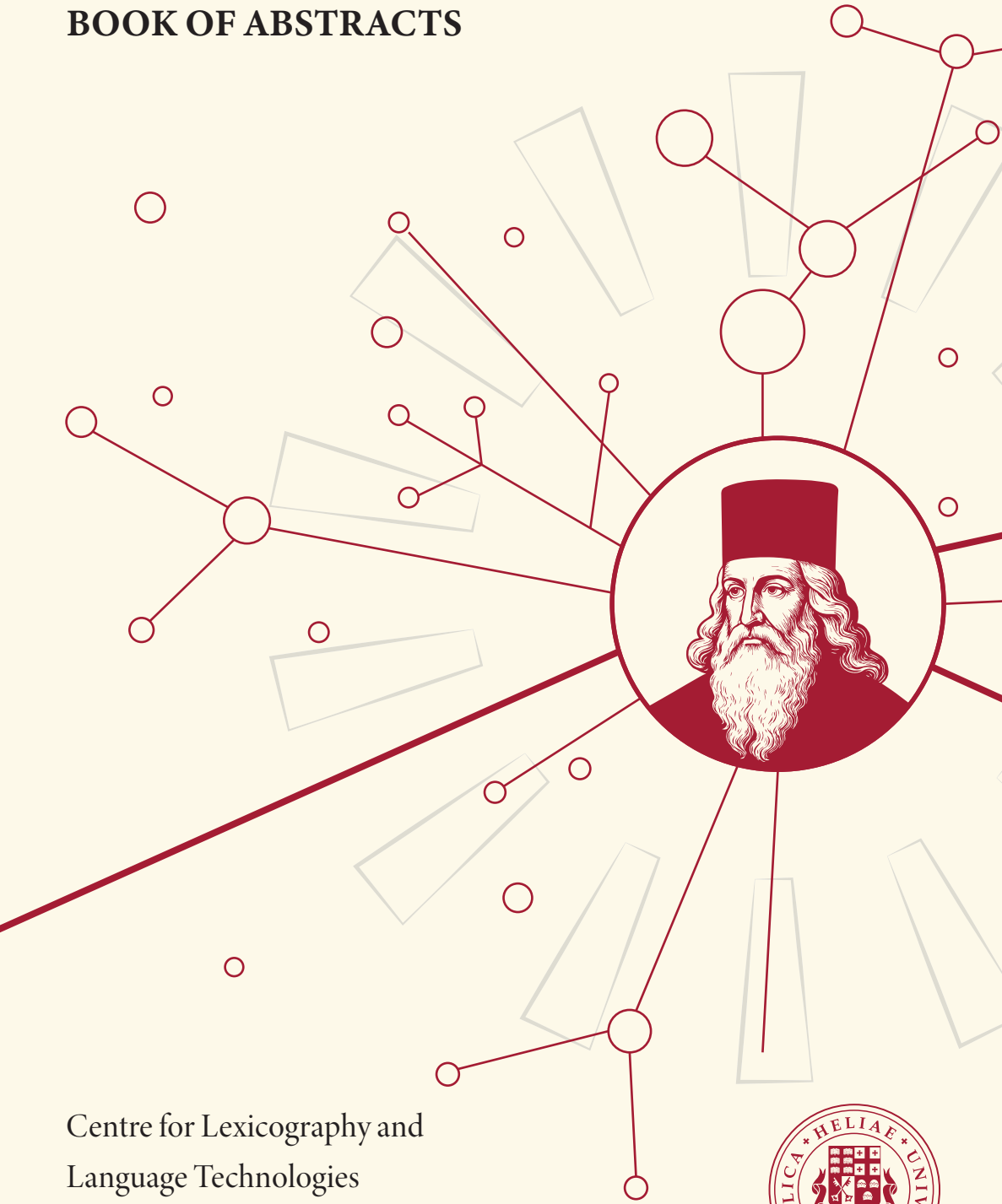
II INTERNATIONAL CONFERENCE

LEXICOGRAPHY IN THE XXI CENTURY

BOOK OF ABSTRACTS



Centre for Lexicography and
Language Technologies
Ilia State University


PUBLICA • HELIAE • UNIVERSITAS •
VIGAMUS

The Conference is dedicated to the 300th anniversary of the death of Sulkhan-Saba Orbeliani, an eminent Georgian lexicographer, writer, diplomat and public figure

კონფერენცია ეძღვნება დიდი ქართველი ლექსიკოგრაფის, მწერლის, დიპლომატისა და სახელმწიფო მოღვაწის, სულხან საბა ორბელიანის გარდაცვალებიდან 300 წლისთავს

II International Conference
Lexicography in the XXI century

Book of Abstracts

10–12 October, 2025

Centre for Lexicography and Language Technologies
Ilia State University

2025

II საერთაშორისო კონფერენცია
ლექსიკოგრაფია 21-ე საუკუნეში

თეზისების კრებული

10–12 ოქტომბერი, 2025 წელი

ლექსიკოგრაფიისა და ენობრივი ტექნოლოგიების ცენტრი
ილიას სახელმწიფო უნივერსიტეტი

2025

Editor Tinatin Margalitadze

რედაქტორი თინათინ მარგალიტაძე

ISBN 978-9941-18-489-5

ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა
ქაქუცა ჩოლოყაშვილის 3/5, თბილისი, 0162, საქართველო

ILIA STATE UNIVERSITY PRESS
3/5 Cholokashvili Ave, Tbilisi, 0162, Georgia

CONTENTS

Abstracts of Plenary Lectures

Abstracts of Papers

სარჩევი

პლენარული მოხსენებების თეზისები

მოხსენებების თეზისები

ABSTRACTS OF PLENARY LECTURES
პლენარული მოხსენებების თეზისები

Between Terminology and Lexicography: when terms serve society

Costa, Rute

CLUNL | Universidade NOVA de Lisboa, Portugal

rute.costa@fcsh.unl.pt

The boundaries between terminology and lexicography are often seen as disciplinary, but at their core lies a shared mission: to make sense of lexical use in linguistic and social context. While lexicography captures the richness of general language, terminology seeks precision in specialised domains. Yet both fields converge in a critical dimension, their service to society.

This talk explores the dynamic interface between terminology and lexicography, focusing on how terms evolve and acquire social relevance. In domains such as medicine, law, technology, and the environment, the careful crafting and dissemination of terms can shape public understanding, policy, and access to knowledge. Terminologists strive for clarity and consistency; lexicographers strive to reflect use. For these purposes, definitions in lexicography and terminology science serve distinct societal roles. We move between specialised, general, and simplified definitions — the latter often functioning more as explanations than definitions.

When terms serve society, they do more than name things — they carry authority, enable inclusion or exclusion, and often reflect power structures. This presentation will examine how the practices of both fields can be harnessed not only to describe language, but also to support communication equity, public engagement, and informed citizenship.

ტერმინოლოგიასა და ლექსიკოგრაფიას შორის: როდესაც ტერმინები საზოგადოებას ემსახურება

რუტე კოსტა

ნოვას უნივერსიტეტი, ლისაბონი, პორტუგალია

rute.costa@fcsh.unl.pt

ტერმინოლოგიასა და ლექსიკოგრაფიას შორის ზღვარი ხშირად დისციპლინურ საკითხად აღიქმება, თუმცა არსებითია მათი საერთო მისია: ლექსიკის გამოყენების გააზრება ენობრივ და სოციალურ კონტექსტში. მაშინ, როდესაც ლექსიკოგრაფია ასახავს ძირითადი ლექსიკური ფონდის სიმდიდრეს, ტერმინოლოგია ესწრაფვის სიზუსტეს სპეციალიზებულ დარგებში. თუმცა, ორივე სფერო ერთ მნიშვნელოვან განზომილებაში იკვეთება — მათი მიზანი საზოგადოების სამსახურია.

ეს მოხსენება განიხილავს ტერმინოლოგიასა და ლექსიკოგრაფიას შორის დინამიკურ ურთიერთკავშირს, განსაკუთრებული აქცენტით იმაზე, თუ როგორ ვითარდება ტერმინები და იძენს სოციალურ მნიშვნელობას. მედიცინის, სამართლის, ტექნოლოგიებისა და გარემოს სფეროებში ტერმინების სწორ შემუშავებასა და გავრცელებას შეუძლია ჩამოაყალიბოს საზოგადოებრივი აზრი, პოლიტიკა და შექმნას ცოდნაზე წვდომის შესაძლებლობა. ტერმინოლოგები ესწრაფვიან სიზუსტესა და თანმიმდევრულობას; ლექსიკოგრაფები კი ცდილობენ ასახონ ენის გამოყენება. ამ მიზნებისთვის, განმარტებები ლექსიკოგრაფიასა და ტერმინოლოგიაში განსხვავებულ საზოგადოებრივ როლს ასრულებენ. ჩვენ ვმოძრაობთ სპეციალიზებულ, ზოგად და გამარტივებულ განმარტებებს შორის — რომელთაგან ეს უკანასკნელნი ხშირად უფრო ახსნის ფუნქციას ასრულებენ, ვიდრე დეფინიციისა.

როდესაც ტერმინები საზოგადოების სამსახურში დგებიან, ისინი უფრო მეტს აკეთებენ, ვიდრე, უბრალოდ, სახელდებაა — მათ გააჩნიათ გარკვეული ძალა და ერთგვარი ავტორიტეტიც, რითაც ხელს უწყობენ ჩართულობას ან გამორიცხვას და ხშირად ასახავენ ძალაუფლების სტრუქტურებს. ეს პრეზენტაცია განიხილავს, თუ როგორ შეიძლება ორივე სფეროს პრაქტიკის გამოყენება არა მხოლოდ ენის აღსაწერად, არამედ სამართლიანობის, საზოგადოების ჩართულობისა და ინფორმირებული მოქალაქეობის მხარდასაჭერად.

References

Costa, R. (2013). "Terminology and Specialised Lexicography: two complementary domains". *Lexicographica*. Volume 29, Issue 1. International Annual of Lexicography / Revue Internationale de Lexicographie / Internationales Jahrbuch für Lexicographie. Ed. by Gouws, Rufus Hjalmar / Heid, Ulrich / Schierholz, Stefan J. / Schweickard, Wolfgand / Wiegand, Heribert Ernst. Berlin, New York: De Gruyter, pp. 29–42, ISSN 0175-6206, e-ISSN 1865-9403. DOI: 10.1515/lexi-2013-0004.

Costa, R., Salgado, A., Ramos, M., Almeida, B., Silva, R., Carvalho, S., Khan, F., Tasovac, T., Khemakhem, M. and L. Romary (2023). A crossroad between lexicography and terminology work: Knowledge organization and domain labelling. *Digital Scholarship in the Humanities*, 38 (Issue supplement 1), pp.i17–i29. e-ISSN 2055-768X. https://doi.org/10.1093/llc/fqad022.

ISO 704. (2022) Terminology Work — Principles and Methods, Geneva, International Organization for Standardization, 2022.

ISO 1087 (2019) Terminology work and Terminology Science — Vocabulary, Geneva, International Organization for Standardization, 2019.

Rey, A. (1979) La terminologie : noms et notions, collection « Que sais-je ? », P.U.F., Paris.

Wiegand, H. E. (1970). Onomasiologie und Semasiologie. *Germanistische Linguistik*. 1969/70. Hft.3, Olms. ISBN: 3487029995, 9783487029993.

Multidimensional Approaches to Lexical Innovation: Insights from the ENEOLI Network

Tallarico, Giovanni
University of Verona, Italy
giovanni.tallarico@univr.it

ENEOLI is the acronym for European Network on Lexical Innovation (2023-2027), COST Action 22126 (www.eneoli.eu). COST (European **Co**operation in **S**cience and **T**echnology) Actions are funding frameworks that support networking activities for researchers and innovators across Europe and beyond. ENEOLI is deeply committed to the main COST Action challenges, namely:

- Building cross-border research networks (through meetings, grants and short-term scientific missions), to create long-term, sustainable collaboration between members;
- Interdisciplinary collaboration, between linguistics, terminology, lexicography, translation studies, computational approaches, sociolinguistics, etc.;
- Capacity building and inclusiveness, ensuring researchers from less research-intensive countries (named "Inclusiveness Target Countries", such as Georgia) can participate fully in the network activities;
- Supporting early-career researchers, through training schools, dedicated grants, poster sessions at events, etc.
- Disseminating knowledge, both within academia and to stakeholders in industry, policy, and civil society.

More generally, ENEOLI's aim is to create an engaged research community on lexical innovation, a collaborative ecosystem aiming to establish this topic at the heart of the sociocultural sphere today.

Neology – the study of new words and meanings – has long been underrepresented in linguistics despite its central role in understanding language evolution and functioning. This paper introduces the European Network on Lexical Innovation (ENEOLI), which aims to address critical gaps in neology research through an ambitious, multilingual, and collaborative approach. ENEOLI currently involves over 370 researchers from 50 countries. The initiative is structured around four key objectives: (1) creating a multilingual glossary of the metaterminology of neology; (2) developing digital resources and tools for neological research; (3) conducting comparative diachronic and synchronic studies; and (4) establishing training initiatives to professionalize neology education.

The paper describes ENEOLI's innovative methodological framework, which integrates traditional lexicographic practices with corpus linguistics, computational tools, and multilingual analyses. A major contribution is the development of *NeoVoc*, a multilingual thesaurus of neology terms, and *NeoCorpus*, a comprehensive corpus supporting term validation and cross-linguistic study. The glossary project, managed via the ENEOLI Wikibase, emphasizes expert validation, structured data management, and interoperability, setting new standards for terminological consistency and accessibility. From a lexicographic perspective,

ENEOLI's work also examines editorial practices for selecting, describing, and tracking neologisms in dictionaries, aiming to produce a publicly available typology of inclusion criteria.

With its collaborative, interdisciplinary, and multilingual approach, ENEOLI sets the stage for a new paradigm in the study of lexical innovation across Europe and beyond.

მრავალგანზომილებიანი მიდგომები ლექსიკური სიახლეებისადმი: ENEOLI-ის ქსელის გამოცდილება

ჯოვანი ტალარიკო
ვერონის უნივერსიტეტი, იტალია
giovanni.tallarico@univr.it

ENEOLI არის აკრონიმი პროექტის სახელწოდებისა *European Network on Lexical Innovation* (ლექსიკური სიახლეების ევროპული ქსელი, 2023–2027), რომელიც ფინანსდება „მეცნიერებისა და ტექნოლოგიის სფეროში ევროპული თანამშრომლობის“ პროგრამის ფარგლებში (*European Cooperation in Science and Technology*, COST Action 22126). აღნიშნული პროგრამა (www.eneoli.eu) ხელს უწყობს მკვლევართა და ნოვატორთა ქსელურ თანამშრომლობას ევროპაში და მის ფარგლებს გარეთ. ENEOLI მტკიცედ უჭერს მხარს COST-ის ძირითად გამოწვევებს, რაც გულისხმობს:

- კვლევითი ქსელების შექმნას საზღვრების გარეშე (შეხვედრების, გრანტებისა და მოკლევადიანი სამეცნიერო მივლინებების მეშვეობით), წევრებს შორის ხანგრძლივი და მდგრადი თანამშრომლობის უზრუნველსაყოფად;
- ინტერდისციპლინურ თანამშრომლობას ლინგვისტიკაში, ტერმინოლოგიაში, ლექსიკოგრაფიაში, თარგმნის კვლევებში, კომპიუტერულ მიდგომებში, სოციოლინგვისტიკასა და სხვა დარგებში;
- პოტენციალის გაძლიერებასა და ინკლუზიურობას, რომლის მიზანია უზრუნველყოს კვლევითი თვალსაზრისით ნაკლებად აქტიური და ნაკლებად ჩართული ქვეყნების (ე.წ. *Inclusiveness Target Countries*, მათ შორის საქართველოს) მკვლევრების სრულფასოვანი ჩართულობა ქსელის საქმიანობაში;
- ახალგაზრდა, კარიერის ადრეულ საფეხურზე მყოფი, მკვლევრების მხარდაჭერას ტრენინგ-სკოლების, სპეციალური გრანტების, პოსტერების სესიებისა და სხვა აქტივობების მეშვეობით;
- ცოდნის გავრცელებას როგორც აკადემიურ წრეებში, ისე ინდუსტრიის, პოლიტიკისა და სამოქალაქო საზოგადოების წარმომადგენელ დაინტერესებულ მხარეებთან.

ზოგადად, ENEOLI-ის მიზანია ლექსიკური სიახლეების თემაზე მომუშავე კვლევითი საზოგადოების ჩამოყალიბება, რაც გულისხმობს თანამშრომლობითი ეკოსისტემის შექმნას, რომელიც ამ თემას დღევანდელი სოციოკულტურული სივრცის ცენტრში მოაქცევს.

მიუხედავად იმისა, რომ ნეოლოგიას (ახალი სიტყვებისა და მნიშვნელობების შესწავლას) დიდი მნიშვნელობა აქვს ენის განვითარებისა და ფუნქციონირების გააზრებაში, მას ნაკლები ყურადღება ეთმობა ენათმეცნიერებაში.

ეს მოხსენება წარმოგიდგენთ *ლექსიკური სიახლეების ევროპულ ქსელს*, ENEOLI-ს, რომლის მიზანია ნეოლოგიის კვლევის სფეროში არსებული ხარვეზების აღმოფხვრა ფართომასშტაბიანი, მრავალენოვანი და თანამშრომლობითი მიდგომებით. ENEOLI-ს ქსელში დღეისათვის გაერთიანებულია 50 ქვეყნის 370-ზე მეტი მკვლევარი. აღნიშნულ პროგრამას აქვს ოთხი ძირითადი მიზანი: (1) ნეოლოგიის მეტატერმინოლოგიის მრავალენოვანი გლოსარიუმის შექმნა; (2) ნეოლოგიის კვლევისთვის ციფრული რესურსებისა და ინსტრუმენტების განვითარება; (3) შედარებითი დიაქრონიული და სინქრონიული კვლევების ჩატარება და (4) ტრენინგების ინიციატივების დანერგვა ნეოლოგიის სფეროში განათლების პროფესიულ დონეზე განვითარებისათვის.

მოხსენებაში წარმოვადგენ ENEOLI-ის ინოვაციურ მეთოდოლოგიურ ჩარჩოს, რომელიც აერთიანებს ტრადიციულ ლექსიკოგრაფიულ პრაქტიკას კორპუსულ ლინგვისტიკასთან, კომპიუტერულ ინსტრუმენტებთან და მრავალენოვანი ანალიზის მიდგომებთან. ENEOLI-ის ერთ-ერთი მნიშვნელოვანი დამსახურებაა *NeoVoc*-ის, ნეოლოგიზმების მრავალენოვანი თეზაურუსის, შექმნა, აგრეთვე *NeoCorpus*-ის, ნეოლოგიზმების ვრცელი კორპუსის, შემუშავება, რომლის მეშვეობითაც შესაძლებელია ტერმინების დადასტურება და ენათშორისი კვლევების ჩატარება. გლოსარიუმის პროექტი, რომელიც ENEOLI Wikibase-ის მეშვეობით იმართება, განსაკუთრებულ ყურადღებას აქცევს ექსპერტების ჩართულობას, სტრუქტურირებული მონაცემების მართვას და ურთიერთთავსებადობას, რაც ტერმინოლოგიური თანმიმდევრულობისა და ხელმისაწვდომობის ახალ სტანდარტებს აწესებს. ლექსიკოგრაფიული პერსპექტივიდან, ENEOLI-ის საქმიანობა გულისხმობს ასევე ლექსიკონებში ნეოლოგიზმების შერჩევის, აღწერისა და აღმოჩენის სარედაქციო პრაქტიკების კვლევას, რომელიც მიზნად ისახავს საჯაროდ ხელმისაწვდომი გახადოს ლექსიკონებში ნეოლოგიზმების შეტანის კრიტერიუმების ტიპოლოგია.

ENEOLI ქმნის საფუძველს ლექსიკური სიახლეების შესწავლის ახალი პარადიგმის შესაქმნელად ევროპასა და მის ფარგლებს გარეთ თანამშრომლობითი, ინტერდისციპლინური და მრავალენოვანი მიდგომით.

