირისპიროდ, ანოტაცია ელექტრონული მონაცემის ლინგვისტური მეტაინფორმაციით გამდიდრების ერთ-ერთი საშუალებაა (თანდაშვილი და ფურცხვანიძე, 2014: 39). კორპუსში შემავალი მასალის ავტომატური დამუშავების დროს თეგირება შეიძლება მოიცავდეს სამ დონეს: მორფოლოგიურ, სინტაქსურ და სემანტიკურ ანალიზს. სასწავლო კორპუსიც, სხვა ტიპის კორპუსების მსგავსად შედგება შეგროვებული დაუმუშავებელი ელექტრონული ტექსტებისაგან, რომელთაც ანოტაციის საფუძურზე ლინგვისტური ღირებულება ემატება.

სასწავლო კორპუსის ანოტაციას და მის ანალიზს უცხოური ენის შესწავლის, ან სასწავლო ლექსიკოგრაფიის კვლევის ტრადიციული მეთოდებისაგან განსხვავებით, სამი უპირატესობა გააჩნია: ზომა, მასალის მრავალფეროვნება და მონაცემთა

ავტომატური დამუშავებით დაზოგილი დრო. შეცდომათა თეგირება (ანოტაცია შეცდომების გამოვლენისათვის) ანოტაციის ერთ-ერთი ყველაზე გავრცელებული ფორმაა სასწავლო კორპუსლინგვისტიკაში. თუმცა, ანალიზის სხვა ფორმები, როგორცაა, მაგალითად, მორფოლოგიური ან სემანტიკური ანოტაცია, კორპუსს უფრო ფასეულს ხდის (გრეინჯერი, 2004: 124).

დღესდღეობით, ანოტაციის ყველაზე გავრცელებული ფორმა არის POS-ანოტაცია, რომლის შემოტანამაც მონაცემების საერთაშორისო ასპარეზზე გამოყენებას ჩაუყარა საფუძველი. ნაშრომში ვიკვლევთ მორფოლოგიური ანოტაციით მიღებულ შედეგებს და ეს გადაწყვეტილება განპირობებულია შემდეგი მიზეზებით: POS-tagging საშუალებას გვაძლევს მეტყველების ნაწილების, ლექსიკური წყვილების, დასტებისა და კოლოკაციების კვლევის საშუალებას, რაც ინგლისურ-ქართული ლექსიკოგრაფიის განვითარებისთვის ხელსაყრელ პირობას შექმნის. ანოტირებისას შესაძლებელია წავაწყდეთ სისტემის ან ენის შემსწავლელთა ენის იდიოსინკრატული თავისებურებებისაგან გამოწვეულ ხარვეზებს. სასწავლო კორპუსებისათვის ყველაზე შედეგიანი თეგერის შერჩევისას, დავეფუძნეთ ბ.ვ. როისა და ლ. შაფერის (2003) კვლევას, რომლის მიხედვითაც ანოტაციის სამი ერთმანეთისაგან განსხვავებული სისტემის (TOSCA-ICLE, Brill, Claws) შედარების შედეგად, CLAWS თეგერს სიზუსტის 96,3% მიენიჭა (როი და შაფერი, 2003: 840).

ჩვენ მიერ შექმნილი ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსი (GLEAN) ანოტირებულია, რომელიც გვებმარება ლინგვისტური კვლევისთვის ენის სხვადასხვა დონეზე. ინგლისური ენის ქართველ შემსწავლელთა კორპუსში ყველა სიტყვაფორმას მიემართება მეტყველების შესაბამისი ნაწილის აღმნიშვნელი მარკერი. ანოტაციისთვის გამოვიყენეთ სტანდარტული თეგირების წყება CLAWS7. ამ თეგების ნაკრების მიხედვით, მაგ., არსებითი სახელი აღინიშნება მარკერით NN, ინფინიტივი – VVI, მაქვემდებარებელი კავშირები – CS, ზმნიზედა – RR, ზედსართავი სახელი – JJ, RRT - აღმატებითი ხარისხის ზმნიზედა და ა. შ. (UCREL, 1993). წარმოვადგენთ ნიმუშს ჩვენ მიერ POS-ანოტირებული სასწავლო კორპუსიდან (იხ. სურათი 4).

Freedom-loving_NN1 and_CC proud_JJ Georgian_JJ people_NN are_VBR not_XX accustomed_JJ to_TO surrender_VVI or_CC acceptance_NN1 of_IO the_AT unwanted_JJ ._.
With_IW our_APPGE new_JJ vision_NN1 ,_, ambitious_JJ ,_, wide-scale_JJ plans_NN2 and_CC projects_NN2 ,_, we_PPIS2 welcome_VV0 2023_MC with_IW greater_JJR daring_JJ in_II anticipation_NN1 of_IO a_AT1 better_JJR ,_, more_RGR successful_JJ ,_, much_RR more_RGR abundant_JJ and_CC peaceful_JJ year_NNT1 " " ,_, Garibashvili_NP1 noted_VVD ._.
In_II his_APPGE address_NN1 ,_, Garibashvili_NP1 congratulated_VVD " " heroic_JJ " " servicemembers_NN2 ,_, police_NN2 officers_NN2 ,_, firefighters_NN2 ,_, doctors_NN2 and_CC teachers_NN2 ,_, as_II31 well_II32 as_II33 sportspersons_NN2 ,_, civil_JJ servants_NN2 ,_, representatives_NN2 of_IO arts_NN2 and_CC culture_NN1 ,_, veterans_NN2 and_CC Georgia_NP1 's_GE " " distinguished_JJ " " elderly_JJ generation_NN1 ._.
He_PPHS1 also_RR wished_VVD Happy_JJ New_JJ Year_NNT1 to_II the_AT Georgian_JJ diaspora_NN1 . .

მაშასადამე, თითოეულ სიტყვაფორმას მიემართება შესაბამისი მეტყველების ნაწილის აღმნიშვნელი მარკერი, ხოლო ამ უკანასკნელის სასწავლო კორპუსში ცალკეულად ან ერთმანეთთან კომბინაციაში მოძიებისას, ყველა არსებულ შემთხვევას მოვიძიებთ, მაგალითისათვის, ჩვენს სასწავლო კორპუსში თეგების აღნიშნული მარკერების ძიებისას *JJ JJ _NN1* შედეგად მივიღებთ *ზედსართავი სახელი + ზედსართავი სახელი + არსებითი სახელი* კომბინაციას (იხ. სურათი 5).

სურათი 5. *_JJ _JJ _NN1* თეგის კომბინაციის შედეგი ინგლისური ენის ქართველ შემსწავლელთა კორპუსში

gning the Association Agreement and for	<u>initial Free Trade</u>	Agreement . The President-elect Marg
ake it stronger and closer . Italy supports	<u>Georgian European orientation</u>	. After a consultation with the Italian a
Murat Burhan . " Turkey always supports	<u>Georgian territorial integrity</u>	and its European integration . Ministry
as and Legal Assistance , considering the	<u>high public interest</u>	, general inspection of the Ministry sta
ip launched in May 2009 with the aim of	<u>accelerating political association</u>	and deepening economic integration w
of accelerating political association and	<u>deepening economic integration</u>	with six Eastern partner countries : Ar
in the first half of 2014 , according to the	<u>Georgian Foreign Minister</u>	Maia Panjikidze . Panjikidze did not sp
vardnadze , former First Secretary of the	<u>Georgian Communist Party</u>	, was appointed as chairman of the Geo
ter . The reason for death is still unclear .	<u>Intensive investigative work</u>	has been ongoing during the last perio
uded the premiere of a new project at the	<u>Georgian National Ballet</u>	Sukhishvili with his family . The prem
Ramishvili is an innovative project of the	<u>Georgian National Ballet</u>	Sukhishvili which is a modernist impro

2.5. შესიტყვებების კვლევის ასოციაციური საზომები

ლინგვისტური დასკვნის სისწორის თეორიული შესაძლებლობა იზრდება ცდაში მოქცეული ენობრივი მონაცემების ზრდასთან ერთად. ამგვარი სიგნიფიკანტური ზრდის ერთადერთი ველი კი იქმნება ენობრივი კორპუსის ფარგლებში. შესიტყვებების მარტივი სიხშირით ძებნა ჩავანაცვლოთ ძებნით, რომელიც ითვალისწინებს ასოციაციურ საზომებს, რადგან სასწავლო კორპუსებში მხოლოდ შესიტყვებების სიხშირის დადგენა მისი ენაში რეალიზების შესწავლისთვის საკმარისი არ არის. არსებობს დამატებითი საზომები, რომლებიც კოლოკაციურ ერთეულებს შორის არსებულ კავშირებზე გვაწვდიან ინფორმაციას. როგორც უკვე განვსაზღვრეთ, კორპუსლინგვისტიკაში სიტყვები კოლოკაციებად, ანუ შესიტყვებებად მიიჩნევიან, როდესაც ისინი ტექსტობრივ მონაცემებში ერთმანეთის კომბინაციაში გვხვდებიან. „მითხარი რა წრეში ტრიალებს სიტყვა და გეტყვი, რა არის ის“ (ფირსი, 1957: 11) დღემდე ამ კოლოკაციური კავშირების არაოფიციალურ დეფინიციას წარმოადგენს. შესიტყვებების კორპუსზე დამყარებული სტატისტიკური განმარტება, რომელიც აგრეთვე ლექსიკოგრაფიულ კვლევებში გამოიყენება, ტრადიციულად მოიცავს კოლოკაციურობის ორ მთავარ კრიტერიუმს: აბსოლუტურ სიხშირესა და სიტყვათა კომბინაციებს შორის არსებულ ასოციაციებს.

აბსოლუტური (დაუმუშავებელი) სიხშირე ტექსტკორპუსში არსებულ ენობრივ ერთეულებს ითვლის და წარმოაჩენს ენობრივ დინამიკურობას. თუმცა, კორპუსლინგვისტიკაში მონაცემთა ანალიზისათვის მხოლოდ აბსოლუტური სიხშირის გამოყენება მართებული არაა რამდენიმე მიზეზის გათვალისწინებით. უპირველეს ყოვლისა, კონტრასტული ანალიზის დროს ორი სხვადასხვა მოცულობის კორპუსის შედარება აბსოლუტური (დაუმუშავებელი) სიხშირის საფუძველზე ალოგიკურია და მცდარ დასკვნამდე მიგვიყვანს (ჰოფმანი და სხვ., 2008: 70). მაგალითად განვიხილოთ შემდეგი შემთხვევა: *happy end* კოლოკაციური წყვილი 100 მილიონიან კორპუსში თუ 50-ჯერ დასტურდება და იგივე კოლოკაციური წყვილი 1 მილიონიან კორპუსში 50-ჯერ გვხვდება, შეცდომა იქნება დავასკვნათ, რომ კონტრასტულმა ანალიზმა *happy end* შესიტყვების თანაბარი რაოდენობა გამოავლინა

ორივე კორპუსში. ამიტომ სწორი შედეგის მისაღებად საჭიროა ნორმალიზებული სიხშირის (normalised frequency) დადგენა და დასკვნის შემდგომ გამოტანა.

ვინაიდან სასწავლო ტიპის კორპუსები მოიცავენ არამშობლიურ ენაზე მოსაუბრეთა მასალას, მხედველობაში უნდა მივიღოთ უცხოური ენის შემსწავლელთა ენობრივი კომპეტენციის საკითხი, მათი ლექსიკური უნიკალურობა, შესიტყვებების კვლევისათვის მნიშვნელოვანია სიღრმისეული სტატისტიკური ანალიზის წარმოება. აბსოლუტური სიხშირე ეფექტურად განსაზღვრავს ლექსიკური ერთეულების განმეორებადობას ტექსტკორპუსში. თუმცა, მისი გამოყენება ენობრივი კანონზომიერების დადგენასა და სიტყვაფორმების წინასწარ პროგნოზირების შესაძლებლობას არ გვაძლევს.

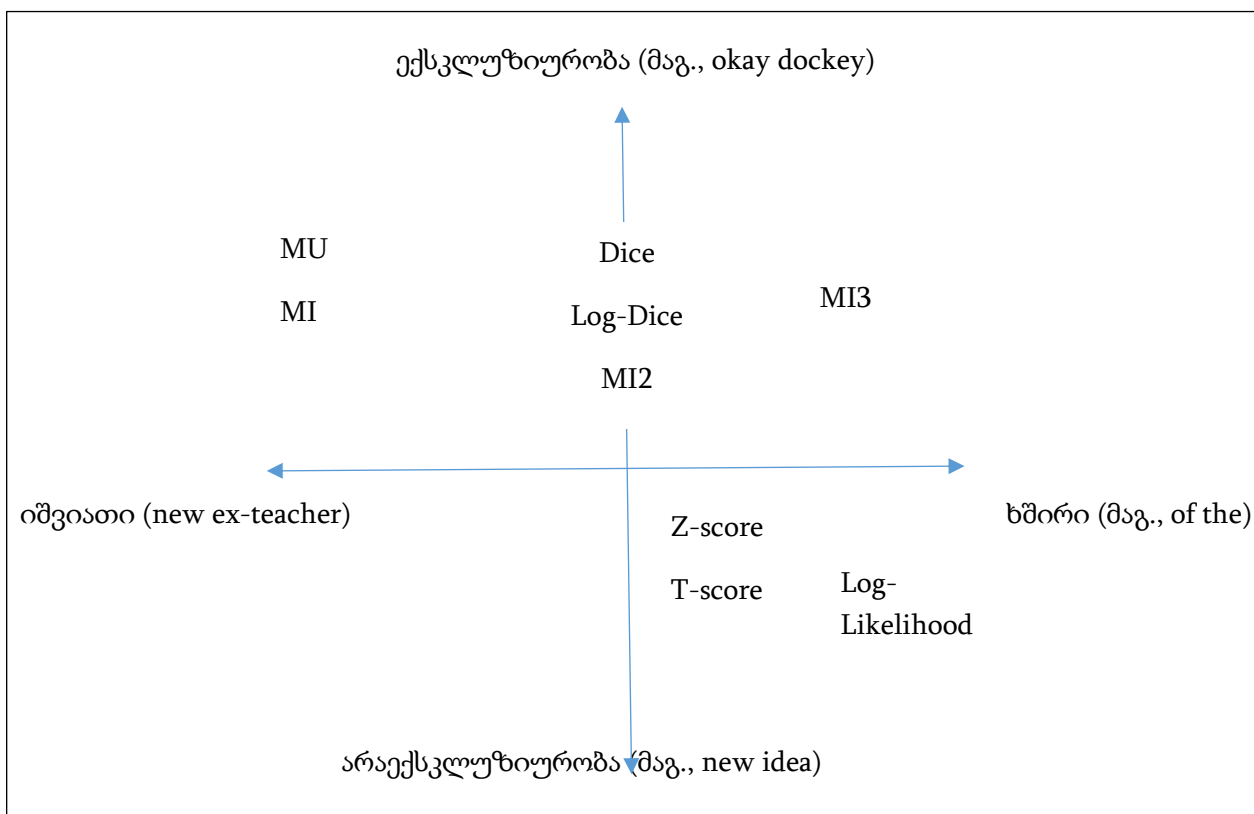
კორპუსლინგვისტიკის კვლევები აჩვენებს, რომ ზოგჯერ საკმაოდ ხშირი სიტყვათა თანხვედრის შემთხვევებიც კი შესაძლოა წარმოებული იყოს ენის ძალიან შეზღუდულ, კონკრეტულ კონტექსტში. მაგალითისათვის, ბრიტანულ ნაციონალურ კორპუსში (BNC) შესიტყვებები risk issues და moral issues თითქმის მსგავსი აბსოლუტური სიხშირე აქვთ მინიჭებული (54 და 51 შემთხვევა, შესაბამისად). თუმცა ამ შემთხვევათა უმრავლესობა ფიქსირდება ერთ კონკრეტულ ტექსტურ ნიმუშში. აქედან გამომდინარე, ლექსიკური ერთეულების გადანაწილების თავისებურება სტატისტიკურად სურათს ცვლის და რეალობას ნაკლები სიზუსტით ასახავს.

შესიტყვებების ცოდნისა და გამოყენების სიხშირის ნათელსაყოფად, ვ. ბრეზინა გამოჰყოფს სამ განზომილებას: გავრცელებულობა, ექსკლუზიურობა, ურთიერთმიმართულებითობა. პირველი მათგანი – კოლოკაციათა გავრცელებულობა, გამოიყენება ენობრივი რეგულარულობის გამოსავლენად, რომელიც გამოხატავს რამდენად (არა)თანაბარზომიერად არის ეს თვისება კორპუსში გამოკვეთილი. მეორე განზომილება მოიცავს შესიტყვებების ექსკლუზიურობას, ანუ როგორი განფენილობისაა, რა მასშტაბით გვხვდება სიტყვაფორმები ტექსტკორპუსში: დამოუკიდებელ წევრებად თუ მეტწილად ერთმანეთთან მიმართებით? ექსკლუზიურობა ის მნიშვნელოვანი ფაქტორია, რომელიც შესიტყვებების თანხვედრის წინასწარ განჭვრეტის შესაძლებლობას იძლევა. მაგალითად, შესიტყვებები zig zag, okey dokey, annus mirabilis ექსკლუზიურად ერთმანეთთან ასოცირებულები არიან, ანუ მარტივად ამოცნობადია რა ენობრივი ერთეული

შეიძლება მოჰყვეს zig-ს, okey-ს და ა.შ. შესიტყვებების შესწავლის მესამე განზომილება მოიცავს ურთიერთმიმართულებითობას, რომელიც ცხადყოფს იმ ფაქტს, რომ შესიტყვებაში შემავალი კომპონენტები თანაბარი სიმდიერით არ იზიდავენ ერთმანეთს და ეს მიზიდულობა ხშირად ასიმეტრიულია, მაგალითად, შესიტყვებაში Christmas decorations, მიზიდულობა decorations-დან Christmas-მდე უფრო ძლიერია, ვიდრე პირიქით, ამ ვარაუდს ამტკიცებს ის ფაქტი, რომ BNC კორპუსში decorations შემთხვევათა 11%-ში, Christmas სიტყვა მას წინ უძღვის (ბრეზინა, 2012: 71-72).

შესიტყვებების იდენტიფიკაციისა და ასოციაციური საზომების გამოყენებისათვის აღნიშნული განზომილებები მხედველობაში უნდა იქნეს მიღებული. შესიტყვებების კვლევისათვის ვიყენებთ სტატისტიკურ საზომებს, რომლებიც სიტყვებს შორის არსებულ ასოციაციურ სიმდიერეს ითვლიან. თანამედროვე კორპუსლინგვისტიკურ კვლევებში სტატისტიკურ მეტრიკებს t-score და MI-score დომინანტური პოზიცია უკავიათ (შმიტი, 2012: 6). თუმცა არ არსებობს უნივერსალური ასოციაციური საზომი, რომელიც ყველა საკვლევ საკითხს თანაბარჯეროვნად მოერგება. კვლევისათვის რომელი კოლოკაციური საზომის გამოყენება არის საუკეთესო, ამ კითხვაზე პასუხის გასაცემად გასათვალისწინებელია კოლოკაციური ურთიერთმიმართების რომელი ასპექტის კვლევა შედის ინტერესის სფეროში. ძირითად შემთხვევებში მათი შერჩევა პირობითია.

სიხშირის, ურთიერთმიმართებულობისა და ექსკლუზიურობის პრინციპებზე დაყრდნობით, ვ. ბრეზინა წარმოგვიდგენს შესაბამის ასოციაციურ საზომებს, რომლის გამოყენებაც შესიტყვებებს შორის არსებულ კავშირებს განსაზღვრავს (ბრეზინა, 2018: 71). MI ასოციაციური საზომი ძირითადად გათვლილია შესიტყვებების იშვიათ ექსკლუზიურობაზე, ასეთი შემთხვევა კორპუსში შესაძლოა იყოს ერთი ან თუნდაც ორი. სხვა მეტრიკა, როგორცაა Dice, log Dice, MI2 ანგარიშობს ისეთ შესიტყვებებს, რომლებიც მხოლოდ ერთმანეთთან თანაარსებობენ, თუმცა იშვიათი რაოდენობით არ არიან გამოვლინებულნი, იხილეთ სქემა 9:



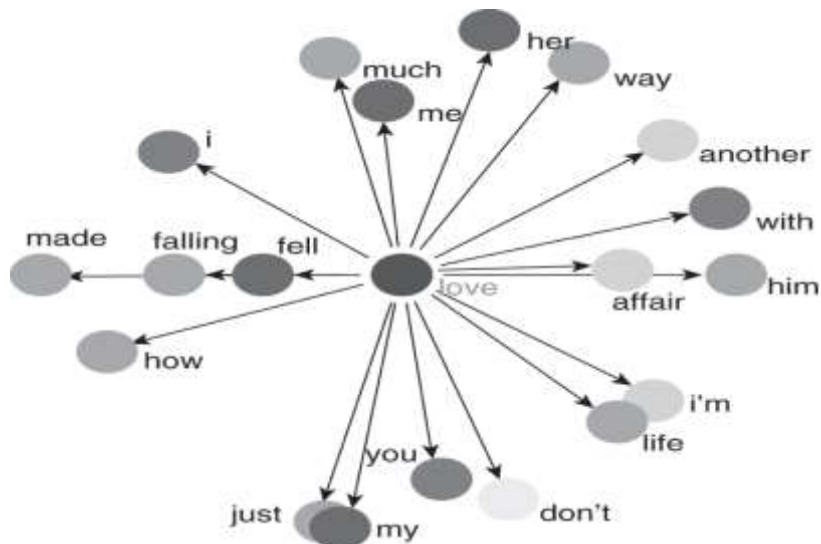
სიხშირის თვალსაზრისით, Log-likelihood და Mutual information (MU) ორი უკიდურესი საზომი ერთეულებია. Z-score მათ შორის არსებულ პოლარულობას აბალანსებს ისევე, როგორც Dice კოეფიციენტი, რომელიც მსგავსი გამაწონასწორებელი ბუნებით ხასიათდება. Log-likelihood საზომი სრულად მნიშვნელობაზე ორიენტირებული ასოციაციური საზომია, რომელიც მაღალ ქულას არ მიაწიჭებს არც შემთხვევით კომბინაციებს მათი მაღალი თანხვედრის სიხშირის მიუხედავად, არც იმ სიტყვათა მიმდევრობას, რომლებშიც ძალიან იშვიათი სიტყვებია ჩართული. იგი უგულებელყოფს ეფექტის ზომის, დამოუკიდებელი ცვლადის ეფექტის დაკვირვებული ობიექტური და ხშირად სტანდარტიზებული სიდიდის კრიტერიუმს, ხოლო ამ დანაკლისს ავსებს MI ასოციაციური საზომი, რომელიც ეფექტურად ფილტრავს ტრივიალურ (თუმცა სტატისტიკურად მნიშვნელოვან) თანმდევ შემთხვევებს (ჰოფმანი და სხვ., 2008: 157).

შესიტყვებების კვლევის სისტემატური და გამჭვირვალე შედეგების მისაღწევად, სასწავლო კორპუსის მონაცემების დამუშავებისას მნიშვნელოვანია შესიტყვებების გრაფებისა და ქსელების ვიზუალიზაცია. კოლოკაციური გრაფები და

ქსელები საკვანძო სიტყვასა და მის კოლოკატებს შორის არსებულ კავშირებს ვიზუალურად წარმოადგენენ. შესიტყვებების სიის სტანდარტულად წარმოდგენის ნაცვლად, გრაფა ლექსიკურ ერთეულებს შორის დამყარებულ კავშირებს სიძლიერის, სიხშირისა და პოზიციის მიხედვით წარმოაჩენს, ხოლო შესიტყვებების ქსელები უფრო მასშტაბური და ფართო სისტემის გრაფებია, რომლებიც ასახავენ უფრო მეტ ასოციაციურ კანონზომიერებას.

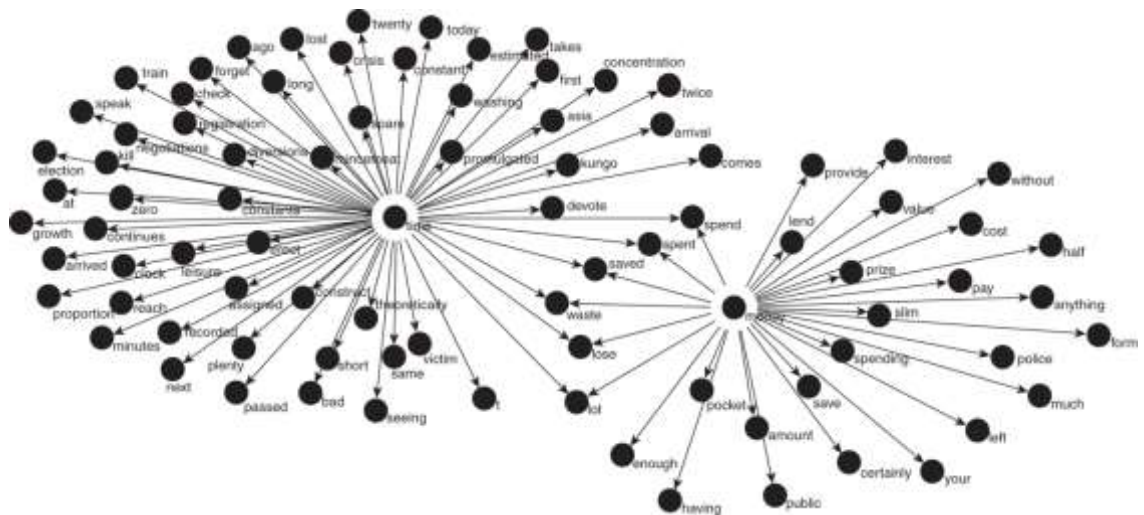
პირველ რიგში, ნიმუშად წარმოვადგენთ საკვანძო სიტყვის *love* შესიტყვების გრაფას, რომელშიც ლექსიკურ ერთეულს შორის არსებული ურთიერთკავშირი ასახულია ვ. ბრეზინას მიერ შემოთავაზებული სამი ძირითადი განზომილების მიხედვით: ა) ასოციაციური სიძლიერე, რომელიც გამოიხატება კოლოკატებსა და სიტყვა „love” შორის არსებული მანძილებით. რაც უფრო ახლოსაა შესიტყვების ერთეული საკვანძო სიტყვასთან, მით უფრო ძლიერია ასოციაციური კავშირი (მაგნიტის პრინციპით); ბ) შესიტყვების სიხშირე, რაც ვიზუალიზებულია წრის ფერის მუქი ან ღია ტონით. რაც უფრო ინტენსიური, მუქია კოლოკატი, მით უფრო ხშირად გვხვდება ის კორპუსში; გ) შესიტყვების პოზიცია ტექსტში, ანუ კოლოკატი საკვანძო სიტყვის წინ გვხვდება თუ მის შემდეგ გამოსახულია გრაფაში წერტილების პოზიციით: მარცხნივ, შუაში, ან მარჯვნივ (ბრეზინა, 2018: 76). მაგალითად, *my* კოლოკატი გვხვდება საკვანძო სიტყვის მარცხნივ, რაც გულისხმობს იმას რომ სახეზეა შესიტყვება *my love*, ხოლო *life* გვხვდება საკვანძო სიტყვის მარჯვნივ, რაც ავტომატურად მიუთითებს *love life* შესიტყვების არსებობაზე (იხ. სურათი 6).

სურათი 6. შესიტყვების გრაფა *love* (ბრეზინა, 2018: 76)



მეორე მხრივ, შესიტყვებების ქსელები, გრაფებთან შედარებით კოლოკატების რეალიზების უფრო ფართო ფართომასშტაბურ სურათს წარმოადგენენ. სამიეზო სიტყვის შორეული კოლოკატიც კი მონაწილეობენ ამ სისტემაში საერთო კონცეპტუალური სივრცის გაზიარებით. მაგალითისათვის, ორი ლექსიკური ერთეული money და time, რომლებიც პირდაპირი გზით შესიტყვებას არ ქმნიან, კოლოკაციური ქსელის მეშვეობით და მათ შრის არსებული საერთო ასოციაციური სიტყვების დახმარებით, როგორებიცაა spend, save, waste, lose, მათ შორის კავშირები ამდაგვარად მყარდება (იხ. სურათი 7):

სურათი 7. საკვანძო სიტყვების *time* და *money* საერთო შესიტყვებების (ბრეზინა, 2012: 78)



შესიტყვების გრაფები და ქსელები ტექსტოროპუსებში სიტყვათა შორის არსებულ სემანტიკურ კავშირებს წარმოაჩენენ. კოლოკაციური გრაფებისა და ქსელების შექმნისათვის საჭიროა პროგრამა #Lancsbox (ბრეზინა და სხვ., 2015), რომელიც უზრუნველყოფს შესიტყვებათა თანხვედრის შედარებებსა და ანალიზს ვიზუალური ფორმით. შესიტყვებების ასოციაციური საზომების, შესიტყვებების გრაფებისა და ქსელების მახასიათებლები შესაძლებელია შემდეგი დასკვნების სახით შევაჯამოთ:

- შესიტყვებების კორპუსზე დამყარებული კვლევის წარმართვისათვის ვითვალისწინებთ ვ. ბრეზინას მიერ შემოთავაზებულ განზომილებებს გავრცელებულობას, ექსკლუზიურობას, მიმართულებითობას.

- ასოციაციური საზომების შერჩევისას ვისარგებლებთ შესიტყვებების ასოციაციური საზომების მრავალფეროვნებით (მაგ., log-likelihood, MI, MI3, log Dice, Z-score და ა.შ.), რათა მივიღოთ სისტემატური და გამჭვირვალე შედეგები.
- სასწავლო კორპუსში შესიტყვებაში შემავალ ელემენტებს შორის ურთიერთკავშირი ასახულია სამი ძირითადი პარამეტრის მიხედვით: ასოციაციური სიძლიერე, შესიტყვების სიხშირე, შესიტყვების სიძლიერე.
- კოლოკაციური გრაფები და ქსელები საკვანძო სიტყვასა და მის კოლოკატებს შორის არსებული კავშირებს ვიზუალურად წარმოადგენენ. გრაფა ლექსიკურ ერთეულებს შორის დამყარებულ კავშირებს სიძლიერის, სიხშირისა და პოზიციის მიხედვით წარმოაჩენს, ხოლო შესიტყვებების ქსელები უფრო მასშტაბური და ფართო სისტემის გრაფებია, რომლებიც ასახავენ უფრო მეტ ასოციაციურ კანონზომიერებას.

2.6. კონტრასტული ანალიზისთვის შერჩეული სასწავლო კორპუსები

სასწავლო კორპუსების კვლევის მეთოდოლოგიაში კონტრასტული ინტერენობრივი ანალიზი (CIA, *Contrastive Interlanguage Analysis*) უცხოური ან მეორე ენის შემსწავლელთა ლინგვისტურ თავისებურებებზე დასაკვირვებლად ერთ-ერთი ყველაზე გავრცელებული მეთოდია. კონტრასტული ანალიზი ორი, შესაძლოა, განხორციელდეს ორი მიმართულებით: 1. როდესაც ერთმანეთს ვადარებთ უცხოური ენის ტექსტობრივ მონაცემებს და მისი ავტორები სხვადასხვა მშობლიურ ენაზე მოსაუბრეები არიან, მაგალითად, გერმანული ენის შემსწავლელები, რომელთაც განსხვავებული ენობრივი ფონი აქვთ (მაგ., მათი მშობლიური ენა არის პოლონური, ფრანგული, ჩინური და სხვ.). 2. ისეთ კონტრასტული ანალიზი, რომელიც ერთმანეთს ადარებს მშობლიურ ენაზე მოსაუბრეთა ტექსტობრივ მონაცემებს იმ ენის შემსწავლელების ენობრივ მასალას. მაგალითად, პირობითად ბრიტანელი სტუდენტების ესეები შეგვიძლია შევადაროთ ინგლისური ენის შემსწავლელი შვედი სტუდენტების ესეებს (გრეინჯერი, 1996).

კონტრასტული ანალიზის აუცილებელ პირობას არ წარმოადგენს შესადარებელი კორპუსების განსხვავებული მოცულობა, რადგან კვლევის ანალიზისას ნორმალიზებულ სიხშირესა და ასოციაციურ საზომებს ვიყენებთ. ჩვენს სადისერტაციო ნაშრომში ასახული კვლევა კონტრასტულ ინტერენობრივი ანალიზის მეთოდოლოგიას ეფუძნება. ერთი მხრივ, შევადარებთ ინგლისური ქართველი შემსწავლელების ენობრივ პროდუქტს ინგლისური ენის აზიელ შემსწავლელთა საერთაშორისო კორპუსს (International Corpus Network of Asian Learners of English, ICNALE), ხოლო, მეორე მხრივ, „საკონტროლო კორპუსად“ შერჩეული გვაქვს მიჩიგანის უნივერსიტეტის ამერიკელი სტუდენტების (Michigan Corpus of Upper-level Student Papers, MICUSP) ინგლისურენოვანი ნაწერების ტექსტები.

2.6.1. ინგლისური ენის აზიელ შემსწავლელთა საერთაშორისო კორპუსი (ICNALE)

კონტრასტული ანალიზისათვის შერჩეული ინგლისური ენის შემსწავლელთა საერთაშორისო კორპუსზე შეჩერებული არჩევანი რამდენიმე მიზეზითაა განპირობებული. აღნიშნული სასწავლო კორპუსი წლების განმავლობაში დაგროვებული გამოცდილებითა და მასშტაბურობით გამოირჩევა და დიდი საერთაშორისო პოპულაციის ინგლისური ენის ცოდნის ამსახველია. ბაზაში თავმოყრილია 10 აზიური ქვეყნის ინგლისური ენის შემსწავლელების ზეპირმეტყველებისა და წერილობითი მონაცემების 1,5 მილიონამდე სიტყვა. ეს ქვეყნებია: ჩინეთი, ჰონკონგი, ინდონეზია, იაპონია, კორეა, პაკისტანი, ფილიპინები, სინგაპური, მალაიზია, ტაივანი და ტაილანდი. საერთაშორისო სასწავლო კორპუსი 4 მოდულისაგან შედგება: წარმოთქმული მონოლოგი, წარმოთქმული დიალოგი, წერილობითი ესეები და რედაქტირებული ნარკვევები.

კონტრასტული ანალიზისათვის ჩამოთვლილთაგან ვიყენეთ ესეების ქვეკორპუსს, რადგან ჩვენ მიერ შექმნილი სასწავლო კორპუსი ზეპირმეტყველების მონაცემს არ მოიცავს. მიუხედავად იმისა, რომ ICNALE კორპუსული მასალა ჟანრული მრავალფეროვნებით არ გამოირჩევა, მასში რამდენიმე დარგის: ჰუმანიტარული, სოციალური, ტექნოლოგიური და სიცოცხლის მეცნიერებების წერილობითი ნამუშევარია თავმოყრილი.

შემსწავლელთა მეტამონაცემების, სხვადასხვა ცვლადების თანმიმდევრულობა აღნიშნულ სასწავლო კორპუსს კონტრასტული ანალიზისათვის შესაფერის კორპუსად აქცევს. ბოლოს, ინგლისური ენის შემსწავლელთა საერთაშორისო კორპუსი იმავე, ლანკასტერის უნივერსიტეტის CQPweb ბაზაშია ატვირთული, რომელშიც ჩვენი სასწავლო კორპუსი და ეს ფაქტორი შედარებით ანალიზს მნიშვნელოვნად ამარტივებს.

სასწავლო კორპუსის უწინდელი კვლევები ძირითადად ევროპელ შემსწავლელებზე იყო ორიენტირებული. თუმცა ეკონომიკურმა, სოციოკულტურულმა და ენობრივმა გლობალიზაციამ აზიის კონტინენტზე ინგლისური ენის შემსწავლელთა რიცხვი გაზარდა. ინგლისური ენის აზიელ შემსწავლელებისთვის უცხოური ენის ათვისების ფაქტორები, მაგალითად, ინგლისური ენის რეალიზების შესაძლებლობებით, მოტივაციით, ლინგვისტური საჭროებებით ევროპისაგან სავსებით განსხვავებულია (იშიკავა, 2014: 95).

ICNALE კორპუსის დავალების ცვლადები მოიცავს ორ პრინციპს ეფუძნება. ერთი კატეგორიის ესეები აბსოლუტურად არაკონტროლირებადია, თავისუფალ სივრცეშია შესრულებული, დამხმარე საშუალებების გამოყენებით. მეორე კი მკაცრად კონტროლირებადია კორპუსის ჰომოგენურობის მაღალი დონის მიღწევის მიზნით. მაგალითად, ენის შემსწავლელებისადმი დარიგებულ საინსტრუქციო ფურცელზე გაწერილი იყო დავალების შესრულების წესები და პირობები („არ გადაიწეროთ, არ გამოიყენოთ ლექსიკონები, გადაიკითხეთ ნაწერები დასრულების შემდეგ“). ამ მხრივ, დავალების ცვლადების პოლარულობა მსგავსია ჩვენი სასწავლო კორპუსის დავალებებისა. ერთ მხრივ, GLEAN-ში გვაქვს სტუდენტთა კონტროლირებად გარემოში შესრულებული ესეები, მეორე მხრივ, კორესპონდენცია და საგაზეთო სტატიები თავისუფალ, ჩაურეველ გარემოში შესრულდა.

2.6.2. მიჩიგანის უნივერსიტეტის ამერიკელი სტუდენტების ესეების სასწავლო კორპუსი (MICUSP)

მიჩიგანის ამერიკელი სტუდენტების ესეების სასწავლო კორპუსის კონტრასტული ანალიზისათვის შერჩევა განპირობებულია ორი მიზეზით.

უპირველეს ყოვლისა, ვინაიდან მასში გაერთიანებულია ამერიკელ სტუდენტთა (ვისთვისაც ინგლისური ენა მშობლიურია) წერილობითი ნამუშევრები, ჩვენთვის მათ მშობლიურენოვან სტრუქტურებზე დაკვირვება და ქართველი შემსწავლელების ენობრივ პროდუქტთან შედარება მნიშვნელოვანია. MICUSP პირველი სასწავლო კორპუსი ჩრდილოეთ ამერიკის კონტინენტზე და მოიცავს 3 მილიონამდე სიტყვას, 830 წერილობით ნამუშევარს. კორპუსის შექმნის მიზანი კორპუსლინგვისტიკის მკვლევრებისა და საგანმანათლებლო სექტორში დასაქმებულთა დახმარების სურვილია და საკვლევი დიაპაზონი მოიცავს როგორც ლექსიკური თავისებურებებისა და წერილობითი დისკურსის კვლევას, ისე ვიწროდარგობრივი ტერმინოლოგიის შესწავლას ამერიკელი სტუდენტების ესეების მაგალითზე.

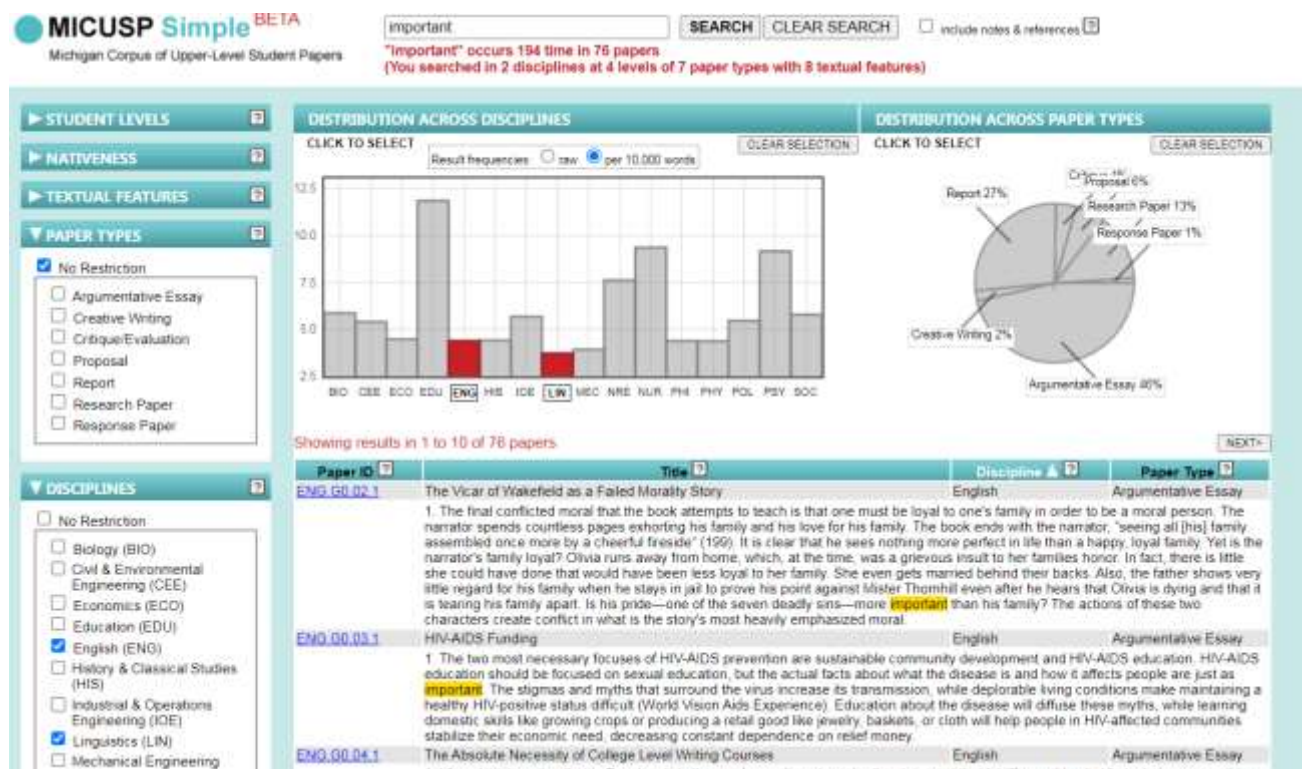
ამავდროულად, სიტყვაფორმების ძიების სიმარტივე და ხელმისაწვდომობა ერთ-ერთი განმაპირობებელი ფაქტორი გახდა MICUSP სასწავლო კორპუსის შედარებითი ანალიზისათვის შერჩევისას. ბაზაში ერთიანდება სხვადასხვა ჟანრისა და დისციპლინის მასალა. ეს დარგებია: ბიოლოგია, სამოქალაქო მშენებლობა, ეკონომიკა, განათლება, ინგლისური ფილოლოგია, ისტორია, ინდუსტრიული ინჟინერია, ლინგვისტიკა, მანქანათმშენებლობა, ბუნებრივი რესურსები და გარემო, საექთნო საქმე, ფილოსოფია, ფიზიკა, პოლიტიკური მეცნიერებები, ფსიქოლოგია, სოციოლოგია. წერილობითი მასალის ტიპებში გაერთიანებულია არგუმენტირებული ესე, შემოქმედებითი წერა, კრიტიკა/შეფასება, ანგარიში, კვლევითი სტატია, კორესპონდენცია. ყველა მათგანი შესრულებულია ბაკალავრიატის საფეხურის მეოთხე კურსის ან მაგისტრატურის საფეხურის სტუდენტების მიერ, ვისი ნაშრომიც ფრიადზე (A) შეფასდა.

კონტრასტული ანალიზისას თანაბარზომიერობის შესანარჩუნებლად MICUSP სასწავლო კორპუსიდან შევარჩევთ მხოლოდ იმ დისციპლინებს, რომელიც ჩვენს მიერ შექმნილ სასწავლო კორპუსს შეესაბამება, ესენია ინგლისური ფილოლოგია და ლინგვისტიკა, ხოლო მასალიდან გამოვიკვლევთ ყველა ტექსტურ ტიპს: არგუმენტირებული ესე, შემოქმედებითი წერა, კრიტიკა/შეფასება, ანგარიში, კვლევითი სტატია, კორესპონდენცია.

საძიებო ველში შესაძლებელი არაა ცალკეული მეტყველების ნაწილის მოძიება, ამიტომ საკვლევ მასალად შევარჩევთ მხოლოდ იმ შესიტყვებებსა და სამ-

ოთხკომპონენტთან ლექსემებს, რომლებიც ინგლისური ენის ქართველ შემსწავლელთა კორპუსში დასტურდება. ბაზაში რაოდენობრივი ანალიზისას შედეგად ვიღებთ როგორც აბსოლუტურ, ნედლ სიხშირეს, ისე ნორმალიზებულს (per 10, 000). MICUSP და GLEAN სასწავლო კორპუსები განსხვავებული მოცულობისაა და შედეგების შესადარებლად მართებულია ნორმალიზებული სიხშირის კვლევა (იხ. სურათი 9).

სურათი 8. MICUSP სასწავლო კორპუსის საძიებო სისტემა (Michigan Corpus of Upper-level Student Papers, 2009. <https://elicorpora.info/main>)



ბაზაში რამდენიმე კომპონენტისაგან შედგენილი ლექსიკური ერთეულების თვისობრივი ანალიზი მათი კონტექსტუალური რეალიზების კვლევით არის შესაძლებელი. ბაზაში ერთ-ერთ ველში მითითებულია ნამუშევრის კოდი, სათაური, დისციპლინა და ტექსტის ტიპი, სადაც ესა თუ ის საძიებო სიტყვა დასტურდება.

2.7. შუალედური დასკვნა

წინამდებარე თავის მიზანი იყო წარმოგვედგინა ჩვენ მიერ შექმნილი ინგლისური ენის ქართველ შემსწავლელთა კორპუსში შემავალი ტექსტების შერჩევის კრიტერიუმები, ჟანრულ კლასიფიკაცია და მონაცემთა მოგროვების დეტალები.

აგრეთვე, კონტრასტული ანალიზისათვის საერთაშორისო სასწავლო კორპუსების შერჩევის კრიტერიუმები. თავში განვიხილეთ ჯ. სინკლერის კორპუსის შექმნის ფუნდამენტური პრინციპების: დაბალანსებულობის, რეპრეზენტაციულობის, ჰომოგენურობისა და მიუკერძოებლობის მნიშვნელობა და მათი გათვალისწინება GLEAN სასწავლო კორპუსში.

მასალის შერჩევის ყველაზე მნიშვნელოვანი საფუძველი იყო, მოგვეგროვებინა მაქსიმალურად დიდი მოცულობის ტექსტობრივი მონაცემები, რომელიც უფრო ნათლად გამოაშკარავებდა ინგლისური ენის ქართველ შემსწავლელთა ენობრივ თავისებურებებს. როგორც აღვნიშნეთ, კორპუსლინგვისტიკაში მასალის სიუხვე პირდაპირპროპორციულია კვლევის შედეგების მასშტაბურობასთან. გადაწყვეტილების მიღებისას, აღმოჩნდა, რომ მხოლოდ სტუდენტთა ესეების შეგროვება გაცილებით მცირე რესურსიანი სასწავლო კორპუსის შექმნის შესაძლებლობას მოგვცემდა. ამიტომ, საჭიროდ ჩავთვალეთ, გაგვეზიარებინა, French Interlanguage Database (FRIDA) სასწავლო კორპუსის გამოცდილება, რომელიც არა მარტო სტუდენტთა ნაწერებს, არამედ პრესის მასალაც მოიცავს. მამასადამე, ჟანრული განაწილების თვალსაზრისით ინგლისური ენის ქართველ შემსწავლელთა კორპუსი სპეციალიზებული ხასიათისაა და მოიცავს დარგობრივი და ჟანრული მრავალფეროვნების მქონე ტექსტებს. სასწავლო კორპუსში გამოვყავით ჟანრთა კატეგორიები და ქვეკატეგორიები. მასალა განეკუთვნება აკადემიურ, ესეისტიკის, პუბლიცისტურ და სასაუბრო ჟანრებს.

როგორც აღვწერეთ, ესეებისა და ელექტრონული მეილების მასალის შეგროვებისათვის კვლევაში მონაწილეობა მიიღეს ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის, ინგლისური ფილოლოგიის მიმართულების ბაკალავრიატისა და მაგისტრატურის საფეხურის სტუდენტებმა. კორესპონდენციის მასალაში ერთიანდება ინგლისურენოვანი ელექტრონული მეილები (საქმიანი, არაოფიციალური და ა.შ.), რომლებიც სტუდენტებმა უკანასკნელი ერთი წლის განმავლობაში გაგზავნეს. მეილებიდან გაფილტრულ იქნა ადრესატისა და გამგზავნის შესახებ არსებული მეტაინფორმაცია. პრესის მასალის წყაროებად გამოყენებულია: ა) ახალი ამბების ონლაინ პლატფორმა agenda.ge, რომელშიც თავს იყრის როგორც პოლიტიკური, ისე ტურისტული, ჯანდაცვის, საგანმანათლებლო,

სპორტული და სასოფლო-სამეურნეო თემატიკის სტატიები; ბ) ქართველ სტუდენტთა ინგლისურენოვანი ბლოგები სამოგზაურო თავგადასავლებისა და პირადი გამოცდილებების შესახებ, აგრეთვე თავმოყრილია მიმოხილვები. შეგროვებული მასალა კომუნიკაციური ფუნქციის მატარებელია და იცავს მიუკერძოებლობის პრინციპს. ჰომოგენურობის პრინციპის მიხედვით კორპუსში არ უნდა იქნეს ისეთი უჩვეულო და საექვო ტექსტები შეტანილი, რომლებიც თავიანთ კატეგორიაში სხვებისგან რადიკალურად განსხვავდებიან.

აღნიშნული თავის ერთ-ერთი მიზანი იყო წარმოგვეჩინა ის ფაქტი, რომ ჩვენ მიერ შექმნილი ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსი (GLEAN) ანოტირებულია და გვებმარება ლინგვისტური კვლევისთვის ენის სხვადასხვა დონეზე. ინგლისური ენის ქართველ შემსწავლელთა კორპუსში ყველა სიტყვაფორმას მიემართება მეტყველების შესაბამისი ნაწილის აღმნიშვნელი მარკერი. ანოტაციისთვის გამოვიყენეთ სტანდარტული თეგირების წყება CLAWS7.

კიდევ ერთი მნიშვნელოვანი განხილული საკითხი ეხება შესიტყვებების კვლევის ასოციაციურ საზომებს. მარტივი სიხშირით ძებნა ჩავანაცვლეთ ძებნით, რომელიც ითვალისწინებს კომპლექსურ კვლევას, რადგან სასწავლო კორპუსებში მხოლოდ შესიტყვებების სიხშირის დადგენა მისი ენაში რეალიზების შესწავლისთვის საკმარისი არ არის. არსებობს დამატებითი საზომები, რომლებიც კოლოკაციურ ერთეულებს შორის არსებულ კავშირებზე გვაწვდიან ინფორმაციას. აბსოლუტური (დაუმუშავებელი) სიხშირე ტექსტკორპუსში არსებულ ენობრივ ერთეულებს ითვლის და წარმოაჩენს ენობრივ დინამიკურობას. თუმცა, კორპუსლინგვისტიკაში მონაცემთა ანალიზისათვის მხოლოდ აბსოლუტური სიხშირის გამოყენება მართებული არაა, რადგან კონტრასტული ანალიზის დროს ორი სხვადასხვა მოცულობის კორპუსის შედარება აბსოლუტური (დაუმუშავებელი) სიხშირის საფუძველზე ალოგიკურია და მცდარ დასკვნამდე მიგვიყვანს. მაშასადამე, შესიტყვებების კორპუსზე დამყარებული კვლევის წარმართვისათვის ვითვალისწინებთ ვ. ბრეზინას მიერ შემოთავაზებულ განზომილებებს გავრცელებულობას, ექსკლუზიურობას, მიმართულებითობას. ასოციაციური საზომების შერჩევისას ვისარგებლებთ შესიტყვებების ასოციაციური საზომების მრავალფეროვნებით (მაგ., log-likelihood, MI, MI3, log Dice, Z-score და ა.შ.), რათა მივიღოთ სისტემატური და გამჭვირვალე შედეგები.

შედგების ვიზუალიზაციისთვის განვიხილეთ კოლოკაციური გრაფები და ქსელები, რომლებიც საკვანძო სიტყვასა და მის კოლოკატებს შორის არსებული კავშირებს წარმოადგენენ. გრაფა ლექსიკურ ერთეულებს შორის დამყარებულ კავშირებს სიმძლიერის, სიხშირისა და პოზიციის მიხედვით წარმოაჩენს, ხოლო შესიტყვებების ქსელები უფრო მასშტაბური და ფართო სისტემის გრაფებია, რომლებიც ასახავენ უფრო მეტ ასოციაციურ კანონზომიერებას.

ჩვენს სადისერტაციო ნაშრომში ასახული კვლევა კონტრასტულ ინტერენობრივი ანალიზის მეთოდოლოგიას ეფუძნება. ერთი მხრივ, შევადარებთ ინგლისურის ქართველი შემსწავლელების ენობრივ პროდუქტს ინგლისური ენის აზიელ შემსწავლელთა საერთაშორისო კორპუსს (International Corpus Network of Asian Learners of English, ICNALE), ხოლო, მეორე მხრივ, „საკონტროლო კორპუსად“ შერჩეული გვაქვს მიჩიგანის უნივერსიტეტის ამერიკელი სტუდენტების (Michigan Corpus of Upper-level Student Papers, MICUSP) ინგლისურენოვანი ნაწერების ტექსტები.

თავი III. ინგლისური ენის ქართველ შემსწავლელთა კორპუსის (GLEAN) ანალიზი

3. ზოგადი მიმოხილვა

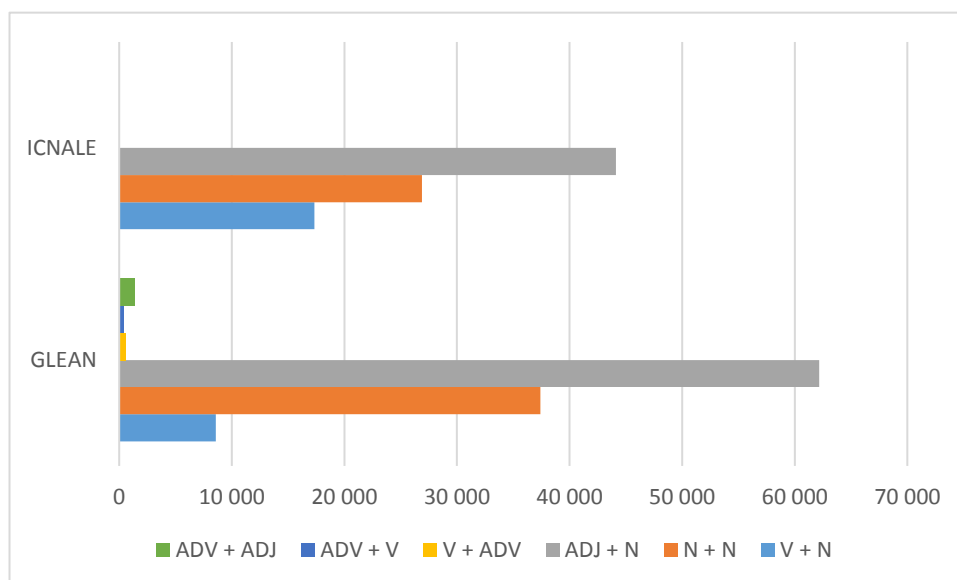
წინამდებარე დისერტაციის ერთ-ერთი მიზანია ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში ორკომპონენტიანი შესიტყვებებისა და სამ-, ოთხკომპონენტიანი ლექსიკური ერთეულების სტრუქტურული, სემანტიკური და დისკურსის ფუნქციური თავისებურებების გამოვლენა და შედარება როგორც ინგლისური ენის აზიელი შემსწავლელების, ისე ამერიკელი სტუდენტების მიერ რეალიზებულ ენასთან. ნებისმიერ ენაში სტრუქტურული და სემანტიკური პარალებელი უკავშირდება გრამატიკური წყობის სხვადასხვა ასპექტს, ამიტომ ლექსიკური კოლოკაციური წყვილებისა და სინტაგმების კვლევა მიზანშეწონილია სხვადასხვა ენობრივი შრის შესწავლის გზით. წინა თავში გამოვყავით ლექსიკური და გრამატიკული შესიტყვებების მოდელები. ლექსიკურ მოდელებში გავაერთიანებთ ისეთ ორკომპონენტიან შესიტყვებებს, რომლის ელემენტებიც შედგება სახელური, ზმნური, ადვერბული კოლოკაციური წყვილებისაგან: 1. ზმნა (V) + არსებითი სახელი (N); 2. არსებითი სახელი (N) + არსებითი სახელი (N); 3. ზედსართავი სახელი (ADJ) + არსებითი სახელი (N); 4. ზმნა (V) + ზმნიზედა (ADV); 5. ზმნიზედა (ADV) + ზმნა (V); 6. ზმნიზედა (ADV) + ზედსართავი სახელი (ADJ). ხოლო სამ-, ოთხკომპონენტიანი ფრაზემები წინდებულს და სხვა ტიპის ერთეულებსაც მოიცავენ და გრამატიკული კოლოკაციების მოდელებში ერთიანდებიან.

3.1. ორკომპონენტიანი შესიტყვებების რაოდენობრივი ანალიზი

კვლევაში შესწავლილია ლექსიკური შესიტყვების ექვსი სტრუქტურული მოდელი და მათი შედეგი როგორც ინგლისური ენის ქართველ შემსწავლელთა კორპუსში, ისე ინგლისური ენის აზიელ სტუდენტთა სასწავლო კორპუსში. ვინაიდან, ამ ორ კორპუსს (GLEAN და ICNALE) განსხვავებული სიტყვაფორმული მოცულობა აქვთ, თითოეული სტატისტიკური სიხშირე ნორმალიზებულია მილიონ სიტყვაზე (pmw) იმისათვის, რომ კოლოკაციათა ხმარების რაოდენობრივი განსხვავებულობის

საერთო სურათი ნათლად მივიღოთ. ყველა ლექსიკური მოდელის მქონე შესიტყვების საერთო რაოდენობა გაყოფილია კორპუსის სიტყვათა ჯამურ ოდენობაზე და გამრავლებულია მილიონზე (იხ. სქემა 10). გარდა ნორმალიზებული სიხშირისა, მოცემული გვაქვს ნედლი (აბსოლუტური) სიხშირეც (იხ. სქემა 11).

სქემა 10. ლექსიკური შესიტყვებების რაოდენობა მილიონ სიტყვაში (pmw)



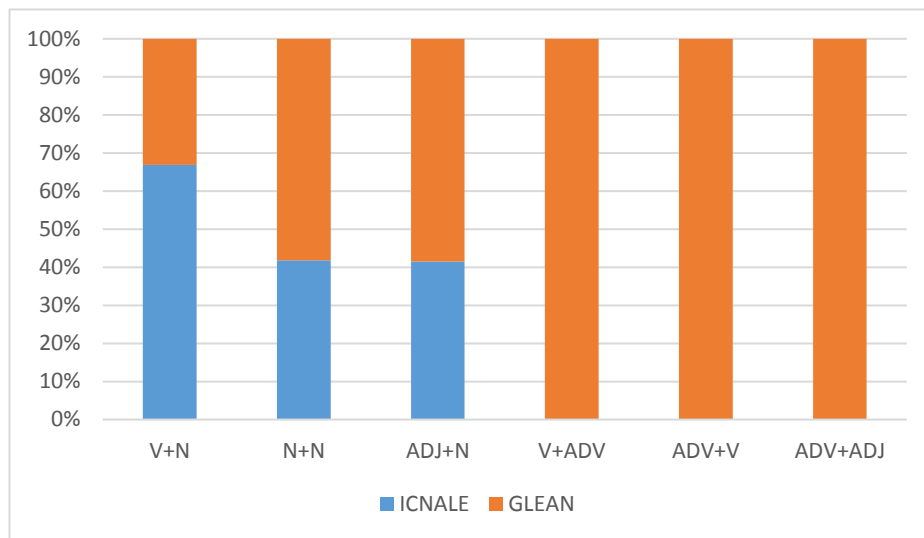
სქემა 11. შესიტყვებების საერთო აბსოლუტური რაოდენობა თითოეული ლექსიკური მოდელის მიხედვით

ლექსიკური მოდელი	GLEAN	ICNALE
ზმნა (V) + არსებითი სახელი (N)	n= 83,935	n= 24,812
არსებითი სახელი (N) + არსებითი სახელი (N)	n= 365,654	n= 38,502
ზედსართავი სახელი (ADJ) + არსებითი სახელი (N)	n= 607,946	n= 63,132
ზმნა (V) + ზმნიზედა (ADV)	n= 5,170	n= 0
ზმნიზედა (ADV) + ზმნა (V)	n= 3,763	n= 0
ზმნიზედა (ADV) + ზედსართავი სახელი (ADJ)	n= 13,408	n= 0

ნორმალიზებული სიხშირის ანალიზის შედეგად დასტურდება, რომ ინგლისური ენის ქართველი შემსწავლელელები უფრო მრავალფეროვან შესიტყვებათა მოდელებს იყენებენ, ვიდრე აზიელი სტუდენტები. გამონაკლისია ზმნა + არსებითი

სახელის კოლოკაციური წყვილი, რომელიც ICNALE სასწავლო კორპუსში ორჯერ მეტად ფიქსირდება. რაც შეეხება ზმნა + ზმნიზედა (ლოკატიური), აგრეთვე ზმნიზედა + ზმნისა და ზმნიზედა + ზედსართავი სახელის ლექსიკური მოდელები საერთოდ არ დასტურდება აზიელი სტუდენტების ტექსტობრივ მასალაში მაშინ, როცა ინგლისური ენის ქართველ შემსწავლელებში შედარებით მცირედ, მაგრამ მაინც მოიპოვება. წარმოვადგენთ კონტრასტული ანალიზით შესიტყვების მოდელების პროცენტულ გაანწილებას ორ კორპუსში (იხ. დიაგრამა 3).