Lexicographic Principles of Sulkhan-Saba Orbeliani

Margalitadze, Tinatini

Ilia State University, Georgia

tinatin.margalitadze@iliauni.edu.ge

The history of Georgian lexicography follows the same stages of development as European lexicography. The earliest period was characterized by glosses, marginal notes or explanations added to manuscripts and glossaries appended to the books of ancient authors. This Greco-Byzantine tradition was known to Georgian scholars who introduced this practice into Georgian reality.

The next stage in the history of European lexicography is the compilation of bilingual dictionaries (15th-17th centuries). Georgian lexicography also underwent this stage of development and the 17th century is characterised by the creation of bilingual dictionaries of Georgian in respect to both Oriental (Persian, Arabic) and European languages (Italian, Dutch).

The third stage of the European lexicography (from the 18th century) is dedicated to monolingual dictionaries and Sulkhan-Saba Orbeliani holds a very special place in the history of Georgian lexicography from this point of view, as his ground-breaking dictionary *Sitqvis Kona* (literally 'a bunch of words') is the first complete explanatory dictionary of Georgian.

This conference is dedicated to the 300th anniversary of his death. Moreover, students of the international lexicographic programme European Master in Lexicography, enrolled in this program in 2025, also bear his name – Sulkhan-Saba Orbeliani cohort and we are very grateful to the Consortium of 8 universities that run this program for honouring the outstanding Georgian lexicographer, writer, diplomat and public figure.

Sulkhan-Saba Orbeliani could be styled the Academy of Georgia, exactly the way Samuel Johnson was called the Academy of the Island by his contemporaries for his *Dictionary of the English Language*. Like Samuel Johnson, Orbeliani compiled his Dictionary alone, toiling over its three editions for almost 40 years. Unlike their French and Italian colleagues, who worked on dictionaries inside academies, they were alone, developing their unique lexicographic principles and laying the foundations of the lexicography of their languages and countries.

It is the fate of a representative of a small nation to remain unnoticed, in the majority of cases, to the outer world. This International Conference is a possibility for us to pay tribute to his name and make his legacy known beyond his homeland.

In my talk, I will examine the lexicographic principles developed by the lexicographer by the end of the 17th century, which places Sulkhan-Saba Orbeliani among the outstanding representatives of world lexicography.

სულხან-საბა ორბელიანის ლექსიკოგრაფიული პრინციპები

თინათინ მარგალიტაძე

ილიას სახელმწიფო უნივერსიტეტი, საქართველო

tinatin.margalitadze@iliauni.edu.ge

ქართულმა ლექსიკოგრაფიამ განვითარების იგივე გზა განვლო, რაც ევროპულმა ლექსიკოგრაფიამ. ლექსიკოგრაფიის განვითარების წინარე პერიოდში (ჩვ. წ-ის XV საუკუნემდე) ძირითადად იქმნებოდა გლოსები და გლოსარიუმები, რომელთა ფუნქცია იყო რთული და გაუგებარი სიტყვების ახსნა. ცნობილი ავტორების წიგნებისათვის გლოსარიუმების დართვის ბერძნულ-ბიზანტიური ტრადიცია კარგად იყო ცნობილი ქართველი მოღვაწეებისათვის, რომლებმაც ეს პრაქტიკა შემოიტანეს და დანერგეს ქართულ რეალობაში.

ევროპული ლექსიკოგრაფიის განვითარების შემდეგი ეტაპი იყო ორენოვანი ლექსიკონების შედგენის პერიოდი (XV-XVII საუკუნეები). ეს პერიოდი ქართული ლექსიკოგრაფიისთვისაც არის დამახასიათებელი. XVII საუკუნეში ქართული ენის არაერთი ორენოვანი ლექსიკონი შეიქმნა როგორც აღმოსავლურ (სპარსული, არაბული), ისე ევროპულ ენებთან (იტალიური, ჰოლანდიური) მიმართებით.

ევროპული ლექსიკოგრაფიის მესამე ეტაპი (მე-18 საუკუნიდან) განმარტებითი ლექსიკონების შექმნის პერიოდია და ამ თვალსაზრისით სულხან-საბა ორბელიანს განსაკუთრებული ადგილი უჭირავს ქართული ლექსიკოგრაფიის ისტორიაში, რადგან მისი ნოვატორული ლექსიკონი *სიტყვის კონა* ქართული ენის პირველი სრული განმარტებითი ლექსიკონია.

ეს კონფერენცია ეძღვნება 300 წლისთავს სულხან-საბა ორბელიანის გარდაცვალებიდან. აღსანიშნავია ისიც, რომ 2025 წელს საერთაშორისო ლექსიკოგრაფიულ პროგრამაზე („ევროპული სამაგისტრო პროგრამა ლექსიკოგრაფიაში“ European Master in Lexicography) ჩარიცხულ სტუდენტებს მისი სახელი მიენიჭა – სულხან-საბა ორბელიანის სახელობის კურსი. ილიას სახელმწიფო უნივერსიტეტი მადლიერია 8 ევროპული უნივერსიტეტის კონსორციუმისა, რომელიც მართავს ამ პროგრამას, იმისათვის, რომ მან პატივი მიაგო ქართველ ლექსიკოგრაფს, მწერალს, დიპლომატსა და სახელმწიფო მოღვაწეს.

სულხან-საბა ორბელიანს შეიძლება ვუწოდოთ „საქართველოს აკადემია“, ზუსტად ისე, როგორც სემიუელ ჯონსონს შეარქვეს „კუნძულის აკადემია“ თანამედროვეებმა მის მიერ შექმნილი *ინგლისური ენის ლექსიკონისთვის*. სემიუელ ჯონსონის მსგავსად, ორბელიანმაც მარტომ შეადგინა თავისი ლექსიკონი და თითქმის 40 წლის განმავლობაში შრომობდა მის სამ რედაქციაზე. მათი ფრანგი და იტალიელი კოლეგებისაგან განსხვავებით, რომლებიც აკადემიებში მუშაობდნენ ლექსიკონებზე, ისინი მარტო იყვნენ, მარტო ავითარებდნენ საკუთარ უნიკალურ ლექ-

სიკოგრაფიულ პრინციპებს და უყრიდნენ საფუძველს თავიანთი ენებისა და ქვეყნების ლექსიკოგრაფიას.

პატარა ერის წარმომადგენლის ბედისწერაა, უმეტეს შემთხვევაში, გარე სამყაროსთვის შეუმჩნეველი დარჩეს. ეს საერთაშორისო კონფერენცია საშუალებას გვაძლევს პატივი მივაგოთ მის სახელს და მისი მემკვიდრეობა გავაცნოთ საქართველოს ფარგლებს გარეთაც.

ჩემს მოხსენებაში დეტალურად განვიხილავ იმ ლექსიკოგრაფიულ პრინციპებს, რომლებიც ლექსიკოგრაფმა XVII საუკუნის ბოლოს შეიმუშავა და რომლებიც სულხან-საბა ორბელიანს მსოფლიო ლექსიკოგრაფიის საუკეთესო წარმომადგენლებს შორის უჩენს ადგილს.

Abstracts of Papers
მოხსენებების თეზისები

Bilingual Experiences of Filipino Learners of German and their Attitudes towards Dictionary Use in the Foreign Language Classroom

Agustin, Adam Rey
Goethe Institut Philippinen
European Master in Lexicography (EMLex) alumni
atagusin@alum.up.edu.ph, adamreyagustin@gmail.com

Preliminary findings suggest that Filipino-English bilingual learners of German show a strong preference for bilingual dictionaries, particularly those offering English-German equivalence, when performing receptive tasks such as reading comprehension. Despite reporting balanced proficiency in both Filipino and English, learners tend to favor English as the intermediary language for decoding German texts. This preference appears to be influenced by their educational exposure to English through formal instruction and media, rather than early home use.

Correlational analysis indicates that individual bilingual profiles —such as age of acquisition, language dominance, and self-rated proficiency—are associated with dictionary preferences. Learners with higher self-reported English proficiency are more likely to perceive English-based dictionaries as more effective and accessible. In contrast, monolingual German dictionaries and those translating into Philippine languages are met with neutral or less favorable attitudes, suggesting a gap in perceived utility or accessibility.

These findings underscore the importance of considering learners' bilingual backgrounds when designing lexicographic tools and instructional strategies in the foreign language classroom.

Keywords: *bilingualism, bilingual lexicography, foreign language learning, attitudes, LEAP-Q*

გერმანული ენის ფილიპინელ შემსწავლელთა ორენოვანი გამოცდილება და მათი დამოკიდებულება ლექსიკონის გამოყენების მიმართ უცხო ენის სწავლის პროცესში

ადამ რეი აგუსტინი
ევროპული მაგისტრი ლექსიკოგრაფიაში
გოეთეს ინსტიტუტი ფილიპინებში
atagusin@alum.up.edu.ph, adamreyagustin@gmail.com

კვლევის წინასწარი შედეგები მიუთითებს, რომ გერმანული ენის შემსწავლელი ორენოვანი ფილიპინელები (მეორე ენა – ინგლისური), ისეთი დავალებების შესრულებისას, როგორიცაა წაკითხულის გააზრება, უპირატესობას ანიჭებენ ორენოვან ლექსიკონებს, განსაკუთრებით ისეთ ლექსიკონებს, რომლებიც გვთავაზობენ ინგლისურ-გერმანულ ეკვივალენტებს. მიუხედავად იმისა, რომ შემსწავლელები თანაბრად ფლობენ როგორც ფილიპინურ, ისე ინგლისურ ენებს, ისინი უპირატესობას ანიჭებენ ინგლისურს, როგორც შუამავალ ენას გერმანული ტექსტების ინტერპრეტირებისას. აღნიშნულ უპირატესობაზე, როგორც ჩანს, გავლენას ახდენს მედია, აგრეთვე ინგლისური ენის დომინანტური მდგომარეობა საგანმანათლებლო პროცესში და არა ინგლისურის გამოყენება ოჯახში ადრეული ასაკიდან.

კორელაციური ანალიზი მიუთითებს, რომ ინდივიდუალური ორენოვანი მონაცემები, როგორიცაა ათვისების ასაკი, ენის დომინანტობა და ენის თვითშეფასებითი ფლობა, დაკავშირებულია ლექსიკონის შერჩევასთან. შემსწავლელები, რომლებმაც ინგლისური ენის ფლობის უფრო მაღალი თვითშეფასება დააფიქსირეს, მიიჩნევენ, რომ ინგლისურ ენაზე დაფუძნებული ლექსიკონები უფრო ეფექტური და ხელმისაწვდომია. ამის საპირისპიროდ, ერთენოვანი გერმანული ლექსიკონები და ფილიპინურ ენებზე ნათარგმნი ლექსიკონები ნეიტრალურ ან ნაკლებად დადებით დამოკიდებულებას იწვევს.

ეს შედეგები ხაზს უსვამს შემსწავლელთა ორენოვანი წარმომავლობის გათვალისწინების მნიშვნელობას ლექსიკოგრაფიული ინსტრუმენტებისა და სასწავლო სტრატეგიების შემუშავებისას უცხო ენის სწავლის პროცესში.

საკვანძო სიტყვები: *ბილინგვიზმი, ორენოვანი ლექსიკოგრაფია, უცხო ენის შესწავლა, დამოკიდებულებები, LEAP-Q*

თეზისი ქართულად თარგმნა თამარ ლალუაშვილმა.

Who Creates New Lithuanian Words: Society or Linguists? The case of *The Database of Lithuanian Neologisms*

Aleksaitė, Agnė
Institute of the Lithuanian Language, Lithuania
agne.aleksaite@lki.lt

Throughout time all languages of the world undergo a number of changes, which is a natural and necessary phenomenon (Crystal 2005: 31) ensuring the vitality of language in the globalized world. The most dynamic part of language, namely lexis may be affected by time and people (Jakaitienė 2018: 3), for example, by creators of neologisms. Currently, the Lithuanian language faces 'the age of neologisms' (Naktinienė 2002: 47; Würschinger 2021: 15) or 'neologism boom' (Girčienė 2005: 78; Miliūnaitė 2020: 82; Nabievna 2023: 289), as new words are being coined all the time. To record new words occurring in the Lithuanian language, in the autumn of 2011, the Institute of the Lithuanian Language initiated its compilation and digitalization by creating The Database of Lithuanian Neologisms (henceforth – ND; see more about the database at https://ekalba.lt/naujazodziai), which contains new units of language that entered the Lithuanian language at the end of the 20th and beginning of the 21st century and became part of common parlance. At the beginning of March 2025, the database comprises more than 10,000 new words from different domains of usage of contemporary Lithuanian. The database provides the definition of a new coinage, indicates the area where it is used, shows its semantic and compositional ties with other new words in the database, illustrates usage with examples, and gives normative comments to language users if necessary.

The selection of data for ND is grounded on one criterion that is recognized in the theory and practice of new coinages, which is the recording of words and their meanings in dictionaries. New coinages are words and meanings thereof that are not featured in the main sources of lexicography. The dictionaries containing lexis originating in the early 21st century that are important for establishing the database are The Contemporary Lithuanian Language Dictionary (2006), The Dictionary of the Lithuanian Language, and The Dictionary of International Words (2001).

The unique feature of ND is that, in addition to neologisms in their traditional sense, it also features lexical innovations (occasionalisms) that have only been used once, situationally, figuratively as often as not, and only time may show whether they will become common words and make their way into dictionaries. Stylistically neutral, nominative neologisms identify new realities and perform a nominative function.

The presentation will focus on the purpose of creating ND, its data, sources, and search options.

In Lithuania, there is a prevalent stereotype that neologisms are created solely by linguists. I will provide examples from the database to demonstrate that new

Lithuanian words are created by the society, including writers, journalists, politicians, influencers, and others. I will also clarify the reasons why this stereotype remains prevalent in Lithuania.

Keywords: *Lithuanian language, neologisms, language creation, database.*

ვინ ქმნის ახალ ლიეტუვურ სიტყვებს: საზოგადოება თუ ლინგვისტები?
ლიეტუვური ნეოლოგიზმების მონაცემთა ბაზა

აგნე ალექსაიტე
ლიეტუვური ენის ინსტიტუტი, ლიეტუვა
agne.aleksaite@lki.lt

დროსთან ერთად მსოფლიოს ყველა ენა განიცდის ცვლილებებს, რაც ბუნებრივი და აუცილებელი მოვლენაა (Crystal 2005: 31) იმისთვის, რომ გლობალიზებულ სამყაროში ენამ შეინარჩუნოს სიცოცხლისუნარიანობა. ენის ყველაზე დინამიკურ დონეზე, კერძოდ, ლექსიკაზე შესაძლოა დრო და ადამიანები ახდენდნენ ზეგავლენას (Jakaitienė 2018: 3), მაგალითად, ნეოლოგიზმების შემქმნელები. დღესდღეობით ლიეტუვური ენა „ნეოლოგიზმების ეპოქის“ (Naktinienė 2002: 47; Würschinger 2021: 15) ან „ნეოლოგიზმების აღმავლობის“ (Girčienė 2005: 78; Miliūnaitė 2020: 82; Nabievna 2023: 289) პერიოდშია, ვინაიდან მუდმივად იქმნება ახალი სიტყვები. ლიეტუვურ ენაში არსებული ახალი სიტყვების დოკუმენტირებისავის 2011 წლის შემოდგომაზე, ლიეტუვური ენის ინსტიტუტმა წამოიწყო პროექტი მათი შედგენისა და გაციფრების მიზნით. ამისათვის შეიქმნა ლიეტუვური ნეოლოგიზმების მონაცემთა ბაზა (ND; იხ. ვრცლად მონაცემთა ბაზის შესახებ https://ekalba.lt/naujazodziai), რომელიც მოიცავს იმ ლექსიკურ ერთეულებს, რომლებიც ლიეტუვურ ენაში შემოვიდა მე-20 საუკუნის მიწურულსა და 21-ე საუკუნის დასაწყისში და დამკვიდრდა საერთო ლექსიკაში. 2025 წლის მარტის დასაწყისისთვის მონაცემთა ბაზა მოიცავს სხვადასხვა დარგის 10,000 ახალ სიტყვას თანამედროვე ლიეტუვურში. მონაცემთა ბაზაში მოცემულია ახალი სიტყვის განმარტება, მითითებულია გამოყენების სფერო, ნაჩვენებია მონაცემთა ბაზაში შეტანილ სხვა ახალ სიტყვებთან არსებული სემანტიკური და კომპოზიციური კავშირები, მოცემულია საილუსტრაციო მაგალითები და საჭიროების შემთხვევაში, ვხვდებით ენის მომხმარებლებისათვის განკუთვნილ ნორმატიული შინაარსის მქონე კომენტარებს.

მონაცემთა ბაზისათვის ნეოლოგიზმების შერჩევა ეფუძნება ახალი სიტყვების თეორიასა და პრაქტიკაში კარგად დამკვიდრებულ კრიტერიუმს, კერძოდ ლექსიკური ერთეულებისა და მათი მნიშვნელობების არსებობას ლექსიკონებში. შესაბამისად, ახალ სიტყვებად მიიჩნევა ის ლექსიკური ერთეულები და მნიშვნელობები, რომლებიც არ დასტურდება ამჟამად არსებულ ძირითად ლექსიკოგრაფიულ წყაროებში, როგორებიცაა: *თანამედროვე ლიეტუვური ენის ლექსიკონი* (2006), *ლიეტუვური ენის ლექსიკონი* და *საერთაშორისო სიტყვების ლექსიკონი* (2001).

ნეოლოგიზმების მონაცემთა ბაზის უნიკალური მახასიათებელია ის, რომ გარდა, ტრადიციული გაგებით არსებული ნეოლოგიზმებისა, ბაზა მოიცავს ლექსიკურ ინოვაციებს (ოკაზიონალიზმებს), რომლებიც ერთხელაა გამოყენებული და მხოლოდ დრო გვიჩვენებს გადაიქცევა თუ

არა ეს სიტყვები საერთო ლექსიკის ნაწილად და დამკვიდრდება ლექსიკონებში. სტილისტურად ნეიტრალური, სახელადი ნეოლოგიზმები ასახავენ ახალ რეალობას და ასრულებენ სახელდების ფუნქციას.

პრეზენტაციაში ყურადღება გამახვილდება ნეოლოგიზმების მონაცემთა ბაზის შექმნაზე, წყაროებსა და ძიების შესაძლებლობებზე.

ლიეტუვაში გავრცელებული სტერეოტიპის თანახმად, ნეოლოგიზმებს ქმნიან მხოლოდ ლინგვისტები. პრეზენტაციაში წარმოვადგენ მაგალითებს მონაცემთა ბაზიდან, სადაც ნათლად ჩანს, რომ ახალი სიტყვების შექმნაში მონაწილეობს საზოგადოება, მათ შორის, მწერლები, ჟურნალისტები, პოლიტიკოსები, ინფლუენსერები და სხვა. აგრეთვე, განვმარტავ, თუ რატომ დომინირებს აღნიშნული სტერეოტიპი ლიეტუვაში.

საკვანძო სიტყვები: *ლიეტუვური ენა, ნეოლოგიზმები, სიტყვათქმნადობა ენაში, მონაცემთა ბაზა.*

თეზისი ქართულად თარგმნა ქეთევან მჭედლიშვილმა.

References:

Crystal, D. (2005). *Kalbos mirtis*, Vilnius: Tyto alba.

Girčienė, J. (2005). Neologizmų integracija į tekstą. – *Žmogus ir žodis* 1, 78–82.

Jakaitienė, E. (2018). Lietuvių kalba per šimtmetį – kaip pakito leksika? – *Gimtoji kalba* 2, 3–16.

Miliūnaitė, R. (2020). Neologijos terminija ir *Lietuvių kalbos naujažodžių duomenyno* praktika. – *Terminologija* 27, 81–107. Available online: https://doi.org/10.35321/term27-04.

Nabievna Ibragimova, Z. (2023). Lexico-grammatical Problems of Translating the English Neologisms to Uzbek. – *EPRA International Journal of Multidisciplinary Research (IJMR)* 9 (3), 289–292. Available online: http://eprajournals.net/index.php/IJMR/article/view/1734.

Naktinienė, G. (2002). Specialiosios leksikos pateikimas *Bendrinės lietuvių kalbos žodyne*. – *Kalbos kultūra* 75, 47–56. Available online: http://www.bendrinekalba.lt/Straipsniai/75/Naktiniene_KK_75_straipsnis.pdf.