დიაგრამა 3. შესიტყვების მოდელების პროცენტული გაანწილება ICNALE და GLEAN სასწავლო კორპუსებში



3.1.1. V+N სტრუქტურის მეორე ორგომპონენტის შესიტყვებები

ინგლისური ენის ქართველ შემსწავლელთა კორპუსში ზმნა + არსებითი სახელის ორგომპონენტის კოლოკაციური მოდელი რამდენიმე ნიშანდობლივი ასპექტით გამოირჩევა. როგორც აღვნიშნეთ, აზიელი სტუდენტების სასწავლო კორპუსთან მიმართებით V+N წყვილის შემთხვევაში, ქართველი შემსწავლელების ენობრივი მონაცემი ორჯერ ჩამოუვარდება აზიელი სტუდენტების შედეგს, რაც დასტურდება ს. ჰოფმანის (2008: 80) შემოთავაზებული კონტრასტული ანალიზის სანდოობის სიგნიფიკანტურობის ტესტით (log-likelihood G²). შედეგად, $G^2 = 8,3$, სიგნიფიკანტურობა არის $p < .001$, ნდობის ინტერვალი 95%. მიზეზ-შედეგობრივი

² ტესტი ჩატარებულია Corpus Frequency Wizard <http://sigil.collocations.de/wizard.html>

კავშირის ძიებისას შეგვიძლია ვივარაუდოთ, რომ თანამედროვე ჩინურ ენაში არსებულ მრავალ-კატეგორიანი სიტყვების (multi-category word) და მათ შორის ყველაზე დომინანტური შესიტყვებითი წყვილის (V+N), ესოდენ გავრცელებულობაში. მშობლიურ ენაში (ჩინურში) ასეთი რთული გრამატიკული ფენომენის არსებობის შემდეგ, ინგლისურ ენაში ჩინელი სტუდენტებისათვის გაცილებით მარტივია ზმნა + არსებითი სახელის კონსტრუქციის ინტერნალიზაცია.

წარმოვადგენთ V+N წყვილის ყველაზე ხშირად გამოყენებული შესიტყვების ანალიზის შედეგს სქემაში (იხ. სქემა 13).

სქემა 12. GLEAN სასწავლო კორპუსში V+N შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები

N.	V+N შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1.	take place	903	1.08%
2.	hold meetings	769	0.92%
3.	discuss issues	769	0.92%
4.	take part	493	0.59%
5.	spent time	343	0.41%
6.	lost control	284	0.34%
7.	raise awareness	237	0.28%
8.	use technology	234	0.28%
9.	stay home	226	0.27%
10.	spark tension	218	0.26%
...	...	...	...
50.	expressed gratitude	59	0.07%

შედეგად, მოცემული კოლოკაციური წყვილებიდან ზოგიერთი მათგანი კონკრეტულად კორესპონდენციის ნიმუშებში იძებნება, მაგალითად *expressed gratitude*, ხოლო *raise awareness*, *lost control* საგანგებოდ მხოლოდ - პრესის მასალაში. დანარჩენი შესიტყვებები ყველა დარგისა და ჟანრული კატეგორიის ტექსტებში იკვეთება. მეორე მხრივ, აღნიშნულ სიაში შესიტყვების ერთ-ერთ კომპონენტად გამოაშკარავებულია *take* ლექსემის სიჭარბე. ამ ტენდენციის სიღრმისეული კვლევისათვის ნორმალიზებული სიხშირის დადგენის გარდა, გამოვიყენებთ მეორე თავში განხილულ ასოციაციურ საზომებს, რომლებიც განისაზღვრება შესიტყვებათა გავრცელებულობის, ექსკლუზიურობისა და ურთიერთმიმართებულობის განზომილებებით. შერჩეულ იქნა რამდენიმე ასოციაციური საზომი, Loglik, MI, DICE, Z-score, რომლის წყვილებიც მოცემულია ცხრილში (იხ. სქემა 14).

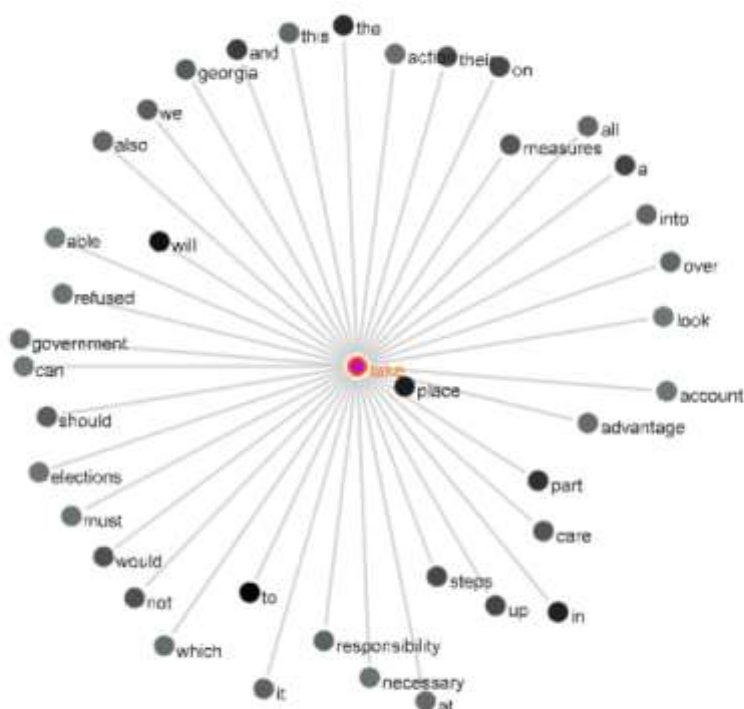
სქემა 13. GLEAN სასწავლო კორპუსში არსებული *take* კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით

N.	პოზიცია	<i>take</i> კოლოკატი	LogLik (Log-Likelihood)	DICE	MI (Mutual Information)	Z-score
1	მარჯვნივ	place	9627.08	28.4	423.7	620.5
2	მარჯვნივ	part	4140.8	25.2	287.7	281.1
3	მარჯვნივ	steps	2229.6	23.9	287.0	258.6
4	მარჯვნივ	care	1454.05	22.7	214.9	232.1
5	მარჯვნივ	measures	1212.8	21.9	198.9	208.2
6	მარჯვნივ	responsibility	941.2	21.2	161.9	165.5
7	მარჯვნივ	advantage	904.7	20.4	153.7	150.2

სქემა 14-ში კოლოკატები დალაგებულია როგორც ასოციაციური საზომების მიხედვით, ისე მათი პოზიციის გათვალისწინებით. „მარჯვნივ“ გულისხმობს სიტყვის მდებარეობას *take* კოლოკატის გვერდით. ამ შემთხვევაში ყველა მათგანი საკვანძო სიტყვის მარჯვნივ მდებარეობს. მაშასადამე, შედეგად ვიღებთ *take place*, *take part*, *take steps*, *take care*, *take measures*, *take responsibility*, *take advantage* შესიტყვებების სიჭარბეს ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში. როგორც ვხედავთ, V+N მოდელის კორპუსული სიხშირული ანალიზი ზედაპირულ შედეგს გვაძლევს, ხოლო ასოციაციური საზომების გამოყენებით შედეგად მივიღეთ კოლოკაციური სიჭრელე.

ვიზუალურად წამოვადგენთ საკვანძო სიტყვასთან „take“ კოლოკაციურად დაკავშირებულ ლემებს (იხ. სურათი 9).

სურათი 9. GLEAN სასწავლო კორპუსში take და მასთან დაკავშირებული ლექსემების შესიტყვებითი გრაფა, საზომი z-score



როგორც ვხედავთ, გამოხატულია ასოციაციური სიძლიერე საკვანძო სიტყვასა და მასთან ყველაზე ახლოს მყოფ ლექსიკურ ერთეულებს შორის, ანუ ყველაზე ძლიერი ასოციაციური კავშირი მყარდება კოლოკაციურ წყვილში take place. შესიტყვების სიხშირე განისაზღვრება წრის ფერის მუქი ან ღია ტონით, ამ შემთხვევაში სიხშირით გამოირჩევა take part, take place შესიტყვებები, ხოლო შესიტყვების პოზიციის გათვალისწინებით ჩვენთვის ნიშანდობლივია ყველა კოლოკატი, რომელიც მარჯვნივ სხივებითაა დაკავშირებული.

3.1.2. N+N სტრუქტურის მქონე ორკომპონენტიანი შესიტყვებები

წარმოდგენილი სტრუქტურის მქონე ორკომპონენტიანი შესიტყვების არსებითი სახელი (N) + არსებითი სახელი (N) სასწავლო კორპუსების ანალიზის შედეგად დასტურდება, რომ ქართველი შემსწავლელები უფრო მეტი სიხშირით იყენებენ აღნიშნულ მოდელს თავიანთ წერილობით მასალაში, ვიდრე ეს აზიელ

სტუდენტებთან შეინიშნება. კონტრასტული ანალიზის სიგნიფიკანტურობის ტესტის (log-likelihood G2) შედეგად, $G2 = 4.2$, სიგნიფიკანტურობა არის $p < .001$ (ნდობის ინტერვალი 95%). წარმოვადგენთ N+N წყვილის ყველაზე ხშირად გამოყენებული 50 შესიტყვების ანალიზის შედეგს სქემაში (იხ. სქემა 15).

სქემა 14. N+N წყვილის ყველაზე ხშირად გამოყენებული 50 შესიტყვებები GLEAN კორპუსში

N.	N+N შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1	parliament speaker	3841	1.05%
2	press office	3267	0.89%
3	growth rate	2883	0.79%
4	time management	1816	0.50%
5	film festival	1743	0.48%
6	work life	1439	0.39%
7	life	1421	0.39%
8	candidate status	1399	0.38%
9	Dream party	1311	0.36%
10	travel ideas	1291	0.35%
...	...	...	...
50	youth program	519	0.14%

აღნიშნული შესიტყვების აბსოლუტური უმრავლესობა პუბლიცისტური ჟანრის მასალაში დასტურდება, შეგვიძლია ვივარაუდოთ, რომ კონტრასტული ანალიზის აღნიშნული შედეგი განპირობებულია ინგლისური ენის ქართველ შემსწავლელთა კორპუსში პრესის მასალის სიუხვით. აზიელი სტუდენტების მასალაში N+N კოლოკაციური წყვილის დეფიციტი გამოწვეულია იმ ფაქტით, რომ ICNALE სასწავლო კორპუსში პრესის მასალა არ შედის.

ყველაზე ხშირად გამოყენებული შესიტყვებებიდან, უმრავლესობა პრესის ჟანრის მასალაში დასტურდება, ესენია: *parliament speaker*, *press office*, *candidate status*, *Dream party*, *growth rate*. პოლიტიკურ ტექსტებში *growth rate* შესიტყვების 3000-მდე შემთხვევა ფიქსირდება. საკვანძო სიტყვის კონტექსტში მოძიებისას შედეგი ასეთია:

- Domestic high economic **growth rate** was expected to be maintained.
- the economic **growth rate** of Georgia was "remarkable".
- We have had double-digit economic **growth rate** over the past two years.

სახელური შესიტყვებების (N+N) დანარჩენი ნიმუშები ნაწილდება ბლოგებსა და ესეებში, რაც სასწავლო კორპუსის რეპრეზენტაციულობას უსვამს ხაზს. სამოგზაურო ჟანრის ბლოგში ხშირია *travel ideas* კოლოკაციური წყვილი, მსგავსი კონტექსტით: *I found a platform and got some **travel ideas** and booking opportunities.*

ვინაიდან სქემა 15-ში მოცემულ შესიტყვებების წყვილებში ერთ-ერთ ერთეულად ღემა *life* რამდენჯერმე მეორდება, სიღრმისეული კვლევისათვის აქაც კვლავ გამოვიყენებთ ასოციაციურ საზომებს Log-likelihood, MI, DICE, Z-score, რომლის წყვილებიც მოცემულია ცხრილში (იხ. სქემა 16).

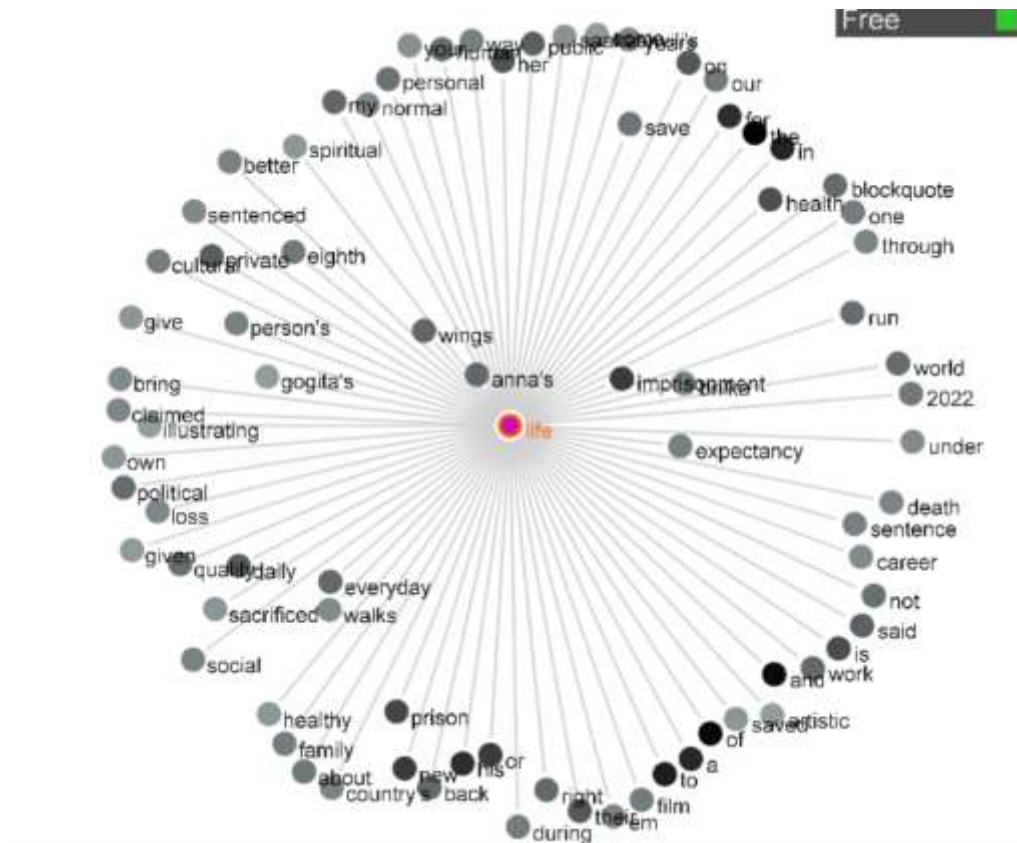
სქემა 15. GLEAN სასწავლო კორპუსში არსებული *take* კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით

N.	პოზიცია	<i>life</i> კოლოკატი	LogLik	DICE	MI	Z-score
			(Log-Likelihood)		(Mutual Information)	
1	მარჯვნივ 3	imprisonment	1616.7	1.93	11.7	337
2	მარჯვნივ 3	expectancy	826	0.94	11.2	279.2
3	მარცხნივ 3	health	569.2	0.91	7.2	122.3
4	მარცხნივ 3	quality	280,8	0.72	6.8	65
5	მარჯვნივ 3	sentence	182.5	0.28	6.6	64
6	მარცხნივ 3	family	42.4	0.09	4.1	45.2
7	მარცხნივ 3	work	28.6	0.007	3.9	29.8

შედეგად, კოლოკატები დალაგებულია როგორც სტატისტიკური საზომების, ისე პოზიციის (მარჯვენა და მარცხენა) მიხედვითაც: life imprisonment, life expectancy, health life, quality life, life sentence, family life, work life. ასოციაციური საზომების

კვლევით კოლოკაციური წყვილების ბევრად მეტი რაოდენობა გამოიკვეთა, თუმცა მანუალური გზით შევარჩიეთ მხოლოდ ისინი, რომლებიც სახელური შესიტყვების სტრუქტურულ მოდელს აკმაყოფილებდა. აღნიშნული კოლოკაციური სიუხვე ნათლად იკვეთება შესიტყვებების გრაფაში (იხ. სურათი 10).

სურათი 10. GLEAN სასწავლო კორპუსში life შესიტყვების გრაფა (Z-score)



3.1.3. ADJ+N სტრუქტურის მექანიზმების აღწერა

ზედსართავი სახელი + არსებითი სახელის სტრუქტურის მქონე ორკომპონენტიანი შესიტყვების ანალიზისას რამდენიმე მნიშვნელოვანი ასპექტი გამოიკვეთა. წარმოდგენილი სტრუქტურის მქონე შესიტყვების კონტრასტული ანალიზის შედეგად დასტურდება, რომ ქართველი შემსწავლელები უფრო მეტი სიხშირით იყენენ აღნიშნულ მოდელს თავიანთ წერილობით მასალაში, ვიდრე აზიელი სტუდენტები. სიგნიფიკანტურობის ტესტის (log-likelihood G2) შედეგად, $G2 = 7,8$, სიგნიფიკანტურობა არის $p < .001$ (ნდობის ინტერვალი 95%). აზიელი

სტუდენტების სასწავლო კორპუსში ზედსართავი სახელი + არსებითი სახელის კოლოკაციური შესიტყვების დეფიციტი შესაძლოა გამოწვეული იყოს ინგლისურსა და აზიურ ენებს შორის არსებული გრამატიკული განსხვავებებით. მაგალითად, ჩინურის მსგავს ენებს ინგლისურთან შედარებით წინადადების სხვადასხვა სტრუქტურა აქვთ. ინგლისური ეყრდნობა სუბიექტი-ზმნა-ობიექტი წყობას, ხოლო ჩინური იყენებს „თემა-კომენტარი“ სტრუქტურას. მრავალ აზიურ ენაში, მათ შორის ჩინურშიც, ზმნებისა და ზედსართავების კლასიკური გაყოფა შეუძლებელია და მეტყველების ნაწილების სიღრმულ პლანში განხილვისას ზედსართავი სახელები უნდა გავაერთიანოთ ზმნებთან. ამრიგად, ინგლისური ენის ჩინელი შემსწავლელებისათვის სირთულეს წარმოადგენს ინგლისურად ჩასწვდენ ზედსართავი სახელების დაწყვილების ბუნებრივ დინებას.

წარმოვადგენთ აღნიშნული ლექსიკური კოლოკაციური მოდელის ყველაზე ხშირად ხმარებულ შესიტყვებას (იხ. სქემა 17).

სქემა 16. GLEAN სასწავლო კორპუსში ADJ+N შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები

N.	ADJ+N შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1	my opinion	10373	1.71%
2	good governance	5327	0.88%
3	traditional food	4682	0.77%
4	Georgian government	4255	0.70%
5	high officials	3155	0.52%
6	ruling party	3104	0.51%
7	visa-free regimen	3013	0.50%
8	good relationship	2668	0.44%
9	domestic violence	2486	0.41%
10	good luck	2343	0.39%
...	...	...	...
50	kind request	710	0.12%

ზედსართავი სახელი + არსებითი სახელის შესიტყვებების სიხშირულმა ანალიზმა რამდენიმე მნიშვნელოვანი შედეგი გვიჩვენა. უპირველეს ყოვლისა, ინგლისური ენის ქართველ შემსწავლელთა კორპუსში გამოიკვეთა ხშირად დაშვებული ლექსიკური შეცდომა *visa-free regimen* ნაცვლად *visa-free regime*-ისა. პირველი მათგანი აღნიშნავს მკაცრად დადგენილ წესს, რეჟიმს, რომელიც

მკურნალობის, კვების, დასვენების მიმართ გამოიყენება, ხოლო სახელმწიფო წყობილება, ან სისტემა, რაიმე დადგენილი წესი არის regime და სასწავლო კორპუსის კონტექსტებში, სავიზო მიმოსვლებთან დაკავშირებით ეს უკანასკნელი უნდა იყოს გამოყენებული.

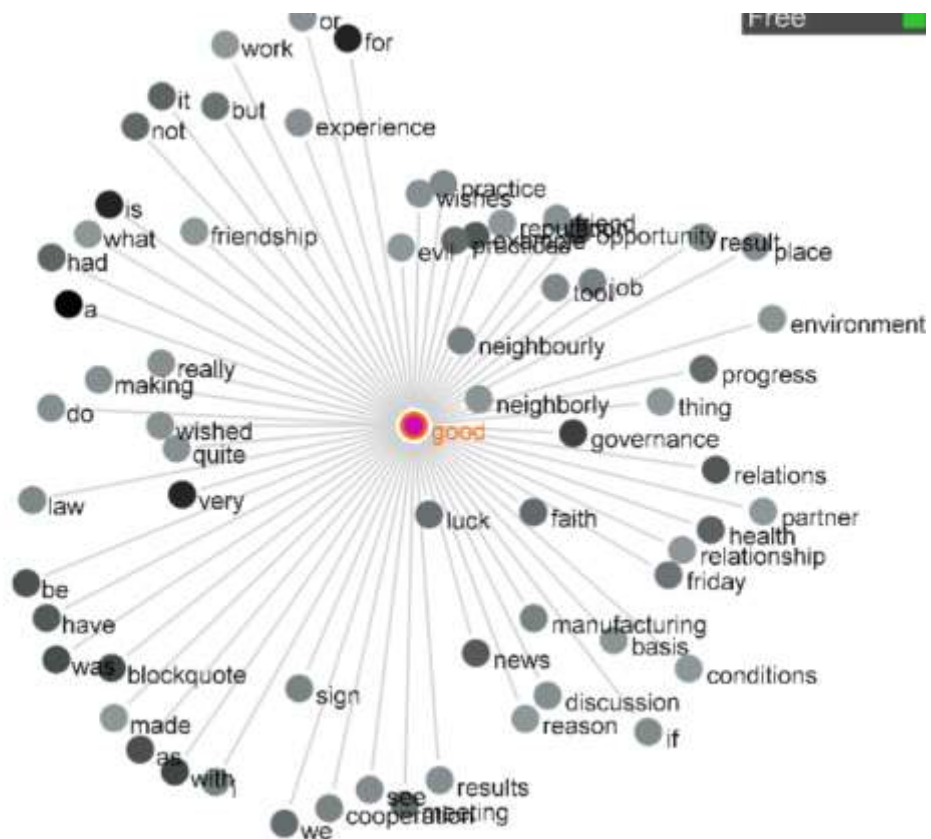
მეორე მხრივ, ზედსართავი სახელი + არსებითი სახელის შესიტყვების უმრავლესობა მხოლოდ კონკრეტულად პოლიტიკურ ტექსტებში გვხვდება, როგორიცაა: *good governance, ruling party, Georgian government, visa-free regimen, domestic violence*. თუმცა, ვხვდებით ორ კოლოკაციურ წყვილს, რომელიც როგორც პოლიტიკურ, ისე კორესპონდენციის ჟანრულ კატეგორიაში შეინიშნება *kind request, good luck*.

ამავდროულად, ჩამოთვლილ წყვილებში შეინიშნება *good* ლექსემის სიჭარბე, რომელსაც ასოციაციური კავშირებით ჩაღრმავებულად გამოვიკვლევთ (იხ. სქემა 18).

სქემა 17. GLEAN სასწავლო კორპუსში არსებული *good* კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით

N.	პოზიცია	<i>good</i> კოლოკატი	LogLik (Log-Likelihood)	DICE	MI (Mutual Information)	Z-score
1	მარჯვნივ	governance	1170	0.084	14.8	290.2
2	მარჯვნივ	luck	434.8	0.081	11.5	189.7
3	მარჯვნივ	faith	399	0.073	11.4	184.5
4	მარჯვნივ	example	350	0.046	11.2	150
5	მარჯვნივ	practices	337.4	0.032	7.2	126
6	მარჯვნივ	news	333.7	0.031	6.8	106.1
7	მარჯვნივ	opportunity	319.2	0.029	6.1	89.9
8	მარჯვნივ	wishes	297.4	0.012	4.9	53.7
9	მარჯვნივ	reputation	285	0.011	3.2	53.2
10	მარჯვნივ	conditions	155.6	0.0087	2.7	44.5
11	მარჯვნივ	dynamics	36.2	0.0053	2.6	23.8
12	მარჯვნივ	spirit	34.7	0.0029	1.5	21.7
13	მარჯვნივ	friday	28.9	0.0026	0.8	20.8
14	მარჯვნივ	partner	25.3	0.0019	0.09	14.71
15	მარჯვნივ	fortune	25.2	0.0012	0.074	10.5

ასოციაციური საზომებით ზედსართავი სახელი (ამ შემთხვევაში *good*) + არსებითი სახელის შესიტყვებათა კვლევამ უფრო ნათელი სურათი მოგვცა, ვიდრე მხოლოდ სიხშირულმა ანალიზმა. ვინაიდან კოლოკაციის წყვილში პირველი კომპონენტი ზედსართავი სახელია და მეორე არსებითი სახელი, *good*-ის კოლოკატებს გასაკვირი არაა, რომ მარჯვენა პოზიცია უკავიათ, რაც გამოიხატება *good wishes, good reputation, good dynamics* და ა.შ. სტრუქტურაში. ADJ+N შესიტყვებათა მოდელს მკვეთრად გამოხატული პოზიცია აქვს კოლოკაციურ გრაფაში და კომპონენტებს შორის არსებულ ასოციაციურ სიმძლიერეს, სიხშირესა და პოზიციონირებას წარმოვადგენთ შესიტყვებების გრაფის ვიზუალიზაციით (იხ. სურათი 11).



როგორც ვხედავთ, *good* ლემის მარჯვნივ მყოფი ლექსიკური ერთეულები უფრო მჭიდროდ და მოკლე მანძილით არიან დაკავშირებულნი საკვანძო სიტყვას, რაც ბუნებრივია ასოციაციური საზომების შედეგების გათვალისწინებით, *good faith*, *good*

luck, good governance და სხვ. ურთიერთმიმართებულობის, სიხშირისა და ერთმანეთისადმი ძლიერი კოლოკაციური კავშირით გამოირჩევიან.

3.1.4. V+ADV სტრუქტურის მქონე ორკომპონენტური შესიტყვებები

ზმნა + ზმნიზედის (ლოკატივი) სტრუქტურის მქონე ორკომპონენტური შესიტყვების კონტრასტული ანალიზის შედეგად დასტურდება, რომ ქართველი შემსწავლელები უფრო ხშირად იყენებენ აღნიშნულ მოდელს თავიანთ წერილობით მასალაში, ვიდრე ეს აზიელ სტუდენტებთან შეინიშნება. კონტრასტული ანალიზის სიგნიფიკანტურობის ტესტის (log-likelihood G2) შედეგად, $G2 = 1,4$, სიგნიფიკანტურობა არის $p < .001$ (ნდობის ინტერვალი 95%).

წარმოვადგენთ GLEAN სასწავლო კორპუსში V+ADV კოლოკაციური მოდელის სიხშირულ სიას (იხილეთ სქემა 19).

სქემა 18. GLEAN სასწავლო კორპუსში V+ADV (ლოკატივი) შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები

N.	V+ADV (ლოკატივი) შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1	work together	224	4.33%
2	looking forward	210	4.06%
3	bringing together	175	3.38%
4	passed away	91	1.76%
5	living abroad	88	1.70%
6	swept away	31	0.60%
7	washed away	20	0.39%
8	pushed forward	19	0.37%
9	click here	15	0.20%
10	standing nearby	5	0.11%
...			
50	known worldwide	4	0.08%

შედეგად ირკვევა, რომ ინგლისური ენის ქართველი შემსწავლელები V+ADV კოლოკაციურ წყვილში ზმნიზედის შერჩევისას უპირატესობას ანიჭებენ ლოკატიურ ზმნიზედას, მაგალითად, ნიმუშებში სჭარბობს ლემები *together*, *forward*, *away*. კონკორდანსში მათი კონტექსტში გამოყენების ნიმუშები ასეთია:

- You have to **work together** and not fight with each other.

- We should cooperate directly, share information, have contact persons and **work together** on common cases.
- Let 's **work together** to make it real.

ზმნა + ზმნიზედის კოლოკაციური წყვილი GLEAN სასწავლო კორპუსის მასალის ყველა ჟანრულ კლასიფიკაციაში - პრესაში, არგუმენტირებულ თუ თავისუფალ ესეებში, კორეპონდენციასა და ბლოგებში მოიპოვება. თუმცა, გამოვყოფთ კონკრეტულად ბლოგების მასალაში აღმოჩენილ *click here* კოლოკაციას.

- **Click here** to see the infographics of Georgian Wine Industry.
- **Click here** to visit our comprehensive interactive map of Tsu library.
- **Click here** to see a photo story of what the city looks like after Tbilisi Open Air.

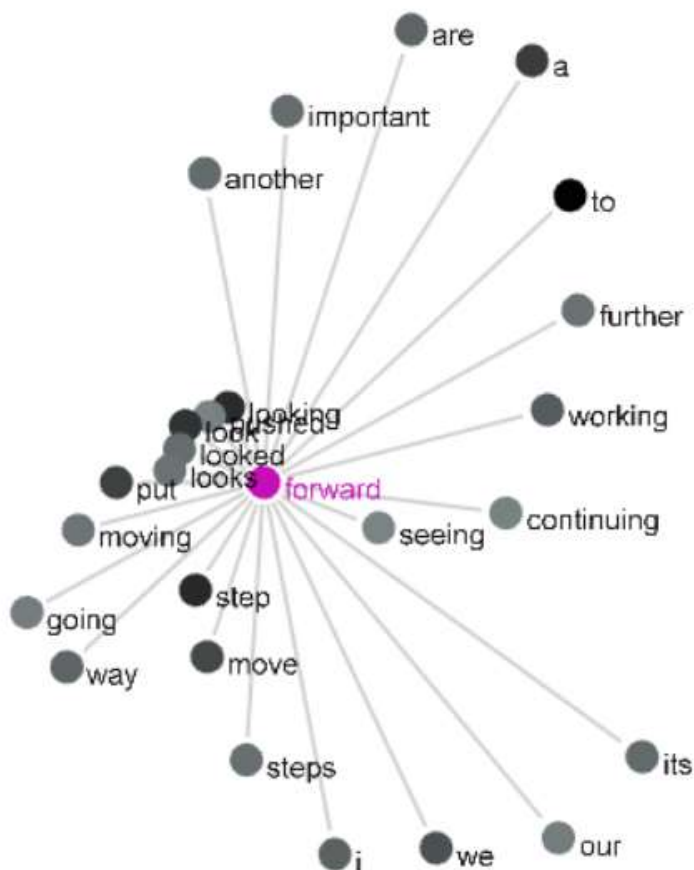
შესიტყვების ასოციაციური საზომების კვლევისათვის საკვანძო სიტყვად ავიღებთ ლოკატიურ ზმნიზედას *forward*, რომელიც სიხშირული კალკულაციის მიხედვით ერთ-ერთი ყველაზე ხშირად ხმარებული შესიტყვებითი ერთეულია. პოზიციისა და სტატისტიკური საზომების გათვალისწინებით წარმოვადგენთ სქემას (იხ. სქემა 20).

სქემა 19. GLEAN სასწავლო კორპუსში არსებული *forward* კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით

N.	პოზიცია	<i>forward</i> კოლოკატი	LogLik (Log- Likelihood)	DICE	MI (Mutual Information)	Z-score
1	მარცხნივ	looking	2842.2	0.19	10.78	611.9
2	მარცხნივ	pushed	1427.5	0.16	10.74	285
3	მარცხნივ	looked	1417	0.15	10.6	239
4	მარცხნივ	looks	1415,8	0.10	10.57	220.9
5	მარცხნივ	look	1295.5	0.07	10.52	211
6	მარცხნივ	seeing	814	0.04	10.03	144.8
7	მარცხნივ	step	794.4	0.02	9.96	78.6

შესიტყვების ასოციაციურმა საზომებმა გამოააშკარაა forward სიტყვასთან დაკავშირებული ლექსემა *look* სხვადასხვა გრამატიკული ვარიაციით, რაც ამ კოლოკაციური წყვილის ძლიერ კავშირს უსვამს ხაზს. ვიზუალურად მათ შორის არსებული კავშირი შემდეგნაირად გამოიყურება (იხ. სურათი 12).

სურათი 12. GLEAN სასწავლო კორპუსში *forward* შესიტყვების გრაფა (MI)



მაშასადამე, კოლოკატები *looking*, *look*, *looked*, *looks* ერთ ველშია შემჭიდროვებული, რაც მათ საკვანძო სიტყვასთან არსებულ ბმას ააშკარავებს. მეტყველების სხვა ნაწილები, მაგალითად, პირის ნაცვალსახელები, წინდებულები მოშორებით მდებარეობენ და ამჟამად ჩვენს კვლევაში მათ შორის კორელაცია ნაკლებ საყურადღებოა.

3.1.5. ADV+V სტრუქტურის მექანე ორკომპონენტიანი შესიტყვებები

ზმნიზედა + ზმნის ორკომპონენტიან შესიტყვებით მოდელს GLEAN ინგლისური ენის ქართველი შემსწავლელების კორპუსში უფრო ხშირად იყენებენ, ვიდრე ICNALE სასწავლო კორპუსში. კონტრასტული ანალიზისას სიგნიფიკანტურობის ტესტის (log-likelihood G2) შედეგად, $G2 = 1,02$, სიგნიფიკანტურობა არის $p < .001$ (ნდობის ინტერვალი 95%).

წარმოვადგენთ ADV+V წყვილის ყველაზე ხშირად გამოყენებული შესიტყვების ანალიზის შედეგს სქემაში (იხ. სქემა 21).

სქემა 20. GLEAN სასწავლო კორპუსში ADV+V შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები

N.	ADV+V შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1	further strengthen	302	8.03%
2	later released	181	4.81%
3	further deepen	136	3.61%
4	further develop	115	3.06%
5	further notes	114	3.03%
6	higher compared	110	2.92%
7	further enhance	75	1.99%
8	less compared	49	1.30%
9	further strengthening	47	1.25%
10	further complicates	43	1.14%
...			
50	further stated	11	0.29%

შედეგად დასტურდება, რომ GLEAN სასწავლო კორპუსში ქართველი შემსწავლელები აქტიურად იყენებენ შედარებითი ხარისხის ზმნიზედას, further, less, more, higher. კონტექსტუალური ანალიზით აღმოჩნდა, რომ ასეთი ტიპის პარადიგმები, კერძოდ further + ზმნა საგაზეთო პოლიტიკური სტატიების, დიპლომატიური დისკურსის მასალაში გვხვდება, მაგალითად:

- This kind of effort will **further strengthen** Georgian democracy;
- Commission of the Council of Europe , on Friday expressed readiness to **further strengthen** cooperation with Georgia.

- Our country is ready and willing to **further strengthen** the security of the Euro-Atlantic area.

ვინაიდან შესიტყვებების წყვილებში ლემა *further* ყველაზე ხშირად მეორდება, სიღრმისეული კვლევისათვის ამ შემთხვევაშიც კვლავ გამოვიყენებთ ასოციაციურ საზომებს Log-likelihood, MI, DICE, Z-score, რომლის შედეგიც მოცემულია სქემაში (იხ. სქემა 22).

სქემა 21. GLEAN სასწავლო კორპუსში არსებული *further* კოლოკატები ასოციაციური საზომების (loglik, MI, DICE, Z-score) მიხედვით

N.	პოზიცია	<i>further</i> კოლოკატი	LogLik (Log-Likelihood)	DICE	MI (Mutual Information)	Z-score
1	მარჯვნივ	strengthen	4627.08	18.4	373.7	344.5
2	მარჯვნივ	deepen	4140.8	5.2	287.7	264.2
3	მარჯვნივ	increase	2729.6	13.9	286.4	193.7
4	მარჯვნივ	contribute	1554	12.7	240.9	188.9
5	მარჯვნივ	spread	1212.8	11.9	176.9	162.9
6	მარჯვნივ	develop	941.2	11.2	161.9	152.2
7	მარჯვნივ	enhance	800.8	10.4	53.7	143.2
8	მარჯვნივ	intensify	127.08	8.4	23.7	122.1
9	მარჯვნივ	improve	40.8	5.2	7.7	65.8

ინგლისური ენის ქართულ შემსწავლელთა კორპუსში კოლოკაციური წყვილი *further* + ზმნა ენობრივი მრავალფეროვნებით ხასიათდება. შესიტყვებაში შემავალი ზმნები დადებითი კონოტაციის მატარებლები არიან, მაგალითად, *further contribute*, *further strengthen*, *further improve*, *further develop*. ვინაიდან ინგლისური ენის აზიელი შემსწავლელების სასწავლო კორპუსში ზმნიზება + ზმნის შესიტყვებათა მოდელის ნიმუშები არ დასტურდება, მიზნად დავისახეთ შედარებითი ხარისხის სემანტიკის მქონე ზმნიზება გადაგვემოწმებინა მიჩიგანის სასწავლო კორპუსში. სქემა 22-ში ასახული შესიტყვებათა წყვილებიდან მხოლოდ *further enhance* დადასტურდა ერთხელ, სხვა დანარჩენი კოლოკაციების სიხშირე 0-ია. შეგვიძლია ვივარაუდოთ, რომ

ინგლისური ენის ქართველ შემსწავლელებს ტრანსფერის გზით მშობლიური ენიდან უცხოურ ენაზე გადააქვთ აღნიშნული სტრუქტურული მოდელი „მეტად“ + ზმნა, მაგალითად, *მეტად განვითარდეს, მეტად გაუმჯობესდეს, მეტად გაღრმავდეს* და მსგავსი შესიტყვებითი პარადიგმა ინგლისური ენისათვის ისეთივე დამახასიათებელი არაა, როგორც ქართული ენისათვის.

3.1.6. ADV + ADJ გამაძლიერებელი კოლოკაციები (*very, utterly, extremely, perfectly, absolutely, totally + ზედსართავი სახელი*)

ზმნიზედა + ზედსართავი სახელის შესიტყვებების სიხშირული კვლევის მიხედვით ინგლისური ენის ქართველ შემსწავლელთა კორპუსში გამოიკვეთა გამაძლიერებელი სემანტიკის მქონე ზმნიზედების შესიტყვების ერთ-ერთ წევრად ყოფნის ტენდენცია. ესენია: *very, utterly, extremely, perfectly, absolutely, totally*. აღნიშნული ზმნიზედების მთავარი პრაგმატიკულ-სემანტიკური ფუნქცია ზედსართავი სახელის შინაარსობრივი გაძლიერებაა.

კვლევაში შეპირისპირებითი ანალიზით გამოვიკვლევთ არა მხოლოდ აზიელი შემსწავლელების სასწავლო კორპუსის შედეგს *adv + adj* მოდელთან მიმართებით, არამედ მიჩიგანის უნივერსიტეტის სასწავლო კორპუსის შედეგებსაც გავანალიზებთ. წარმოდგენილი სტრუქტურის მქონე შესიტყვების კონტრასტული ანალიზის შედეგად დასტურდება, რომ ქართველი შემსწავლელები უფრო მეტი სიხშირით იყენებენ აღნიშნულ მოდელს თავიანთ წერილობით მასალაში, ვიდრე აზიელი სტუდენტები. სიგნიფიკანტურობის ტესტის (*log-likelihood G2*) შედეგად, $G2 = 3,7$, სიგნიფიკანტურობა არის $p < .001$ (ნდობის ინტერვალი 95%).

წარმოვადგენთ *ADV+ADJ* წყვილის ყველაზე ხშირად გამოყენებული შესიტყვებების ანალიზის შედეგს სქემაში (იხ. სქემა 22).

სქემა 22. GLEAN სასწავლო კორპუსში ADV+ADJ შესიტყვების ყველაზე ხშირად გამოყენებული ნიმუშები

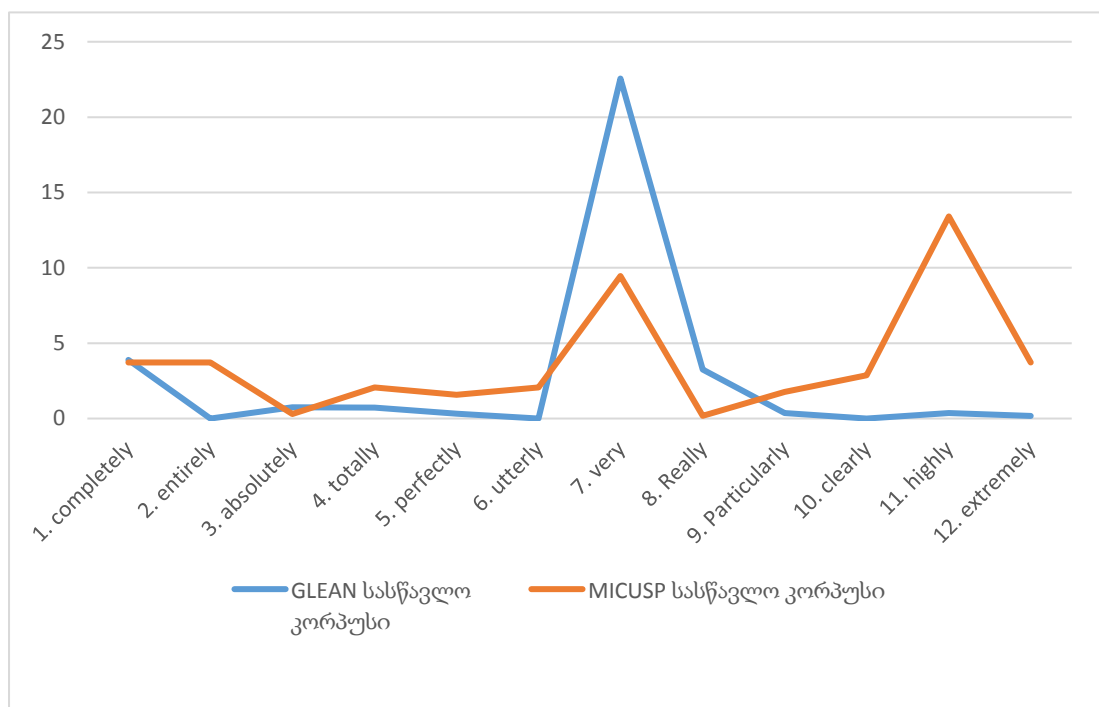
N.	ADV+ADJ შესიტყვება	შემთხვევების რაოდენობა n.	პროცენტი
1	very important	1181	8.81%
2	most difficult	672	5.01%
3	very good	194	1.45%
4	very recent	192	1.43%
5	more effective	177	1.32%
6	very difficult	137	1.02%
7	very fruitful	136	1.01%
8	very positive	123	0.92%
9	very interesting	119	0.89%
10	very strong	118	0.88%
11	very happy	115	0.86%
12	really hard	110	0.82%
13	totally dependent	105	0.78%
14	extremely inefficient	100	0.75%
15	totally high	99	0.74%
16	perfectly invisible	96	0.72%
17	highly important	94	0.70%
18	particularly attractive	93	0.69%
19	extremely prestigious	93	0.69%
20	Completely private	91	0.68%

სიხშირული კვლევის შედეგად გამოიკვეთა, რომ ქართველი შემსწავლელები მნიშვნელოვნად დაბალი სიხშირით იყენებენ -ly ზმნიზედებს, ვიდრე სხვა მორფოლოგიური სტრუქტურის მქონე ზმნიზედებს. როგორც ვხედავთ *very, more, most* სჭარბობს -ly სუფიქსიან ზმნიზედებს. ქანრული მიკუთვნილობისას ზმნიზედა + ზედსართავი სახელის ხვედრითი წილი ესეებში, კორესპონდენციასა და ბლოგებში ნაწილდება, ვიდრე პრესის მასალაში. ინგლისური ენის ქართველი შემსწავლელების მიერ ენობრივად რეალიზებული კოლოკაციები ხშირად არამშობლიურად ჟღერს. შეინიშნება ენობრივი ელემენტების დივერსიფიკაციის ნაკლებობა:

- Your support is **really important** for me in collecting the documents for Ca'foscari.
- I have attended a conference recently and had a **very fruitful** meeting with them.
- This is a **very significant** and positive tradition when little angles, our kids visit our workplace.

ამავდროულად, მნიშვნელოვანია დაკვირვება თუ რამდენად ხშირად/იშვიათად იყენებენ ამერიკელი სტუდენტები გამაძლიერებელ ზმნიშედებისა და ზედსართავი სახელების კოლოკაციურ წყვილებს თავიანთ ნამუშევრებში. მიჩიგანის სასწავლო კორპუსზე დაკვირვებამ რამდენიმე ანგარიშგასაწევი შედეგი გამოავლინა. ვინაიდან ორივე სასწავლო კორპუსი სხვადასხვა რაოდენობის ენობრივ მასალას მოიცავს, მონაცემთა სიხშირული დამუშავებისას სანდოობის გათვალისწინებით შედარებულია არა აბსოლუტური სიხშირე, არამედ ნორმალიზებული (იხ. დიაგრამა 4). ვერტიკალზე ასახული 0-დან 25-მდე ცვლადი ასახავს შესიტყვებების დისტრიბუციულ გაანწილებას ყოველ 10,000 სიტყვაში.

დიაგრამა 4. GLEAN და MICUSP სასწავლო კორპუსებში გამაძლიერებელი ზმნიშედებისა+ADJ შესიტყვებების რაოდენობრივი გადანაწილება



კოლოკატების ანალიზისას ვასკვნით, რომ ქართველი სტუდენტები ჭარბად გამოიყენებენ *very* და *really* ზმნიშედებს მაშინ, როცა ამერიკელ სტუდენტთა შორის *very* და *highly* + ზედსართავი სახელი ყველაზე მაღალი სიხშირით გამოირჩევა. ეს უკანასკნელი მოდელი GLEAN სასწავლო კორპუსში საერთოდ არ დასტურდება. კორპუსების მონაცემებს შორის დრამატული სხვაობა სახეზეა *entirely* + ზედსართავი

სახელის კოლოკაციური წყვილის შედარებისას, რადგან ინგლისური ენის ქართველ შემსწავლელებში ეს კონკრეტული შესიტყვებითი პარადიგმა მნიშვნელოვნად ჩამოუვარდება MICUSP სასწავლო კორპუსის შედეგს. მკვეთრი განსხვავება შეინიშნება utterly + ზედსართავი სახელის შესიტყვებითი წყვილის შედარებისას, კერძოდ, GLEAN სასწავლო კორპუსში ასეთი სტრუქტურა საერთოდ არ ფიქსირდება მაშინ, როდესაც ინგლისურ ენის, როგორც მშობლიურ ენაზე მოსაუბრეთა ენობრივ რეპერტუარში utterly + ზედსართავი სახელი ხშირად გამოიყენება. ორ სასწავლო კორპუსს შორის თითქმის თანაბარი სიხშირული შედეგი მივიღეთ completely + ზედსართავი სახელის კონტრასტული ანალიზისას.

მეორე მხრივ, კვლევისას ზოგადი სემანტიკური და მორფოლოგიური თავისებურებები გამოიკვეთა. ინგლისური ენის ქართველ შემსწავლელთა კორპუსულ მასალაში *completely* დაკავშირებულია ისეთ ზედსართავ სახელებთან, რომლებიც ნეიტრალური კონოტაციის მატარებლები არიან. მაშინ, როცა ამერიკელი სტუდენტები აღნიშნულ ზმნიზედას უარყოფითი სემანტიკური ასოციაციით გამოიყენებენ (ზედსართავ სახელებთან პრეფიქსით un-, in-). ზმნიზედა *perfectly* სრულიად დადებითი კონოტაციური შინაარსი ჰქონდა ამერიკელი სტუდენტების ესეებში, გასხვავებით ქართველი სტუდენტების სასწავლო კორპუსის მონაცემებისა. ამ უკანასკნელში, *perfectly* უარყოფითი კონოტაციითაა წარმოდგენილი კონტექსტებში, კერძოდ, *perfectly unimportant*, *perfectly invisible*. ზმნიზედა *extremely*-ს მიზიგანის სასწავლო კორპუსში დადებითი ასოციაციების მქონე ზედსართავი სახელები მოსდევს, როგორიცაა „დიდსულოვანი“, „კაშკაშა“ და შედეგად ვიღებთ ამდაგვარ შესიტყვებებს: *extremely illuminating*, *extremely magnanimous*, *extremely feminine*. ხოლო ამის საპირისპიროდ, ქართველი შემსწავლელების კორპუსში *extremely* მოსდევს მორფოლოგიურად და სემანტიკურად უარყოფითი ასოციაციის მქონე ზედსართავი სახელები „არაეფექტური“, „არაჯანსაღი“, „უპასუხისმგებლო“ და შედეგად ვიღებთ ასეთ კოლოკაციურ წყვილებს: *extremely inefficient*, *extremely unhealthy*, *extremely irresponsible*.

3.2. სამ- და ოთხკომპონენტური ლექსიკური ფრაზების რაოდენობრივი ანალიზი

კორპუსული ძიების დროს ტექსტის დანაწევრება ხდება ფრაგმენტებად, რომელიც მეორე ეტაპზე ელემენტებად, ენ-გრამებად (n-gram) ჯგუფდება. ისინი სამ-, ოთხ-, ხუთკომპონენტური ლექსიკური ფრაზების (კონების) მოძიებისა და დადგენის შესაძლებლობას გვაძლევს. ასეთი სახის ენობრივ ერთეულებს სიტყვის კლასტერებს ან ლექსიკურ კონებსაც უწოდებენ, რომლებიც ენის ბუნებრივ იერარქიაში სიტყვებზე მაღლა დგანან. ისინი ერთმანეთთან მიმდევრობით, სხვა ლემის ჩარევის გარეშე ფუნქციონირებენ (კენჭიაშვილი, 2020: 162).

ჩვენს კვლევაში ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში მრავალსიტყვიანი ლექსიკური ერთეულების გამოსავლენად ვიყენებთ Antconc (ენტონი, 2011) პროგრამას. დაჯგუფებული კომპონენტები ძირითადად ორიდან ხუთ ლექსიკურ ელემენტამდე მერყეობს. ჩვენ აქცენტს გავაკეთებთ სამ- და ოთხკომპონენტურ ლექსიკურ კონებზე, რადგან დისკურსში მათი ფუნქციური მახასიათებლები მკვეთრად გამოხატულია.

ვინაიდან, GLEAN კორპუსი ანოტირებულია და თითოეულ მეტყველების ნაწილს საგანგებოდ განკუთვნილი მარკერი აქვს, მასში ცალკეული მეტყველების ნაწილების მოძიება ენ-გრამების მეშვეობით არის შესაძლებელი. საკვანძო სიტყვასა და მასთან მიმდევრობაში მყოფი კომპონენტების ძიებისათვის გამოვიყენეთ რამდენიმე ტეგი: ა) *_APPGE* კუთვნილებითი ნაცვალსახელის აღმოსაჩენად; ბ) *_CS* დაქვემდებარებული კავშირის მოსაძიებლად; გ) *_II* წინდებულისათვის; დ) *_DDQ* Wh-შემცველი მსაზღვრელი სიტყვის მოსაძიებლად; ე) *_IO* კონკრეტულად (of) წინდებულისათვის.

წარმოვადგენთ მოძიებული ლექსიკური ფრაზების კონტექსტუალურ რეალიზაციას (იხ. სურათი 13).

სურათი 13. სამ- და ოთხკომპონენტიანი ლექსიკური ფრაზების ძიება AntConc-ით (ენტონი, 2011)

	File	Left Context	Hit	Right Context
11	Agenda_PO...	Izoria said . He further noted that " in the	context of	enhancing the security environment , we believe it
12	Agenda_PO...	principle was " all the more important " in the	context of	the ongoing Russian military aggression against U
13	Agenda_PO...	, " resolve " and " vision " in the	context of	Russia 's ongoing invasion of Ukraine and a
14	Agenda_PO...	e occupation was a " special circumstance " in the	context of	the support for Ukraine . Everyone - including
15	Agenda_PO...	on Monday , was " especially important " in the	context of	the country 's anticipation of the EU decision
16	Agenda_PO...	's participation was " especially important " in the	context of	the country 's anticipation of the EU decision
17	Agenda_PO...	cological , urban , and demographic shifts " in the	context of	political transformation and climate change , its or
18	Agenda_PO...	t Georgia 's " particularly difficult situation " in the	context of	the war in Ukraine , Davitashvili said the
19	Agenda_PO...	also noted Germany 's " important role " in the	success of	Georgia 's Euro-Atlantic integration process . Phot
20	Agenda_PO...	one of the most " tragic days " in the	history of	the country . The 70-year rule of the
21	Agenda_PO...	Azerbaijan had played a " huge role " in the	history of	Dinamo so it was an " honour "
22	Agenda_PO...	distinguished by " unprecedented building " in the	history of	the Georgian church . Hundreds of churches have

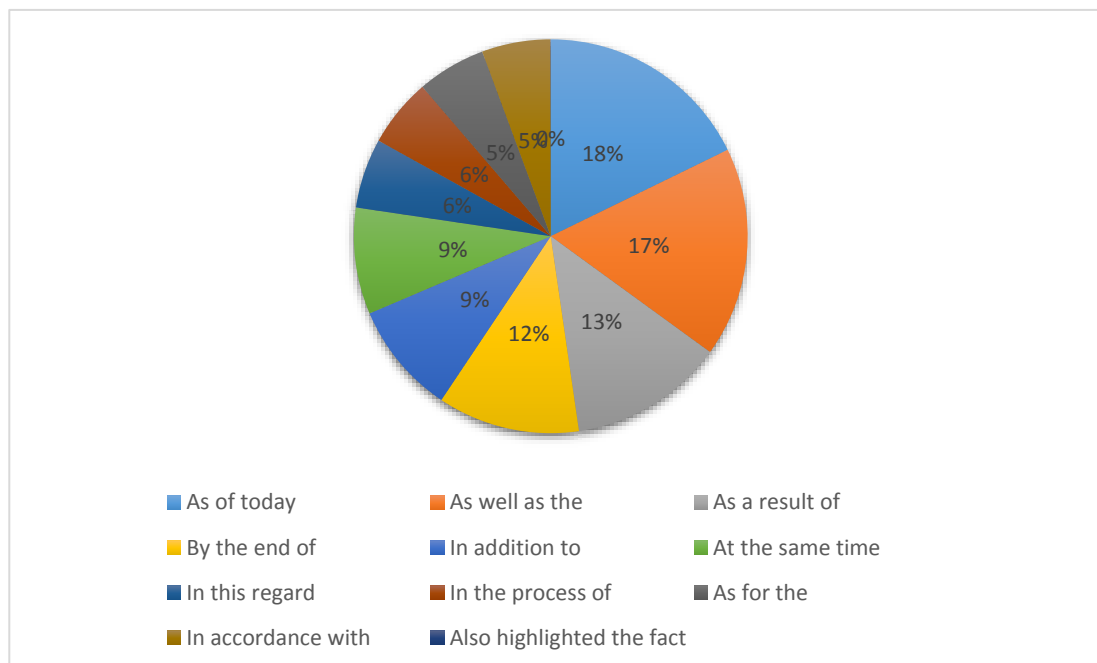
როგორც ვხედავთ, ზემოთ განხილული ტეგების მითითების შემდეგ დავალაგეთ ლექსიკურ ფრაზებში არსებული კომპონენტების პოზიცია და მიმდევრობა. საბოლოოდ, ენ-გრამების ანალიზის გზით სიხშირის პრინციპზე დამყარებით გამოვყავით 42 სამ- ან ოთხკომპონენტიანი ლექსიკური ფრაზა, რომლებსაც ინგლისური ენის ქართველ შემსწავლელები თავიანთ რეპერტუარში იყენებენ (იხ. სქემა 23).

ყველაზე ხშირად ხმარებული ლექსემატა კლასტერების ინგლისური ენის ქართველ შემსწავლელთა კორპუსში პროცენტული წილი ნაჩვენებია დიაგრამაში (იხ. დიაგრამა 5). სიხშირულმა ანალიზმა აჩვენა, რომ ყველაზე ხშირად ხმარებული სამკომპონენტიან ლექსემატა კლასტერი არის *as of today*, ნორმალიზებული სიხშირით 123.6 (pmw), რომელიც ძირითადად პრესის მასალაში დასტურდება. *also highlighted the fact* ფრაზა კი ნორმალიზებული სიხშირით 0.205 ჩვენ მიერ მანუალურად შერჩეულ ლექსიკურ კონებს შორის ბოლო პოზიციას იკავებს.

სქემა 23. GLEAN სასწავლო კორპუსში გამოვლენილი სამ- და ოთხკომპონენტიანი ლექსიკური ფრაზები

<i>as of today</i>	<i>as part of the</i>	<i>as a result of</i>
<i>by the end of</i>	<i>one of the most</i>	<i>as a consequence of</i>
<i>at the beginning of</i>	<i>as one of the</i>	<i>by the fact that,</i>
<i>in the middle of</i>	<i>in a number of</i>	<i>due to the fact</i>
<i>at the heart of</i>	<i>in the context of</i>	<i>in terms of the</i>
<i>on top of that</i>	<i>the nature of the</i>	<i>as noted earlier</i>
<i>at the same time</i>	<i>in the case of</i>	<i>in all likelihood</i>
<i>on the backdrop of</i>	<i>the relationship between the</i>	<i>no doubt that</i>
<i>in the process of</i>	<i>as well as the</i>	<i>it is clear that</i>
<i>in the wake of</i>	<i>as for the</i>	<i>on the off chance</i>
<i>the extent to which</i>	<i>on the contrary</i>	<i>it is possible that</i>
<i>in addition to</i>	<i>on the one hand</i>	<i>of the utmost importance</i>
<i>in an attempt to</i>	<i>in comparison to</i>	<i>goes without saying that</i>
<i>in comparison to the</i>	<i>on the grounds of</i>	
<i>in honour of</i>		

დიაგრამა 5. GLEAN კორპუსში ყველაზე ხშირად ხმარებული სამ- ან ოთხკომპონენტიანი ლექსიკური კონეზი



3.3. სამ- და ოთხკომპონენტიანი ლექსიკური ფრაზების ფუნქციურ-დისკურსული ანალიზი

მრავალკომპონენტიანი ლექსიკური ფრაზები, რომელიც ერთი შეხედვით არასპეციალიზებულ, ყოველდღიურ ლექსიკას განეკუთვნებიან, პარადოქსულად, აკადემიური ტექსტის შემადგენელი ნაწილები არიან. ასეთ ფართო ლექსიკას და მისი სისტემის კომპონენტებს შორის თვალსაჩინოდ ფიგურირებენ ლექსიკური ფრაზები, რომლებიც სამეტყველო აქტის პრაგმატულ ჩონჩხს ქმნიან. რაც ყველაზე მნიშვნელოვანია, მრავალკომპონენტიან ლექსემათა კლასტერები დანაწევრებული, მაგრამ ამავედროულად ერთი მთლიანობაა, რომელიც ხშირად გაუცნობიერებლად შეიმცნება და რეალიზდება მშობლიური ენის მატარებლებში. უცხოური ენის შემსწავლელთათვის კი სამ- ან ოთხკომპონენტიანი ლექსიკური ფრაზების დასწავლა რთულ კოგნიტურ პროცესებს მოიცავს. ამის დასტურია, კ. გლედჰილის ცნება „ფრაზეოლოგიური აქცენტი“, რომელიც მან სპეციალიზებული წერილობითი მასალის საფუძველზე დაამკვიდრა. გერმანელი აკადემიკოსებისა და ჟურნალისტების ტექსტებში ინგლისურისათვის უჩვეულო მრავალსიტყვიანი ლექსიკური ელემენტების გამოყენების სიუხვე გამოიკვეთა (გლედჰილი, 2000: 204). ასეთი ტიპის სინტაგმებს, მაგალითად, *it is argued that* ან *the same goes for* „მეორე დონის დისკურსის მარკერებადაც მოიხსენიებენ“ (სიპმანი, 2004: 104).