ND – *The Database of Lithuanian Neologisms*. Rita Miliūnaitė and Agnė Aleksaitė, Vilnius: Institute of the Lithuanian Language. Available online: https://ekalba.lt/naujazodziai/. ISBN 978-609-411-147-1. https://doi.org/10.35321/neol.

Würschinger, Q. (2021). Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter. – *Frontiers in Artificial Intelligence* 4, 1–20. Available online: https://doi.org/10.3389/frai.2021.648583.

Shaping the Future of Neology Repositories:
Insights from the ENEOLI Survey & Strategic Recommendations

Asadpour, Hiwa
Goethe University Frankfurt, Germany
asadpour@lingua.uni-frankfurt.de

The rapid development of digital technology has transformed lexicography, requiring modernized solutions for tracking and managing neologisms. Effective repository design is crucial to make sure about the quality, accessibility and inclusivity in the storage and dissemination of neological data. This presentation analyses the needs of different stakeholders involved in neology research – including lexicographers, terminologists, translators, teachers, repository managers and software developers – based on the results of the ENEOLI[1] COST[2] Action survey. Existing repository models vary widely in terms of functionality and usability. Institutional repositories provide structured and validated data but may not effectively serve minority languages or specialized domains. National language portals focus on official terminology, but often lack dynamic user contributions. Specialized corpora provide high-quality domain-specific data, but their limited scope limits wider linguistic applications. Open access platforms encourage collaboration but pose challenges in terms of data reliability. Crowdsourcing tools leverage community participation, but require continuous quality control. AI-powered repositories and Linked Open Data (LOD) technologies offer promising approaches, but raise concerns about automation bias, implementation complexity, and interoperability. This presentation evaluates three proposed repository models to address these challenges:

1. Hybrid model – combining institutional oversight with community-driven contributions to balance reliability and innovation.
2. Multilingual LOD platform – using semantic web technologies to link different linguistic resources, ensuring wider accessibility and cross-domain compatibility.
3. AI-enabled repository – integrating machine learning and large language models (LLMs) for automated neologism detection and semantic annotation.

Real-world examples will illustrate the strengths and weaknesses of different repository approaches. These include TermPortal's success in promoting indigenous Icelandic terminology (Steingrímsson et al., 2020), Observatorio Lázaro's analysis of anglicisms in the Spanish media (Observatorio Lázaro, 2024), and Sketch Engine's corpus-based lexicographic applications. I will also discuss how LOD technologies can improve semantic interoperability and facilitate multilingual research. Key takeaways for conference participants will include practical recommendations for designing repositories that effectively accommodate emerging linguistic trends and support collaborative innovation. The findings

1 ENEOLI is the acronym for European Network on Lexical Innovation (www.eneoli.eu).

2 COST is the acronym for European **Co**operation in **S**cience and **T**echnology.

emphasize the need for repositories that are adaptable, scalable and capable of integrating diverse data sources while maintaining high quality standards. The session will conclude with strategic recommendations for future repository development, incorporating feedback from lexicographers and stakeholders. By addressing the intersection of lexicography and technology, this study contributes to the ongoing discourse on modern lexicographic practices. It demonstrates how well-structured repositories can improve research capabilities, continuing terminology management, and developing interdisciplinary collaboration in neology studies. The findings will be of value to researchers, educators and technology developers seeking to optimize digital linguistic resources for the 21st century.

Keywords: *Neology tracking, Lexicographic repositories, Terminology management, Linked Open Data (LOD), AI in lexicography, Language technology*

ნეოლოგიზმების ბაზების (საცავების) შექმნის სამომავლო პერსპექტივები: ENEOLI-ის პროექტის ფარგლებში ჩატარებული გამოკითხვის შედეგები და სტრატეგიული რეკომენდაციები

ჰივა ასადპური

ფრანკფურტის გოეთეს უნივერსიტეტი, გერმანია

asadpour@lingua.uni-frankfurt.de

ციფრული ტექნოლოგიების სწრაფმა განვითარებამ შეცვალა ლექსიკოგრაფია, რამაც ნეოლოგიზმების აღმოჩენის, მონიტორინგისა და მართვის მიმართულებით მოდერნიზებული გადაწყვეტილებების საჭიროებაც წარმოშვა. ნეოლოგიზმების ბაზების დიზაინის ეფექტურად შემუშავება გადამწყვეტი მნიშვნელობისაა, რათა ნეოლოგიზმების თავმოყრისა/შეგროვებისა და გავრცელების პროცესში უზრუნველყოფილ იქნეს ხარისხი, ხელმისაწვდომობა და ინკლუზიურობა. წინამდებარე პრეზენტაციაში, ENEOLI COST Action-ის გამოკითხვის შედეგებზე დაყრდნობით, გაანალიზებულია ნეოლოგიზმების კვლევაში ჩართული სხვადასხვა დაინტერესებული მხარის – მათ შორის ლექსიკოგრაფების, ტერმინოლოგიის სპეციალისტების, მთარგმნელების, მასწავლებლების, ბაზების მენეჯერებისა და პროგრამული უზრუნველყოფის დეველოპერების – საჭიროებები. არსებული ბაზების მოდელები მნიშვნელოვნად განსხვავდება ფუნქციებისა და გამოყენების თვალსაზრისით. ინსტიტუციურ ბაზებში წარმოდგენილია სტრუქტურირებული და შემოწმებული/დადასტურებული მონაცემები, თუმცა, ისინი ხშირად არ გამოდგება უმცირესობათა ენებთან თუ სპეციალიზებულ დარგებთან მიმართებით. ენათა ეროვნული პორტალები ორიენტირებულია ოფიციალურ ტერმინოლოგიაზე, მაგრამ ხშირად დინამიკურად ვერ ასახავს მომხმარებელთა წვლილს. სპეციალიზებულ კორპუსებში წარმოდგენილია მაღალი ხარისხის, დარგისთვის სპეციფიკური მონაცემები, თუმცა მათი მასშტაბი შეზღუდულია, რაც თავის მხრივ ლინგვისტური გამოყენების სპექტრსაც ზღუდავს. ღია წვდომის მქონე პლატფორმები თანამშრომლობას ეფუძნება, თუმცა მონაცემების სანდოობის თვალსაზრისით არაერთი გამოწვევის წინაშეც დგას. საზოგადოების ფართო ფენების დახმარებით შექმნილი (ე.წ. „ქრაუდსორსინგის“) ინსტრუმენტები საზოგადოებრივი ჯგუფების ჩართულობას უზრუნველყოფს, თუმცა ის ასევე ხარისხის უწყვეტად კონტროლს საჭიროებს. ხელოვნური ინტელექტზე დაფუძნებული ბაზები და „დაკავშირებული ღია მონაცემების“ (Linked Open Data (LOD)) ტექნოლოგიები პერსპექტიულ მიდგომებს გვთავაზობს, თუმცა ისინი პრობლემურია ავტომატიზაციის მიკერძოების, განხორციელების სირთულისა და ოპერაციული თავსებადობის თვალსაზრისით. ამ გამოწვევებთან გამკლავების მიზნით, წინამდებარე პრეზენტაციაში შეფასებულია ბაზების სამი შემოთავაზებული მოდელი:

1. ჰიბრიდული მოდელი – ინსტიტუციური ზედამხედველობისა და საზოგადოების წვლილის კომბინირება სანდოობისა და ინოვაციის დასაბალანსებლად.
2. მრავალენოვანი LOD პლატფორმა – სემანტიკური ქსელის ტექნოლოგიების გამოყენება სხვადასხვა ენობრივი რესურსის ერთმანეთთან დასაკავშირებლად, რაც ფართო ხელმისაწვდომობასა და სხვადასხვა დარგთან თავსებადობას უზრუნველყოფს.
3. ხელოვნურ ინტელექტზე დაფუძნებული ბაზები – მანქანური სწავლებისა და დიდი ენობრივი მოდელების (LLMs) ინტეგრირება ნეოლოგიზმების ავტომატური აღმოჩენისა და სემანტიკური ანოტირებისთვის.

რეალური მაგალითები ბაზების სხვადასხვა მოდელების ძლიერ და სუსტ მხარეებს წარმოაჩენს. მათ შორისაა: TermPortal-ის წარმატება ისლანდიური ტერმინოლოგიის პოპულარიზაციაში (Steingrímsson et al., 2020), Observatorio Lázaro-ს მიერ ესპანურ მედიაში გამოყენებული ანგლიციზმების ანალიზი (Observatorio Lázaro, 2024), და Sketch Engine-ის კორპუსზე დაფუძნებული ლექსიკოგრაფიული პროექტები. ასევე განვიხილავ, როგორ შეუძლია LOD ტექნოლოგიებს ოპერაციული თავსებადობის გაუმჯობესება სემანტიკურ დონეზე, ასევე მრავალენოვანი კვლევების ხელშეწყობა. კონფერენციის მონაწილეები მიიღებენ პრაქტიკულ რეკომენდაციებს ბაზების ისეთ დიზაინთან დაკავშირებით, რომლებიც ეფექტურად ერგება ახალ ენობრივ ტენდენციებს და ხელს უწყობს თანამშრომლობაზე დაფუძნებულ ინოვაციას. ეს კვლევა ხაზს უსვამს ისეთი ბაზების შექმნის საჭიროებას, რომლებიც არის ადაპტირებადი, რომლის მასშტაბის გაფართოება შესაძლებელია და აქვს სხვადასხვა მონაცემთა წყაროს ინტეგრირების უნარი და ამასთანავე შენარჩუნებულია მაღალი ხარისხის სტანდარტები. მოხსენება დასრულდება სტრატეგიული რეკომენდაციებით ბაზების სამომავლო განვითარებასთან დაკავშირებით, სადაც გათვალისწინებული იქნება ლექსიკოგრაფებისა და დაინტერესებული მხარეების უკუკავშირი. ლექსიკოგრაფიისა და ტექნოლოგიის კვეთაზე, ეს კვლევა წვლილს შეიტანს თანამედროვე ლექსიკოგრაფიული პრაქტიკის შესახებ მიმდინარე დისკუსიაში. ის აჩვენებს, თუ როგორ შეუძლია კარგად სტრუქტურირებულ ბაზებს გააუმჯობესოს კვლევის შესაძლებლობები, ტერმინოლოგიის მართვა და ნეოლოგიზმების კვლევებში ინტერდისციპლინური თანამშრომლობის განვითარება. კვლევის შედეგები ღირებული იქნება მკვლევრებისთვის, პედაგოგებისა და ტექნოლოგიების დეველოპერებისთვის, რომლებიც 21-ე საუკუნის ციფრული ენობრივი რესურსების ოპტიმიზებაზე მუშაობენ.

საკვანძო სიტყვები: *ნეოლოგიზმების აღმოჩენა და მონიტორინგი, ლექსიკოგრაფიული ბაზები, ტერმინოლოგიის მართვა, Linked Open Data (LOD), ხელოვნური ინტელექტი ლექსიკოგრაფიაში, ენობრივი ტექნოლოგიები*

თეზისი ქართულად თარგმნა ქეთევან ჭიღლაძემ.

References

Observatorio Lázaro (2024). *Observatory of anglicism usage in the Spanish press*. Retrieved from https://observatoriolazaro.es/en/

Steingrímsson, S., Þorbergsdóttir, A., Danielsson, H. and G. Th. Ornolfsson. (2020). TermPortal: A Workbench for Automatic Term Extraction from Icelandic Texts. In *Proceedings of the 6th International Workshop on Computational Terminology*, pages 8–16, Marseille, France. European Language Resources Association.

On the Issue of Borrowings in the Georgian Business Setting

Baratashvili, Zurabi
Ivane Javakhishvili Tbilisi State University
Georgian National Academy of Sciences, Georgia

Buskivadze, Khatuna
Caucasus University, Georgia
zurabbaratashvili85@gmail.com, khbuskivadze@cu.edu.ge

In recent years, we have seen a growing trend of voluntarily asserting business terminology in the spoken and written Georgian language. Yet, little empirical research has systematically examined the motivations behind using these lexical items. This research aims to study the nature of this business terminology and the frequency of its use. Moreover, the research aims to classify the lexical items structurally in the Georgian Business Press and among Georgians working in the field of business. The methodology deals with the qualitative method (documenting 50 business e-articles to grab the lexical items on BMG – Business Media in Georgia) and quantitative (using Google Forms to assess the frequency of using the business terminology in the spoken discourse, among Georgians working in the business organizations).

The study aims: 1. to measure adaptation/accommodation quality of borrowings; 2. to determine the correlation between respondents' age, sex, living and working places, and frequency of using Business borrowings; 3. to structurally classify the collected lexical borrowings in the Georgian business setting. Out of 50 business e-articles, we collected business terms, which were grouped accordingly: 1. Those adopted/accommodated in the written and spoken Georgian language in the 20th century (e.g., *transformatsia*, "transformation", *sanktsia*, "sanction", etc.); 2. The rest were adopted/accommodated primarily in the spoken Georgian language (e.g., *autsorsi* "outsource", *keshbeki* "cashback", *startapi* "start-up", *timi* "team", *dedlaini* "deadline", etc.). Based on the quantitative research, we collected 107 responses with the help of Google Forms. Then, we analyzed them with the JASP program to measure the frequency (Lakert's scale from 1 – never to 5 always) of usage and its connection with respondents' age, sex, current working status, and living places. The quantitative data shows that female (74) respondents use most of the terms more frequently than males (34), e.g., *ტენდენცია/tendentsia* "tendency" (F (median – 4, mean – 3.699) M (median – 3, mean – 3.382); *გენერირება/generireba* "to generate" (F (median – 3, mean – 3.110) M (median – 2, mean – 2.676), *თიმი/timi timi* "team" (F (median – 4, mean – 3.740), (M (median – 4, mean – 3.324). Moreover, the younger the respondents are, the more frequently they use business borrowings. Respondents who spent their lifetime in Tbilisi and other urban areas use borrowed lexical items less regularly than those from rural areas, e.g., *ბოსი/bosi* "boss", *სტარტაპი/startapi* "start-up" (rural – median – 5, while in Tbilisi and other urban areas – median is 4). Surprisingly, there was almost no difference in frequency regarding the re-

spondents' working areas; people working in business sectors use the business borrowings as frequently as those from other fields, e.g., *ტარიფი/taripi* – "tariff", *სტარტაპი/startapi* – "start-up" (means in business and other fields are the same – 4). Regarding structural classification, most lexical borrowings are nouns, while adjectives are in the minority.

Analysis reveals that lexical borrowings are not merely gap-fillers but serve discourse-level functions such as indexing modernity and signaling group identity. The findings challenge the traditional assumptions about language purity and suggest that lexical borrowing is a dynamic, context-sensitive process that reflects broader social negotiations of identity and globalization, especially for such a field as Business.

Keywords*: Borrowings, Business Setting, Spoken Georgian, Press.*

ლექსიკური სესხების საკითხი ქართულ ბიზნესკონტექსტში

ზურაბ ბარათაშვილი

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი, საქართველო

საქართველოს მეცნიერებათა ეროვნული აკადემია,

ხათუნა ბუსკივაძე

კავკასიის უნივერსიტეტი, საქართველო

zurabbaratashvili85@gmail.com, khbuskivadze@cu.edu.ge

ბოლო წლებში თავი იჩინა ბიზნესტერმინოლოგიის ნებაყოფლობითი დამტკიცების მზარდმა ტენდენციამ სალაპარაკო და წერილობით ქართულ ენაში. მიუხედავად ამისა, მცირეა ემპირიული კვლევები, რომლებიც რეგულარულად იკვლევენ ამ ლექსიკური ერთეულების გამოყენების მოტივაციებს. წინამდებარე კვლევა მიზნად ისახავს ქართულ ენაში ნასესხები ბიზნესტერმინოლოგიის ბუნებისა და მისი გამოყენების სიხშირის შესწავლას. უფრო მეტიც, კვლევა მიზნად ისახავს ლექსიკური ერთეულების სტრუქტურულ კლასიფიკაციას ქართულ ბიზნესპრესასა და ბიზნესის სფეროში მოღვაწე ქართველებში. კვლევა ეყრდნობა თვისებრივ (50 ელექტრონული ბიზნესსტატიის დოკუმენტირება ლექსიკური ერთეულების შესაგროვებლად BMG – ბიზნესმედია საქართველოში) და რაოდენობრივ (Google Forms-ის გამოყენებით ბიზნესის ტერმინოლოგიის გამოყენების სიხშირის შესაფასებლად სალაპარაკო ქართულში, ბიზნესორგანიზაციებში მომუშავე ქართველებში) მეთოდებს.

კვლევის მიზანია: 1. გაზომოს სესხების ადაპტაციის/აკომოდაციის ხარისხი; 2. დაადგინოს ურთიერთკავშირი რესპონდენტთა ასაკს, სქესს, საცხოვრებელ და სამუშაო ადგილსა და ბიზნესსესხების გამოყენების სიხშირეს შორის; 3. მოახდინოს შეგროვებული ლექსიკური ნასესხობების ქართულ ბიზნესსფეროში სტრუქტურული კლასიფიკაცია. 50 ელექტრონული ბიზნესსტატიიდან შევკრიბეთ ბიზნესტერმინები, რომლებიც დაჯგუფდა შემდეგნაირად: 1. მე-20 საუკუნეში წერილობით და სალაპარაკო ქართულში ნასესხები ტერმინები (მაგ., *ტრანსფორმაცია* “transformation”, *სანქცია* “sanction”, და სხვა); 2. ტერმინები, რომლებიც ძირითადად სალაპარაკო ქართულში გამოიყენება (მაგ., *აუთსორსი* “outsource”, *ქეშბექი* “cashback”, *სტარტაპი* “start-up”, *თიმი* “team”, დედლაინი “deadline” და სხვა). რაოდენობრივი კვლევის საფუძველზე, Google Forms-ის დახმარებით შევაგროვეთ 107 პასუხი. გავაანალიზეთ ისინი JASP პროგრამით, რათა გაგვეზომა გამოყენების სიხშირე (ლაკერტის სკალით 1-დან (არასდროს) 5-მდე (ყოველთვის)) და მისი კავშირი რესპონდენტთა ასაკთან, სქესთან, სამუშაო და საცხოვრებელ ადგილებთან.

რაოდენობრივი მონაცემები აჩვენებს, რომ მდედრობითი (74) რესპონდენტები ტერმინების უმეტესობას უფრო ხშირად იყენებენ, ვიდრე

მამრობითი (34), მაგ., *ტენდენცია* “tendency” (მდედრობითი (მედიანა – 4, საშუალო არითმეტიკული – 3,699) მამრობითი (მედიანა – 3, საშუალო არითმეტიკული – 3,382); *გენერირება* “to generate” (მდედრობითი (მედიანა – 3, საშუალო არითმეტიკული – 3,110) M (მედიანა – 2, საშუალო არითმეტიკული – 2,676), *თიმი* “team“ (მდედრობითი (მედიანა – 4, საშუალო არითმეტიკული – 3,740), (მამრობითი (მედიანა – 4, საშუალო არითმეტიკული – 3,324). უფრო მეტიც, რაც უფრო ახალგაზრდაა რესპონდენტი, მით უფრო ხშირად იყენებს ნასესხებ ბიზნესტერმინებს. რესპონდენტები, რომლებიც მთელი ცხოვრება თბილისსა და სხვა საქალაქო დასახლებებში ცხოვრობდნენ, ნასესხებ ლექსიკურ ერთეულებს ნაკლებად რეგულარულად იყენებენ, ვიდრე სასოფლო ტიპის დასახლებების მაცხოვრებლები, მაგ., *ბოსი* “boss”, *სტარტაპი* “start-up” (სასოფლო ტიპის დასახლებების მცხოვრებლები – მედიანა – 5), ხოლო თბილისსა და სხვა საქალაქო დასახლებებში – მედიანა არის 4). გასაკვირია, რომ რესპონდენტებში სამუშაო სფეროების მიხედვით სიხშირეში თითქმის არანაირი განსხვავება არ დაფიქსირებულა; ბიზნესსექტორში დასაქმებული ადამიანები ნასესხებ ბიზნესსიტყვებს ისევე ხშირად იყენებენ, როგორც სხვა სფეროებში დასაქმებულები, მაგალითად, *ტარიფი* “tariff”, *სტარტაპი* “start-up” (ბიზნესსა და სხვა სფეროებში ნასესხები სიტყვები გამოყენების თანაბარი სიხშირით ხასიათდება, საშუალო არითმეტიკული – 4). სტრუქტურული კლასიფიკაციის თვალსაზრისით, ლექსიკური ნასესხობების უმეტესობა არსებითი სახელია, ხოლო ზედსართავი სახელები უმცირესობაშია.