არსებობს მთელი რიგი დისკურსის მარკერები, რომლებიც გვხვდება როგორც ზეპირმეტყველებაში, ისე წერილობით მასალაში გვხვდება. მაგალითად ინგლისური მარკერები, როგორიცაა *of course* (რა თქმა უნდა), *in fact* (ფაქტობრივად), *in any case* (ნებისმიერ შემთხვევაში), ფრანგული - *justement* (ზუსტად), *franchement* (გულწრფელად რომ ვთქვათ) ან *en fait* (ფაქტობრივად), გერმანული *etwa* (დაახლოებით), *jedenfalls* (ყოველ შემთხვევაში) ან *immerhin* (სულ მცირე) როგორც მეტყველებაში, ისე წერილობით ტექსტებშიც საერთოა. ამიტომ, ლოგიკური იქნება თუ დისკურსის მარკერებს მხოლოდ ზეპირმეტყველების ენაში რეალიზებით არ შევზღუდავთ. დ. სიპმანის განმარტებით, „დისკურსის მარკერები სხვადასხვა სიგრძის ენობრივი გამოხატულებებია, რომლებიც პრაგმატულ ან/და პროპოზიციულ

მნიშვნელობას ატარებენ, როგორც მეტყველებაში, ისე წერილობით მასალაში და უფრო ხელს უწყობენ დისკურსის უწყვეტობას“ (სიპმანი, 2004: 43).

სასწავლო კორპუსში სიხშირული გზით მოპოვებული სამ- და ოთხკომპონენტური ლექსიკური კონების ფუნქციურ-დისკურსული ანალიზისათვის ვეფუძნებით ჯ. ო'ფლინის (2022: 82) მიერ შემოთავაზებულ ფუნქციურ ტაქსონომიას. გამოვყავით ლექსიკურ ფრაზათა რამდენიმე ფუნქციური კატეგორია: კვლევაზე ორიენტირებული, ტექსტზე ორიენტირებული, მონაწილეზე ორიენტირებული ლექსიკური კონები.

3.3.1. კვლევაზე ორიენტირებული ლექსიკური ფრაზები

სასწავლო კორპუსიდან ამოკრებილი 42 ლექსიკური ფრაზა მოვაქციეთ დისკურსულ-პრაგმატიკული ფუნქციების შესაბამის კატეგორიაში. კლასიფიკაციის პირველი ჯგუფი ისეთ ლექსიკურ კონებს მოიცავს, რომლებიც თავისი ფუნქციით ამა თუ იმ იდეის წარმოდგენასა და განხილვაში იღებს მონაწილეობს. კვლევისას ვეფუძნებით ჯ. ო'ფლინის ფუნქციურ ტაქსონომიას, რომელიც მორგებულია არა ზეპირმეტყველების დისკურსზე, არამედ მთავარ აქცენტს წერილობით ჟანრებზე აკეთებს (ო'ფლინი, 2022: 82).

სამიზნე ლექსიკური ფრაზების კონტექსტი ინგლისური ენის ქართულ შემსწავლელთა კორპუსის კონკორდანსის მეშვეობით გავაანალიზეთ და დავადგინეთ მათი პრაგმატული ფუნქციები. ლოკაციის ფუნქციით კორპუსში წარმოდგენილი კვლევაზე ორიენტირებული ლექსიკური კონები შემდეგია: *as of today, by the end of, at the beginning of, in the middle of, at the heart of, at the same time, on the backdrop of*. თითოეული მათგანი დისკურსში გამოიყენება როგორც დროის (1), ისე ადგილის (2) ფუნქციით. ქვემოთ მოცემულ კონტექსტებს შესაბამისი ფუნქციის პირობითად აღმნიშვნელ რიცხვს მივუწერთ.

(1) On one hand they protected Georgia's national interests and **at the same time** defended the rights of foreign investors [პრესა].

(ერთი მხრივ, მათ მხარი დაუჭირეს საქართველოს ეროვნულ ინტერესებს და, ამავდროულად, დაიცვეს უცხოელი ინვესტორების უფლებები).

(1) I will be able to turn in the assignment **by the end of** the day, apologies [კორესპონდენცია].

(დავალების შესრულებას დღის ბოლომდე შევძლებ, ბოდიში).

(2) Professor found the manuscript in the Lilly Library last year that included specially illustrated letters **at the beginning of** each paragraph [პრესა].

(პროფესორმა გასულ წელს იპოვა ხელნაწერი ლილის ბიბლიოთეკაში, რომელშიც თითოეული აბზაცის დასაწყისში საგანგებოდ ილუსტრირებული წერილები შედიოდა).

(2) We only learn her real last name as Link does **in the middle of** the novel also uses her car to escape from home and its implied class, racial and gender values [მიმოხილვითი ესე].

(მის ნამდვილ გვარს ლინკის მსგავსად მხოლოდ მაშინ ვიგებთ, როდესაც რომანის შუაში იგი სახლიდან თავისივე მანქანით გარბის და მიაწვდის კლასობრივ, რასობრივ და გენდერულ ფასეულობებზე).

(1) **As of today** the city does not have a proper water supply system. [პრესა]

(დღევანდელი მდგომარეობით ქალაქს არ აქვს წყლის მიწოდების სათანადო სისტემა).

(2) Polyphonic singing tradition is a tradition **at the heart of** Georgian cultural identity [ბლოგი].

(პოლიფონიური სიმღერის ტრადიცია ქართული კულტურული იდენტობის წიაღში არსებული ტრადიციაა)

(1) Moll's occupation as a thief necessitates her isolation from society, yet **at the same time**, her bond with her husband seems to be one of the most genuine in the book [მიმოხილვითი ესე].

(მოლის, როგორც ქურდის საქმიანობა, საზოგადოებისგან იზოლირების აუცილებლობას წარმოადგენს, თუმცა, იმავდროულად, მისი კავშირი ქმართან, როგორც ჩანს, ერთ-ერთი ყველაზე გულწრფელია წიგნში).

(2) The plot focused on his long-standing and new relationships with locals **on the backdrop of** a conflict [ბლოგი].

(სიუჟეტი ყურადღებას ამახვილებდა ადგილობრივებს შორის კონფლიქტის ფონზე არსებულ ძველ და ახალ ურთიერთობებზე).

დროისა ფუნქციის მქონე ლექსიკური ფრაზები ძირითადად პრესის მასალასა და კორესპონდენციაში დასტურდება, ხოლო ადგილის ფუნქციის მატარებელი მრავალსიტყვიანი ლექსიკური ერთეულები ძირითადად არგუმენტირებულ და აღწერით ესეებში გვხვდება.

ტაქსონომიაში კვლევაზე ორიენტირებული ფუნქციის შემდეგი კატეგორია პროცედურული (3) ლექსიკური ფრაზებია, მაგალითად *in the process of*, *in the wake of*, რომელიც, ძირითადად, პოლიტიკურ საგაზეთო სტატიებში დადასტურდა.

(3) Russian officials called for a boycott of Georgian resorts and trade goods **in the wake of** the protests [საინფორმაციო ქრონიკა].

(რუსი თანამდებობის პირები საპროტესტო აქციების კვალდაკვალ საქართველოს კურორტებისა და სავაჭრო საქონლის ბოიკოტისკენ მოუწოდებდნენ).

(3) Contemporary art begins to unexpectedly flourish **in the wake of** the country's freeing itself from the Soviet Union [რეპორტაჟი].

(თანამედროვე ხელოვნება მოულოდნელად აყვავებას იწყებს ქვეყნის კვალდაკვალ, საბჭოთა კავშირისგან გათავისუფლების შემდეგ).

(3) This University performed, performs and will play a key role **in the process of** establishing the identity of our country [რეპორტაჟი].

(ეს უნივერსიტეტი ჩვენი ქვეყნის თვითმყოფადობის დამკვიდრების პროცესში საკვანძო როლს ასრულებდა, ასრულებს და შეასრულებს მომავალშიც).

რაოდენობის ამსახველი ფუნქციის მატარებელი ლექსიკური კონები კონტექსტებში როგორც აბსტრაქტული (4) შინაარსით გვხვდებიან, ისე ზუსტად განსაზღვრული საზომით ან რაოდენობით (5). რაოდენობის ასახვის აბსტრაქტული სემანტიკის ლექსიკური ფრაზები *the extent to which*, *as part of the* ძირითადად

კვლევით ნაშრომებსა და არგუმენტირებულ ესეებში შეინიშნება, ხოლო უფრო კონკრეტული ოდენობის აღმნიშვნელი ლექსიკური ფრაზა პრესის მასალაში დასტურდება: *one of the most, as one of the, in a number of*. აბსტრაქტული შინაარსის მქონე ლექსიკური კონების სიხშირე კორპუსში იმით აიხსნება, რომ ინგლისური ენის ქართველი შემსწავლელების ესეები და კვლევითი ნაშრომები ჰუმანიტარულ მეცნიერებებს განეკუთვნება და არა ეკონომიკას ან სხვა ტექნიკურ დარგებს.

(4) Thousands of archived items have been restored within the past year **as part of the** National Archives of Georgia project [საინფორმაციო ქრონიკა].

(ათასობით დაარქივებული ნივთი რესტავრირდა ბოლო წლის განმავლობაში, საქართველოს ეროვნული არქივის პროექტის ფარგლებში).

(4) This revelation on James's part forecasts **the extent to which** her tragic story will come to affect James [არგუმენტირებული ესე].

(ეს გამოცხადება ჯეიმსზე ნაწილობრივ იმის მაუწყებელია, რამდენად იმოქმედებს მასზე მისი ტრაგიკული ამბავი).

(4) For example, Astin (1985) attributes student development to **the extent to which** students are involved in activities [სამეცნიერო სტატია].

(მაგალითად, ასტინი (1985) მიაწერს სტუდენტის განვითარებას იმას, თუ რამდენად არიან ისინი ჩართულნი აქტივობებში).

(5) Three Georgian films will be shown at **one of the most** prestigious film events in Central and Eastern Europe [საინფორმაციო ქრონიკა].

ცენტრალური და აღმოსავლეთ ევროპის ერთ-ერთ ყველაზე პრესტიჟულ კინო ღონისძიებაზე სამ ქართულ ფილმს აჩვენებენ.

(5) Between Georgia and the US, and bilateral ties could be found **in a number of** ways [პრესა].

(საქართველოსა და აშშ-ს შორის ორმხრივი კავშირები არაერთი გზით შეიძლება დადგინდეს).

ტაქსონომიის კვლევაზე ორიენტირებული ფუნქციის ბოლო კატეგორია აბსტრაქტული აღწერაა (6), რომელსაც მივაკუთვნეთ შემდეგი ლექსიკური ფრაზები –

in the context of, the nature of the, in the case of. კვლავ, კონკრეტიკას მოკლებული ლექსიკური კლასტერები მიაწოდებს ესეთა და პრესის ჟანრში აბსტრაქტული კონსტრუქციების არსებობაზე.

(6) The photograph realizes **the nature of the** original bond between the man and the woman [აღწერიითი ესე].

(ფოტო ასახავს ქალსა და მამაკაცს შორის არსებულ პირველქმნილი კავშირის ბუნებას).

(6) More importantly, **in the context of** Antin's book, the presence of the letters creates a physical connection between Antin, the author, and the domestic [მიმოხილვითი ესე].

(რაც უფრო მნიშვნელოვანია, ანტინის წიგნის კონტექსტში ასოების არსებობა ქმნის ფიზიკურ კავშირს ანტინს, ავტორსა და შინაურს შორის).

3.3.2. ტექსტზე ორიენტირებული ლექსიკური ფრაზები

კლასიფიკაციის მეორე ჯგუფი ისეთ ლექსიკურ ფრაზებს მოიცავს, რომლებიც ტექსტის ორგანიზებაზეა (7) პასუხისმგებელი, უფრო ზუსტად კი მათში ერთიანდება ტრანზიციული მარკერები, რომლებიც აზრთა ლოგიკური ხაზის გადასვლას ერთი პოზიციიდან, მდგომარეობიდან, სუბიექტიდან, მოსაზრებიდან მეორეზე გადატანას უზრუნველყოფენ. ამ კლასში საკმაოდ დიდი რაოდენობის 3- ან ოთხკომპონენტური ლექსემები ერთიანდება, რაც ლოგიკურია იქიდან გამომდინარე, რომ აკადემიური წერა, განსაკუთრებით ლიტერატურასა და ლინგვისტიკაში, რიტორიკული მოქმედებაა და მოითხოვს მკაფიო და ჩამოყალიბებულ არგუმენტებს. აღნიშნულ კატეგორიაში ერთიანდება მომდევნო ფრაზები: *the relationship between the, as well as the, as for the, on the contrary, on the one hand, in comparison to.*

(7) The last sentence of the piece, as quoted below, conveys **the relationship between the** sitter and painter [მიმოხილვითი ესე].

(ნაწარმოების ბოლო წინადადება, როგორც ქვემოთ იყო მოყვანილი, გადმოსცემს კრიტიკოსსა და ფერმწერს შორის ურთიერთობას).

(7) Higher pay for teachers comes at a price for the country **as well as** for the teachers themselves [არგუმენტირებული ესე].

(მასწავლებლების მაღალი ანაზღაურებით ქვეყანაც ისეთივე ფასს იხდის, როგორც თავად მასწავლებელი).

(7) **On the contrary**, no matter how obscure his thoughts become, Stephen's musings always channel themselves back into the anxieties he feels [მიმოხილვითი ესე].

(პირიქით, როგორი ბუნდოვანიც არ უნდა გახდეს მისი აზრები, სტივენის ნააზრევის მოყვარულები ყოველთვის ისევ მის მიერ განცდილ შფოთვებში ინთქმებიან).

(7) **In comparison to** the first hour of the lectures, my Business English students in my second hour demonstrate a better understanding of the content I instruct [არგუმენტირებული ესე].

ლექციების პირველ საათთან შედარებით, ჩემი ბიზნეს ინგლისურის სტუდენტები ლექციის მეორე საათში მასალის უკეთ გაგებას ახდენენ.

ტექსტზე ორიენტირებული ლექსიკური ფრაზების მეორე ჯგუფი მიზეზ-შედეგობრივ (8) კავშირებს მოიცავს, მაგალითად, *on the grounds of, as a result of, as a consequence of*. რეზულტატური კავშირების კონტექსტების უმრავლესობა პოლიტიკურ სტატიებში გვხვდება. ინგლისური ენის ქართველი შემსწავლელები აღნიშნული ფუნქციის მქონე ლექსიკურ ფრაზებს სოციალური, პოლიტიკური ან ისტორიული მოვლენების შედეგებზე მსჯელობისას გამოიყენებენიან.

(8) This decision may be considered discrimination **on the grounds of** nationality, citizenship, origin or sexuality [რეპორტაჟი].

(ეს გადაწყვეტილება შესაძლოა მიჩნეულ იქნას დისკრიმინაციად ეროვნების, მოქალაქეობის, წარმოშობის ან სექსუალობის ნიადაგზე).

(8) **As a result of** an increased awareness of Georgian brands, demand for Georgian jewellery has increased on international markets [საინფორმაციო ქრონიკა].

(ქართული ბრენდების უფრო გაზრდილი ცნობადობის შედეგად საერთაშორისო ბაზრებზე ქართულ ძვირფასეულობაზე მოთხოვნა გაიზარდა).

(8) Two new pro-opposition TV channels emerged **as a consequence of** the conflict [რეპორტაჟი].

(კონფლიქტის შედეგად ოპოზიციის მომხრე ორი ახალი ტელეარხი გამოჩნდა).

ტექსტზე ორიენტირებული ლექსიკური ფრაზების მესამე ჯგუფი ამხსნელ-დამაზუსტებელი (9) კავშირებს მოიცავს და ვხვდებით ძირითადად კორესპონდენციისა და პოლიტიკური პრესის კონტექსტებში, მაგალითად, *by the fact that, due to the fact, in terms of the, as noted earlier*:

(9) The importance of the document is also underlined **by the fact that** it served as the basis for the modern-day constitution of Georgia [რეპორტაჟი].

(დოკუმენტის მნიშვნელობას ისიც ხაზს უსვამს, რომ ის საქართველოს თანამედროვე კონსტიტუციის საფუძველია).

(9) Foreign Ministry had achieved significant progress **in terms of the** country's Euro-Atlantic integration [რეპორტაჟი].

(საგარეო საქმეთა სამინისტრომ მნიშვნელოვან პროგრესს მიაღწია ქვეყნისა და ევროატლანტიკური ინტეგრაციის თვალსაზრისით).

(9) Unfortunately sent files don't open perfectly on iOS **due to the fact** that it is written in my phone [კორესპონდენცია].

(სამწუხაროდ გამოგზავნილი ფაილები iOS-ზე უპრობლემოდ არ იხსნება, რადგან ჩემს მობილურ ტელეფონშია შესრულებული).

(9) Much like Clytemnestra, the Chorus does not believe her ominous words that a great evil is being plotted in this house **due to the fact** that she is merely a slave [სამეცნიერო სტატია].

(ისევე, როგორც კლიტემნესტრას, ქოროსაც არ სჯერა ავისმომასწავებელი სიტყვების, რომ ამ სახლში დიდი ბოროტება იგეგმება იმის გამო, რომ თავად უბრალოდ მონაა).

(9) I was wondering if you could provide me with more details **in terms of** the conference fees and procedures [კორესპონდენცია].

(მაინტერესებდა, კონფერენციის საფასურისა და პროცედურების კუთხით უფრო მეტ დეტალს თუ მომაწვდიდით).

(9) **As noted earlier** this was a very fluid period in planning [კორესპონდენცია].

(როგორც ადრე აღვნიშნეთ, დაგეგმვის დროს ძალიან ცვალებადი პერიოდი იყო).

3.3.3 მონაწილეზე ორიენტირებული ლექსიკური ფრაზები

ფუნქციური ტაქსონომიის ბოლო ჯგუფი, მონაწილეზე ორიენტირებული მარკერები, ისეთ ლექსიკურ ფრაზებს მოიცავს, რომლებიც დისკურსში საკუთარი პოზიციის, დამოკიდებულების გამოხატვის ფუნქციას ასრულებენ. ამ ჯგუფის პირველ კატეგორიაში მოსაზრების გამომხატველი კავშირები ერთიანდება, რომლებიც დარწმუნების (10), შესაძლებლობისა (11) და მნიშვნელობის (12) ხარისხის მატარებლები არიან. ასეთი ტიპის ლექსიკური ფრაზები უნივერსალურია და ისეთ ჟანრებში გვხვდება, როგორიცაა ნარატიული, არგუმენტირებული ესეები და მიმოხილვები. ამას გარდა, ბლოგებსა და დღიურებშიც ხშირად გვვლინება. ინგლისური ენის ქართულ შემსწავლელთა კორპუსში ამ ფუნქციის მქონე შემდეგი ლექსიკური კონები გამოვლინდა: *in all likelihood, no doubt that, it is clear that, on the off chance, it is possible that, of the utmost importance, goes without saying that*.

(11) **In all likelihood** these will be delivered this year [კორესპონდენცია].

(დიდი ალბათობით წელს ესენი იქნება მოტანილი).

(10) I continue to have absolutely **no doubt that** writing courses should be nothing less than compulsory at universities and colleges [კორესპონდენცია].

(სავსებით არ მეპარება ეჭვი, რომ წერის კურსები უნივერსიტეტებსა და კოლეჯებში სავალდებულო უნდა იყოს).

(10) **It is clear that** love is not really a contributor to their marriage [მიმოხილვითი ესე].

(ნათელია, რომ სიყვარული მათი ქორწინების მიზეზი ნამდვილად არ არის).

(11) I was not in the mood to sit around for half the day **on the off chance** that I might meet someone half-way interesting [ნარატიული ესე].

(ისეთ ხასიათზე არ ვიყავი, ნახევარი დღე ვმჯდარიყავი იმ იმედით, რომ ვინმე ოდნავ საინტერესოს მაინც შევხვდებოდი).

(11) **It is possible that** the owners would not be entitled to the same rights if they engaged in illegal business actions [არგუმენტირებული ესე].

(შესაძლებელია, მფლობელებს არ ჰქონოდათ იგივე უფლებები, თუ უკანონო ქმედებებს ჩაიდენდნენ).

(12) It is **of the utmost importance** for the Georgians to become assimilated to the European culture as quickly as possible [არგუმენტირებული ესე].

(უაღრესად მნიშვნელოვანია, რომ ქართველები რაც შეიძლება სწრაფად ასიმილირდნენ ევროპულ კულტურაში).

(12) **It goes without saying** that medical expenses can be quite high [არგუმენტირებული ესე].

(თავისთავად ცხადია, რომ სამედიცინო ხარჯები შეიძლება საკმაოდ მაღალი იყოს).

მონაწილეზე ორიენტირებული მარკერების მეორე კატეგორიაში ერთიანდება ჩართულობის გამომხატველი კავშირები (13), რომლის ფუნქციაც პირდაპირ მკითხველთან კავშირის დამყარება და მიმართვაა, მაგალითად, consider this, let us imagine that, as you know, take a moment to:

(13) **Consider this** scenario: you are in a casino with a friend and suddenly, near the slot machines, bells and whistles and lights go off [თავისუფალი კომპოზიციის ესე].

(განვიხილოთ ეს სცენარი: მეგობართან ერთად კაზინოში ხარ და უცებ, სათამაშო ავტომატებთან, ზარები, სირენები და ნათურები ირთვება).

(13) **As you know**, my groupmates arrived on Monday afternoon [კორესპონდენცია].

(როგორც იცით, ჩემი ჯგუფელები ორშაბათს, ნაშუადღევს ჩამოვიდნენ).

(13) Let's **take a moment to** unpack these texts and see whether they survive the fire [მიმოხილვითი ესე].

(მოდით, ერთი წუთით გავანალიზოთ ეს ტექსტები და ვნახოთ გადაურჩებიან თუ არა ხანძარს).

ფუნქციურ ტაქსონომიაში კვლევაზე ორიენტირებული ლექსიკური ფრაზები, ლოკაციის, აბსტრაქტული აღწერისა და პროცედურული ფუნქციებით, ყველაზე ხშირად გვხვდება ინგლისური ენის ქართველ შემსწავლელთა კორპუსში. შემდეგ მოდის ტექსტზე ორიენტირებული ლექსიკური ფრაზები, მიზეზ-შედეგობრივი, ამხსნელ-დამაზუსტებელი და ტექსტის მაორგანიზებელი ფუნქციებით. ჩვენს სასწავლო კორპუსში ყველაზე ნაკლებად მონაწილეზე ორიენტირებული მარკერები გვხვდება და ამ ქვეჯგუფში გაერთიანებული ლექსიკური ფრაზები კონკრეტულად კორესპონდენციისა და თავისუფალი კომპოზიციის ესესთვისაა დამახასიათებელი.

3.4. ლექსიკური სიმდიდრის განსაზღვრა

3.4.1. „ფრაზეოლოგიური სათამაშო დათუნები“ GLEAN სასწავლო კორპუსში

კვლევაში კონტრასტული ანალიზის გზით შევისწავლით სამ- ან ოთხკომპონენტიანი ლექსიკური ფრაზების განაწილებას ინგლისური ენის ქართველ შემსწავლელთა და ამერიკელ სტუდენტთა (MICUSP) სასწავლო კორპუსში. ეს გადაწყვეტილება განპირობებულია იმ მიზნით, რომ განვსაზღვროთ რომელი ლექსიკური კლასტერებია ენის შემსწავლელთათვის ყველაზე ნაცნობი და მოსახერხებელი დისკურსის წარმოებისას. ამგვარ ფრაზებს ჰ. ჰასელგარდი „ფრაზეოლოგიურ სათამაშო დათუნებს“ უწოდებს (ჰასელგარდი, 2019: 339).

ჩვენ მიერ გამოყოფილი ლექსიკური ფრაზების სიხშირის შედარებამ სხვა პერსპექტივით დაგვანახა მათი ჭარბი ან იშვიათი ხმარება. ვინაიდან ერთმანეთს ვადარებთ ორი განსხვავებული მოცულობის სასწავლო კორპუსს, კონტრასტული ანალიზისას სიგნიფიკანტურობა შემოწმებულ იქნა log-likelihood სანდოობის ტესტით³ (იხ. სქემა 24).

³ [SIGIL: Corpus Frequency Wizard \(collocations.de\)](http://SIGIL: Corpus Frequency Wizard (collocations.de))

საინტერესოა, რომ ამერიკელ სტუდენტთა სასწავლო ლექსიკონის მასალასთან მიმართებით, ქართველ შემსწავლელებში იშვიათად ხმარებული ლექსიკური ფრაზების რაოდენობა მეტია ჭარბად ხმარებული ფრაზების რაოდენობაზე. კონტრასტული ანალიზის მიხედვით, ქართველი შემსწავლელები ხშირად მოიხმარენ კვლევაზე ორიენტირებულ ლექსიკური ფრაზებს, განსაკუთრებით ლოკაციისა და აბსტრაქტული აღწერის ფუნქციის მქონე ლექსიკურ კონებს.

სქემა 24. ლექსიკური ფრაზების ჭარბი/იშვიათი მოხმარება GLEAN სასწავლო კორპუსში log-likelihood სანდოობის ტესტით

	ლექსიკური ფრაზების ჭარბი მოხმარება	ლექსიკური ფრაზების იშვიათი მოხმარება
p≤0.0001		
p≤0.001	as of today by the end of on the backdrop of in the process of in the wake of in honour of as part of the one of the most as for the in response to the	At the beginning of On top of that At the same time The extent to which In addition to In an attempt to In comparison to The nature of the In the context of The relationship between It is clear that In terms of the No doubt that Due to the fact
p≤0.01	As well as the	In the middle of At the heart of Goes without saying that of the utmost importance
p≤0.5	As a result of	When it comes to

მაშასადამე, ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსში „ფრაზეოლოგიური სათამაშო დათუნიებად“ გვევლინება შემდეგი ლექსიკური ფრაზები: *as of today, by the end of, on the backdrop of, in the process of, in the wake of, in honour of, as part of the, one of the most, as for the, in response to the, as well as the, as a result of*. შედეგი ცხადყოფს, რომ ინგლისური ენის ქართულ შემსწავლელებს მკაფიო ფრაზეოლოგიური ფავორიტები აქვთ, რომლებსაც დისკურსში ყველაზე კომფორტულად იყენებენ და რომლებთანაც თავს „უსაფრთხოდ გრძნობენ“.

3.4.2. Type/Token Ratio და სიხშირული სია ინგლისური ენის ქართულ შემსწავლელთა კორპუსში

ლექსიკური მრავალფეროვნების განსაზღვრისათვის სასწავლო კორპუსში ერთმანეთს შეფაფარდეთ სხვადასხვა სიტყვების (ტიპების) რაოდენობა სიტყვათა (ტოკენების) საერთო რიცხვთან, ანუ ლექსიკური სიმდიდრე განისაზღვრა ტიპ-ტოკენის მიმართებით (Type/Token Ratio), სხვა სიტყვებით რომ ვთქვათ, კორპუსული მონაცემების ტიპებისა და ტოკენების რაოდენობების ურთიერთშესაბამისობის ინდექსით. მოცემული ცვლადი გამოითვლება შემდეგნაირად:

$$\text{Type/token ratio} = \frac{\text{ტიპების რაოდენობა კორპუსში}}{\text{ტოკენების რაოდენობა კორპუსში}}$$

ინგლისური ენის ქართულ შემსწავლელთა კორპუსში ტოკენების რაოდენობა არის 9,776,931, ხოლო ტიპების – 118,947. შედეგად, მივიღეთ 0.0122 ტიპი თითოეულ ტოკენში. აღნიშნული სიდიდე რაც უფრო ახლოსაა 1-თან, მით უფრო მრავალფეროვანია სასწავლო კორპუსის ლექსიკა. ლექსიკური სიმდიდრის ინდექსის დადგენისას ვიყენებთ შეფასების კიდევ ერთ საზომს ს. გრეინჯერის მიერ შემოთავაზებულ lemma/Token Ratio, ანუ სიტყვათფორმების საწყისის, ძირითადი ფორმების მიხედვით სიდიდის დადგენას. მოცემული ცვლადი კი გამოითვლება შემდეგნაირად:

$$\text{lemma/token ratio} = \frac{\text{ლემების რაოდენობა კორპუსში}}{\text{ტოკენების რაოდენობა კორპუსში}}$$

ინგლისური ენის ქართველ შემსწავლელთა კორპუსში ლემათა რაოდენობა არის 105,846, ხოლო ტოკენების 9,776,931. შედეგად მივიღეთ 0.0108 ლემა თითოეულ ტოკენში. აღნიშნული სიდიდე უფრო ზუსტ სურათს გვაძლევს, როგორია ინგლისური ენის ქართველ შემსწავლელთა ლექსიკური მრავალფეროვნება. როგორც ვხედავთ, TTR მეტია LTR-ზე, და განპირობებულია იმით, რომ სიტყვების ფლექსიური ფორმები პროგრამულად ერთი ტიპად არის აღქმული.

სურათი 14. ყველაზე ხშირად განმეორებადი „სიტყვათა ღრუბელი“ GLEAN კორპუსში (log-ratio)

3.5. შუალედური დასკვნა

მესამე თავის მიზანი იყო ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში ორკომპონენტიანი შესიტყვებებისა და სამ-, ოთხკომპონენტიანი ლექსიკური ერთეულების სტრუქტურული, სემანტიკური და დისკურსის ფუნქციური თავისებურებების გამოვლენა და შედარება როგორც ინგლისური ენის აზიელი შემსწავლელების, ისე ამერიკელი სტუდენტების მიერ რეალიზებულ ენასთან.