კვლევა აჩვენებს, რომ ლექსიკური ნასესხობები არა მხოლოდ შესაბამისი ლექსიკური ერთეულების სიმწირის შემავსებლებია, არამედ ემსახურება დისკურსის დონის ფუნქციებს, როგორიცაა თანამედროვეობის ჩვენება და ჯგუფური იდენტობის გამოხატვა. დასკვნები ეჭვქვეშ აყენებს ენის სიწმინდის შესახებ ტრადიციულ დაშვებებს და მიუთითებს, რომ ლექსიკური სესხება არის დინამიკური, კონტექსტით განპირობებული პროცესი, რომელიც ასახავს იდენტობისა და გლობალიზაციის უფრო ფართო სოციალურ მახასიათებელს, განსაკუთრებით ისეთი სფეროსთვის, როგორიცაა ბიზნესი.

საკვანძო სიტყვები*: სესხება, ბიზნესგარემო, სალაპარაკო ქართული, პრესა*

From Corpus to Purpose: A Teleological Framework for Legal Translation

Beridze, Khatuna
Batumi Shota Rustaveli State University
khatuna.beridze@bsu.edu.ge

Translation of the Acquis communautaire was followed in almost every acceding EU state with debates about the quality, multiple cases of requests for corrigenda, and controversies regarding the terminology "consistence." Insignificant in one respect or undetected in another, translational mistakes incurred costs as well as additional efforts on behalf of the EU and national institutions (Beridze, Diasamidze, 2025). The paper examines the consistency of terminology of the EU–Georgian Association Agreement, Title IV, articles 1–276. The analyses were conducted using a custom-built bilingual corpus platform and software. The goal of this research was to continue the development of the CT analyses and methodology for the translated Acquis, as well as legal texts in general. We analyzed legal collocations or "terminological collocations" that are translated with the functional equivalents, causing inconsistencies and errors in the teleological interpretation of the law, as well as term variation. Beckner (1959) characterized the teleological approach as purposive behavior. The Court of Justice of the European Union CJEU prioritizes the aim and purpose of EU law (teleological interpretation) over merely the wording of the provisions (linguistic interpretation). This approach is driven by factors such as the open-ended and policy-oriented nature of EU Treaties and the principle of EU legal multilingualism, which ensures that all language versions of EU law are equally authentic. As a result, the Court cannot rely on the wording of a single version, unlike national courts. To interpret a legal provision, the Court considers its purpose (teleological interpretation) and its context (systemic interpretation); Nonetheless, as noted with criticism, legal terminology is still translated with the "key adequacy requirements" (Ramos, 2023:377 qtd. Stefaniak 2017, Šarčević 2018) per ISO 17100 instructions, which prioritizes semantic accuracy to upkeep the conformity with the terminological univocity in the multilingual law. We discuss potential of the teleological interpretation, or goal-oriented translation of legal terms and legal texts within the domain of Cognitive Linguistics (Lakoff and Jonson 1980, 1999; Johnson, 1987; Lakoff, 1987; Lakoff, 1993; Fauconnier and Turner, 2002; Lakoff & Turner 2006) since it deals with how meaning is shaped by mental processes, goals, and contexts. Beyond linguistic domain, the interpretation of legal discourse, including conceptual and contextual meanings of a term, should take into account the social and cultural contexts "in which it takes shape, is interpreted, used and exploited to achieve socio-political ends" (Bhatia and Bhatia, 2011). We also interconnect teleological principle of interpretation legal terminology to the Vermeer's Skopos and Commission in Translational Action (1989/2000), Toury's Descriptive Translation Studies (1995/2012) and Nord's Functional Translation approach, as well as text analysis principles in translation (1997, 2005). Application of the teleological interpretation leads to the Bajčić's (2017) stance, that the most significant innovation in contemporary legal lexicography lies in the

adoption of cognitive principles for concept organization. In legal contexts, this means that every concept must be understood within its broader legal framework, considering factors such as jurisdictional differences, historical development, and practical application. Arguably, main purpose of a legal dictionary, is to enable users to learn about legal concepts in order to understand the law. This approach mirrors teleological interpretation's focus on understanding legal provisions through their underlying purposes rather than just their literal text and prioritize functional understanding of a term and its context, rather than finding functional equivalents. Another value of the research is that it ascertains and promotes the significance of the corpus linguistic methodology for the development of the CTS of legal texts.

Keywords: *legal terminology, corpus linguistics, translation.*

კორპუსიდან მიზნამდე: იურიდიული თარგმანის ტელეოლოგიური საკითხი

ხათუნა ბერიძე

ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტი

khatuna.beridze@bsu.edu.ge

ასოცირების ხელშეკრულების თარგმანს ევროკავშირში გაწევრებულ თითქმის ყველა სახელმწიფოში მოჰყვა კრიტიკა თარგმანის ხარისხის შესახებ. ტერმინოლოგიის უსისტემო გამოყენება ჯერ სამეცნიერო მსჯელობისა და დაპირისპირებების საგნად იქცა, მოგვიანებით კი ევროკომისიის თარგმანის გენერალურ დირექტორატში შესწორებების მოთხოვნით არაერთი მიმართვის მიზეზად. ერთი შეხედვით უმნიშვნელო, ან შეუმჩნეველი თარგმნითი შეცდომები ეროვნული ინსტიტუტების მხრიდან ხარჯებს, ხოლო ევროკომისიის თარგმნის გენერალური დირექტორატისგან დამატებით ძალისხმევას მოითხოვდა (ბერიძე, დიასამიძე, 2025). წინამდებარე კორპუსლინგვისტიკური კვლევა წარმოადგენს იურიდიული კოლოკაციების, ან ტერმინოლოგიური კოლოკაციების ფუნქციური ეკვივალენტებით თარგმანთა ანალიზს. გაანალიზდა საქართველო-ევროკავშირის ასოცირების შეთანხმების ნაწილი IV, მუხლები 1-276. ტერმინოლოგიის კვლევის ერთ-ერთი მიზანი იყო 2018 წელს დაწყებული ტერმინოლოგიის კორპუსული ანალიზისა და იურიდიული ტექსტების ლინგვისტური ინტერპრეტაციის მეთოდოლოგიის განვითარების გაგრძელება. კვლევა ჩატარდა ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტში სპეციალურად შექმნილი ორენოვანი კორპუსის პლატფორმის: corpus.bsu.edu.ge-სა და პროგრამული უზრუნველყოფის გამოყენებით. კვლევით დასტურდება, რომ რიგ შემთხვევებში ფუნქციური ეკვივალენტებით თარგმანი იწვევს ტერმინთა ვარიაციას, ტერმინების არათანმიმდევრულ გამოყენებასა და შეცდომებს კანონის ტელეოლოგიურ ინტერპრეტაციაში. ბექნერმა (1959) ტელეოლოგიური მიდგომა დაახასიათა როგორც მიზნობრივი ქცევა. ევროკავშირის სასამართლო CJEU პრიორიტეტს ანიჭებს ევროკავშირის სამართლის მიზანსა და დანიშნულებას, ანუ ტელეოლოგიურ ინტერპრეტაციას და არა დებულებათა ლინგვისტურ ინტერპრეტაციასა და ფორმულირებას. თარგმნისას სამართლის მიზნისა და დანიშნულების წვდომის მოთხოვნა განპირობებულია ისეთი ფაქტორებით, როგორიცაა ღია ფორმულირების ტიპის ევროკავშირის ხელშეკრულებები, იურიდიული ტექსტების პოლიტიკაზე ორიენტირებულობა და ევროკავშირის სამართლებრივი მრავალენოვნების პრინციპი, რომელიც უზრუნველყოფს ევროკავშირის სამართლის ყველა ენობრივი ვერსიის თანაბრად ავთენტიკურობას. შედეგად, სასამართლოს არ შეუძლია დაეყრდნოს ერთი ვერსიის ფორმულირებას, ეროვნული სასამართლოებისგან განსხვავებით. სამართლებრივი დებულების ინტერპრეტაციისთვის, სასამართლო ითვალისწინებს მის მიზანს (ტელეოლოგიური ინტერპრეტაცია) და კონტექსტს (სისტემური ინტერპრეტაცია); მიუხედავად ამისა, თეორეტიკოსები კრიტიკუ-

ლად შენიშნავენ, რომ იურიდიული ტერმინოლოგია კვლავ ითარგმნება მხოლოდ „ადეკვატურობის მოთხოვნით“ (Ramos, 2023:377 ციტ. Stefaniak 2017, Šarčević 2018) ISO 17100 ინსტრუქციების შესაბამისად, რომელიც პრიორიტეტს ანიჭებს სემანტიკურ სიზუსტეს, რათა შენარჩუნდეს მრავალენოვან კანონში ტერმინოლოგიური ერთგვაროვნების შესაბამისობა. ჩვენ განვიხილავთ იურიდიული ტერმინებისა და ტექსტების მიზანზე ორიენტირებული თარგმნის, ანუ ტელეოლოგიური ინტერპრეტაციის შესაძლებლობას კოგნიტიური ლინგვისტიკის სფეროში (Lakoff and Jonson 1980, 1999; Johnson, 1987; Lakoff, 1987; Lakoff, 1993; Fauconnier and Turner, 2002; Lakoff & Turner 2006). იურიდიული დისკურსის ინტერპრეტაციისას, ტერმინის კონცეპტუალური და კონტექსტუალური მნიშვნელობები სცდება ლინგვისტურ სფეროს; მთარგმნელმა უნდა გაითვალისწინოს სოციალური და კულტურული კონტექსტები, რომლებშიც იურიდიული ტექსტი ყალიბდება, რომლის მიხედვითაც ექვემდებარება ინტერპრეტაციასა და რომლის ფარგლებშიც მისით ხელმძღვანელობენ სოციალურ-პოლიტიკური მიზნების მისაღწევად (Bhatia and Bhatia, 2011). იურიდიული ტერმინოლოგიის ინტერპრეტაციის ტელეოლოგიურ პრინციპს ასევე ვუკავშირებთ ვერმეერის (1989/2000) „სკოპოსის“ და თაურის „აღწერილობითი თარგმანის თეორიას“ (1995/2012), ნორდის „ფუნქციური თარგმანის“ მიდგომასა და „ტექსტის ანალიზის პრინციპებს თარგმანში“ (1997, 2005). ტელეოლოგიური ინტერპრეტაციის პრინციპის გამოყენებისას ვიზიარებთ ბაიჩიჩის (2017) პოზიციას, რომ თანამედროვე იურიდიულ ლექსიკოგრაფიაში ყველაზე მნიშვნელოვანი სიახლეა კოგნიტიური პრინციპების შემოღება ცნებების ორგანიზებისთვის. იურიდიულ კონტექსტში ეს ნიშნავს, რომ ყველა ცნება უნდა იქნეს გაგებული მის უფრო ფართო სამართლებრივ ჩარჩოში, ისეთი ფაქტორების გათვალისწინებით, როგორიცაა იურისდიქციული განსხვავებები, ისტორიული განვითარება და პრაქტიკული გამოყენება. კონცეპტუალური ორგანიზების იურიდიული ლექსიკონის მთავარი მიზანია, მომხმარებლებს მისცეს საშუალება, გაეცნონ იურიდიულ ცნებებს კანონის გასაგებად. კოგნიტიურ პრინციპებზე დაფუძნებულ იურიდიულ ლექსიკოგრაფიას ფოკუსი გადააქვს ტელეოლოგიურ ინტერპრეტაციაზე, ანუ სამართლებრივი დებულებების გაგებაზე მათი ძირითადი მიზნების მხედველობაში მიღებით. ასეთ ლექსიკონში უპირატესობა ენიჭება ტერმინის კონტექსტურ, ფუნქციურ და არა სიტყვასიტყვით გაგებასა და ფუნქციური ეკვივალენტების მოძიებას. კვლევის ღირებულებათა შორის უნდა აღინიშნოს, რომ იგი წარმოაჩენს კორპუსლინგვისტური მეთოდოლოგიის მნიშვნელობას იურიდიული ტექსტების ტერმინოლოგიური კვლევების განვითარებისთვის.

საკვანძო სიტყვები*: იურიდიული ტერმინოლოგია, კორპუსული ლინგვისტიკა, თარგმანი*

References:

Bajčić, M. and K. Dobrić Basaneže. (2021). Considering foreignization and domestication in EU legal translation: a corpus-based study. *Perspectives* 29, no. 5 : 706-721.

Bajčić, M. and A. Martinović. (2018). A mutual learning exercise in terminology and multilingual law. In *Language and Law: The Role of Language and Translation in EU Competition Law*, pp. 207-223. Cham: Springer International Publishing.

Bajčić, M. (2017). *New insights into the semantics of legal concepts and the legal dictionary*.

Beckner, M. (1959). *The Biological Way of Thought*. New York, Columbia University Press.

Beridze, Kh. and Kh. Diasamidze. (2025). Acquis terminology in English–Georgian translation. *Babel* (2025).

Bhatia, Vijay K., and A. Bhatia. (2011). Legal discourse across cultures and socio-pragmatic contexts. *World Englishes* 30, no. 4 : 481-495.

Fauconnier, G. and M. Turner. (2003). Conceptual blending, form and meaning. *Recherches en communication* 19 : 57-86.

Johnson, M. and G. Lakoff. (2002). Why cognitive linguistics requires embodied realism. *Cognitive linguistics* 13, no. 3 : 245-264.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Lakoff, G. and M. Turner. (2006). *The Way We Think: Conceptual Metaphor and the Mind*. New York: Basic Books.

Lakoff, G. (1993). The contemporary theory of metaphor.

Lakoff, G. and M. Johnson. (1980). The metaphorical structure of the human conceptual system. *Cognitive science* 4, no. 2 : 195-208.

Lakoff, G. (1987). Image metaphors. *Metaphor and Symbol* 2(3). 219-222.

Margalitadze, T., Meladze, G. and Z. Pourtskhvanidze. (2022). English–Georgian Parallel Corpus and Its Application in Georgian Lexicography. *Lexikos* 32 (2): 158–176.

Nord, Ch. (1997). Defining translation functions. The translation brief as a guideline for the trainee translation. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies* 33: 039-054.

Nord, Ch. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (No. 94). Rodopi.

Stefaniak, K. (2017). Terminology work in the European Commission: Ensuring high-quality translation in a multilingual environment. *Quality aspects in institutional translation* 8, no. 109 : 1.

Toury, G. (2012). *Descriptive Translation Studies — and Beyond*. Revised ed. Amsterdam: John Benjamins.

Vermeer, H.J. (2021). Skopos and Commission in Translational Action. In *The Translation Studies Reader*, edited by Lawrence Venuti, 221-232; 227–238. 4th ed. London: Routledge.

Forming a Theory for Words: Georgian Grammar and the Underpinnings of Part-of-speech Annotation

Daraselia, Sophiko
University of Leeds, UK
s.daraselia@leeds.ac.uk

Hardie, Andrew
Lancaster University
a.hardie@lancaster.ac.uk

Part-of-speech (POS) tagging is the assignment to every word token in a running text of a tag indicating its grammatical category. This process therefore presupposes some exhaustive schema of such categories. But the theoretical model of grammar that underpins such a schema may not adhere fully to all precepts of models devised for non-POS-tagging purposes. In this paper, we illustrate this issue for Georgian, whose complexities raise several problems for POS tagging.

We argue that POS tagging, as a categorisation task, requires a grammar model tailored to annotation needs rather than those of traditional Georgian grammar. Thus, our tagset-focused grammar model (TFGM) diverges from classical Georgian grammatical constructs to better serve the needs of corpus annotation. We discuss three points of grammar which illustrate the need for such divergence, whether by drawing additional distinctions, discarding a traditional distinction, or narrowing the scope of categorisation. All these, hinge on matters of annotation by form versus function.

The TFGM requires additional distinctions: Georgian nominals distinguish a nominative-absolutive case in *-i* from a zero or unmarked case. In traditional Georgian grammar, this distinction is restricted to consonant-final bases. An unsuffixed vowel-final nominal is considered to be the normal nominative-absolutive, e.g. *dghe* 'day', with no separate zero-case form. But this sits poorly with POS tagging's need for consistency of application of categories, as well as its need to handle unrestricted text. Forms which *do* suffix *-i* to vowels (e.g. *dghei*) are known to occur in non-standard dialects. The TFGM model must therefore create a class of zero-case for vowel-final nominal bases like *brma* 'blind', contrary to the traditional analysis.

The TFGM must disregard a distinction: like many languages, Georgian uses its demonstratives as third person pronouns. Traditional accounts (e.g. Shanidze, 1980, 41-43) describe first and second person pronouns as "genuine" pronouns and the third person pronouns as a function of certain demonstratives, thus preserving the category of third person pronouns for Georgian. This approach is tailored to support human understanding and it is problematic for the POS tagset; there is no clear annotatable distinction between a demonstrative used as a pronoun and a demonstrative that is not. Thus, in the TFGM only the first and second person pronouns exist as categories within the personal pronouns; demonstratives are not counted as contributing a category of third person pronouns.

The TFGM narrows the scope of categorisation: Typical accounts of the Georgian verb are detailed and exhaustive regarding the morphological categories that it expresses (including features such as the preverb, the version marker, and numerous tense/aspect/modality and voice formants, as well as extensive agreement), as well as factors of semantics and argument valence. But at the POS level, such exhaustiveness defeats the objective of a usable abstraction across the variety of verbal forms. Thus, our TFGM limits verb categorisation to screeve and agreement patterns. We introduce a 19-pattern agreement schema based on v-/m-agreement which moves away from the traditional characterisation of alignment-driven subject/object agreement target roles, and demonstrate how this abstraction supports practical tagging without sacrificing grammatical insight.

სიტყვებისათვის თეორიის შექმნა: ქართული გრამატიკა და მორფოსინტაქსური ანოტირება

სოფიკო დარასელია
ლიდზის უნივერსიტეტი, გაერთიანებული სამეფო
s.daraselia@leeds.ac.uk

ენდრიუ ჰარდი
ლანკასტერის უნივერსიტეტი, გაერთიანებული სამეფო
a.hardie@lancaster.ac.uk

მორფოსინტაქსური ანოტირება ნიშნავს ამა თუ იმ ტექსტში თითოეული სიტყვისათვის გრამატიკული კატეგორიის იდენტიფიცირებასა და მინიჭებას. ეს პროცესი მოითხოვს შესაბამისი გრამატიკული კატეგორიების დაწვრილებითი სქემის არსებობას. აღსანიშნავია ის, რომ გრამატიკის თეორიული მოდელი, რომელსაც ეს სქემა ეფუძნება, სრულად ვერ/არ დაექვემდებარება იმ გრამატიკის მოდელის პრინციპებს, რომელშიც მორფოსინტაქსური ანოტირება არაა გათვალისწინებული.

წინამდებარე ნაშრომში ჩვენ ვამტკიცებთ, რომ მორფოსინტაქსური ანოტირება საჭიროებს ანოტირებაზე მორგებულ გრამატიკის მოდელს და არა ტრადიციულ გრამატიკულ მოდელს. ამგვარად, ჩვენ შევქმენით *ტაგსეტზე დაფუძნებული გრამატიკის მოდელი* (TFGM), რომელიც სცილდება ქართული გრამატიკის ტრადიციულ პრინციპებსა და ნორმებს, რათა უკეთ მოერგოს კორპუსის ანოტირების საჭიროებებს. აღნიშნულ მოხსენებაში ჩვენ განვიხილავთ ისეთ საკითხებს, რომლებიც წარმოაჩენენ ტრადიციული გრამატიკის ნორმებიდან გადახვევის აუცილებლობას. კერძოდ, აქ საუბარია შემდეგ სამ საკითხზე, რომელიც ეხება დამატებითი კატეგორიის შემოღებას, ტრადიციული კატეგორიის ამოღებასა და არსებული კატეგორიის შემცირებას. ყოველივე ეს კი თავის მხრივ დამოკიდებულია იმაზე, *ფორმის* მიხედვით ხდება ანოტირება თუ *შინაარსის* მიხედვით.

<u>ტაგსეტზე დაფუძნებულ გრამატიკის მოდელში (TFGM) დამატებითი კატეგორიის შემოღება</u>: როგორც ვიცით, სახელები ქართულ ენაში განარჩევენ სახელობით *-ი* სუფიქსიან ბრუნვას ნულოვანი/არამარკირებული ბრუნვისაგან. ტრადიციულ ქართულ გრამატიკაში ეს ეხება თანხმოვანფუძიან სახელებს. ხმოვნიან ფუძესთან სახელობითი ფორმის ალომორფად მიჩნეულია არამარკირებული ფორმა, მაგ.: *დღე*. ამგვარი ანალიზი ვერ გამოდგება მორფოსიტაქსური ანოტირებასთვის, ვინაიდან აქ დარღვეულია კატეგორიის თანმიმდევრულობა. ამასთან, გასათვალისწინებელია სხვადასხვა ტიპის/ჟანრის ტექსტი, სადაც მოსალოდნელია ისეთი ფორმები, რომლებიც *-ი* სუფიქსს დაირთავენ (მაგ.: *დღეი*), მაგალითად დიალექტურ მეტყველებაში. შესაბამისად, ჩვენს გრამატიკულ მოდელში შემოვიღეთ დამატებითი კატეგორია ნულოვანი ბრუნვისა, რომ ცალკე მოინიშნოს ამგვარი არამარკირებული ხმოვანფუძიანი სახელები.

ტაგსეტზე დაფუძნებულ გრამატიკის მოდელში (TFGM) კატეგორიის ამოღება: ქართულ ენაში, ისევე როგორც სხვა არაერთ ენაში, ჩვენებითი ნაცვალსახელი გამოიყენება როგორც მესამე პირის ნაცვალსახელი. ქართულ გრამატიკაში (შანიძე, 1980), მხოლოდ პირველი და მეორე პირის ნაცვალსახელი მიიჩნევა „ნამდვილ“ პირის ნაცვალსახელად. ხოლო რაც შეეხება მესამე პირის ნაცვალსახელებს, აქ ჩვენებითი ნაცვალსახელები ასრულებენ მესამე პირის ნაცვალსახელის ფუნქციას. ასეთი ანალიზი ადამიანისთვის/ლინგვისტისთვის არის განკუთვნილი და არა მორფოსინტაქსური ანალიზისათვის. მორფოსინტაქსური ანოტირებისათვის აუცილებელია კატეგორიებს შორის არსებობდეს მკაფიო და პრაქტიკული განსხვავება. შესაბამისად, ჩვენი გრამატიკის მოდელი განასხვავებს მხოლოდ პირველ და მეორე პირის ნაცვალსახელებს და ჩვენებითი ნაცვალსახელები მასში არ არის მოცემული.

ტაგსეტზე დაფუძნებულ გრამატიკის მოდელში (TFGM) კატეგორიების შეზღუდვა: ზოგადად, ქართულ გრამატიკებში ზმნის აღწერისას დაწვრილებითაა მოცემული ზმნის არამარტო მორფოლოგიური კატეგორიები, არამედ სემანტიკური მახასიათებლებიც. მორფოსინტაქსური ანოტირების დონეზე კი ზმნის ასეთი მრავალმხრივი ანალიზი ვერ გამოდგება. ტაგსეტზე დაფუძნებული გრამატიკის მოდელი (TFGM) ზმნის მხოლოდ ამ ორ კატეგორიას ეყრდნობა: მწკრივი და პირთა შეთანხმება. შესაბამისად, ჩვენ შემოვიღეთ 19-პირთა კომბინაციაზე დაფუძნებული *ვ-/მ*-იანი შეთანხმებათა სისტემა, რომელიც თავისი არსით ტრადიციული სუბიექტური/ობიექტური პირის სისტემისაგან განსხვავდება. წინამდებარე მოხსენებაში ჩვენ განვიხილავთ ამ სისტემის პრაქტიკულ უპირატესობებს მოფროსინტაქსურ ანოტირებაში და ვაჩვენებთ, რომ შესაძლებელია ქართული ზმნის მდიდარი მორფოლოგიური კატეგორიების შენარჩუნება მიუხედავად ამგვარი შეზღუდვისა.

Emmanuele da Iglesias' Dictionary (a Concise Grammar Overview) and the Georgian Grammatical and Lexicographical Heritage

Doborjginidze Nino; Damenia Maia; Cheishvili Tamari; Chkuaseli Ana

Ilia State University, Georgia

nino_doborjginidze@iliauni.edu.ge, maia.damenia.1@iliauni.edu.ge
tamar.cheishvili@iliauni.edu.ge, ana.chkuaseli.1@iliauni.edu.ge

With the support of the Shota Rustaveli National Scientific Foundation of Georgia, a research project (HE-21-977) is being carried out at Ilia State University, focusing on the study of previously unknown lexicographical sources of European religious missions. These sources describe the multifaceted process of Catholic inculturation in Georgia from various perspectives, including creation and dissemination of Georgian-language literature of various kinds, as well as bilingual and explanatory dictionaries by foreign missionaries.

As part of the project, the research team examined the Georgian-Italian and Italian-Georgian dictionary, along with the Georgian grammar, compiled by the 19th-century missionary and superior of the Catholic Church in Gori (Viareggio 1845, 268), Emmanuele da Iglesias (*Grammatica & Dizionario Italiano-Georgiano e Georgiano-Italiano*). The incomplete Georgian-Italian part of the dictionary has been published (Doborjginidze et al. 2023: 019^{r}-025^{v} fol.), and the major Italian-Georgian part (026^{r}-128^{r} fol.) along with the grammatical overview (001^{r}-018^{r} fol.) is being prepared for publication.

Emmanuele da Iglesias' dictionary is particularly noteworthy as it represents the first presentation of a Georgian lexicographical work from the early modern period to the Western world. Research indicates that, in compiling the Georgian-Italian section, the scholar follows Sulkhan-Saba Orbeliani's dictionary, retaining the lemmas unchanged and, in many cases, translating Saba's explanations of the dictionary entries.

According to the European lexicographical tradition, Emmanuele da Iglesias appended a grammatical section to the bilingual dictionary. It is known that missionaries in Georgia independently created such grammatical overviews at the early stage (17th-18th centuries) of missionary activity, and in the later period they relied on Anton I's (1720-1788) Georgian grammar (Uturgaidze 1999, 28-29). The grammatical overviews of previously known and studied missionary dictionaries (primarily Paolini and Irbachi's dictionary, 1629) belong to the first period of the missionary tradition. Among the overviews known today, da Iglesias' grammar is the only text created in the second period where a clear trace of the Georgian linguistic tradition is evident (Chkuaseli 2021). Thus, it represents the projection of traditional Georgian grammar (τέχνη γραμματική) into the Western (Italian) sphere during the early modern period.

The paper will show the research results, clearly demonstrating how distinctly the trace of the Georgian lexicographical and grammatical tradition is reflected in the work of the Italian missionary.

ემანუელე და იგლესიასის ლექსიკონი (გრამატიკის მოკლე მიმოხილვით) და ქართული გრამატიკული და ლექსიკოგრაფიული მემკვიდრეობა

**ნინო დობორჯგინიძე, მაია დამენია,
თამარ ჭეიშვილი, ანა ჭკუასელი**

ილიას სახელმწიფო უნივერსიტეტი, საქართველო

nino_doborjginidze@iliauni.edu.ge, maia.damenia.1@iliauni.edu.ge,
tamar.cheishvili@iliauni.edu.ge, ana.chkuaseli.1@iliauni.edu.ge

ილიას სახელმწიფო უნივერსიტეტში შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის მხარდაჭერით ხორციელდება კვლევითი პროექტი (HE-21-977), რომლის ერთ-ერთი ფოკუსი არის ევროპულ რელიგიურ მისიათა დღემდე უცნობი ლექსიკოგრაფიული წყაროების შესწავლა. აღნიშნული წყაროები სხვადასხვა პერსპექტივიდან აღწერს საქართველოში კათოლიკური ინკულტურაციის მრავალპლანიან პროცესს, მათ შორის უცხოელი მისიონერების მიერ სხვადასხვა ხასიათის ქართულენოვანი ლიტერატურის, ორენოვანი და განმარტებითი ლექსიკონების შექმნასა და გავრცელებას.

პროექტის ფარგლებში კვლევითმა ჯგუფმა შეისწავლა XIX საუკუნის I ნახევარში საქართველოში მოღვაწე მისიონერის, გორის კათოლიკური ეკლესიის წინამძღოლის (Viareggio 1845, 268), ემანუელე და იგლესიასის ქართულ-იტალიური და იტალიურ-ქართული ლექსიკონი და ქართული ენის გრამატიკა (“Grammatica & Dizionario Italiano-Georgiano e Georgiano-Italiano”). ლექსიკონის არასრული ქართულ-იტალიური ნაწილი გამოცემულია (დობორჯგინიძე და სხვ. 2023: 019^{r}-025^{v} fol.), გამოსაცემად მზადდება ლექსიკონის ძირითადი, იტალიურ-ქართული ნაწილი (026^{r}-128^{r} fol.) და გრამატიკული მიმოხილვა (001^{r}-018^{r} fol.).

ემანუელე და იგლესიასის ლექსიკონი განსაკუთრებით საინტერესოა იმ თვალსაზრისით, რომ იგი წარმოადგენს ადრეული მოდერნულობის პერიოდის პირველი ქართული ლექსიკოგრაფიული ნაშრომის პრეზენტაციას დასავლურ სამყაროში. კვლევა აჩვენებს, რომ ემანუელეს ლექსიკონის ქართულ-იტალიური ნაწილის შედგენისას უცვლელად გადმოაქვს სულხან-საბა ორბელიანის ლექსიკონის სიტყვანი და, ხშირ შემთხვევაში, თარგმნის სალექსიკონო ლემათა საბასეულ განმარტებებსაც.

ევროპული ლექსიკოგრაფიული ტრადიციის თანახმად, ემანუელე და ეგლესიასი ორენოვან ლექსიკონს დაურთავს გრამატიკულ ნაწილსაც. ცნობილია, რომ მისიონერი ავტორები საქართველოში ამგვარ გრამატიკულ მიმოხილვებს მისიონერული საქმიანობის საწყის ეტაპზე (XVII-XVII ს.) დამოუკიდებლად ქმნიდნენ, მომდევნო პერიოდში კი ეყრდნობოდნენ ანტონ I-ის (1720-1788) ქართულ გრამატიკას (უთურგაიძე 1999, 28-29). აქამდე ცნობილი და შესწავლილი მისიონერული ლექსიკონების გრამატიკული მიმოხილვები (პირველ რიგში, პაოლინისა და ირბახის ლექსიკონი (1629)) მისიონერული ტრადიციის პირველ პერიოდს

განეკუთვნება. დღეისათვის ცნობილი მიმოხილვებიდან, და იგლესიასის გრამატიკა მეორე პერიოდში შექმნილი ერთადერთი ტექსტია, რომელშიც მკაფიოდ იკვეთება ქართული ლინგვისტური ტრადიციის მნიშვნელოვანი კვალი (ჭკუასელი 2021). ამდენად, იგი წარმოადგენს ადრეული მოდერნულობის პერიოდის ტრადიციული ქართული გრამატიკის (გრამატიკული τέχνη-ს) პროექციას დასავლურ (იტალიურ) სივრცეში.

მოხსენებაში წარმოდგენილი იქნება კვლევის შედეგები, რომლებიც ნათლად აჩვენებს, თუ რამდენად მკაფიოდ აისახა იტალიელი მისიონერის ნაშრომში ქართული ლექსიკოგრაფიული და გრამატიკული ტრადიციის კვალი.

References:

Doborjginidze, N., Damenia, M. Cheishvili, T. Chkuaseli, A. (2023). *Lexicographical Heritage of Georgian Catholic Missions. Unknown Author. 'Georgian Language Grammar and Dictionary.' Emmanuele and Iglesias. Georgian-Italian Dictionary.* Tbilisi: Ilia State University (in Georgian).

Uturgaidze, T. (1999). *History of the Study of the Georgian Language.* Tbilisi: Georgian Language (*in Georgian*).

Georgian Grammar, compiled by Anton I. 1885. Tbilisi: Ekvtime Kheladze Printing House (in Georgian).

Chkuaseli, A. (2021). *From the History of Georgian Philology: Catholic Inculturation and Unknown Dictionaries of the 18th-19th Centuries.* PhD thesis. Tbilisi: Ilia State University (in Georgian).

Paolini, Stefano, and Niceforo Irbachi. 1629. *Dittionario giorgiano e italiano.* Rome: Congregatio de Propaganda Fide.

Viareggio, Damien J. (Capuchin, Late Prefect-Apostolic of Georgia). 1845. "Letter from the Reverend Father Damian de Viareggio, Capuchin and Prefect-Apostolic of Georgia, to the President of the Central Council of Lyons. Trebizond, February 1-13, 1845." In *Annals of Propaganda Fide.* Vol. VIII, No. XLIII: 258-273. Dublin: W. Powell.

Advertisements in Early 20th-Century US Immigrant Dictionaries

Farina, Donna M. T. Cr., New Jersey City University, USA
Vrbinc, Marjeta, University of Ljubljana, Slovenia
Vrbinc, Alenka, University of Ljubljana, Slovenia

dfarina@njcu.edu
marjeta.vrbinc@ff.uni-lj.si
alenka.vrbinc@ef.uni-lj.si

Immigrant dictionaries are bilingual reference works designed for immigrants in their new cultural context. They are not the same as learner's dictionaries (Xue and Tarp 2018), not only because they are not monolingual but because of the tailored information and specialized lemmata they contain.

The term *immigrant dictionary* was first introduced by Pálfi and Tarp (2009). The existing literature allows for an enumeration of some of the types of appropriate culture-specific information that can be found in immigrant dictionaries. Grønvik (2016) mentions the necessity of including concepts that are non-existent in the immigrant's home culture; an example of this from Fjeld (2012) is information about parliamentary rules in a late 19th-century Norwegian immigrant dictionary published in Minneapolis, Minnesota. Immigrant dictionaries can contain information to fulfill communicative needs both in official and non-official social settings; an example of an official setting is applying for a green card (Vacalopoulou et al. 2011). An example of less official communication comes from an early 20th-century English–Slovak dictionary (Kadak 1905b) that includes a lengthy question-and-answer section designed to help Slovak immigrants explain who they are, where they are from, etc. to people in the US and Canada. Kavanagh (2000: 101) addresses words and meanings that convey "attitudes, manners, and social norms"; another example from Kadak (1905b) is a section showing immigrants how to communicate in a workplace: "Good morning, boss, I am looking for work; do you need help?" (ibid.: 66); and "I like to dig bituminomus [recte bituminous] coal best" (ibid.: 68).

Vacalopoulou et al. (2011) suggest the importance of including in immigrant dictionaries proper nouns covering geographical entities and names of official offices or organizations along with their acronyms. While such information can appear outside of the dictionary proper, it can also be incorporated into a dictionary's lemma list. If we accept that *cultural information* is conveyed in multiple ways—before, within, and after the body of the dictionary itself—then clearly the immigrant dictionary's lemma list should have a sharp focus on conveying culture. Vrbinc et al. (in press, 2025) discuss criteria for lemma selection in immigrant dictionaries. Likewise, Xue and Tarp (2018) comment on how lemma selection criteria differ in immigrant dictionaries as compared to other learner's resources.

Information from advertisements should be added to the above list for inclusion in immigrant dictionaries. Just like other forms of helpful culture-specific information, advertisements address immigrants' reality at the same time that

they influence and shape their new cultural identity. To our knowledge, no studies have analyzed advertisements in immigrant dictionaries, and this is a necessary step if the field of lexicography is to further solidify its theoretical understanding of the genre *immigrant dictionary.*

The aim of this paper is to investigate the role of advertisements in US-published immigrant dictionaries—mostly English–Slovenian dictionaries—of the early 1900s (K. As will be discussed, the advertisements in some US immigrant dictionaries of the early 20th century are just as "individually targeted" and "personalized" (Dziemianko 2020) as modern online dictionary ads are today.

While we present here Slovenian and English dictionaries (Košutnik 1912, Kubelka 1904 & 1912), we also examine for comparison a bilingual dictionary with another language of the Austro-Hungarian Empire, Slovak (Kadak 1905a). From 1880 to 1918, more than 3.9 million Austro-Hungarians arrived in the US (Granatir Alexander 2009), so our target dictionaries were circulating broadly in the early 20th century.

Generally speaking, when advertisements appear in immigrant dictionaries of this time period, they follow similar patterns across dictionaries and across languages, because the immigrants for whom they were made had comparable lifestyles across ethnic groups. In addition, the different immigrant groups experienced similar trajectories of acculturation. The advertisements, as we will see, reflect this acculturation process.

Keywords: *immigrant dictionary, advertisements, cultural identity, immigrant identity, culture-specific information.*

სარეკლამო განცხადებები მე-20 საუკუნის დასაწყისის აშშ-ის იმიგრანტთა ლექსიკონებში

დონა ფარინა, ნიუ-ჯერზის უნივერსიტეტი, აშშ
მარიეტა ვერბინცი, ალენკა ვერბინცი,
ლიუბლიანას უნივერსიტეტი, სლოვენია

dfarina@njcu.edu
marjeta.vrbinc@ff.uni-lj.si
alenka.vrbinc@ef.uni-lj.si

იმიგრანტთა ლექსიკონები არის ორენოვანი ცნობარები, რომლებიც შექმნილია იმიგრანტებისთვის მათი ახალი კულტურული გარემოს გათვალისწინებით. ისინი სასწავლო ლექსიკონებისგან განსხვავდება (Xue and Tarp 2018) არა მხოლოდ იმით, რომ ორენოვანია, არამედ იმითაც, რომ შეიცავს სპეციალურად შერჩეულ ინფორმაციასა და სპეციფიკურ ლემებს.

ტერმინი *იმიგრანტთა ლექსიკონი* პირველად პალფიმ და ტარპმა (2009) შემოიღეს. არსებული ლიტერატურა საშუალებას გვაძლევს გამოვყოთ იმიგრანტთა ლექსიკონებში წარმოდგენილი კულტურულად სპეციფიკური ინფორმაციის რამდენიმე კატეგორია. გრონვიკი (2016) ამგვარი ტიპის ლექსიკონში აუცილებლად მიიჩნევს ისეთ ცნებების შეტანას, რომლებიც არ არსებობს იმიგრანტის მშობლიურ კულტურაში; ფიელდს (2012) ამის მაგალითად მოჰყავს მე-19 საუკუნის ბოლოს მინესოტას შტატში, მინეაპოლისში გამოცემულ ნორვეგიულ იმიგრანტთა ლექსიკონში პარლამენტის წესებზე არსებული ინფორმაცია.

იმიგრანტთა ლექსიკონებში შეიძლება შევხვდეთ ინფორმაციას, რომელიც საჭიროა, როგორც ოფიციალურ, ისე არაოფიციალურ საკომუნიკაციო გარემოში. მაგალითად, ოფიციალურ კონტექსტში მოყვანილია „გრინ ქარდზე“ განაცხადის შევსების მაგალითი (Vacalopoulou et al. 2011). არაოფიციალური კომუნიკაციის მაგალითი კი გვხვდება მე-20 საუკუნის დასაწყისის ინგლისურ-სლოვაკურ ლექსიკონში (Kadak 1905b), სადაც მოცემულია ვრცელი კითხვა–პასუხის ველი, რათა სლოვაკმა იმიგრანტებმა აშშ-სა და კანადაში მცხოვრებ ხალხს აუხსნან, ვინ არიან, საიდან მოდიან და ა.შ. კავენაჰი (2000: 101) განიხილავს სიტყვებსა და მნიშვნელობებს, რომლებიც გამოხატავს „დამოკიდებულებებს, მანერებსა და სოციალურ ნორმებს“; კადაკთან (1905b) ერთ-ერთ პარაგრაფში წარმოდგენილია მაგალითი, თუ როგორ უნდა იურთიერთონ იმიგრანტებმა სამუშაო გარემოში. „დილა მშვიდობისა, ბატონო, სამსახურს ვეძებ; გჭირდებათ დახმარება?“ (ასევე: 66); და „ყველაზე მეტად ბიტუმიანი ნახშირის გათხრა მიყვარს“ (ასევე: 68).