გამოვყავით ლექსიკური და გრამატიკული შესიტყვებების მოდელები. ლექსიკურ მოდელებში ვაერთიანებთ ისეთ ორკომპონენტიან შესიტყვებებს, რომლის ელემენტებიც შედგება სახელური, ზმნური, ადვერბული კოლოკაციური წყვილებისაგან: 1. ზმნა (V) + არსებითი სახელი (N); 2. არსებითი სახელი (N) + არსებითი სახელი (N); 3. ზედსართავი სახელი (ADJ) + არსებითი სახელი (N); 4. ზმნა (V) + ზმნიზედა (ADV); 5. ზმნიზედა (ADV) + ზმნა (V); 6. ზმნიზედა (ADV) + ზედსართავი სახელი (ADJ). ხოლო სამ-, ოთხკომპონენტიანი ფრაზემები წინდებულს და სხვა ტიპის ერთეულებსაც მოიცავენ და გრამატიკული კოლოკაციების მოდელებში ერთიანდებიან. კონტრასტულმა ანალიზმა, ნორმალიზებული სიხშირის ანალიზის შედეგად გვიჩვენა, რომ ინგლისური ენის ქართველი შემსწავლელები უფრო მრავალფეროვან შესიტყვებათა მოდელებს იყენებენ, ვიდრე აზიელი სტუდენტები. გამონაკლისია ზმნა + არსებითი სახელის კოლოკაციური წყვილი, რომელიც ICNALE სასწავლო კორპუსში ორჯერ მეტად ფიქსირდება. მიზეზ-შედეგობრივი კავშირის ძიებისას შეგვიძლია ვივარაუდოთ, რომ თანამედროვე ჩინურ ენაში არსებულ მრავალკატეგორიანი სიტყვების (multi-category word) და მათ შორის ყველაზე დომინანტური შესიტყვებითი წყვილის (V+N), ესოდენ გავრცელებულობაში. მშობლიურ ენაში (ჩინურში) ასეთი რთული გრამატიკული ფენომენის არსებობის შემდეგ, ინგლისურ ენაში ჩინელი სტუდენტებისათვის გაცილებით მარტივია ზმნა + არსებითი სახელის კონსტრუქციის ინტერნალიზაცია.

კიდევ ერთი ტენდენცია, რომელიც კონტრასტული ანალიზის დროს შეინიშნება დაკავშირებულია აზიელი სტუდენტების სასწავლო კორპუსში ზედსართავი სახელი + არსებითი სახელის, ზმნა + ზმნიზედისა და ზმნიზედა + ზედსართავი სახელის კოლოკაციური შესიტყვების დეფიციტთან. შესაძლოა, ეს გარემოება გამოწვეული იყოს ინგლისურსა და აზიურ ენებს შორის არსებული გრამატიკული განსხვავებებით.

მაგალითად, ჩინურის მსგავს ენებს ინგლისურთან შედარებით წინადადების სხვადასხვა სტრუქტურა აქვთ. ინგლისური ეყრდნობა სუბიექტი-ზმნა-ობიექტი წყობას, ხოლო ჩინური იყენებს თემა-სტრუქტურას. მრავალ ენაში, მათ შორის ჩინურშიც, ზმნებისა და ზედსართავების კლასიკური გაყოფა შეუძლებელია და მეტყველების ნაწილების სიღრმულ პლანში განხილვისას ზედსართავი სახელები უნდა გავაერთიანოთ ზმნებთან. ამრიგად, ინგლისური ენის ჩინელი შემსწავლელებისათვის სირთულეს წარმოადგენს ინგლისურად ჩასწვდენ ზედსართავი სახელების დაწყვილების ბუნებრივ დინებას.

თითოეული კოლოკაციური მოდელის კვლევისას გამოვიყენეთ მეორე თავში განხილული ასოციაციური საზომები. შერჩეულ იქნა რამდენიმე ასოციაციური საზომი, Loglik, MI, DICE, Z-score. GLEAN სასწავლო კორპუსში სახელური შესიტყვებები (N+N) ყველა ჟანრული კატეგორიაშია წარმოდგენილი: პოლიტიკურ სტატიებში, ბლოგებსა და ესეებში. შესიტყვებების წყვილებში ერთ-ერთ ერთეულად ლემა *life* რამდენჯერმე მეორდება. კორპუსში ADJ+N წყვილებში შეინიშნება *good* ლექსემის სიჭარბე. V+ADV კოლოკაციურ წყვილში ზმნიზედის შერჩევისას უპირატესობას ანიჭებენ ლოკატიურ ზმნიზედას, მაგალითად, ნიმუშებში სჭარბობს ლემები *together, forward, away*. ADV+V კორპუსული ანალიზის შედეგად დასტურდება, რომ GLEAN სასწავლო კორპუსში ქართველი შემსწავლელები აქტიურად იყენებენ შედარებითი ხარისხის ზმნიზედას, *further, less, more, higher*. კონტექსტუალური ანალიზით აღმოჩნდა, რომ ასეთი ტიპის პარადიგმები, კერძოდ *further + ზმნა* საგაზეთო პოლიტიკური სტატიების, დიპლომატიური დისკურსის მასალაში გვხვდება.

ვინაიდან ინგლისური ენის აზიელი შემსწავლელების სასწავლო კორპუსში ზმნიზედა + ზმნის შესიტყვებათა მოდელის ნიმუშები არ დასტურდება, მიზნად დავისახეთ შედარებითი ხარისხის სემანტიკის მქონე ზმნიზედა გადაგვემოწმებინა მიჩიგანის სასწავლო კორპუსში. *further enhance* დადასტურდა ერთხელ, სხვა დანარჩენი კოლოკაციების სიხშირე 0-ია. შეგვიძლია ვივარაუდოთ, რომ ინგლისური ენის ქართველ შემსწავლელებს ტრანსფერის გზით მშობლიური ენიდან უცხოურ ენაზე გადააქვთ აღნიშნული სტრუქტურული მოდელი „მეტად“ + ზმნა, მაგალითად, *მეტად განვითარდეს, მეტად გაუმჯობესდეს, მეტად გაღრმავდეს* და მსგავსი

შესიტყვებითი პარადიგმა ინგლისური ენისათვის ისეთივე დამახასიათებელი არაა, როგორც ქართული ენისათვის.

ზმნიზედა + ზედსართავი სახელის შესიტყვებების სიხშირული კვლევის მიხედვით ინგლისური ენის ქართველ შემსწავლელთა კორპუსში გამოიკვეთა გამაძლიერებელი სემანტიკის მქონე ზმნიზედების შესიტყვების ერთ-ერთ წევრად ყოფნის ტენდენცია. ესენია: *very, utterly, extremely, perfectly, absolutely, totally*. სიხშირული კვლევის შედეგად გამოიკვეთა, რომ ქართველი შემსწავლელები მნიშვნელოვნად დაბალი სიხშირით იყენებენ *-ly* ზმნიზედებს, ვიდრე სხვა მორფოლოგიური სტრუქტურის მქონე ზმნიზედებს.

ქართველი სტუდენტები ჭარბად გამოიყენებენ *very* და *really* ზმნიზედებს მაშინ, როცა ამერიკელ სტუდენტთა შორის *very* და *highly* + ზედსართავი სახელი ყველაზე მაღალი სიხშირით გამოირჩევა. კორპუსების მონაცემებს შორის დრამატული სხვაობა სახეზეა *entirely* + ზედსართავი სახელის კოლოკაციური წყვილის შედარებისას, რადგან ინგლისური ენის ქართველ შემსწავლელებში ეს კონკრეტული შესიტყვებითი პარადიგმა მნიშვნელოვნად ჩამოუვარდება MICUSP სასწავლო კორპუსის შედეგს. მკვეთრი განსხვავება შეინიშნება *utterly* + ზედსართავი სახელის შესიტყვებითი წყვილის შედარებისას, კერძოდ, GLEAN სასწავლო კორპუსში ასეთი სტრუქტურა საერთოდ არ ფიქსირდება მაშინ, როდესაც ინგლისურ ენის, როგორც მშობლიურ ენაზე მოსაუბრეთა ენობრივ რეპერტუარში *utterly* + ზედსართავი სახელი ხშირად გამოიყენება. ორ სასწავლო კორპუსს შორის თითქმის თანაბარი სიხშირული შედეგი მივიღეთ *completely* + ზედსართავი სახელის კონტრასტული ანალიზისას.

მეორე მხრივ, კვლევისას ზოგადი სემანტიკური და მორფოლოგიური თავისებურებები გამოიკვეთა. ინგლისური ენის ქართველ შემსწავლელთა კორპუსულ მასალაში *completely* დაკავშირებულია ისეთ ზედსართავ სახელებთან, რომლებიც ნეიტრალური კონოტაციის მატარებლები არიან. მაშინ, როცა ამერიკელი სტუდენტები აღნიშნულ ზმნიზედას უარყოფითი სემანტიკური ასოციაციით გამოიყენებენ (ზედსართავ სახელებთან პრეფიქსით *un-, in-*). ზმნიზედა *perfectly* სრულიად დადებითი კონოტაციური შინაარსი ჰქონდა ამერიკელი სტუდენტების ესეებში, გასხვავებით ქართველი სტუდენტების სასწავლო კორპუსის მონაცემებისა. ამ

უკანასკნელში, *perfectly* უარყოფითი კონოტაციითაა წარმოდგენილი კონტექსტებში, კერძოდ, *perfectly unimportant, perfectly invisible*. ზმნიზედა *extremely*-ს მიზიგანის სასწავლო კორპუსში დადებითი ასოციაციების მქონე ზედსართავი სახელები მოსდევს, როგორიცაა „დიდსულოვანი“, „კაშკაშა“ და შედეგად ვიღებთ ამდაგვარ შესიტყვებებს: *extremely illuminating, extremely magnanimous, extremely feminine*. ხოლო ამის საპირისპიროდ, ქართველი შემსწავლელების კორპუსში *extremely* მოსდევს მორფოლოგიურად და სემანტიკურად უარყოფითი ასოციაციის მქონე ზედსართავი სახელები „არაეფექტური“, „არაჯანსაღი“, „უპასუხისმგებლო“ და შედეგად ვიღებთ ასეთ კოლოკაციურ წყვილებს: *extremely inefficient, extremely unhealthy, extremely irresponsible*.

სასწავლო კორპუსში სიხშირული გზით მოპოვებული სამ- და ოთხკომპონენტიანი ლექსიკური კონების ფუნქციურ-დისკურსული ანალიზისათვის ვეფუძნებით ჯ. ო' ფლინის (2022: 82) მიერ შემოთავაზებულ ფუნქციურ ტაქსონომიას. გამოვყავით ლექსიკურ ფრაზათა რამდენიმე ფუნქციური კატეგორია: კვლევაზე ორიენტირებული, ტექსტზე ორიენტირებული, მონაწილეზე ორიენტირებული ლექსიკური კონები. ფუნქციურ ტაქსონომიაში კვლევაზე ორიენტირებული ლექსიკური ფრაზები, ლოკაციის, აბსტრაქტული აღწერისა და პროცედურული ფუნქციებით, ყველაზე ხშირად გვხვდება ინგლისური ენის ქართველ შემსწავლელთა კორპუსში. შემდეგ მოდის ტექსტზე ორიენტირებული ლექსიკური ფრაზები, მიზეზ-შედეგობრივი, ამხსნელ-დამაზუსტებელი და ტექსტის მაორგანიზებელი ფუნქციებით. ჩვენს სასწავლო კორპუსში ყველაზე ნაკლებად მონაწილეზე ორიენტირებული მარკერები გვხვდება და ამ ქვეჯგუფში გაერთიანებული ლექსიკური ფრაზები კონკრეტულად კორესპონდენციისა და თავისუფალი კომპოზიციის ესესთვისაა დამახასიათებელი.

სიხშირულმა ანალიზმა აჩვენა, რომ ყველაზე ხშირად ხმარებული სამკომპონენტიან ლექსებათა კლასტერი არის *as of today*, ნორმალიზებული სიხშირით 123.6 (pmw), რომელიც ძირითადად პრესის მასალაში დასტურდება. *also highlighted the fact* ფრაზა კი ნორმალიზებული სიხშირით 0.205 ჩვენ მიერ მანუალურად შერჩეულ ლექსიკურ კონებს შორის ბოლო პოზიციას იკავებს.

ჩვენ მიერ გამოყოფილი ლექსიკური ფრაზების სიხშირის შედარებამ სხვა პერსპექტივით დაგვანახა მათი ჭარბი ან იშვიათი ხმარება. ვინაიდან ერთმანეთს ვადარებთ ორი განსხვავებული მოცულობის სასწავლო კორპუსს, კონტრასტული ანალიზისას სიგნიფიკანტურობა შემოწმებულ იქნა log-likelihood სანდოობის ტესტით. მაშასადამე, ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში „ფრაზეოლოგიური სათამაშო დათუნიებად“ გვევლინება შემდეგი ლექსიკური ფრაზები: *as of today, by the end of, on the backdrop of, in the process of, in the wake of, in honour of, as part of the, one of the most, as for the, in response to the, as well as the, as a result of*. შედეგი ცხადყოფს, რომ ინგლისური ენის ქართველ შემსწავლელებს მკაფიო ფრაზეოლოგიური ფავორიტები აქვთ, რომლებსაც დისკურსში ყველაზე კომფორტულად იყენებენ და რომლებთანაც თავს „უსაფრთხოდ გრძნობენ“.

დაბოლოს, ლექსიკური სიმდიდრის გამოსათვლელად გამოვიყენეთ, როგორც TTR საზომი, ისე ს. გრეინჯერის მიერ შემოთავაზებული LTR სიდიდე. პირველ შემთხვევაში, შედეგად მივიღეთ 0.0122 ტიპი თითოეულ ტოკენში. ხოლო LTR შედეგი აღმოჩნდა 0.0108 ლემა თითოეულ ტოკენში. აღნიშნული სიდიდე რაც უფრო ახლოსაა 1-თან, მით უფრო მრავალფეროვანია სასწავლო კორპუსის ლექსიკა. ლოგიკურია დავასკვნათ, რომ ინგლისური ენის ქართველ შემსწავლელთა ლექსიკა მდიდარია.

IV თავი. ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსი (GLEAN), როგორც წყარო ინგლისურ-ქართულ სპეციალიზებული (EAP) ლექსიკონის სიტყვა-სტატიის შესადგენად

4. ზოგადი მიმოხილვა

კორპუსის, როგორც ენობრივი მონაცემების ამსახველი რესურსის, გამოყენება ნებისმიერ ლინგვისტურ თეორიასთან მიმართებით სავსებით შეთავსებადია. სხვა პერსპექტივით რომ შევხედოთ, მაგალითად, ლექსიკონების მხოლოდ კორპუსლინგვისტიკისა და კომპიუტერიზებული მეთოდებით შექმნას გარდაუვალი საფრთხეები სდევს თან. კობილდის შესიტყვებითი ლექსიკონის (სინკლერი, 1995) გამოცდილებამ, რომელიც უმეტესწილად ლექსიკოგრაფისგან ჩაურეველი მეთოდოლოგიით შეიქმნა, მრავალი ხარვეზი გვიჩვენა. შესიტყვებითი ლექსიკონი მოიცავს კოლოკაციურ წყვილებს (disaster relief, schools inspector), თავისუფალ კომბინაციებს (new მინიჭებული აქვს კოლოკატი gallery, nature შესიტყვებად მიჩნეულია because, ან such + fiasco და ა.შ.). როგორც აღნიშნული მაგალითებიდან ჩანს, კომპიუტერულ მეთოდებზე გადამეტებულად დაყრდნობა შესიტყვებების ასახვის დამახინჯებულ სურათს მოგვცემს და „ადამიანურ განსჯას მოკლებული და ჩაურეველი ლექსიკოგრაფიული მონაცემები ინგლისური ენის შემსწავლელებს შეცდომაში შეიყვანს.“ (სიპმანი, 2004: 8). მაშასადამე, ლექსიკოგრაფიული საქმიანობის წარმართვისათვის მხოლოდ სტატისტიკური რაოდენობრივი, მეთოდოლოგია საკმარისი არაა, მნიშვნელოვანია ლექსიკური სტრუქტურების თვისებრივი ანალიზიც.

ამ თავში წარმოვადგენთ ჩვენ მიერ შექმნილ სასწავლო კორპუსზე დაყრდნობით შესრულებულ რამდენიმე სამ- ან ოთხკომპონენტიანი ლექსიკური ფრაზების სალექსიკონო სიტყვა-სტატიას. აღნიშნული მასალა შესაძლებელია გახდეს სპეციალიზებული, აკადემიურ წერაზე ორიენტირებული (EAP) ლექსიკონის შემადგენელი ნაწილი. თითოეული სიტყვა-სტატიის მიკროსტრუქტურა მცირე მოდიფიკაციებით ეფუძნება დ. სიპმანის მიერ შემოთავაზებულ სამეთაურო ფრაზების ლექსიკონში ასახვის პრინციპებს. ლექსიკონის მაკროსტრუქტურის ჩამოყალიბებისას ვიყენებთ კონცეპტუალურ, თემატურ მეთოდს, რომელიც ემყარება როგორც ჯ.

ო'ფლინის ფუნქციურ ტაქსონომიას, ისე დ. სიპმანის დისკურსის მარკერების თემატურ-ფუნქციურ მწკრივებსაც.

4.1. ინგლისური ენის ქართულ შემსწავლელთა სასწავლო კორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები

ლექსიკოგრაფიული სამუშაოს პირველ ეტაპზე შერჩეულ იქნა რამდენიმეკომპონენტისანი ლექსიკური ფრაზების თემატურ-ფუნქციური ჯგუფები, როგორცაა კვლევაზე ორიენტირებული, ტექსტზე ორიენტირებული, მონაწილეზე ორიენტირებული ლექსიკური კონები და მათი ქვეჯგუფები: კონტრასტული, აღიარებითი, მაგალითით ილუსტრირების, ამხსნელი, განმმარტებელი, მთვლელი, შემაჯამებელი, მიზეზ-შედეგობრივი, მიზნობრივი, (თემის) მაინიცირებელი/შემცვლელი, თემიდან გადახვევის, კითხვა-პასუხის, აზრის აქცენტირებისთვის, ინფორმაციის მიმწოდებელი, დამაზუსტებელი, მიმანიშნებელი, ჰიპოთეზის, შემზღუდავი, ვისამე მიმაკუთვნებელი, ხელახლა ფორმულირების დისკურსის მარკერები. მას შემდეგ, რაც GLEAN სასწავლო კორპუსიდან მოპოვებული ლექსიკური ფრაზები მივაკუთვნეთ აღნიშნულ კატეგორიებსა და ქვეკატეგორიებს, (EAP) აკადემიურ წერაზე ორიენტირებული სპეციალიზებული ინგლისურ-ქართული ლექსიკონის მიკროსტრუქტურა შედგება ამ კომპონენტებისაგან:

1. თემატიკა (რომელ დისკურსის მარკერსაც განეკუთვნება მეთაური ფრაზა);
2. მეთაური ფრაზა (ინგლისური ლექსიკური ფრაზა);
3. მეთაური ფრაზის ფუნქციური აღწერა;
4. ინგლისური მეთაური ფრაზის ქართული ეკვივალენტი;
5. ინგლისური ლექსიკური ფრაზის ილუსტრაცია (GLEAN სასწავლო კორპუსიდან ამოღებული) ტექსტის ჟანრული კატეგორიის მითითებით;
6. ინგლისური ილუსტრაციის ქართულად თარგმნილი ვარიანტი;
7. ტექსტური გრაფა – შესაძლო შეცდომათა გამაფრთხილებლად (“Get it right box”)

ინგლისური ენის ქართველ შემსწავლელთა კორპუსის მიხედვით შედგენილი სიტყვა-სტატიის ნიმუში *as for the* ავტვირთეთ Lexonomy⁴-ზე (მეჩურა, 2017) (იხ. სურათი 15).

სურათი 15. სიტყვა-სტატია *as for the* სპეციალიზებული (EAP) ინგლისურ-ქართული ლექსიკონისათვის

თემის შემცველი მარკერი **as for the**

ტექსტზე ორიენტირებული, სტრუქტურირებისთვის
დამხმარე ლექსიკური ფრაზა

prep

— რაც შეეხება

- As for the annual inflation rate, it was mainly influenced by price changes in transport and housing [პრესა]
რაც შეეხება ყოველწლიური ინფლაციის მაჩვენებელს, მასზე ძირითადად გავლენას ახდენდა ტრანსპორტისა და საბინაო ფასების ცვლილება.
- As for the geopolitical situation, the report pointed out Georgia was exposed to geopolitical risks [პრესა]
რაც შეეხება გეოპოლიტიკურ ვითარებას, ანგარიშში აღნიშნული იყო საქართველოს წინაშე არსებული გეოპოლიტიკური რისკები.

► Do not use "what about" as an equivalent for "as for the". You can use the preposition as for for introducing a subject that is related to what you have just been discussing. This preposition often occurs at the beginning of a sentence.

⁴ <https://www.lexonomy.eu/>

სიტყვა-სტატია 2.

აზრის აქცენტირების მარკერი

goes without saying

მონაწილეზე ორიენტირებული, მოსაზრების გამომხატველი ლექსიკური ფრაზა

idm

— თავისთავად ცხადია

1. It goes without saying that we will miss him terribly. [სასაუბრო, კორესპონდენცია]

თავისთავად ცხადია, ის საშინლად მოგვენატრება.

2. It goes without saying that enhancing democracy in Georgia is a cornerstone of success [პუბლიცისტური, რეპორტაჟი].

თავისთავად ცხადია, საქართველოში დემოკრატიის გაძლიერება წარმატების ქვაკუთხედი.

□ You can use this lexical bundle in a context where sth. is unambiguous, perfectly clear, or self-evident; to be already acknowledged, established, or accepted. You can use "that is a matter of course" as an equivalent for "it goes without saying".

სიტყვა-სტატია 3.

მიმანიშნებელი მარკერი

in all likelihood

მონაწილეზე ორიენტირებული, ალბათობის, შესაძლებლობის გამომხატველი ლექსიკური ფრაზა

phrase

— ეტყობა, დიდი ალბათობით

1. In all likelihood, that document was considerably more massive than average [სასაუბრო, კორესპონდენცია].

ეტყობა, ის დოკუმენტი საშუალოზე საგრძნობლად დიდი იყო.

2. In all likelihood, the existing banking crisis will be defused [პუბლიცისტური, საინფორმაციო ქრონიკა].

დიდი ალბათობით, არსებული საბანკო კრიზისი განიმუხტება.

□ "In all likelihood" strengthens the idea that follows it, expressing a high degree of probability. You can use "in all probability" as an equivalent form for this lexical bundle.

Do not use an article in between the words "~~in all the likelihood~~".

სიტყვა-სტატია 4.

მიმანიშნებელი მარკერი

of the utmost importance

მონაწილეზე ორიენტირებული, ალბათობის, შესაძლებლობის გამომხატველი
ლექსიკური ფრაზა

phrase

— უაღრესად/განსაკუთრებით მნიშვნელოვანია

1. It is **of the utmost importance** for the immigrants to become assimilated to the European culture as quickly as possible [ესეისტიკა, არგუმენტირებული ესე].

უაღრესად მნიშვნელოვანია, რომ იმიგრანტები რაც შეიძლება სწრაფად
ასიმილირდნენ ევროპულ კულტურაში.

სიტყვა-სტატია 5.

მიმანიშნებელი მარკერი

one thing is certain

მონაწილეზე ორიენტირებული, ალბათობის, შესაძლებლობის გამომხატველი
ლექსიკური ფრაზა

phrase

— ერთი რამ აშკარაა/ცხადია

1. **One thing is certain**, we are entering a new age of environmental thought and businesses need to move constantly to keep up with the ever-changing market [ესეისტიკა, არგუმენტირებული ესე].

ერთი რამ ცხადია, ჩვენ ეკოლოგიური აზროვნების ახალ ხანაში შევდივართ და ბიზნესს განუწყვეტელი მოძრაობა სჭირდება, რათა მუდმივად ცვალებად ბაზარს არ ჩამორჩეს.

სიტყვა-სტატია 6.

მიმანიშნებელი მარკერი

on the off chance

მონაწილეზე ორიენტირებული, ალბათობის, შესაძლებლობის გამომხატველი ლექსიკური ფრაზა

phrase colloquial

— ყოველი შემთხვევისათვის, იმ იმედით რომ, ვინიგობაა

1. I was not in the mood to sit around for half the day on the off chance that I might meet someone half-way interesting [ესეისტიკა, ნარატიული ესე].

ისეთ ხასიათზე არ ვიყავი, ნახევარი დღე ვმჯდარიყავი იმ იმედით, რომ ვინმე ოდნავ საინტერესოს მაინც შევხვედროდი.

2. But on the off chance that someone else might find it useful for something I want to make it (a resource) available [სასაუბრო, კორესპონდენცია].

ვინიგობაა ვინმეს რაიმეში გამოადგეს, მინდა (რესურსი) ხელმისაწვდომი გავხადო.

სიტყვა-სტატია 7.

ჰიპოთეზის მარკერი

let us imagine that

მონაწილეზე ორიენტირებული, ჩართულობის გამომხატველი ლექსიკური ფრაზა

phrase

— წარმოვიდგინოთ, ვთქვათ

1. Let us imagine that the world decided today to abandon its appetite for sausages, rib and roast lamb [ესეისტიკა, თავისუფალი კომპოზიციის ესე]

წარმოვიდგინოთ, რომ მსოფლიომ დღეს გადაწყვიტა, უარი ეთქვა ძეხვის, ნეკნისა და შემწვარი ცხვრის ხორცის ჭამაზე.

2. Let us imagine that any victim of trauma would prefer to live in an imaginary world, better than live with reality [ესეისტიკა, არგუმენტირებული ესე]

წარმოვიდგინოთ, რომ ტრავმის ნებისმიერი მსხვერპლი წარმოსახვით სამყაროში ცხოვრებას ამჯობინებდა, რეალობაში ცხოვრებას ეს აჯობებდა.

სიტყვა-სტატია 8.

ჩარჩოში მომქცევი მარკერი

against the backdrop of

კვლევაზე ორიენტირებული, აბსტრაქტული აღწერის გამომხატველი ლექსიკური ფრაზა

phrase

— რისამე ფონზე

1. The classics were composed **against the backdrop of** growing civil war and political breakdown. [არგუმენტირებული ესე, MICUSP]

ეს კლასიკური ტექსტები შედგენილი იყო მზარდი სამოქალაქო ომებისა და პოლიტიკური კრიზისის ფონზე.

□ “Against the backdrop of” sets the context or background for the information that follows. You can use “in the context of” as an equivalent for this lexical bundle. Do not use ~~on the backdrop of~~.

სიტყვა-სტატია 9.

ჩარჩოში მომქცევი მარკერი

in the context of

კვლევაზე ორიენტირებული, აბსტრაქტული აღწერის გამომხატველი ლექსიკური ფრაზა

phrase

— რისამე კონტექსტში/ფარგლებში

1. Political polarizations and radicalisation were particularly palpable in the context of the self-government elections [პუბლიცისტური, საინფორმაციო ქრონიკა].

განსაკუთრებით საგრძნობი იყო პოლიტიკური პოლარიზაცია და რადიკალიზაცია თვითმმართველობის არჩევნების კონტექსტში.

2. Differences in terminology are especially obvious in the context of roads [აკადემიური, კვლევითი სტატია].

ტერმინოლოგიის განსხვავებები განსაკუთრებით აშკარაა გზების კონტექსტში.

სიტყვა-სტატია 10.

ჩარჩოში მომქცევი მარკერი

in the case of

კვლევაზე ორიენტირებული, აბსტრაქტული აღწერის გამომხატველი ლექსიკური ფრაზა

phrase

— რისამე შემთხვევაში

1. In the case of insolvency, in order to satisfy the creditors' requirements the legislation and enforcement mechanisms will also be improved [პუბლიცისტური, საინფორმაციო ქრონიკა].

გადახდისუნარიობის შემთხვევაში კრედიტორების დასაკმაყოფილებლად კანონმდებლობისა და სამართალდამცავი მექანიზმების გაუმჯობესებაც მოხდება.

2. Often, in the case of women, the trauma are associated with sexual violations perpetrated by men [ესეისტიკა, არგუმენტირებული ესე].

ხშირად, ქალების შემთხვევაში, ტრავმა მამაკაცების მიერ ჩადენილ სექსუალურ დარღვევებს უკავშირდება.

3. The situation was equally bad in the case of leisure facilities, like swimming pool, kids' space... [სასაუბრო, კორესპონდენცია].

ასეთივე ცუდი მდგომარეობა იყო დასასვენებელი ობიექტების შემთხვევაშიც, მაგალითად საცურაო აუზის, საბავშვო სივრცის...

სიტყვა-სტატია 11.

ჩარჩოში მომქცევი მარკერი

the nature of the

კვლევაზე ორიენტირებული, აბსტრაქტული აღწერის გამომხატველი ლექსიკური ფრაზა

phrase

— რისამე ხასიათი/ბუნება

1. The play between the sounds of the words, highlights the nature of the bond between the man and the woman [ესეისტიკა, აღწერითი ესე]

სიტყვების ბგერების თამაში, ხაზს უსვამს მამაკაცსა და ქალს შორის არსებული კავშირის ბუნებას.

2. The exhibition focused on exploring the nature of the alphabet system which, while based on adherence to rules, features the innate freedom of the language [ესეისტიკა, აღწერითი ესე]

გამოფენა ორიენტირებულია ანბანის სისტემის ბუნების შესწავლაზე, რომელიც წესებზე დაყრდნობისდა მიუხედავად, ენის თანდაყოლილ თავისუფლებას ასახავს.

სიტყვა-სტატია 12.

მიზეზ-შედეგობრივი მარკერი

as a result of

ტექსტზე ორიენტირებული, სტრუქტურირებისთვის დამხმარე ლექსიკური ფრაზა

phrase

— რისამე შედეგად

1. In particular, as a result of our joint effort, alcoholic drinks, mineral waters and agricultural products from Georgia exported to Japan [პუბლიცისტური, რეპორტაჟი].

კერძოდ, ჩვენი ერთობლივი ძალისხმევის შედეგად საქართველოდან იაპონიაში გატანილ იქნა ალკოჰოლური სასმელები, მინერალური წყლები და სასოფლო-სამეურნეო პროდუქცია.

2. As a result of feeling isolated, I feels meaningless, insignificant, but learn to embrace the reality. [სასაუბრო, ბლოგი].

იზოლირებულად ყოფნის შედეგად თავს ვგრძნობ უფუნქციოდ, უმნიშვნელოდ, მაგრამ ვსწავლობ ამ რეალობის გათავისებას.

სიტყვა-სტატია 13.

მიზეზ-შედეგობრივი მარკერი

as a consequence of

ტექსტზე ორიენტირებული, სტრუქტურირებისთვის დამხმარე ლექსიკური ფრაზა
phrase

— რისამე შედეგად, გამომდინარე

1. Illegal border crossing without any legal basis has become more frequent as a consequence of borderisation [პუბლიცისტური, საინფორმაციო ქრონიკა].

ბორდერიზაციის შედეგად, საზღვრის გადაკვეთა ყოველგვარი სამართლებრივი საფუძვლის გარეშე გახშირდა.

2. As a consequence of this, business owners must raise prices to compensate for the increased investment in security [ესეისტიკა, არგუმენტირებული ესე].

აქედან გამომდინარე, ბიზნესის მფლობელებმა უნდა გაზარდონ ფასები უსაფრთხოებაში გაზრდილი ინვესტიციის საკომპენსაციოდ.

სიტყვა-სტატია 14.

მიზეზ-შედეგობრივი მარკერი

on the grounds of

ტექსტზე ორიენტირებული, დასაბუთებისთვის განკუთვნილი ლექსიკური ფრაზა
phrase

— რისამე საფუძველზე/ნიადაგზე/მიზეზით

1. There must be no discrimination on the grounds of sexual orientation and gender identity. [ესეისტიკა, არგუმენტირებული ესე].

სექსუალური ორიენტაციისა და გენდერული იდენტობის ნიადაგზე დისკრიმინაცია არ უნდა არსებობდეს.

2. None of the disqualifications was on the grounds of vote buying, intimidation, violence, illegal promises, or other chicanery [პუბლიცისტური, პოლიტიკური სტატია].

არც ერთი დისკვალიფიკაცია არ იყო ხმების მოსყიდვის, დაშინების, ძალადობის, უკანონო დაპირებების ან სხვა მაქინაციის ნიადაგზე.

□ “On the grounds of” introduces the reason or justification for a statement that has just been made, explaining why the speaker believes something to be true. You can use “by reason of” as an equivalent.

სიტყვა-სტატია 15.

მიზეზ-შედეგობრივი მარკერი

when it comes to

ტექსტზე ორიენტირებული, დასაბუთებისთვის განკუთვნილი ლექსიკური ფრაზა
phrase

— როდესაც საქმე მიდგება/ეხება, რაც შეეხება

1. There must be no discrimination **on the grounds of** sexual orientation and gender identity [ესეისტიკა, არგუმენტირებული ესე].

სიტყვა-სტატია 16.

ადიტიური მარკერი

on top of that

ტექსტზე ორიენტირებული, მისი სტრუქტურირებისთვის დამხმარე ლექსიკური ფრაზა
phrase

— ამას გარდა, ამას ისიც ემატება

1. Georgian Dream already holds a parliamentary majority and on top of that, has an advantage in almost all the self governing bodies in the country [პუბლიცისტური, რეპორტაჟი].

ქართულ ოცნებას უკვე უჭირავს საპარლამენტო უმრავლესობა და ამას გარდა, უპირატესობა აქვს ქვეყნის თვითმმართველობის თითქმის ყველა ორგანოში.

2. Visitors will have to go into 14-day quarantine at their own expense, on top of that, they will also need to obtain travel insurance valid for six months [პუბლიცისტური, საინფორმაციო ქრონიკა].

ვიზიტორებს საკუთარი ხარჯით 14 დღიანი კარანტინის გავლა მოუწევთ, ამას ისიც დაემატება, რომ ექვსი თვით მგზავრობის დაზღვევის გადასახადის მოპოვებაც დასჭირდებათ.

□ “On top of that“ introduces additional information that is related to the point just mentioned, often emphasizing a negative consequence or burden.

სიტყვა-სტატია 17.

დროისმიერი მარკერი

at the same time

კვლევაზე ორიენტირებული, დროის განმსაზღვრელი ფუნქციის ლექსიკური ფრაზა
phrase

— ამავე დროს, ამასთან ერთად, ამასთანავე

1. You are serving the most difficult and a very responsible job at the same time [სასაუბრო, კორესპონდენცია]

თქვენ ერთდროულად ყველაზე რთულ და ძალიან საპასუხისმგებლო საქმეს ემსახურებით.

2. We made a firm decision to sign the Association Agreement with the European Union, at the same time we are trying to normalize relations with Russia [პუბლიცისტური, პოლიტიკური სტატია].

ევროკავშირთან ასოციაციის ხელშეკრულების გაფორმების მტკიცე გადაწყვეტილება მივიღეთ, ამავე დროს რუსეთთან ურთიერთობის ნორმალიზებას ვცდილობთ.

3. While reading Brooks’s letters in the archives, I was struck by the apparent and at the same time dubious statements about her own life and writing [ესეისტიკა, თავისუფალი კომპოზიციის ესე].

არქივებში ბრუქსის წერილების კითხვისას, გამოაცა საკუთარი ცხოვრებისა და მწერლობის მიმართ მისი გამონათქვამების სიცხადემ და ამავე დროს საეჭვოობამ.

□ “at the same time” is used to introduce a statement that slightly changes or contradicts the previous statement.

სიტყვა-სტატია 18.

დროისმიერი მარკერი

as of today

კვლევაზე ორიენტირებული, დროის განმსაზღვრელი ფუნქციის ლექსიკური ფრაზა
phrase

— დღეიდან, დღის მდგომარეობით

1. **As of today** Radisson operates all across the country [პუბლიცისტური, რეპორტაჟი].

დღევანდელი მდგომარეობით რედისონი ოპერირებს მთელი ქვეყნის მასშტაბით.

□ You can use “as of now” as a synonymous form for this lexical bundle.

სიტყვა-სტატია 19.

ადგილისმიერი მარკერი

at the heart of

კვლევაზე ორიენტირებული, ადგილის განმსაზღვრელი ფუნქციის ლექსიკური ფრაზა
phrase

— რისამე შუაგულში/ცენტრში; ეფუძნება

1. Georgian documentaries will be at the heart of an international film festival held in Switzerland next year [პუბლიცისტური, რეპორტაჟი].

ქართული დოკუმენტური ფილმები მომავალ წელს შვეიცარიაში გასამართ საერთაშორისო კინოფესტივალის შუაგულში იქნება წარმოდგენილი.

2. The author describes the aspect of Crusoe's personality that lies at the heart of who he actually is [ესეისტიკა, მიმოხილვითი ესე].

ავტორი აღწერს კრუზოს ხასიათის იმ ასპექტებს, რომელიც ეფუძნება მის ნამდვილ სახეს.

სიტყვა-სტატია 20.

პროცედურული მარკერი

in the wake of

კვლევაზე ორიენტირებული, პროცესის ამსახველი ფუნქციის მქონე ლექსიკური ფრაზა

phrase

— რისამე კვალდაკვალ, რისამე შედეგად

1. Russia continues violating the Russia-Georgia ceasefire deal signed in the wake of the Russia-Georgia 2008 war [პუბლიცისტური, რეპორტაჟი].

რუსეთი აგრძელებს რუსეთ-საქართველოს ცეცხლის შეწყვეტის შეთანხმებას, რომელსაც ხელი მოეწერა 2008 წლის ომის კვალდაკვალ.

2. Sukhishvilebi is embarking on the trip in the wake of successful performances in Israel and Bulgaria over the past two weeks [პუბლიცისტური, საინფორმაციო ქრონიკა].

სუხიშვილები ბოლო ორი კვირის განმავლობაში ისრაელსა და ბულგარეთში წარმატებული წარმოდგენების კვალდაკვალ მოგზაურობას იწყებენ.

სიტყვა-სტატია 21.

რაოდენობის მარკერი

as one of the

კვლევაზე ორიენტირებული, რაოდენობის განმსაზღვრელი ფუნქციის ლექსიკური ფრაზა

phrase

— ერთ-ერთი...

1. Bloom's admission of guilt to Martha serves as one of the most honest dialogues they have ever had [ესეისტიკა, მიმოხილვითი ესე]

ბლუმის მიერ დანაშაულის აღიარება მართას მიმართ წარმოადგენს ერთ-ერთ ყველაზე გულწრფელ დიალოგს, რაც კი მათ ოდესმე ჰქონიათ.

2. This manuscript was considered as one of the most important pieces in historic Western literature [აკადემიური, კვლევითი სტატია].

ეს ხელნაწერი ისტორიულ დასავლურ ლიტერატურაში ერთ-ერთ უმნიშვნელოვანეს ნაწილად მიიჩნეოდა.

□ Do not use singular noun after “as one of the member” , use a plural noun.

სიტყვა-სტატია 22.

რაოდენობის მარკერი

the extent to which

კვლევაზე ორიენტირებული, რაოდენობის განმსაზღვრელი ფუნქციის ლექსიკური ფრაზა

phrase

— თუ რამდენად...

1. The extent to which statistical, mathematical and computer-assisted techniques are used. [აკადემიური, კვლევითი სტატია].

რა ზომით გამოიყენება სტატისტიკური, მათემატიკური და კომპიუტერული მეთოდები.

2. The extent to which the researcher and researched can participate and collaborate in planning the research [აკადემიური, კვლევითი სტატია].

რამდენად შეუძლიათ მკვლევარსა და ცდის პირს მონაწილეობა და თანამშრომლობა კვლევის დაგეგმვისას.

სიტყვა-სტატია 23.

ხელახლა ფორმულირების მარკერი

as noted earlier

ტექსტზე ორიენტირებული, ამხსნელ-დამაზუსტებელი ლექსიკური ფრაზა

phrase

— როგორც უკვე აღინიშნა, როგორც ადრე აღინიშნა

1. **As noted earlier**, black women face steeper financial disadvantages and responsibility than their white counterparts [ესეისტიკა, არგუმენტირებული ესე].

როგორც ადრე აღინიშნა, შავკანიანი ქალები უფრო მკვეთრ ფინანსურ ზარალსა და პასუხისმგებლობას აწყდებიან, ვიდრე მათი თეთრკანიანი ქალები.

2. **As noted earlier** in this paper, some individuals felt that people would not be motivated to participate in a movement, research has shown that this is generally not the case [ესეისტიკა, არგუმენტირებული ესე].

როგორც უკვე ნაშრომში აღინიშნა, ზოგიერთი ინდივიდი ფიქრობდა, რომ ადამიანები მოტივირებული არ იქნებოდნენ, მონაწილეობა მიეღოთ მოძრაობაში, თუმცა კვლევებმა საპირისპირო აჩვენა.

სიტყვა-სტატია 24.

ხელახლა ფორმულირების მარკერი

in other words

ტექსტზე ორიენტირებული, ამხსნელ-დამაზუსტებელი ფუნქციის ლექსიკური ფრაზა
phrase

— ანუ; სხვა სიტყვებით რომ ვთქვათ, სხვაგვარად რომ ვთქვათ

1. In other words, perfect love does not actually exist, like a pair of parallel lines or two polarities, some things simply cannot be conjoined [სასაუბრო, ბლოგი].

ანუ სრულყოფილი სიყვარული სინამდვილეში არ არსებობს; როგორც წყვილი პარალელური ხაზი ან ორი პოლარულობა, ზოგიერთი რამ უბრალოდ ვერ შეერთდება.

2. This mass, in other words Chacha, was a source of secondary production, or wine falsification for many years [პუბლიცისტური, საინფორმაციო ქრონიკა].

ეს მასა, სხვაგვარად რომ ვთქვათ, ჭაჭა, მრავალი წლის განმავლობაში მეორადი წარმოების, ან ღვინის გაყალბების წყარო იყო.

სიტყვა-სტატია 25.

ხელახლა ფორმულირების მარკერი

put another way

ტექსტზე ორიენტირებული, ამხსნელ-დამაზუსტებელი ფუნქციის ლექსიკური ფრაზა
phrase

— სხვაგვარად, სხვა სიტყვებით რომ ვთქვათ

1. **Put another way**, no team is permitted to pick their opposition [ესეისტიკა, ნარატიული ესე].

სხვაგვარად რომ ვთქვათ, არც ერთ გუნდს არ აქვს უფლება, აირჩიოს მათი მოწინააღმდეგე.

სიტყვა-სტატია 26.

კონტრასტული მარკერი

on the contrary

ტექსტზე ორიენტირებული, ტექსტის ორგანიზების ფუნქციის ლექსიკური ფრაზა
phrase

— პირიქით, ამის საპირისპიროდ

1. Compromise is not a sign of weakness, on the contrary, the democracies that have been able to work together have turned out to be the strongest [პუბლიცისტური, საგაზეთო სტატია].

კომპრომისის სისუსტის ნიშანი არ არის, პირიქით, დემოკრატიულმა სახელმწიფოებმა, რომლებმაც ერთად მუშაობა შეძლეს, ყველაზე ძლიერები აღმოჩნდნენ.

2. On the contrary, no matter how obscure his thoughts become, Stephen's musings always channel themselves back into the anxieties he feels [ესეისტიკა, მიმოხილვითი ესე].

პირიქით, როგორი ბუნდოვანიც არ უნდა გახდეს მისი აზრები, სტივენისა ნააზრევის მოყვარულები ყოველთვის მის მიერვე განცდილ შფოთვებში ინთქმებიან.

3. National currency faces no fear in value, on the contrary, in the coming days it will become even stronger [პუბლიცისტური, საგაზეთო სტატია].

ეროვნული ვალუტის ღირებულებას საფრთხე არ ემუქრება, პირიქით, უახლოეს დღეებში ის კიდევ უფრო გაძლიერდება.

□ You can use “quite the contrary” as a synonymous form for this lexical bundle.

სიტყვა-სტატია 27.

კონტრასტული მარკერი

in comparison to

ტექსტზე ორიენტირებული, ტექსტის ორგანიზების ფუნქციის ლექსიკური ფრაზა
phrase

— რაიმესთან შედარებით, განსხვავებით

1. First, I talk about what I expected to find based on a sampling of the literature regarding fricatives in English, including what phonemes the language includes and my data **in comparison to** those expectations [აკადემიური, კვლევითი სტატია].

პირველ რიგში, ვსაუბრობ იმაზე, რის პოვნასაც ველოდი ინგლისურ ენაში ფრიკანტებთან დაკავშირებული ლიტერატურის ნიმუშის მიხედვით, მათ შორის იმაზე, თუ რა ფონემებს მოიცავს ენა და ჩემი მონაცემები ამ მოლოდინებთან შედარებით.

2. Revenue increased by 91 percent **in comparison to** January-February of 2022 [ესეისტიკა, აღწერითი ესე].

2022 წლის იანვარ-თებერვალთან შედარებით შემოსავლები 91 პროცენტით გაიზარდა.

3. The number of visitors has increased twice **in comparison to** the corresponding data for the second quarter of the previous year [ესეისტიკა, აღწერითი ესე].

წინა წლის მეორე მეოთხედის შესაბამის მონაცემებთან შედარებით ვიზიტორთა რაოდენობა ორჯერ გაიზარდა.

სიტყვა-სტატია 28.

კონტრასტული მარკერი

on the one hand

ტექსტზე ორიენტირებული, ტექსტის ორგანიზების ფუნქციის ლექსიკური ფრაზა
phrase

— ერთი მხრივ

1. My concern is that, **on the one hand**, the population should have an income, but this income should be valuable [ესეისტიკა, არგუმენტირებული ესე].

ჩემი სადარდებელი ისაა, რომ, ერთი მხრივ, მოსახლეობას შემოსავალი ჰქონდეს, თანაც ეს შემოსავალი ღირებული უნდა იყოს.

2. **On the one hand**, the educational system needs to deal with the diversity of languages, on the other hand, it must do this while working with limited financial resources [ესეისტიკა, არგუმენტირებული ესე].

ერთი მხრივ, საგანმანათლებლო სისტემას სჭირდება ენების მრავალფეროვნებასთან გამკლავება, მეორე მხრივ, ეს შეზღუდული ფინანსური რესურსების ფონზე უნდა შეძლოს.

□ “On (the) one hand”, “on the other hand”, are used to indicate two contrasted sides of a subject, circumstances, considerations, points of view, etc.

სიტყვა-სტატია 29.

შემზღუდავი მარკერი

in terms of the

ტექსტზე ორიენტირებული, ამხსნელ-დამაზუსტებელი ლექსიკური ფრაზა, რომელიც ზოგად მსჯელობას კონკრეტულ ცნებამდე ავიწროვებს

phrase

— რისამე თვალსაზრისით, რაიმესთან მიმართებით

1. In terms of the global energy, participants discussed what was the cheapest and the cleanest form of energy [ესეისტიკა, მიმოხილვითი ესე].

გლობალურ ენერგიასთან მიმართებით, მონაწილეებმა განიხილეს რა იყო ენერგიის ყველაზე იაფი და სუფთა ფორმა, ყველაზე ეფექტიანი ენერგია, ასევე, როგორ გაემრავალფეროვნებინათ ენერგომომარაგების მარშრუტები.

2. They agreed on full coordination in terms of the trade-economic ties between the countries [პუბლიცისტური, საგაზეთო სტატია].

ქვეყნებს შორის სავაჭრო-ეკონომიკური კავშირების თვალსაზრისით სრულ კოორდინაციაზე შეთანხმდნენ.

სიტყვა-სტატია 30.

ილუსტრირების მარკერი

as an example

ტექსტზე ორიენტირებული, ამხსნელ-დამაზუსტებელი ლექსიკური ფრაზა

phrase

— რისამე მაგალითად, მაგალითის სახით

1. Using Oedipus Rex as an example, the plot follows from the murder of Laius [აკადემიური, კვლევითი სტატია].

ოიდიპოს მეფის მაგალითზე, სიუჟეტური ქარგა იწყება ლაიუსის მკვლელობიდან.

2. I will take here as an example the noun “sky” [აკადემიური, კვლევითი სტატია]

ამის მაგალითად არსებითი სახელი „ცა“ გამოდგება.

4.2. შუალედური დასკვნა

GLEAN სასწავლო კორპუსიდან მოპოვებული ლექსიკური ფრაზები მივაკუთვნეთ პრაგმატიკულ-ფუნქციურ კატეგორიებსა და ქვეკატეგორიებს: კონტრასტული, აღიარებითი, მაგალითით ილუსტრირების, ამხსნელი, განმმარტებელი, მთვლელი, შემაჯამებელი, მიზეზ-შედეგობრივი, მიზნობრივი, (თემის) მაინიცირებელი/შემცვლელი, თემიდან გადახვევის, კითხვა-პასუხის, აზრის აქცენტირებისთვის, ინფორმაციის მიმწოდებელი, დამაზუსტებელი, მიმანიშნებელი, ჰიპოთეზის, შემზღუდავი, ვისამე მიმაკუთვნებელი, ხელახლა ფორმულირების დისკურსის მარკერები. შევარჩიეთ 30 ლექსიკური ფრაზა და (EAP) აკადემიურ წერაზე ორიენტირებული სპეციალიზებული ინგლისურ-ქართული ლექსიკონისათვის შევქმენით სიტყვა-სტატიები. მათი მიკროსტრუქტურა შედგება ამ კომპონენტებისაგან:

1. თემატიკა (რომელ დისკურსის მარკერსაც განეკუთვნება მეთაური ფრაზა);
2. მეთაური ფრაზა (ინგლისური ლექსიკური ფრაზა);
3. მეთაური ფრაზის ფუნქციური აღწერა;
4. ინგლისური მეთაური ფრაზის ქართული ეკვივალენტი;
5. ინგლისური ლექსიკური ფრაზის ილუსტრაცია (GLEAN სასწავლო კორპუსიდან ამოღებული) ტექსტის ჟანრული კატეგორიის მითითებით;
6. ინგლისური ილუსტრაციის ქართულად თარგმნილი ვარიანტი.

სალექსიკონო სიტყვა-სტატიებზე მუშაობამ გვიჩვენა, რომ ისეთ ლექსიკურ ფრაზებს, როგორებიცაა: as for the, it goes without saying, in all likelihood, against the backdrop of, on the grounds of, on top of that, at the same time, as of today, as one of the, on

the contrary, on the one hand, დასჭირდათ ტექსტური გრაფების მითითებაც, რომელიც გამაფრთხილებელი ფუნქციის მატარებელია. სასწავლო კორპუსში ზემოთ ჩამოთვლილ ლექსიკური კონების ხმარების დროს შეინიშნებოდა ენობრივი შეცდომა და ამიტომაც ასახვა ჰპოვა მიკროსტრუქტურაში. მაგალითად, against the backdrop of ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში შეცდომით დასტურდება on the backdrop of ფორმით, ამიტომ საილუსტრაციო მასალად მოვიხმეთ MICUSP სასწავლო კორპუსის, ამერიკელ სტუდენტთა კონტექსტი. ტექსტობრივ ყუთში კი მივუთითეთ შესაძლო ენობრივი ხარვეზის შესახებ. მაშასადამე, მაკროსტრუქტურა ეფუძნება დ. სიპმანის დისკურსის მარკერების ფუნქციურ ტაქსონომიას, ხოლო მიკროსტრუქტურულ ველებში მოცემული ფუნქციები და ამხსნელი მითითებები ასახავს ჯ. ო'ფლინისეულ ფუნქციურ ტაქსონომიას.

შეიძლება ითქვას, რომ ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსის მასალის ლექსიკოგრაფიული დამუშავება და მისი სპეციალიზებულ EAP ინგლისურ-ქართულ ლექსიკონში ასახვა მნიშვნელოვანია ინტერდისციპლინური კვლევების განსახორციელებლად. კორპუსი საშუალებას მოგვცემს თეორიული და პრაქტიკული ასპექტით სრულფასოვნად შევისწავლოთ და გამოვიკვლიოთ ინგლისური ენის შემსწავლელთა სიტყვათხმარებისა და ენობრივი სიმდიდრის თავისებურებები.

4.3. სამომავლო პერსპექტივები

ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსის კვლევის განვითარების მხრივ მნიშვნელოვნად მიგვაჩნია რამდენიმე ქმედითი ნაბიჯის გადადგმა. უპირველესად, ინგლისურ-ქართული ორენოვანი სასწავლო ლექსიკოგრაფიის განვითარებისათვის ღირებული იქნება ჩვენი სასწავლო კორპუსისათვის შეცდომათა ტეგერის შექმნა (error-tagger) და ქართველი ინგლისური ენის შემსწავლელების ენობრივი ლაფსუსების ამ გზით კლასიფიკაცია. როგორც წესი, შეცდომათა ანალიზის სქემა შედგება შეცდომათა ტაქსონომიებისგან, რომელიც ეფუძნება სტრუქტურულ-ენობრივ კატეგორიებს მართლწერას, გრამატიკას, ლექსიკასა თუ სტილს.

მეორე მხრივ, როგორც აღვნიშნეთ, დღესდღეობით, ჩვენი სასწავლო კორპუსი შედგება 10 მილიონამდე სიტყვისაგან, რომლის მუდმივად შევსება როგორც წერილობითი, ისე ზეპირმეტყველების მასალით უცხოური ენის კვლევის მიზნებისათვის პერსპექტიულად გვესახება. სავსებით შესაძლებელია სასწავლო კორპუსს დაემატოს ინგლისურენოვანი საგამოცდო ნაწერები, უფრო მეტი რაოდენობის კვლევითი სტატია, იმ მიზნით, რომ GLEAN-მა ენის შემსწავლელთა უფრო ფართო აუდიტორია მოიცვას. სასწავლო კორპუსის კონკორდანსი, რომელიც სიტყვების უბრალო ჩამონათვალი კი არაა, არამედ გვიჩვენებს მის ადგილს იმ კონტექსტში, პედაგოგიური მიმართულებით სარგებლის მომტანი შეიძლება გახდეს. შესაძლებელია სამომავლოდ ინგლისური ენის შემსწავლელებისათვის კონკორდანსზე დაფუძნებული სავარჯიშების შექმნა, რომელიც შეცდომათა აღმოჩენასა და გასწორებაზე იქნება ორიენტირებული და უცხოური ენის შემსწავლელებს წერილობით უნარებს გაუმჯობესებს.

დისერტაციაში წარმოდგენილი 30 სიტყვა-სტატიის მაგალითზე სამომავლოდ სპეციალიზებული (EAP) ტიპის ინგლისურ-ქართული ლექსიკონის შექმნა მნიშვნელოვნად გვესახება. აღნიშნული ლექსიკონი სრულად დაეფუძნება ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსს და კონკრეტულად აკადემიური წერის უნარების განვითარებაზე იქნება გათვლილი.

ნაშრომის ძირითადი დასკვნები და დებულებები

ჩვენ მიერ გამოკვლეული და შესწავლილი საკითხები ამგვარად შეიძლება ჩამოყალიბდეს და განზოგადდეს: სასწავლო კორპუსები თავისი არსით ენის შემსწავლელთა შემოქმედებითი პოტენციალის რეალიზებისა და მასზე დაკვირვების შესაძლებლობას იძლევა. დღევანდელ ეპოქაში ინგლისური ენის შესწავლისადმი ასეთი ინტერესის, საჭიროებისა და მოტივაციის და საკითხის ამ კუთხით დანახვამ წინა პლანზე წამოსწია არა მხოლოდ მშობლიური ენის, როგორც ნიშანთა სისტემის დოკუმენტირების, არამედ ინგლისური ენის შემსწავლელთა ლინგვისტური თავისებურებების დისციპლინათაშორისი კვლევის აუცილებლობა. ენობრივი სიმდიდრის, სიტყვათხმარების, ენობრივი ქცევის ნორმების შესწავლის მიზნით,

შექმენით 10 მილიონამდე სიტყვისგან შემდგარი ინგლისური ენის ქართველი შემსწავლელების სასწავლო კორპუსი (GLEAN), რომელიც მოიცავს ინგლისური ენის ქართველი შემსწავლელების მიერ შესრულებულ სხვადასხვა ჟანრის მქონე ინგლისურენოვან წერილობით მონაცემს.

GLEAN სასწავლო კორპუსის შექმნისას დავეფუძნეთ კორპუსის დიზაინის ჯ. სინკლერისეულ ფუნდამენტურ პრინციპებს: დაბალანსებულობას, რეპრეზენტაციულობას, ჰომოგენურობასა და მიუკერძოებლობას. ინგლისური ენის ქართველ შემსწავლელთა კორპუსი **სპეციალიზებული ხასიათისაა** და მოიცავს დარგობრივი და ჟანრული მრავალფეროვნების მქონე ტექსტებს. სასწავლო კორპუსში გაერთიანებულია ისეთი ჟანრებული ტექსტები, როგორიცაა ესეისტიკა, აკადემიური, პუბლიცისტური და სასაუბრო სტილის. ესეისტიკის ჟანრის ნაწერები აერთიანებს არგუმენტირებულ, აღწერით, თავისუფალი კომპოზიციის ესეებს, აკადემიური ჟანრი - დარგობრივ კვლევით სტატიებს ლიტერატურასა და ლინგვისტიკაში. მეორე მხრივ, პუბლიცისტისტიკის ჟანრულ კატეგორიაში ერთიანდება საინფორმაციო ქრონიკა, რეპორტაჟი, საგაზეთო სტატიები, ხოლო ბლოგისა და კორესპონდენციის (არაოფიციალური მეილები) სასაუბრო ჟანრში გავაერთიანეთ.

ჩვენ მიერ შექმნილი ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსი (GLEAN) ანოტირებულია და გვეხმარება ლინგვისტური კვლევისთვის ენის სხვადასხვა დონეზე. ინგლისური ენის ქართველ შემსწავლელთა კორპუსში ყველა სიტყვაფორმას მიემართება მეტყველების შესაბამისი ნაწილის აღმნიშვნელი მარკერი. ანოტაციისთვის გამოვიყენეთ სტანდარტული თეგირების წყება CLAWS7.

შევისწავლეთ ენაში მიმდინარე ცვლილებები და გამოვიკვლიეთ ენობრივი ქცევის ნორმები, რათა სწორად დავგვემოთ და განვახორციელოთ უცხოური ენის სწავლება ლექსიკოგრაფიული სამუშაოს ჩატარება. ჩვენი საკვლევი კითხვა მდგომარეობდა იმის გარკვევაში თუ რა სახის ენობრივი თავისებურებებით რეალიზდება ინგლისური ენის ქართველ შემსწავლელთა დისკურსი. ეს საკითხი რამდენიმე მიმართულებით განვიხილეთ: ა) შესიტყვებითი მოდელების როგორც რაოდენობრივი, ისე სტრუქტურულ-სემანტიკური ანალიზით; ბ) მრავალკომპონენტური, კერძოდ სამ- და ოთხკომპონენტური, ლექსიკური ფრაზების სიხშირული და ფუნქციურ-დისკურსული ანალიზით. ორივე საკვლევი

კითხვისათვის პასუხის გაცემა კონტრასტული ანალიზის გზით მოხერხდა; გ) ყველაზე ხშირად გამოყენებული, სტატისტიკურად დაანგარიშებადი ლექსიკური ელემენტები კოლოკაციების ასოციაციური საზომების გამოყენებით; დ)

ტიპ/ტოკენებისა და ლემა/ტოკენების შეფარდების სიდიდეზე დაყრდნობით დადგინდა ინგლისური ენის ქართულ შემსწავლელთა ენობრივი სიმდიდრე, როგორც მრავალგანზომილებიანი ფენომენი; ე) წარმოვადგინეთ ჩვენ მიერ შექმნილ სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტიანი ლექსიკური ფრაზის სალექსიკონო სიტყვა-სტატია. აღნიშნული მასალა შესაძლებელია გახდეს სპეციალიზებული, აკადემიურ წერაზე ორიენტირებული (EAP) ლექსიკონის შემადგენელი ნაწილი.