ვაკალოპულუ და სხვები (2011) აღნიშნავენ, რომ იმიგრანტთა ლექსიკონებში მნიშვნელოვანია ტოპონიმების, ოფიციალური დაწესებულებების სახელებისა და მათი აბრევიატურების შეტანა. ასეთი ინფორმაცია შეიძლება გაჩნდეს როგორც ლექსიკონის ტექსტის მიღმა, ასევე თავად

ლემების სიაშიც. თუ დავუშვებთ, რომ კულტურასთან დაკავშირებული ინფორმაცია მრავალი გზით გადმოიცემა— ლექსიკონის შესავალში, ძირითად ნაწილში, ან ბოლოში — მაშინ ცხადია, რომ იმიგრანტულ ლექსიკონში მკაფიო აქცენტი უნდა გაკეთდეს ამგვარ ინფორმაციაზე. ვერბინცი და სხვები (2025) განიხილავენ იმიგრანტთა ლექსიკონებისთვის ლემათა შერჩევის კრიტერიუმებს. ასევე, შუე და ტარპი (2018) აღნიშნავენ, რომ იმიგრანტთა ლექსიკონებში ლემათა შერჩევის კრიტერიუმები განსხვავდება სხვა სასწავლო რესურსების კრიტერიუმებისაგან.

ზემოთ ჩამოთვლილ ინფორმაციას უნდა დაემატოს სარეკლამო განცხადებებიც. როგორც სხვა სახის კულტურასთან დაკავშირებული სასარგებლო ინფორმაცია, სარეკლამო განცხადებებიც ასახავს იმიგრანტების რეალობას და ამავდროულად გავლენას ახდენს მათი ახალი კულტურული იდენტობის ჩამოყალიბებაზე. რამდენადაც ჩვენთვის ცნობილია, ჯერჯერობით არ ჩატარებულა კვლევა იმიგრანტთა ლექსიკონებში წარმოდგენილი სარეკლამო განცხადებების შესახებ, თუმცა ეს აუცილებელი ნაბიჯია, თუ გვინდა, რომ ლექსიკოგრაფიის დარგმა გაამყაროს თავისი თეორიული ცოდნა *იმიგრანტთა ლექსიკონის* ჟანრის შესახებ.

წინამდებარე სტატიის მიზანია გამოიკვლიოს სარეკლამო განცხადებების როლი იმიგრანტთა ლექსიკონებში, რომლებიც აშშ-ში გამოქვეყნდა — ძირითადად ინგლისურ-სლოვენურ ლექსიკონებში — მე-20 საუკუნის დასაწყისში. როგორც ქვემოთაა განხილული, ზოგიერთ ამერიკულ იმიგრანტთა ლექსიკონში წარმოდგენილი სარეკლამო განცხადებები ისეთივე „ინდივიდუალურად მიზანმიმართული“ და „პერსონალიზებული“ (Dziemianko 2020) იყო, როგორც დღეს თანამედროვე ონლაინლექსიკონების რეკლამებია.

ჩვენ წინამდებარე სტატიაში ვიკვლევთ სლოვენიურ–ინგლისურ ლექსიკონებს (Košutnik 1912; Kubelka 1904 & 1912), მაგრამ ასევე ვადარებთ ავსტრია-უნგრეთის იმპერიის სხვა ენის, სლოვაკურის, ორენოვან ლექსიკონს (Kadak 1905a). 1880-1918 წლებში 3.9 მილიონზე მეტი ავსტრია-უნგრელი ჩავიდა აშშ-ში (Granatir Alexander 2009), შესაბამისად ჩვენი სამიზნე ლექსიკონები ფართოდ იყო გავრცელებული მე-20 საუკუნის დასაწყისში.

ზოგადად, იმ პერიოდის იმიგრანტთა ლექსიკონებში სარეკლამო განცხადებები ერთგვაროვანი იყო, როგორც სხვადასხვა ლექსიკონში, ისე სხვადასხვა ენაზე. მიზეზი ისაა, რომ იმიგრანტები, რომლებზეც ეს ლექსიკონები იყო გათვლილი, ერთმანეთის მსგავს სოციალურ-კულტურულ ცხოვრებას ეწეოდნენ სხვადასხვა ეთნიკურ ჯგუფში. გარდა ამისა, სხვადასხვა იმიგრანტული ჯგუფი ერთსა და იმავე აკულტურაციულ გზას გადიოდა. სარეკლამო განცხადებები ასახავს ამ აკულტურაციულ პროცესს.

საკვანძო სიტყვები: *იმიგრანტთა ლექსიკონი, სარეკლამო განცხადებები, კულტურული იდენტობა, იმიგრანტის იდენტობა, კულტურასთან დაკავშირებული ინფორმაცია.*

References

A. Dictionaries

Kadak, Paul K. (1905a). *Praktičný slovensko-anglický tlumač. The practical Slovak American interpreter*. New York: Slovák v Amerike. [bidirectional dictionary, first Slovak-English, followed by English-Slovak, front matter + dictionary pp. 1–263, then 4 advertisements]

Kadak, Paul K. (1905b). *Praktičný slovensko-anglický tlumač. The practical Slovak American interpreter*. New York: G. Klein & Son. [Monodirectional English–Slovak dictionary, front matter + dictionary, pp. 1–204, back cover has a list of books, no advertisements]

Košutnik, Silvester. (1912). *Ročni slovensko-angleški in angleško slovenski slovar: Zlasti namenjen izseljencem v Ameriko*. Ljubljana: Anton Turk. (Košutnik 1912)

Kubelka, Victor J. (1904). *Slovenian-English Pocket Dictionary to Facilitate the Study of Both Languages. Slovensko-angleški žepni rečnik v olajšavo naučenja obeh jezikov.* New York.
https://babel.hathitrust.org/cgi/pt?id=loc.ark:/13960/t4q-j8p026&view=1up&seq. (Kubelka 1904)

Kubelka, Viktor J. (1912. *Slovensko-angleška Slovnica, Tolmač, Spisovnik in Navodilo za Naturalizacijo. Angleško-Slovenski in Slovensko-Angleški Slovar. Slovenian-English Grammar Interpreter, Letterwriter and Information on Naturalization. English-Slovenian and Slovenian-English Dictionary*. Prva izdaja/ First edition. New York: Viktor J. Kubelka. (Kubelka 1912b)

B. Other Literature

Dziemianko, A. (2020). Smart advertising and online Dictionary usefulness. *International Journal of Lexicography* 33 (4). 377–403. doi: 10.1093/ijl/ecaa017

Fjeld, R. V. (2012). Et forsømt kapittel i leksikografiens historie: Leachman's Norsk-Engelsk Lomme-Ordbog/Norwegian-English pocket dictionary. *Norsk Lingvistisk Tidsskrift, 30*(2). 151–164.
https://ojs.novus.no/index.php/NLT/article/view/205

Granatir Alexander, J. (2009). *Daily Life in Immigrant America, 1870–1920: How the Second Great Wave of Immigrants Made Their Way in America*. Chicago: Ivan R. Dee.

Grønvik, O. (2016). The lexicography of Norwegian. In P. Hanks, & G.-M. de Schryver (Eds.), *International handbook of modern lexis and lexicography* (pp. 1–34). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-45369-4_44-1

Kavanagh, K. (2000). Words in a cultural context?. *Lexikos*, *10*. 99–118. https://doi.org/10.5788/10-0-889

Pálfi, L.-L. And S. Tarp (2009). Lernerlexikographie in Skandinavien – Entwicklung, Kritik und Vorschläge. *Lexicographica, 25*. 135–154. https://doi-org.nukweb.nuk.uni-lj.si/10.1515/9783484605787.135

Vacalopoulou, A., Gioulu, V., Giagkou, M. and Efthimiou, E. (2011). Online dictionaries for immigrants in Greece: Overcoming the communication barriers. In I. Kosem, & K. Kosem (Eds.), *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, 10–12 November 2011, Bled, Slovenia* (pp. 274–279). Ljubljana: Trojina, Institute for Applied Slovene Studies.

Xue, M. and S. Tarp (2018). Towards Chinese learner's dictionaries for foreigners living in China: Some problems related to lemma selection. *Lexikos, 28*(1), 384–404.
https://doi.org/10.5788/28-1-1470

Vrbinc, A., Vrbinc, M. and D. Farina (in press, 2025). "For the Benefit of my Countrymen": Cultural Context and Lemma Choice in Early 20th-Century Dictionaries for Immigrants, *International Journal of Lexicography.*

Code-Mixing Patterns in a Corpus-Based Megrelian–English Dictionary

Gersamia, Rusudani
Lobzhanidze, Irina
Ilia State University, Georgia
rgersamia@iliauni.edu.ge, irina_lobzhanidze@iliauni.edu.ge

The documentation of code-mixing in endangered languages is an important area of linguistic and lexicographic research. This paper presents the methods, challenges, and preliminary findings from the development of a corpus-based Megrelian dictionary, with a particular focus on the documentation and treatment of code-mixing phenomena. Megrelian, a Kartvelian language spoken primarily in Western Georgia, is characterized by extensive contact with Georgian, and, to some extent, Russian, resulting in rich patterns of bilingual and multilingual interaction. Although Megrelian has no official status and is largely unwritten, speakers frequently mix elements from these contact languages into their everyday speech, producing a wide range of mixed utterances and hybrid lexical items.

This paper examines the challenges of integrating code-mixing phenomena into a lexicographic resource such as the Megrelian dictionary. Analysis of the data reveals that code-mixing is not only a matter of loanwords or calques but represents a dynamic process with its own lexical and grammatical regularities, often expressed at the morphemic level. Accordingly, the corpus-based Megrelian–English dictionary and the Megrelian morpheme lexicon include, as an essential part of their comprehensive description, the identification of cross-linguistic contact phenomena that characterize contemporary Megrelian speech.

To achieve this, the paper describes the documentation of code-mixing, drawing on a digitized corpus of spontaneous speech, recorded conversations, narratives, and interviews collected in Megrelian-speaking communities in the Samegrelo region over the past three years (2021–2024). The dictionary includes mixed forms at multiple levels, including insertional code-mixing (e.g., Georgian words embedded in Megrelian text), alternative code-switching (switches between Megrelian and another language at clause or sentence boundaries), and lexicalization (shared morphosyntactic patterns across languages).

The paper outlines the methodological framework used for identifying, annotating, and representing code-mixed material in the dictionary. This includes decisions regarding inclusion criteria (which mixed forms to document), tagging and metadata conventions in the corpus, and strategies for providing definitions, etymologies, and usage notes in the dictionary entries. Furthermore, it explores the main characteristics of loanwords, including their frequency of use, and other factors that influence which code-mixed forms are included and how they are represented.

And finally, the paper outlines the principles for the lexicographic treatment of code-mixing cases in Megrelian, which include:

- Criteria for the selection and inclusion of code-mixed forms in the dictionary;
- Strategies for representing cross-linguistic collocations and hybrid compounds;
- Corpus-based approaches to documenting the frequency and distribution of code-mixed items.

The paper consists of several key sections. The introduction provides an overview of the Megrelian language, its historical development, current sociolinguistic context, and the importance of documenting its lexical and contact phenomena. The theoretical section includes an examination of code-mixing and code-switching within the relevant linguistic and lexicographic context, as well as a detailed description of the corpus, including the sources of data, methods of collection, and types of texts. The methodology section explains the framework used for identifying, annotating, and analyzing code-mixed material, including decisions about inclusion criteria, tagging conventions, and dictionary entry design. The analysis and findings section presents observations and discussion on cases of code-mixing. The conclusion outlines key summary statements that highlight the importance of technology-based lexicography for endangered languages.

კოდის შერევის ნიმუშები კორპუსზე დაფუძნებულ მეგრულ-ინგლისურ ლექსიკონში

რუსუდან გერსამია
ირინა ლობჟანიძე
ილიას სახელმწიფო უნივერსიტეტი, საქართველო
rgersamia@iliauni.edu.ge, irina_lobzhanidze@iliauni.edu.ge

ბოლო პერიოდში საფრთხეში მყოფი ენების შესწავლისას სულ უფრო და უფრო მნიშვნელოვანი ხდება კოდის შერევის ლინგვისტური და ლექსიკოგრაფიული ასპექტების კვლევა. მოხსენებაში განიხილება კორპუსზე დაფუძნებული მეგრული ენის ლექსიკონის შემუშავების ძირითადი მეთოდოლოგია, პრობლემატიკა და წინასწარი შედეგები კოდის შერევის შემთხვევების დამუშავებისა და დოკუმენტირების თვალსაზრისით.

მეგრული – ქართველური ენაა, რომელზეც, ძირითადად, დასავლეთ საქართველოში მეტყველებენ. მეგრულის კონტაქტი ქართულთან და, გარკვეულწილად, რუსულთან ორენოვანი და მრავალენოვანი ურთიერთქმედების განსაზღვრულ მაგალითს წარმოადგენს. მეგრული უმწერლობო ენაა და, მეგრულად მოსაუბრეები ყოველდღიურ მეტყველებაში ხშირად ურევენ საკონტაქტო ენებს, ქმნიან შერეულ შესიტყვებებს და ჰიბრიდულ ლექსიკურ ერთეულთა ფართო სპექტრს.

მოხსენებაში განიხილება კოდის შერევის შემთხვევები და მათი ინტეგრაცია ლექსიკოგრაფიულ რესურსში, როგორიცაა მეგრულის ლექსიკონი. მონაცემთა ანალიზი აჩვენებს, რომ კოდის შერევა მხოლოდ ნასესხები სიტყვებისა თუ კალკების გამოყენებას არ ნიშნავს, არამედ წარმოადგენს დინამიკურ პროცესს, რომელსაც აქვს საკუთარი ლექსიკური და გრამატიკული წესები, რაც გულისხმობს ნასესხები ენობრივი ელემენტების (ძირი, ფუძე, ზოგჯერ-აფიქსი) მიმღები ენის მორფოსტრუქტურაში ჩასმას. შესაბამისად, კორპუსზე დაფუძნებული მეგრულ-ინგლისური ლექსიკონისა და მეგრული მორფემების ლექსიკონის სრულყოფილი აღწერის ნაწილს წარმოადგენს იმ ენათაშორისი კონტაქტების გამოვლენა, რომლებიც მეგრულ ენაზე თანამედროვეთა მეტყველებას ახასიათებს.

სწორედ ამ მიზნით მოხსენებაში განიხილება კოდის შერევის შემთხვევები, რომლებიც ეფუძნება ზეპირი მეტყველების ციფრულ კორპუსს, დიალოგების ჩანაწერებს, მონათხრობებს და ინტერვიუირების გზით ბოლო სამი წლის მანძილზე (2021–2024) სამეგრელოში მოპოვებულ მასალას. ლექსიკონი მოიცავს სხვადასხვა დონეზე შერეულ ფორმებს, მათ შორის: კოდის შერევის (მეგრულ ტექსტში ჩართულ ქართულ სიტყვებს), კოდის გადართვის (მეგრულსა და სხვა ენას შორის გადართვის შემთხვევებს წინადადებებისა ან მათი საზღვრების დონეზე) და ლექსიკალიზაციის შემთხვევებს (მორფოსინტაქსური ნიშნების გაზიარებას ენებს შორის).

მოხსენება ასახავს მეთოდოლოგიურ ჩარჩოს, რომელიც გამოიყენება ლექსიკონში კოდის შერევის შემთხვევების იდენტიფიცირებისთვის, ანოტირებისა და წარმოჩენისთვის. იგი მოიცავს გადაწყვეტილებებს შესაბამისი ერთეულების შერჩევის კრიტერიუმებთან (მაგ. რომელი შერეული ფორმა უნდა იყოს დოკუმენტირებული), კორპუსში ტეგირებისა და მეტამონაცემების ასახვასთან დაკავშირებით; ასევე სტრატეგიებს ლექსიკური ერთეულების განმარტებების, ეტიმოლოგიებისა და მაგალითების გამოყენებისთვის. მოხსენებაში განიხილება ნასესხები სიტყვების ძირითადი მახასიათებლები, მათ შორის გამოყენების სიხშირეც და სხვა ფაქტორები, რომლებიც ზეგავლენას ახდენენ იმაზე, თუ რომელი შერეული ფორმა შევა ლექსიკონში და როგორ იქნება წარმოდგენილი.

და ბოლოს, მოხსენებაში განისაზღვრება მეგრულში კოდის შერევის შემთხვევების ლექსიკოგრაფიული დამუშავების პრინციპები, რომლებიც მოიცავს:

- შერეული ფორმების შერჩევისა და ლექსიკონში ჩასმის კრიტერიუმებს;
- ენებს შორის კოლოკაციებისა და ჰიბრიდული ფორმების წარმოჩენის სტრატეგიებს;
- კორპუსზე დაფუძნებულ მიდგომებს შერეული ფორმების სიხშირისა და გავრცელების დოკუმენტირებისთვის.

მოხსენება რამდენიმე ნაწილისაგან შედგება. შესავალი წარმოადგენს მეგრულის ენის ისტორიული განვითარებისა და მისი დღევანდელი სოციოლინგვისტური კონტექსტის მიმოხილვას, მისი ლექსიკური მახასიათებლებისა და კონტაქტების დოკუმენტირების მნიშვნელობის აღწერას. თეორიული ნაწილი მოიცავს კოდის შერევის და კოდის გადართვის კვლევას შესაბამის ლინგვისტურ და ლექსიკოგრაფიულ კონტექსტში და კორპუსის დეტალურ აღწერას, მონაცემთა წყაროებისა და შეგროვების მეთოდების, ტექსტების ტიპების დახასიათებას. მეთოდოლოგიის ნაწილი განმარტავს ჩარჩოს, რომელიც გამოიყენება შერეული მასალის იდენტიფიცირებისთვის, ანოტირებისა და ანალიზისთვის, მათ შორის, ფორმების შერჩევის კრიტერიუმებს, ტეგირების პრინციპებსა და ლექსიკური ერთეულების სტრუქტურას. ანალიზისა და მიგნებების ნაწილი წარმოადგენს დაკვირვებას და მსჯელობას კოდის შერევის შემთხვევებზე. დასკვნაში წარმოდგენილი შემაჯამებელი დებულებები ხაზს უსვამს ტექნოლოგიაზე დაფუძნებული ლექსიკოგრაფიის მნიშვნელობას საფრთხეში მყოფი ენებისათვის.

How New Terms Are Created and Introduced into Legislative and Other Institutional Texts

Gogia, Nana

LEPL Bureau of Translation of International Agreements of Georgia
Ministry of Foreign Affairs of Georgia

nanagogia22@gmail.com

The process of Georgia's European integration, in accordance with the Association Agreement, requires the approximation of Georgian legislation with EU legislation. High-quality translation of legal documents and unification of the relevant terminology are crucial for the success of this large-scale process.

The report discusses the steps taken by the State in response to this challenge. In particular, official translation agencies have been established in Georgia, such as the "Bureau for Translation of International Agreements of Georgia" within the Ministry of Foreign Affairs of Georgia and the "Translation Center under the Georgian Legislative Herald", operating within the Ministry of Justice of Georgia. In addition, in 2018, the State Language Department began its work, which developed a "Unified Programme (Strategy) for the State Language". This document highlights the national importance of standardizing terminology and the need for state involvement.

Despite these initiatives, the problem of coordinating the translation process and proper terminological work remains relevant in government agencies involved in legislative activities related to European integration. These agencies, within the scope of their competence, introduce new specialized terms. Since most of these institutions do not have a service or a staff unit responsible for terminology management, the standardization and introduction of new terms are often handled by lawyers and field experts employed by the respective agencies, without coordination with language experts. This often leads to inconsistencies. A clear example of this problem is the terminological discrepancies in maritime vocabulary, particularly the incorrect translation of English maritime terms *Marine, Maritime* and *Nautical,* which have been used for many years as benchmarks in the regulatory documents of higher education.