ა) შესიტყვებითი მოდელების სტრუქტურული ანალიზისას ერთმანეთს შევადარეთ GLEAN და ICNALE სასწავლო კორპუსები. ჩატარებულ იქნა სანდოობის, loglikelihood სიგნიფიკანტურობის ტესტი და შესწავლილ იქნა კოლოკაციური წყვილები: 1. ზმნა (V) + არსებითი სახელი (N); 2. არსებითი სახელი (N) + არსებითი სახელი (N); 3. ზედსართავი სახელი (ADJ) + არსებითი სახელი (N); 4. ზმნა (V) + ზმნიზედა (ADV); 5. ზმნიზედა (ADV) + ზმნა (V); 6. ზმნიზედა (ADV) + ზედსართავი სახელი (ADJ). შედეგად, ინგლისური ენის ქართული შემსწავლელები უფრო მრავალფეროვან შესიტყვებათა მოდელებს იყენებენ, ვიდრე აზიელი სტუდენტები. გამონაკლისია ზმნა + არსებითი სახელის კოლოკაციური წყვილი, რომელიც ICNALE სასწავლო კორპუსში ორჯერ მეტად ფიქსირდება. გამოყავით ლექსიკური და გრამატიკული შესიტყვებების მოდელები. ლექსიკურ მოდელებში ვაერთიანებთ ისეთ ორკომპონენტიან შესიტყვებებს, რომლის ელემენტებიც შედგება სახელური, ზმნური, ადვერბული კოლოკაციური წყვილებისაგან: 1. ზმნა (V) + არსებითი სახელი (N); 2. არსებითი სახელი (N) + არსებითი სახელი (N); 3. ზედსართავი სახელი (ADJ) + არსებითი სახელი (N); 4. ზმნა (V) + ზმნიზედა (ADV); 5. ზმნიზედა (ADV) + ზმნა (V); 6. ზმნიზედა (ADV) + ზედსართავი სახელი (ADJ). ხოლო სამ-, ოთხკომპონენტიანი ფრაზემები წინდებულს და სხვა ტიპის ერთეულებსაც მოიცავენ და გრამატიკული კოლოკაციების მოდელებში ერთიანდებიან. კონტრასტულმა ანალიზმა, ნორმალიზებული სიხშირის ანალიზის შედეგად გვიჩვენა, რომ ინგლისური ენის

ქართველი შემსწავლელები უფრო მრავალფეროვან შესიტყვებათა მოდელებს იყენებენ, ვიდრე აზიელი სტუდენტები. თუმცა, აქ შეინიშნება რამდენიმე ტენდენცია:

I. გამონაკლისია ზმნა + არსებითი სახელის კოლოკაციური წყვილი, რომელიც ICNALE სასწავლო კორპუსში ორჯერ მეტად ფიქსირდება. მიზეზ-შედეგობრივი კავშირის ძიებისას შეგვიძლია ვივარაუდოთ, რომ თანამედროვე ჩინურ ენაში არსებულ მრავალ-კატეგორიანი სიტყვების (multi-category word) და მათ შორის ყველაზე დომინანტური შესიტყვებითი წყვილის (V+N), ესოდენ გავრცელებულობაში. მშობლიურ ენაში (ჩინურში) ასეთი რთული გრამატიკული ფენომენის არსებობის შემდეგ, ინგლისურ ენაში ჩინელი სტუდენტებისათვის გაცილებით მარტივია ზმნა + არსებითი სახელის კონსტრუქციის ინტერნალიზაცია;

II. აზიელი სტუდენტების სასწავლო კორპუსში ზედსართავი სახელი + არსებითი სახელის, ზმნა + ზმნიზედისა და ზმნიზედა + ზედსართავი სახელის კოლოკაციური შესიტყვების დეფიციტის გამომწვევი რამდენიმე მიზეზი შეიძლება იყოს. პირველ რიგში, ამ შედეგის სათავე ინგლისურ და აზიურ ენებს შორის არსებულ გრამატიკულ განსხვავებებში შესაძლოა მდგომარეობდეს. მაგალითად, ჩინურის მსგავს ენებს ინგლისურთან შედარებით წინადადების სხვადასხვა სტრუქტურა აქვთ. ინგლისური ეყრდნობა სუბიექტი-ზმნა-ობიექტი წყობას, ხოლო ჩინური იყენებს თემა-სტრუქტურას. მრავალ ენაში, მათ შორის ჩინურშიც, ზმნებისა და ზედსართავების კლასიკური გაყოფა შეუძლებელია და მეტყველების ნაწილების სიღრმულ პლანში განხილვისას ზედსართავი სახელები უნდა გავაერთიანოთ ზმნებთან. ამრიგად, ინგლისური ენის ჩინელი შემსწავლელებისათვის სირთულეს წარმოადგენს ინგლისურად ჩასწვდენ ზედსართავი სახელების დაწყვილების ბუნებრივ დინებას.

ჩვენთვის განსაკუთრებით საინტერესო აღმოჩნდა ADV+ADJ შესიტყვებათა მოდელის შეპირისპირებითი ანალიზი ინგლისური ენის ქართველ შემსწავლელთა და ამერიკელ სტუდენტთა სასწავლო კორპუსში. ანალიზის დროს გამოიკვეთა გამაძლიერებელი სემანტიკის მქონე ზმნიზედების შესიტყვების ერთ-ერთ წევრად ყოფნის ტენდენცია. ესენია: very, utterly, extremely, perfectly, absolutely, totally. რაოდენობრივი კვლევის შედეგად გამოაშკარავდა, რომ ქართველი შემსწავლელები მნიშვნელოვნად დაბალი სიხშირით იყენებენ -ly ზმნიზედებს, ვიდრე სხვა მორფოლოგიური სტრუქტურის მქონე ზმნიზედებს.

ქართველი სტუდენტები ჭარბად გამოიყენებენ *very* და *really* ზმნიზედებს მაშინ, როცა ამერიკელ სტუდენტთა შორის *very* და *highly* + ზედსართავი სახელი ყველაზე მაღალი სიხშირით გამოირჩევა. კორპუსების მონაცემებს შორის დრამატული სხვაობა სახეზეა *entirely* + ზედსართავი სახელის კოლოკაციური წყვილის შედარებისას, რადგან ინგლისური ენის ქართველ შემსწავლელებში ეს კონკრეტული შესიტყვებითი პარადიგმა მნიშვნელოვნად ჩამოუვარდება MICUSP სასწავლო კორპუსის შედეგს. მკვეთრი განსხვავება შეინიშნება *utterly* + ზედსართავი სახელის შესიტყვებითი წყვილის შედარებისას, კერძოდ, GLEAN სასწავლო კორპუსში ასეთი სტრუქტურა საერთოდ არ ფიქსირდება მაშინ, როდესაც ინგლისურ ენის, როგორც მშობლიურ ენაზე მოსაუბრეთა ენობრივ რეპერტუარში *utterly* + ზედსართავი სახელი ხშირად გამოიყენება. ორ სასწავლო კორპუსს შორის თითქმის თანაბარი სიხშირული შედეგი მივიღეთ *completely* + ზედსართავი სახელის კონტრასტული ანალიზისას.

მეორე მხრივ, კვლევისას ზოგადი სემანტიკური და მორფოლოგიური თავისებურებები გამოიკვეთა. ინგლისური ენის ქართველ შემსწავლელთა კორპუსულ მასალაში *completely* დაკავშირებულია ისეთ ზედსართავ სახელებთან, რომლებიც ნეიტრალური კონოტაციის მატარებლები არიან. მაშინ, როცა ამერიკელი სტუდენტები აღნიშნულ ზმნიზედას უარყოფითი სემანტიკური ასოციაციით გამოიყენებენ (ზედსართავ სახელებთან პრეფიქსით *un-*, *in-*). ზმნიზედა *perfectly* სრულიად დადებითი კონოტაციური შინაარსი ჰქონდა ამერიკელი სტუდენტების ესეებში, გასხვავებით ქართველი სტუდენტების სასწავლო კორპუსის მონაცემებისა. ამ უკანასკნელში, *perfectly* უარყოფითი კონოტაციითაა წარმოდგენილი კონტექსტებში, კერძოდ, *perfectly unimportant*, *perfectly invisible*. ზმნიზედა *extremely*-ს მიჩიგანის სასწავლო კორპუსში დადებითი ასოციაციების მქონე ზედსართავი სახელები მოსდევს, როგორიცაა „დიდსულოვანი“, „კაშკაშა“ და შედეგად ვიღებთ ამდაგვარ შესიტყვებებს: *extremely illuminating*, *extremely magnanimous*, *extremely feminine*. ხოლო ამის საპირისპიროდ, ქართველი შემსწავლელების კორპუსში *extremely* მოსდევს მორფოლოგიურად და სემანტიკურად უარყოფითი ასოციაციის მქონე ზედსართავი სახელები „არაეფექტური“, „არაჯანსაღი“, „უპასუხისმგებლო“ და შედეგად ვიღებთ ასეთ კოლოკაციურ წყვილებს: *extremely inefficient*, *extremely unhealthy*, *extremely irresponsible*.

ბ) მრავალკომპონენტური ლექსიკური ფრაზების სიხშირულმა კვლევამ გამოაშკარავა GLEAN სასწავლო კორპუსში ლექსიკურ კომპონენტთა ჭარბად ან იშვიათად გამოყენების ტენდენცია ამერიკელ სტუდენტთა სასწავლო კორპუსის მასალასთან მიმართებაში. გამოვყავით 42 ცალი სამ- და ოთხკომპონენტური ლექსიკური ერთეული, რომელიც კონტრასტული ანალიზის გზით შევადარეთ MICUSP სასწავლო კორპუსს. შედეგების მიხედვით ქართველი შემსწავლელები ხშირად მოიხმარენ კვლევაზე ორიენტირებულ ლექსიკური ფრაზებს, განსაკუთრებით ლოკაციისა და აბსტრაქტული აღწერის ფუნქციის მქონე ლექსიკურ კონებს, ვიდრე მშობლიურ ენაზე მოსაუბრე ამერიკელი სტუდენტები.

ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსში „ფრაზეოლოგიური სათამაშო დათუნიებად“, ანუ ყველაზე ხშირად გამოყენებულ ლექსიკური ფრაზებად გვევლინება შემდეგი სინტაგმები: *as of today, by the end of, on the backdrop of, in the process of, in the wake of, in honour of, as part of the, one of the most, as for the, in response to the, as well as the, as a result of*. შემდეგი ცხადყოფს, რომ ინგლისური ენის ქართველ შემსწავლელებს მკაფიოდ გამოხატული სასურველი ფრაზეოლოგიური არჩევანი აქვთ, რომლებსაც დისკურსში ყველაზე კომფორტულად იყენებენ. მომდევნო ლექსიკურ ფრაზებს კი ნაკლები სიხშირით, ვიდრე ამერიკელი სტუდენტები: *at the same time, the extent to which, in addition to, in an attempt to, in comparison to, the nature of the, on top of that*.

სასწავლო კორპუსში სიხშირული გზით მოპოვებული სამ- და ოთხკომპონენტური ლექსიკური კონების ფუნქციურ-დისკურსული ანალიზისათვის დავეფუძნეთ ჯ. ო' ფლინის (2022: 82) მიერ შემოთავაზებულ ფუნქციურ ტაქსონომიას. გამოვყავით ლექსიკურ ფრაზათა რამდენიმე ფუნქციური კატეგორია: კვლევაზე ორიენტირებული, ტექსტზე ორიენტირებული, მონაწილეზე ორიენტირებული ლექსიკური კონები. ფუნქციურ ტაქსონომიაში კვლევაზე ორიენტირებული ლექსიკური ფრაზები, ლოკაციის, აბსტრაქტული აღწერისა და პროცედურული ფუნქციებით, ყველაზე ხშირად გვხვდება ინგლისური ენის ქართველ შემსწავლელთა კორპუსში. შედარებით მცირე რაოდენობით გვხვდება ტექსტზე ორიენტირებული ლექსიკური ფრაზები, მიზეზ-შედეგობრივი, ამხსნელ-დამაზუსტებელი და ტექსტის მაორგანიზებელი ფუნქციებით. ჩვენს სასწავლო კორპუსში ყველაზე ნაკლებად

მონაწილეზე ორიენტირებული მარკერები გვხვდება და ამ ქვეჯგუფში გაერთიანებული ლექსიკური ფრაზები კონკრეტულად კორესპონდენციისა და თავისუფალი კომპოზიციის ესესთვისაა დამახასიათებელი.

გ) კვლევაში გამოყავით კოლოკაციების ასოციაციური საზომების გამოყენებით (log-likelihood, MI, MI3, log Dice, Z-score) ყველაზე ხშირად გამოყენებული, სტატისტიკურად დაანგარიშებადი ლექსიკური ელემენტები. სახელურ შესიტყვებებში (N+N) ერთ-ერთ ერთეულად ლემა *life* ყველაზე ხშირად მეორდება. კორპუსში ADJ+N წყვილებში შეინიშნება *good* ლექსემის სიჭარბე. V+ADV კოლოკაციურ წყვილში ზმნიზედის შერჩევისას უპირატესობას ანიჭებენ ლოკატიურ ზმნიზედას, მაგალითად, ნიმუშებში სჭარბობს ლემები *together, forward, away*. ADV+V კორპუსული ანალიზის შედეგად დასტურდება, რომ GLEAN სასწავლო კორპუსში ქართველი შემსწავლელები აქტიურად იყენებენ შედარებითი ხარისხის ზმნიზედას, *further, less, more, higher*. კონტექსტუალური ანალიზით აღმოჩნდა, რომ ასეთი ტიპის პარადიგმები, კერძოდ *further* + ზმნა საგაზეთო პოლიტიკური სტატიების, დიპლომატიური დისკურსის მასალაში გვხვდება. გამოვავლინეთ შესიტყვებების ვიზუალური გრაფების ანალიზის შედეგები, რომელშიც ნათლად იკვეთება ასოციაციური სიმდიდრე, შესიტყვების სიხშირე და შესიტყვების პოზიცია.

დ) ენობრივი სიმდიდრე განვსაზღვრეთ როგორც ტიპ/ტოკენების, ისე ლემა/ტოკენების შეფარდების სიდიდეზე დაყრდნობით და დადგინდა, რომ ინგლისური ენის ქართველ შემსწავლელთა ენობრივი ქმედება, როგორც მრავალგანზომილებიანი ფენომენი, ენობრივი მრავალფეროვნებით გამოირჩევა. გამოვიყენეთ როგორც TTR საზომი, ისე ს. გრეინჯერის მიერ შემოთავაზებული LTR სიდიდე. პირველ შემთხვევაში, შედეგად მივიღეთ 0.0122 ტიპი თითოეულ ტოკენში. ხოლო LTR შედეგი აღმოჩნდა 0.0108 ლემა თითოეულ ტოკენში. აღნიშნული სიდიდე რაც უფრო ახლოსაა 1-თან, მით უფრო მრავალფეროვანია სასწავლო კორპუსის ლექსიკა.

ე) წარმოვადგინეთ ჩვენ მიერ შექმნილ სასწავლო კორპუსზე დაყრდნობით შესრულებული 30 მრავალკომპონენტიანი ლექსიკური ფრაზის სალექსიკონო სიტყვა-სტატია. აღნიშნული მასალა შესაძლებელია გახდეს სპეციალიზებული, აკადემიურ წერაზე ორიენტირებული (EAP) ლექსიკონის შემადგენელი ნაწილი. მისი

მაკროსტრუქტურა ეფუძნება დ. სიპმანის დისკურსის მარკერების ფუნქციურ ტაქსონომიას, ხოლო მიკროსტრუქტურულ ველებში მოცემული ფუნქციები და ამხსნელი მითითებები ასახავს ჯ. ო'ფლინისეულ ფუნქციურ ტაქსონომიას. სიტყვა-სტატია შედგება შემდეგი ელემენტებისაგან: თემატიკა (რომელ დისკურსის მარკერსაც განეკუთვნება მეთაური ფრაზა); მეთაური ფრაზა (ინგლისური ლექსიკური ფრაზა); მეთაური ფრაზის ფუნქციური აღწერა; ინგლისური მეთაური ფრაზის ქართული ეკვივალენტი; ინგლისური ლექსიკური ფრაზის ილუსტრაცია (GLEAN სასწავლო კორპუსიდან ამოღებული) ტექსტის ჟანრული კატეგორიისა და ქვეკატეგორიის მითითებით; ინგლისური ილუსტრაციის ქართულად თარგმნილი ვარიანტი; ტექსტური გრაფა – შესაძლო შეცდომათა გამაფრთხილებლად.

დასასრულ, შევხებით სამომავლო კვლევის პერსპექტივას, რომელიც დაკავშირებულია სასწავლო კორპუსის შეცდომათა თეგერის შექმნასთან (error-tagger). მნიშვნელოვანია, ქართველი ინგლისური ენის შემსწავლელების ენობრივი ლაფსუსების ამ გზით კლასიფიკაცია. ჩვენს კვლევაში ენობრივი ხარვეზების აღმოჩენა მხოლოდ მანუალური გზით მოვახერხეთ. როგორც წესი, შეცდომათა ანალიზის სქემა შედგება შეცდომათა ტაქსონომიებისგან, რომელიც ეფუძნება სტრუქტურულ-ენობრივ კატეგორიებს მართლწერას, გრამატიკას, ლექსიკასა თუ სტილს.

სასწავლო კორპუსის კონკორდანსი, რომელიც სიტყვების უბრალო ჩამონათვალი კი არაა, არამედ გვიჩვენებს მის ადგილს იმ კონტექსტში, პედაგოგიური მიმართულებით სარგებლის მომტანი შეიძლება გახდეს. შესაძლებელია სამომავლოდ ინგლისური ენის შემსწავლელებისათვის კონკორდანსზე დაფუძნებული სავარჯიშების შექმნა, რომელიც შეცდომათა აღმოჩენასა და გასწორებაზე იქნება ორიენტირებული და უცხოური ენის შემსწავლელებს წერილობით უნარებს გაუუმჯობესებს.

დისერტაციაში წარმოდგენილი 30 სიტყვა-სტატიის მაგალითზე სამომავლოდ სპეციალიზებული (EAP) ტიპის ინგლისურ-ქართული ლექსიკონის შექმნა აკადემიური წერის შემსწავლელებისთვის საჭირო რესურსად გვესახება. აღნიშნული ლექსიკონი სრულად დაეფუძნება ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსს და კონკრეტულად წერის უნარების განვითარებაზე იქნება გათვლილი. ამას გარდა, სასწავლო კორპუსის ლემათა კონტექსტუალური

რეალიზების მოძიებისას შესაძლებელია სამომავლოდ ინგლისური ენის შემსწავლელებისათვის კონკორდანსზე დაფუძნებული სავარჯიშების შექმნა, რომელიც შეცდომათა აღმოჩენასა და გასწორებაზე იქნება ორიენტირებული. ამდენად, ინგლისური ენის ქართველ შემსწავლელთა სასწავლო კორპუსის არსებობა მნიშვნელოვანია ენის რეალიზების ასპექტების, ლინგვისტურ მონაცემთა თანმიმდევრული და მრავალმიზნობრივი (უმეტესად ლექსიკოგრაფიული) შეგროვება-დამუშავების განსახორციელებლად.

ბიბლიოგრაფია

1. არნარდოთირი, 2022: Arnardóttir, Þórunn, Isidora Glisic, Annika Simonsen, Lilja Stefánsdóttir, and Anton Ingason. "Error corpora for different informant groups: Annotating and analyzing texts from L2 speakers, people with dyslexia and children." In *Proceedings of the 19th International Conference on Natural Language Processing (ICON)*, pp. 245-252. 2022;
2. ბარნარდი, 2005: Burnard, Lou. "Metadata for corpus work." *Developing linguistic corpora: A guide to good practice* (2005): 30-46;
3. ბელცი და ვიატკინა, 2008: Belz, Julie Anne, and Nina Vyatkina. "The pedagogical mediation of a developmental learner corpus for classroom-based language instruction." (2008);
4. ბენსონი და სხვ., 1997: Benson, Morton, Evelyn Benson, and Robert Ilson. "The BBI dictionary of English word combinations." (1997);
5. ბენსონი, 1985: Benson, Morton. "A combinatory dictionary of English." *Dictionaries: Journal of the Dictionary Society of North America* 7, no. 1 (1985): 189-200.
6. ბერგენჰოლცი და გოუვსი, 2014: Bergenholtz, Henning, and Rufus Gouws. "A lexicographical perspective on the classification of multiword combinations." *International Journal of Lexicography* 27, no. 1 (2014): 1-24;
7. ბიბერი, 2006: Biber, Douglas. A corpus-based study of spoken and written registers. John Benjamins. 2006. <https://doi.org/10.1075/scl.23>;
8. ბრეზინა და სხვ., 2015: Brezina, Vaclav, Tony McEnery, and Stephen Wattam. "Collocations in context: A new perspective on collocation networks." *International journal of corpus linguistics* 20, no. 2 (2015): 139-173;
9. ბრეზინა, 2018: Brezina, Vaclav. *Statistics in corpus linguistics: A practical guide*. Cambridge University Press, 2018;
10. ბრეზინა, 2021: Brezina, Vaclav, Pierre Weill-Tessier, and Anthony McEnery. "LancsBox v. 6. x.[software package]." *Computer software*. Lancaster University (2021).
11. გილქინი, 2008a: Gilquin, Gaëtanelle. "Hesitation markers among EFL learners: Pragmatic deficiency or difference." *Pragmatics and corpus linguistics: A mutualistic entente* 2 (2008): 119-149.
12. გლედჰილი, 2000: Gledhill, Christopher J. *Collocations in science writing*. Vol. 22. Gunter Narr Verlag, 2000.

13. გოუვსი და პრინსლო, 2010: Gouws, Rufus Hjalmar, and Daniel Jacobus Prinsloo. *Principles and practice of South African lexicography*. African Sun Media, 2010.
14. გოუვსი და სხვ., 2010: Gouws, Rufus H., Danie J. Prinsloo, A. Dykstra, and T. Schoonheim. "Thinking out of the box—perspectives on the use of lexicographic text boxes." In *Proceedings of the*. 2010.
15. გრეინჯერი და ვინი, 2000: Granger, Sylviane, and Martin Wynne. "Optimising measures of lexical variation in EFL learner corpora." In *Corpora galore*, pp. 249-257. Brill, 2000.
16. გრეინჯერი და სხვ., 2008: Granger, Sylviane, and Magali Paquot. "Disentangling the phraseological web." *S. Granger & Meunier (Eds.)* (2008).
17. გრეინჯერი და სხვ., 2015: Granger, Sylviane, Gaëtanelle Gilquin, and Fanny Meunier, eds. *The Cambridge handbook of learner corpus research*. Cambridge University Press, 2015.
18. გრეინჯერი და ფაქო, 2013: Granger, Sylviane, and Magali Paquot. "Language for specific purposes learner corpora." *The encyclopedia of applied linguistics* (2013).
19. გრეინჯერი, 1996: Granger, Sylviane. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora." (1996).
20. გრეინჯერი, 2004: Granger, Sylviane. "Computer learner corpus research: Current status and future prospects." *Applied Corpus Linguistics* (2004): 123-145.
21. გრეინჯერი, 2008a: Granger, Sylviane. "Learner corpora in foreign language education." (2017).
22. გრეინჯერი, 2013a: Granger, Sylviane. 'Learner corpora', in Chapelle, C. A. (ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley, pp. 3235–42.
23. გრეინჯერი, 2015: Granger, Sylviane. "Contrastive interlanguage analysis: A reappraisal." *International Journal of Learner Corpus Research* 1, no. 1 (2015): 7-24.
24. ევერტი, 2004: Evert, Stefan. "The statistics of word cooccurrences: word pairs and collocations." (2005).
25. ენტონი, 2024: Anthony, Laurence. "AntConc (Windows, Macintosh OS X, and Linux)." *Recuperado de: http://www.antlab.sci.waseda.ac.jp/software/README_AntConc3* 2 (2024).

26. თანდაშვილი და ფურცხვანიძე, 2014: თანდაშვილი, მანანა, ფურცხვანიძე, ზაქარია. კორპუსლინგვისტური პარადიგმა ენათმეცნიერებაში, ფრანკფურტის ლინგვისტური წრე; 2014
27. იესპერსენი, 1924: Jespersen, Otto. "The teaching of grammar." *The English Journal* 13, no. 3 (1924): 161-176.
28. იოჰანსონი, 2008: Johansson, Stig. "Contrastive analysis and learner language: A corpus-based approach." *University of Oslo* 9 (2008);
29. იშიკავა, 2014: Ishikawa, Shin'ichiro. "Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis." *Learner corpus studies in Asia and the world* 2 (2014): 63-76;
30. კენჭიაშვილი, 2020: კენჭიაშვილი, ნათია. "ქართული მჭევრმეტყველების ტექსტების ლინგვისტური ანალიზი: ტროპის სახეები." (2020).
31. კიელმერი, 2014: Kjellmer, Göran. "A mint of phrases." *English corpus linguistics* (2014): 111-127.
32. კობი და ჰორსტი, 2015: Cobb, Tom, and Marlise Horst. "Learner corpora and lexis." *The Cambridge handbook of learner corpus research* (2015): 185-206.
33. კობი, 2002: Cobb, Tom. "Analyzing late interlanguage with learner corpora: Quebec replications of three European studies." *Canadian Modern Language Review* 59, no. 3 (2003): 393-424.
34. ლემანი და სხვ., 2000: Lehmann, Hans Martin, Peter Schneider, and Sebastian Hoffmann. "BNCweb." In *Corpora galore*, pp. 259-266. Brill, 2000.
35. ლიზი, 1997a: Leech, Geoffrey. "Introducing corpus annotation. Dlm. R. Garside, G. Leech & T. McEnery, *Corpus annotation: Linguistics information from computer text corpora*, 1-18." (1997).
36. ლუ, 2016: Lu, Yuanwen. *A corpus study of collocation in Chinese learner English*. Routledge, 2016.
37. მაკენერი და სხვ., 2006: McEnery, Tony, Richard Xiao, and Yukio Tono. *Corpus-based language studies: An advanced resource book*. Taylor & Francis, 2006.
38. მაკკარტი, 1990: McCarthy, Michael. *Vocabulary*. Oxford University Press, 1990.
39. მალვერნი და სხვ., 2004: Malvern, David, Brian Richards, Ngoni Chipere, and Pilar Durán. *Lexical diversity and language development*. Palgrave Macmillan UK, 2004.

40. მერვი 2019: Murphy, Sean. "Shakespeare and his contemporaries: Designing a genre classification scheme for Early English Books Online 1560–1640." *ICAME Journal* 43, no. 1 (2019): 59-82.
41. მეჭურა, 2017: Měchura, Michal Boleslav. "Introducing Lexonomy: an open-source dictionary writing and publishing system." In *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference*, pp. 19-21. 2017.
42. მიულერი, 2005: Müller, Simone. "Discourse markers in native and non-native English discourse." (2005): 1-310.
43. ნესელჰაუფი, 2004: Nesselhauf, Nadja. "Learner corpora and their potential for language teaching." *How to use corpora in language teaching* 12 (2004): 125-156.
44. ნესელჰაუფი, 2005: Nesselhauf, Nadja. "Collocations in a learner corpus." (2005): 1-344.
45. ო'ფლინის 2022: O'Flynn, James. "Lexical bundles in the academic writing of the Arts and Humanities: from corpus to CALL." *Yearbook of Phraseology* 13, no. 1 (2022): 81-108.
46. პალმერი, 1981: Palmer, Frank Robert. *Semantics*. Cambridge university press, 1981.
47. ჟანგი და სხვ., 2021: Zhang, Huiping, Ming Chen, and Xuelan Li. "Developmental features of lexical richness in English writings by Chinese beginner learners." *Frontiers in Psychology* 12 (2021): 665988.
48. რანდელი და გრეინჯერი, 2007: Rundell, Michael, and Sylviane Granger. "From corpora to confidence." *English Teaching Professional* 50, no. 1 (2007): 15-18.
49. რამირეს-ესპარზა და სხვ., 2012: Ramírez-Esparza, Nairán, Kathryn Harris, John Hellermann, Clemence Richard, Patricia K. Kuhl, and Steve Reder. "Socio-Interactive practices and personality in adult learners of english with little formal education." *Language Learning* 62, no. 2 (2012): 541-570.
50. როი და შაფერი, 2009: Van Rooy, Bertus, and Lande Schäfer. "Automatic POS tagging of a learner corpus: the influence of learner error on tagger accuracy." In *Proceedings of the corpus linguistics 2003 conference*, pp. 835-844. Lancaster, England: UCREL, Lancaster University, 2003.
51. სინკლერი, 1991: Sinclair, John. "Corpus, concordance, collocation." (*No Title*) (1991);
52. სინკლერი, 2005: Sinclair, John. "Corpus and Text-Basic Principles in Developing Linguistic Corpora: a Guide to Good Practice, ed. M. Wynne." (2005): 1-16;

53. სინკლერი, 2010: Sinclair, John. "Defining the Definiendum." In *A Way with Words*, pp. 37-48. 2010;
54. სიპმანი, 2004: Dirk, Siepmann. *Discourse markers across languages: A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. Routledge, 2004.
55. სტეიპლსი და სხვ., 2013: Staples, Shelley, Jesse Egbert, Douglas Biber, and Alyson McClair. "Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section." *Journal of English for academic purposes* 12, no. 3 (2013): 214-225.
56. სტეფანოვიჩი და გრიესი, 2003: Stefanowitsch, Anatol, and Stefan Th Gries. "Collostructions: Investigating the interaction of words and constructions." *International journal of corpus linguistics* 8, no. 2 (2003): 209-243.
57. ტონინი-ბონელი, 2001: Tognini-Bonelli, Elena. "Corpus linguistics at work." (2001): 1-236.
58. ფირსი, 1957: Firth, John. "A synopsis of linguistic theory, 1930-1955." *Studies in linguistic analysis* (1957): 10-32.
59. ქოუი, 1994: Cowie, A. P. 1994. 'Phraseology', in Asher, R. E. (ed.), *The Encyclopedia of Language and Linguistics*. Oxford University Press, pp. 3168-71.
60. ჯაკერი, 2003: Jucker, Andreas H., Sara W. Smith, and Tanja Lüdge. "Interactive aspects of vagueness in conversation." *Journal of pragmatics* 35, no. 12 (2003): 1737-1769.
61. ჰაილენდი, 1998: Hyland, Ken. "Persuasion and context: The pragmatics of academic metadiscourse." *Journal of pragmatics* 30, no. 4 (1998): 437-455.
62. ჰაილენდი, 2008: Hyland, Ken. "As can be seen: Lexical bundles and disciplinary variation." *English for specific purposes* 27, no. 1 (2008): 4-21.
63. ჰალიდეი, 1985: Halliday, Michael Alexander Kirkwood. "Spoken and written language." (*No Title*) (1989).
64. ჰასელგარდი, 2019: Hasselgård, Hilde. "Phraseological teddy bears." *Corpus linguistics, context and culture* 15 (2019): 339-362.
65. ჰასელგრენი, 1994: Hasselgren, Angela. "Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary." *International journal of applied linguistics* 4, no. 2 (1994): 237-258.
66. ჰაუსმანი, 1989: Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta. "1991." *Wörterbücher. Ein internationales Handbuch zur*

- Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie* 3 (1989).
67. ჰაუსმანი, 1993: Hausmann, Franz Josef. "Was ist eigentlich Wortschatz." *Wortschatz und Wortschatzerwerb. Bochum, S* (1993): 2-21.
68. ჰერბსტი, 2011: Herbst, Thomas, Susen Faulhaber, and Peter Uhrig, eds. *The phraseological view of language: A tribute to John Sinclair*. Walter de Gruyter, 2011.
69. ჰოფმანი და სხვ., 2008: Sebastian Hoffmann, Smith, Nicholas, and Paul Rayson. "Corpus tools and methods, today and tomorrow: Incorporating linguists' manual annotations." *Literary and linguistic computing* 23, no. 2 (2008): 163-180.