The recommendations of the State Language Department presented in the report on the issues raised by the Batumi State Maritime Academy aim to introduce correct Georgian equivalents of the above-mentioned terms.

To address these issues, it is essential to strengthen translation centres and establish mechanisms to ensure coordination between government agencies. This will create a unified terminology base and ensure high-quality translations, which is an important prerequisite for Georgia's full integration into the European family.

Keywords: *maritime terminology, unified use of terminology*

როგორ იქმნება და ინერგება ახალი ტერმინები საკანონმდებლო და სხვა სახის სამართლებრივ ტექსტებში

ნანა გოგია

სსიპ საქართველოს საერთაშორისო ხელშეკრულებების თარგმნის ბიურო, საქართველოს საგარეო საქმეთა სამინისტრო

nanagogia22@gmail.com

საქართველოს ევროინტეგრაციის პროცესი, ასოციირების შესახებ შეთანხმების თანახმად, გულისხმობს ქართული კანონმდებლობის ევროკავშირის სამართალთან დაახლოებას. ამ მასშტაბური პროცესის წარმატებისთვის უმნიშვნელოვანესია იურიდიული დოკუმენტების მაღალხარისხიანი თარგმნა და შესაბამისი ტერმინოლოგიის უნიფიცირება.

მოხსენებაში საუბარია იმ ნაბიჯებზე, რომელიც გადადგა სახელმწიფომ ამ გამოწვევის საპასუხოდ. კერძოდ, საქართველოში შეიქმნა ოფიციალური მთარგმნელობითი უწყებები, როგორიცაა „საქართველოს საერთაშორისო ხელშეკრულებების თარგმნის ბიურო“ საქართველოს საგარეო საქმეთა სამინისტროში და „მთარგმნელობითი ცენტრი საქართველოს საკანონმდებლო მაცნეში,“ რომელიც ფუნქციონირებს იუსტიციის სამინისტროში. ამას გარდა, 2018 წლიდან საქმიანობას შეუდგა სახელმწიფო ენის დეპარტამენტი, რომელმაც შეიმუშავა „სახელმწიფო ენის ერთიანი პროგრამა (სტრატეგია).“ ამ დოკუმენტში ხაზგასმულია ტერმინოლოგიის სტანდარტიზაციის ეროვნული მნიშვნელობა და მასზე სახელმწიფო ზრუნვის აუცილებლობა.

მიუხედავად ამ ინიციატივებისა, კვლავაც აქტუალურია მთარგმნელობითი პროცესის კოორდინაციისა და გამართული ტერმინოლოგიური მუშაობის პრობლემა საჯარო უწყებებში, რომლებიც ჩართული არიან ევროინტეგრაციასთან დაკავშირებულ კანონშემოქმედებით საქმიანობაში და თავიანთი უფლებამოსილების ფარგლებში ნერგავენ ახალ დარგობრივ ტერმინებს. ვინაიდან ამ უწყებებში არ არსებობს ტერმინოლოგიაზე პასუხისმგებელი სამსახური ან საშტატო ერთეული, ცალკეულ დარგში გაჩენილი ტერმინოლოგიური სიახლეების სტანდარტიზაციასა და დანერგვაზე ოპერატიულად მუშაობენ შესაბამის უწყებაში დასაქმებული იურისტები და დარგის ექსპერტები ენის სპეციალისტებთან შეუთანხმებლად, რაც ხშირად იწვევს შეუსაბამობებს. ამ პრობლემის ნათელი მაგალითია უმაღლესი განათლების მარეგულირებელ დოკუმენტებში ინგლისური საზღვაო ტერმინების *Marine*, *Maritime* და *Nautical* არასწორი და არაერთგვაროვანი შესატყვისები.

მოხსენებაში მოცემული რეკომენდაციები, რომლებიც გასცა სახელმწიფო ენის დეპარტამენტმა ბათუმის სახელმწიფო საზღვაო აკადემიის მიერ დასმულ საკითხებზე, მიზნად ისახავს ზემოხსენებული ტერმინების სწორი ქართული შესატყვისების დანერგვას.

ამ პრობლემების აღმოსაფხვრელად, სასიცოცხლოდ მნიშვნელოვანია მთარგმნელობითი ცენტრების გაძლიერება და მექანიზმების შექმნა სახელმწიფო უწყებებს შორის კოორდინინაციის უზრუნველსაყოფად. ამ გზით შესაძლებელია ერთიანი ტერმინოლოგიური ბაზის ჩამოყალიბება და ხარისხიანი თარგმანის უზრუნველყოფა, რაც საქართველოს ევროპულ ოჯახში სრულფასოვანი ინტეგრაციის მნიშვნელოვანი წინაპირობაა.

საკვანძო სიტყვები: *საზღვაო ტერმინოლოგია, ტერმინოლოგიის ერთგვაროვანი გამოყენება*

The Pedagogical Significance of Lexicography in Linguistics Education: A Multidimensional Analysis of Benefits and Applications

Gvarishvili, Zeinabi; Gumbaridze, Zhuzhuna; Chkonia, Liana

Batumi Shota Rustaveli State University

zeinab.gvarishvili@bsu.edu.ge, zhuzhuna.gumbaridze@bsu.edu.ge, lia.chkonia@bsu.edu.ge

The paper investigates the transformative role of lexicography within undergraduate linguistics education, presenting evidence that systematic instruction in its principles and methodologies yields substantial cognitive, methodological, and professional advantages for linguistics students. Despite lexicography's established significance within the global linguistic discipline, its pedagogical value has been systematically marginalized in Georgian Higher Educational Institutions' linguistics curricula. This marginalization creates a problematic disconnect between theoretical linguistic knowledge and its practical application in lexical documentation. For this research, a mixed-methods sequential explanatory design is used that includes both quantitative outcomes (student performance, engagement rates) and qualitative aspects (pedagogical value, cognitive development). This approach allows us to collect quantitative data first, then use qualitative methods to explain and contextualize the findings. Through a methodical analysis of educational outcomes from a three-year (2022-2025) pedagogical incorporation at Batumi Shota Rustaveli State University (BSU), this study documents significant improvements in student performance and engagement following the integration of lexicography as a compulsory component within the English Philology Bachelor educational program.

The research identifies five interconnected domains where lexicographical training demonstrably enhances undergraduate linguistics education. Firstly, lexicography functions as an integrative framework, synthesizing diverse linguistic subfields—phonology, morphology, syntax, semantics, and pragmatics—enabling students to conceptualize language as an interconnected system rather than isolated components. This integration facilitates a deeper comprehension of how theoretical principles manifest in authentic linguistic documentation. Secondly, students engaged in lexicographical projects develop sophisticated methodological competencies in corpus analysis, field methods, data organization, and definitional precision—transferable research skills applicable across various linguistic specializations. Thirdly, lexicography effectively bridges the theoretical-practical divide, addressing the growing need for linguists capable of producing functional resources for both academic and non-academic contexts, which is particularly relevant for Georgian language documentation. Fourthly, the evidence demonstrates that lexicographical education substantially enhances students' metalinguistic awareness and critical thinking capacities. Confronting the complexities of word meaning, usage contexts, and register variations develops sophisticated analytical capabilities. The necessity of explicit definitional choices encourages

a deeper engagement with fundamental questions about language and meaning that underpin linguistic inquiry. Finally, BSU philology students with lexicographical training tend to be equipped with expanded professional trajectories beyond traditional academic pathways, which contributes to their higher competitiveness in fields such as computational linguistics, language technology development, specialized publishing, and cultural heritage documentation.

The systematic implementation of lexicography as a compulsory study course in the English Philology curriculum at BSU since 2022 has generated measurable improvements in educational outcomes. Comparative analysis reveals significantly higher rates of student engagement (increased participation in research initiatives), enhanced research productivity (as evidenced by undergraduate conference presentations and publications), and improved post-graduation academic placement. These empirically-supported findings strongly suggest that integrating lexicography into undergraduate linguistics curricula across Georgian Higher Educational Institutions would substantially enhance educational outcomes and professional relevance in the contemporary linguistic landscape, potentially serving as a model for similar curricular innovations throughout the region.

ლექსიკოგრაფიის სწავლების მნიშვნელობა ლინგვისტურ განათლებაში: გამოცდილებისა და შედეგების კომპლექსური ანალიზი

ზეინაბ გვარიშვილი
ჟუჟუნა გუმბარიძე
ლიანა ჭყონია

ბათუმის შოთა რუსთაველის სახელობის სახელმწიფო უნივერსიტეტი

zeinab.gvarishvili@bsu.edu.ge, zhuzhuna.gumbaridze@bsu.edu.ge, lia.chkonia@bsu.edu.ge

წინამდებარე ნაშრომი განიხილავს ლექსიკოგრაფიის ფუნდამენტურ როლს საბაკალავრო ლინგვისტურ განათლებაში და ემყარება ჰიპოთეზას, რომ მისი პრინციპებისა და მეთოდოლოგიების სისტემატური სწავლება სტუდენტებისთვის მნიშვნელოვან კოგნიტიურ, მეთოდოლოგიურ და პროფესიულ უპირატესობებს ქმნის. მიუხედავად იმისა, რომ ლექსიკოგრაფია ენათმეცნიერების ერთ-ერთ საკვანძო დისციპლინადაა აღიარებული, მისი პედაგოგიური ღირებულება ხშირად მარგინალიზებულია საქართველოს უმაღლეს საგანმანათლებლო დაწესებულებებში, კერძოდ, ფილოლოგიური მიმართულების სასწავლო საგანმანათლებლო პროგრამებში. ეს დამოკიდებულება გარკვეულ დისონანსს იწვევს თეორიულ ცოდნასა და მისი პრაქტიკული გამოყენების შესაძლებლობებს შორის.

კვლევა ეფუძნება რაოდენობრივ (აკადემიური მოსწრება, ჩართულობის მაჩვენებლები) და თვისებრივ (პედაგოგიური დაკვირვება, კოგნიტიური განვითარება) მეთოდებს. ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტში 2022–2025 წლებში ლექსიკოგრაფიის სწავლების სამწლიანი მეთოდური ანალიზის საფუძველზე, კვლევა ადასტურებს, რომ ლექსიკოგრაფიის სავალდებულო კურსად ინტეგრაცია ინგლისური ფილოლოგიის საბაკალავრო პროგრამაში მნიშვნელოვნად აძლიერებს სტუდენტთა აკადემიურ მიღწევებსა და აქტიურობას.

მოხსენებაში განხილულია ლექსიკოგრაფიული კომპეტენციის გავლენა ხუთ ურთიერთდაკავშირებულ სფეროზე: პირველი – ინტეგრაციული ჩარჩო: ლექსიკოგრაფია აერთიანებს ფონოლოგიას, მორფოლოგიას, სინტაქსს, სემანტიკასა და პრაგმატიკას, რაც ხელს უწყობს ენის სისტემურად გააზრებას; მეორე — მეთოდოლოგიური უნარები: სტუდენტები იძენენ კორპუსული კვლევისთვის აუცილებელ ისეთ პრაქტიკულ და თეორიულ უნარებს, როგორებიცაა მონაცემთა ორგანიზება და განმარტების სიზუსტე; მესამე, თეორიისა და პრაქტიკის სინთეზი: ლექსიკოგრაფია უზრუნველყოფს რესურსების შექმნის უნარს როგორც აკადემიური, ასევე პროფესიული კონტექსტებისთვის — რაც განსაკუთრებით მნიშვნელოვანია ქართული ენის დოკუმენტირებისთვის; მეოთხე, კრიტიკული და მეტალინგვისტური ცნობიერება: სიტყვების მნიშვნელობის, კონტექსტისა და რეგისტრის ვარიაციების ანალიზი ამაღლებს სტუდენტთა კრიტიკულ და ანალიტიკურ აზროვნებას და ბოლოს, მე-

ხუთე — პროფესიული შესაძლებლობები: ლექსიკოგრაფიული ცოდნით აღჭურვილი კურსდამთავრებულები კონკურენტუნარიანნი არიან ისეთ სფეროებში, როგორიცაა კომპიუტერული ლინგვისტიკა, ენობრივი ტექნოლოგიები, სპეციალიზებული გამომცემლობა და კულტურული მემკვიდრეობის დოკუმენტირება.

კვლევის შედეგები მიუთითებს საგანმანათლებლო მიღწევების გაუმჯობესების დინამიკის ტენდენციაზე. კერძოდ, გაზრდილია სტუდენტთა ჩართულობა კვლევით აქტივობებში, მონაწილეობა სამეცნიერო კონფერენციებში, აქტიურობა სამეცნიერო სტუდენტურ პროექტებში და სხვა. ასევე, იკვეთება დასაქმების მაჩვენებლების პოტენციური ზრდის შესაძლებლობებიც. შესაბამისად, კვლევა ხაზს უსვამს ლექსიკოგრაფიის ინტეგრაციის დადებით გავლენას არა მხოლოდ ლინგვისტური განათლების ხარისხზე, არამედ სტუდენტთა პროფესიული განვითარებისა და კარიერული პერსპექტივების განმტკიცებაზე.

Danish Orthography 2.0
(or: How we learned to love the bot)

Henrichsen, Peter Juel
Danish Language Council, Denmark
pjh@dsn.dk

The title of this paper, *Danish orthography 2.0*, refers to a lexicographic framework developed by the Danish Language Council (DSN, dsn.dk) to make the rules and definitions of the official orthography more accessible and useful for NLP (natural language processing). This overview paper presents the basic components and motivations.

Introduction

The official rules for Danish spelling and punctuation are set by DSN and published in Retskrivningsordbogen (RO, ro.dsn.dk). In November 2024, the new edition of RO was – for the first time – released simultaneously as a dictionary (app, website and book) and as a machine-readable database, called COR1 (short for *the Central Danish Word Register, level I*). Unlike the dictionary, COR1 is a full-form repository. Each of its 534,771 inflected wordforms has an id in this format:

COR . <*LIX*> . <*BIX*> . <*VIX*>

, where **<*LIX*>** is the lemma index, **<*BIX*>** the inflexion index, while **<*VIX*>** (variant index) acts as a counter for alternative spellings, e.g. "ressource"N COR.95344.110.01 and "resurse"N COR.95344.110.02 (VIX is omitted in this paper). BIX values correspond to RO's traditional PoS labels, e.g. 127 for nouns in the neuter-plural-definite-genitive form (such as "kyssenes", *the kisses*'). COR1 and related materials are available as open source at ordregister.dk.

A lexicographic ecosystem

Any well-structured dictionary can be included in the COR framework (as a level II or III resource), provided that it has a valid name and that all lexical entries have COR ids incorporating that name. As an example, Kemisk Ordbog, the authoritative dictionary of Danish chemical nomenclature (Damhus 2008) is currently being transformed into COR.KEMI. This procedure is not trivial, as chemical nomenclature and official orthography are not always in sync. The element Cl (*chlorine*) is thus spelled "chlor" by chemists, where RO has "klor". Also, the nomenclature requires neuter gender while most Danes (and in consequence RO) accept both genders, "kloren"COMMON.SG.DEF as well as "kloret"NEUTER.SG.DEF. Both standards can coexist within the framework, with COR.KEMI selectively inheriting information from COR1 (cf. fig.1).

FIGURE 1. Lexical entries from COR1 and COR.KEMI

Id	Form	PoS	Category
COR.86598.110	klor	NOUN.common.sg.indef	110
COR.86598.111	kloren	NOUN.common.sg.def	111
COR.86598.120	klor	NOUN.neuter.sg.indef	120
COR.86598.121	kloret	NOUN.neuter.sg.def	121
COR.KEMI.086598.120	chlor	NOUN.neuter.sg.indef	120
COR.KEMI.086598.121	chloret	NOUN.neuter.sg.def	121

The current COR library includes dictionaries COR.ROHIST (historical orthography; Widman 2023a, 2023b), COR.MORF (subverbal morphs; Henrichsen 2024a, Nguyen 2024), COR.TALE (phonetic forms; *currently in beta*), COR.SEM (word semantics; Pedersen 2023), and more. See Dideriksen (2022) for a general introduction to COR. Debess (2023) presents OTAL, a COR-clone for Faroese.

Preparing text for NLP

CLINK (*DSN's COR-linker*) is an application for annotating text with COR ids. Ambiguous lexemes and tokens out-of-vocabulary are analyzed with rule-based and statistical methods (Bick 2023, Henrichsen 2023). For compound analysis, CLINK uses categorial grammar (Lambek calculus, Morrill 2010, Henrichsen 2024b) as illustrated here by the case of "hydrochlorering" (*hydro chlorination*).

SEGMENTS:	[hydro]	[chlor]	[er]	[ing]	
SEQUENT:	C/C	129	1.9\219	219\110	==> 110

Slash operators (/ and \) are used for functor-argument structure: functor *x*/*a* (*a**x*) looks for its argument *a* to the right (left). The sequent above thus has proofs for *C*=129 and for *C*=110, both consistent with the consequent category 110 (i.e. "hydrochlorering" as NOUN.common.sg.indef). Notice how the sequent combines information from COR1 (prefix "hydro~"), COR.KEMI ("chlor") and COR.MORF (suffixes "~er" and "~ing"). COR-linked text is highly suitable as input for NLP, such as spelling and grammar checkers, and LLM training. See fig.2 below for an overview over the entire COR framework.

Perspectives

COR and CLINK were designed to make essential Danish language resources more accessible to a broad range of NLP developers, so we decided to keep the axiomatization as formally transparent as possible. As a corollary, the minimalistic category definitions and the language independent proof system of the Lambek calculus relieves the lexicographer of the unpleasant choice between a morphological typology derived entirely from the target language and one imported from a culturally dominant foreign language.

We find that a widely shared lexical framework stimulates the development of high-quality language products. The COR user community thus includes not only dictionary and NLP providers, but also AI manufacturers and game developers – in short, bots prefer COR over RO.

დანიური ორთოგრაფია 2.0 (ანუ: როგორ შევიყვარეთ ბოტი)

პეტერ იუელ ჰენრიკსენი
დანიის ენობრივი საბჭო, დანია
pjh@dsn.dk

წინამდებარე ნაშრომის სახელწოდება, *დანიური ორთოგრაფია 2.0*, უკავშირდება დანიის ენობრივი საბჭოს (DSN, dsn.dk)[3] მიერ შემუშავებულ ლექსიკოგრაფიულ სტრუქტურას, რომლის დანიშნულებაცაა ოფიციალური ორთოგრაფიის წესებისა და დეფინიციების უფრო ხელმისაწვდომად გახდა NLP-*სათვის* (ბუნებრივი ენის დამუშავებისათვის). მიმოხილვითი ხასიათის წინამდებარე ნაშრომში წარმოდგენილია ამ სტრუქტურის ძირითადი კომპონენტები და მოტივაციები.

შესავალი

დანიური მართლწერისა და პუნქტუაციის ოფიციალურ წესებს ადგენს DSN-*ი*[4] და აქვეყნებს Retskrivningsordbogen-*ში*[5] (RO, ro.dsn.dk). 2024 წლის ნოემბერში RO („მართლწერის ლექსიკონი“) – პირველად ისტორიაში – გამოვიდა ერთდროულად როგორც ლექსიკონი (აპლიკაცია, ვებ-საიტი და წიგნი) და როგორც მანქანით წაკითხვადი მონაცემთა ბაზა, რომელსაც COR1 ეწოდება („Det Centrale Ordregister,[6] დონე I-*ის*“ შემოკლება: „სიტყვათა ცენტრალური რეესტრი დანიური ენისათვის, დონე I“). ლექსიკონისგან განსხვავებით, COR1 სრულფასოვან მონაცემთა საცავს წარმოადგენს. თითოეულს მის 534.771 ფლექსიურ სიტყვაფორმათაგან აქვს შემდეგი ფორმატის საიდენტიფიკაციო ინდექსი:

COR . <*LIX*> . <*BIX*> . <*VIX*>

, სადაც <***LIX***> ლემის ინდექსს ნიშნავს, <***BIX***> ფლექსიის ინდექსს, ხოლო <***VIX***> (ვარიანტების ინდექსი) დაწერილობის ალტერნატიული ვარიანტების მთვლელის ფუნქციას ასრულებს, *მაგ.* “ressource”N COR.95344.110.01 და “resurse”N COR.95344.110.02 (VIX-*ი* წინამდებარე ნაშრომში შეტანილი არ გვაქვს). BIX-*ის* მნიშვნელობები შეესატყვისება RO-*სეულ* ტრადიციულ მეტყველების ნაწილების (PoS) აღმნიშვნელ კვალიფიკაციებს, *მაგ.* 127 აღნიშნავს არსებითებს საშუალო სქესის-მრავლობითი რიცხვის-განსაზღვრულ[7]-ნათესაობითი ბრუნვის ფორმებში

3 Dansk Sprognævn https://dsn.dk/ [გ.მ.]

4 „დანიის ენობრივი საბჭო“ (ან „“კომისია; Dansk Sprognævn)

5 სიტყვასიტყვით: „მართლწერის ლექსიკონი“

6 https://dsn.dk/forskning/sprogteknologi-og-fagsprog/cor/

7 ე.ი. განსაზღვრული პოსტპოზიტიური არტიკლის (**-ene** *საშ. სქესის მრ. რიცხვის არსებითების, როგორც მაგ.* kys *„კოცნა“; შემთხვევაში*) მქონე.

(როგორც *მაგ.* "kyssenes", *„კოცნებისა“* [8]). COR_1 და მასთან დაკავშირებული მასალები ღიად ხელმისაწვდომია ordregister.dk-ზე.

ლექსიკოგრაფიული ეკოსისტემა

ნებისმიერი კარგად სტრუქტურირებული ლექსიკონი შეიძლება იქნეს ჩართული COR-*ის* სტრუქტურაში (როგორც II ან III დონის რესურსი), იმ პირობით, რომ მას აქვს ვალიდური სახელწოდება და რომ ყოველ მის ცალკეულ ლექსიკურ ჩანაწერს აქვს COR საიდენტიფიკაციო ინდექსები, რომლებშიც ეს სახელწოდება შედის. ასე მაგალითად, ამჟამად მიმდინარეობს Kemisk Ordbog-*ის*,[9] დანიური ქიმიური ნომენკლატურის ავტორიტეტული ლექსიკონის (დამჰუსი 2008), ტრანსფორმირება COR.KEMI-დ. ეს მარტივი პროცედურა არ გახლავთ, რამდენადაც ქიმიური ნომენკლატურა და ოფიციალური ორთოგრაფია ერთმანეთთან ყოველთვის როდია შესაბამისობაში. მაგალითისთვის, ქიმიურ ელემენტს Cl (*ქლორი*) დანიელი ქიმიკოსები წერენ როგორც "chlor", მაშინ, როდესაც RO-*ში* (მართლწერის ლექსიკონში) "klor" წერია. გარდა ამისა, ნომენკლატურა ამ სიტყვისთვის საშუალო გრამატიკულ სქესს მოითხოვს, მაშინ როდესაც დანიელთა უმეტესი ნაწილისთვის (და შესაბამისად RO-*ს თვის*) მისაღებია ორივე გრამატიკული სქესი: "kloren"COMMON.SG.DEF და აგრეთვე "kloret"NEUTER.SG.DEF.[10] სტრუქტურის ფარგლებში ორივე სტანდარტს შეუძლია თანაარსებობა, თუმცა COR.KEMI-*ს* სელექტიურად გადმოაქვს ინფორმაცია COR1-*დან* (*შდრ.* ცხრილი 1).

ცხრილი 1. *ლექსიკური ჩანაწერები* COR_1*-დან და COR.KEMI-დან*

Id-ინდექსი	*ფორმა*	*PoS (მეტყვ. ნაწილი)*	*კატეგორია*
COR.86598.110	klor	NOUN.common.sg.indef	110
COR.86598.111	kloren	NOUN.common.sg.def	111
COR.86598.120	klor	NOUN.neuter.sg.indef	120
COR.86598.121	kloret	NOUN.neuter.sg.def	121
COR.KEMI.086598.120	chlor	NOUN.neuter.sg.indef	120
COR.KEMI.086598.121	chloret	NOUN.neuter.sg.def	121

ამჟამად COR-*ის* ბიბლიოთეკა შეიცავს შემდეგ ლექსიკონებს: COR.ROHIST (ისტორიული ორთოგრაფია; ვიდმანი 2023a, 2023b), COR.MORF (სუბვერბალური მორფები; ჰენრიკსენი 2024a, ნგუენი 2024), COR.TALE (ფონეტიკური ფორმები; *ამ დროისათვის ბეტა ვერსიაში*), COR.SEM (სიტყვის სემანტიკა; პედერსენი 2023) და *სხვ.* COR-*ის* შესახებ ზოგადი საწყისი ინფორმაციისთვის *იხ.* დიდერიკსენი (2022). დებესი (2023) წარმოადგენს OTAL-*ს*, COR-*ის* კლონს ფარერული ენისათვის.

8 https://en.wiktionary.org/wiki/kyssenes#Danish

9 სიტყვასიტყვით „ქიმიური ლექსიკონი“.

10 "kloren" ე.წ. საერთო სქესია, "kloret" კი – საშუალო. გრამატიკულ სქესს პოსტპოზიტიური განსაზღვრული არტიკლები -en (საერთო სქესის) და -et (საშუალო სქესის) განაპირობებენ.

ტექსტის მომზადება NLP-*სთვის* [11]

CLINK-*ი* (*DSN's COR-linker*) წარმოადგენს აპლიკაციას, რომელიც განკუთვნილია ტექსტის ანოტირებისათვის COR-*ის* საიდენტიფიკაციო ინდექსებით. ბუნდოვანი ლექსემები და არადოკუმენტირებული (out-of-vocabulary) ტოკენები ანალიზდება წესებზე დაფუძნებული და სტატისტიკური მეთოდებით (ბიკი 2023, ჰენრიკსენი 2023). რთულ სიტყვათა ანალიზისთვის CLINK-*ი* იყენებს კატეგორიალურ გრამატიკას[12] (ლამბეკის აღრიცხვა,[13] მორილი 2010, ჰენრიკსენი 2024b), როგორც ეს აქ გვაქვს ილუსტრირებული დანიური სიტყვის "hydrochlorering" (*ჰიდროქლორირება*) მაგალითზე.

სეგმენტები:	[hydro]	[chlor]	[er]	[ing]	
სეკვენცია:	C/C	129	1.9\219	219\110	==> 110

სლეშ-ოპერატორები (/ და \) გამოიყენება ფუნქტორ-არგუმენტის სტრუქტურისათვის: ფუნქტორი *x/a* (*a\x*) ეძებს თავის არგუმენტ *a-ს* თავისგან მარჯვნივ (მარცხნივ). ამდენად, ზემომოყვანილ სეკვენციას აქვს მტკიცებულებანი იმისა, რომ *C*=129 და იმისა, რომ *C*=110, რომელთაგანაც ორივე შესაბამისობაშია ლოგიკურად გამომდინარე კატეგორიასთან 110 (ე.ი. "hydrochlorering" განისაზღვრება, როგორც NOUN. common.sg.indef[14]). ყურადღება მიაქციეთ, როგორ ახდენს სეკვენცია ინფორმაციის კომბინირებას COR_1-*დან* (პრეფიქსი "hydro~"), COR.KEMI-*დან* ("chlor") და COR.MORF-*იდან* (სუფიქსები "~er" და "~ing"). COR-*თან* დაკავშირებული ტექსტი სრულიად გამოსადეგია, როგორც სისტემაში შესაყვანი ინფორმაცია NLP-აპლიკაციებისთვის, მაგალითად მართლწერისა და გრამატიკის შესამოწმებელი პროგრამებისთვის და მანქანის LLM[15]-წვრთნისათვის. COR-*ის* მთლიანი სტრუქტურის მიმოხილვა *იხ.* ქვემოთ, ცხრილი 2.

პერსპექტივები

COR-*ი* და CLINK-*ი* დანიური ენის ძირითადი რესურსების NLP-დეველოპერების ფართო წრისათვის უფრო ხელმისაწვდომად გახდის მიზნით შეიქმნა, ამდენად ჩვენ გადავწყვიტეთ, მისი აქსიომატიზაცია ფორმალურად რაც შეიძლება გამჭვირვალე ყოფილიყო. ამის უშუალო შედეგი გახდა ის, რომ კატეგორიათა მინიმალისტური დეფინიციები და ლამბეკის აღრიცხვის ენისგან დამოუკიდებელი მტკიცებულებათა სისტემა ლექსიკოგრაფს ათავისუფლებს უსიამოვნო აუცილებლობისგან, არჩევანი გააკეთოს მთლიანად სამიზნე ენიდან აღებულ ტიპოლო-

11 natural language processing (ბუნებრივი ენის დამუშავება).

12 https://en.wikipedia.org/wiki/Categorial_grammar

13 https://en.wikipedia.org/wiki/Joachim_Lambek

14 არსებითი სახელი, საერთო სქესი, მხოლობითი რიცხვი, განუსაზღვრელი ფორმა (დანიური ენის კონტექსტში).

15 Large Language Model („დიდი ენობრივი მოდელი") მანქანური წვრთნის მეთოდი დიდი ოდენობის მონაცემების საფუძველზე.

გიასა და კულტურულად დომინანტური უცხო ენიდან იმპორტირებულ ტიპოლოგიას შორის.

ჩვენ მიგვაჩნია, რომ ფართოდ გავრცელებული ლექსიკური სტრუქტურა მაღალი ხარისხის ენობრივი პროდუქტების შექმნას უწყობს ხელს. ამდენად, COR-*ის* მომხმარებელთა სათემოში შედიან არა მარტო ლექსიკონებსა და ბუნებრივი ენის დამუშავებაზე (NLP) მომუშავენი, არამედ აგრეთვე ხელოვნური ინტელექტის აპლიკაციების მწარმოებლები და კომპიუტერული თამაშების შემქმნელები. ანუ, მოკლედ რომ ვთქვათ, ბოტებს COR-*ი* [16] უფრო უყვართ, ვიდრე RO.[17]

თეზისი ქართულად თარგმნა გიორგი მელაძემ.

References

Act on the Danish Language Council (2014). Retsinformation.dk, LOV nr.517 26/05/2014

Bick, E. (2023). Linking Danish Parser Output to a Central Word Repository. In: *Proceedings of KONVENS-19*. ACL. 93-101.

Damhus, T. (2008). Kemisk Ordbog. *Nyt Teknisk Forlag*, 3.udgave. ISBN: 9788757126723

Debess, I.N., Lamhauge, S.S., Simonsen, A., Henrichsen, P.J., Hofgaard, E., Johannesen, U., Hammer, P.M.J., Gunnvør, H.B., Thomsen, E.M.D. and B. Poulsen (2023). *Basic Language Resource Kit 1.0 for Faroese.* University of the Faroe Islands Press.

Dideriksen, Ch., Henrichsen, P.J. and W. Thomas (2022). Det Centrale Ordregister. *Nyt fra Sprognævnet* 22/3; Danish Language Council Press.

Henrichsen, P. J. (2024a). Make Each Morph Count – A New Approach to Computational Lexicography for Text Processing. In: *Proceedings of the XXI EURALEX International Congress*, 329-335. Institut za hrvatski jezik.

Henrichsen, P. J. (2024b). Tekstordet som grammatisk domæne. In: *Ny Forskning i Grammatik*, Danish Language Council Press 1(31), ISSN: 2446-1709, 53–70.

Henrichsen, P. J. (2023). Diktaroriske beføelelser. Om ord og uord i Det Centrale Ordregister. In: *Proceedings efter 19. Møde om Udforskningen af Dansk Sprog.* Kirstine Boas, Inger Schoonderbeek Hansen, Tina Thode Hougaard og Ea Lindhardt Overgaard (red.), Århus University Press 2023, p. 131-146.

Morill, G.V. (2010). *Categorial Grammar: Logical Syntax, Semantics, and Processing*. Oxford Linguistics.

Nguyen, M. and P. J. Henrichsen (2024). The locum hyphen. A formal approach to the lexicalization of multiword expressions with rich internal semantics. In: *Proceedings of the XXI EURALEX International Congress*, 113–121. Institut za hrvatski jezik.

Pedersen, B.S., S. Nimb.; N.C.H. Sørensen; S. Olsen; I. Flörke and T. Troelsgaard (2023). Reusing the Danish WordNet for a New Central Word Regis-

16 „სიტყვათა ცენტრალური რეესტრი“ (Det Centrale Ordregister).

17 „მართლწერის ლექსიკონი“ (Retskrivningsordbogen).

ter for Danish. In: *Proceedings of Global WordNet Conference 2023*, Donostia-San Sebastian. 8 p.
Widmann, Th. (2023a). The Central Word Register of the Danish Language. In: *Proceedings of eLex 2023* (The eighth eLex conference Electronic lexicography in the 21st century, Brno), 91-103.
Widmann, Th. (2023b). Det Centrale Ordregister og ROHIST. In: *Proceedings of NFL16 (Nordiska Forening i Lexikografi*), Lund Universitet.

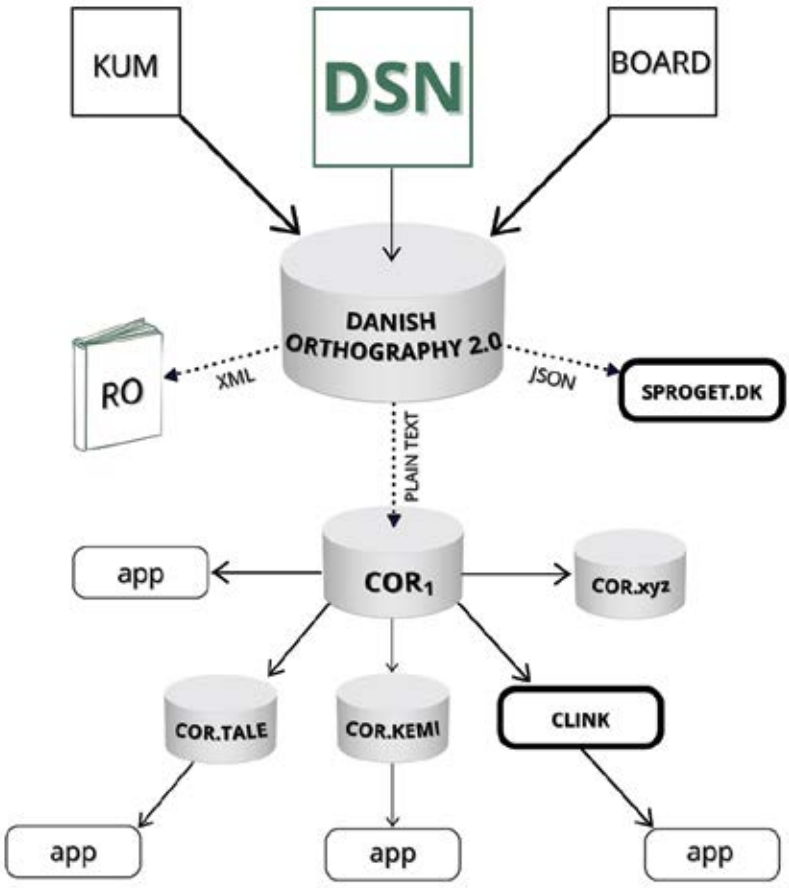


*FIGURE 2. COR (the Central Danish Word Register). High-level principles and requirements are passed from the decision-making bodies (***DSN***=the Danish Language Council;* **KUM***=Ministry of Culture;* **BOARD***=Board of Representatives, see Act 2014) to the developers of DSN's fundamental database ("Danish Orthography 2.0"). The base provides the source files for RO (the book), for DSN's online services (e.g. SPROGET.DK) and for COR1 (discussed in this paper). Dotted arrows stand for unsupervised data transfer, solid arrows are used less formally. Boxes with "app" and "COR.xyz" represent projects not attended by DSN.*

Experimental Lexicography: A Study of Dictionary Use Strategies at Schools

Khuskivadze, Elene
Ivane Javakhishvili Tbilisi State University
ekhuskivadze@gmail.com

My research falls within the subfield of theoretical lexicography which is known as the study of dictionary use and dictionary users. My interest in these issues as a public-school English teacher was determined by the following circumstances:

In 2022, the National Center for Educational Quality Enhancement designated the use of dictionaries, particularly bilingual dictionaries, as a mandatory component of foreign language instruction, within the broader Sectoral Benchmarks of language acquisition. This decision aimed to address challenges arising from the adoption of new foreign language teaching methodologies that have become widespread in Georgian schools and higher education institutions over the past 10–15 years. These methodologies have largely excluded the use of the native language in the process of teaching foreign languages and, consequently, have discouraged the use of bilingual dictionaries in favor of monolingual explanatory dictionaries.

Italian lexicographer Martina Nied, in her plenary speech delivered at the 20th Congress of the European Association for Lexicographers in Germany, observed that the prohibition of bilingual dictionaries in foreign language instruction has significantly diminished interest in dictionaries. This is particularly concerning given that dictionaries – especially bilingual ones – have traditionally been most frequently used in the context of foreign language teaching (Nied, 2022).

Many experiments conducted in the direction of researching dictionary users have revealed the tendency of losing dictionary culture in many countries and its accompanying negative processes. In response, many language communities have begun acknowledging this problem and undertaking measures to counteract it. The aforementioned decision by the National Center for Educational Quality Enhancement should be seen as one such measure. However, introducing this change in the Sectoral Benchmarks of language acquisition will not, by itself, resolve the issue – especially considering that the problem has become deeply entrenched in the Georgian educational system. The complexity of the issue is further underscored by a point made by Martina Nied in the same report. She notes that many teachers are unprepared to reintroduce the culture of dictionary use into schools and higher education institutions; in fact, they often fail to recognize the necessity of dictionary use altogether (Nied, 2022; Margalitadze & Meladze, 2023). As an English teacher, I am acutely aware of these challenges and of teachers' prevailing attitudes toward them. This awareness led me to the conclusion that a more in-depth investigation of this issue was needed at the school level.

For this purpose an extensive study was conducted between December 2024 and January 2025 and involved a survey of both teachers and students concerning the role of dictionaries in the educational process. The questionnaire was based

on the study carried out in Hungary, which examined the dictionary usage habits of foreign language learners, particularly Hungarian students learning English and German. It also addressed issues related to dictionary use by teachers in teaching these languages (P. Markus, 2023). The questions from the Hungarian study were carefully reviewed, translated into Georgian and adapted to meet the needs of Georgian schools and their users.

The empirical component of the research involved teachers of the Georgian language and various foreign languages (English, German, French, Spanish, Russian) working in public schools of Tbilisi and other regions. Additionally, interviews were conducted with teachers of Oriental languages, including Korean, Japanese, Persian, and Chinese. As for the student participants, the survey was conducted among students from public schools in Tbilisi and various regions, as well as among learners of Oriental languages.

In total, 68 teachers and 104 students participated in the study. Among the teachers, 38 were from public schools in Tbilisi, and 20 from regional public schools. Furthermore, interviews were held with 3 teachers of Korean, 2 of Japanese, 1 of Persian, and 4 of Chinese. With regard to the students, the sample included 58 students from public schools in Tbilisi, 14 from regional schools, and 32 learners of Oriental languages.

In my talk, I will present in detail the survey conducted and the obtained results and recommendations. The experiment I conducted is one of the first attempts to explore the use of dictionaries in language learning at the school level in Georgia.

ექსპერიმენტული ლექსიკოგრაფია: ლექსიკონის გამოყენების სტრატეგიების კვლევა სკოლებში

ელენე ხუსკივაძე
ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი
ekhuskivadze@gmail.com

ჩემი კვლევა განეკუთვნება თეორიული ლექსიკოგრაფიის იმ მიმართულებას, რომელსაც ეწოდება ლექსიკონის გამოყენების კვლევა, ლექსიკონის მომხმარებელთა კვლევა (Research on Dictionary Use). ჩემი, როგორც საჯარო სკოლის ინგლისური ენის მასწავლებლის დაინტერესება ამ მიმართულებით, განაპირობა შემდეგმა გარემოებებმა:

2022 წელს განათლების ხარისხის განვითარების ეროვნულმა ცენტრმა ენის დაუფლების დარგობრივ მახასიათებელში უცხოური ენების სწავლების სავალდებულო მეთოდად შეიტანა ლექსიკონების, განსაკუთრებით, ორენოვანი ლექსიკონების გამოყენება. ეს გადაწყვეტილება იყო იმ ვითარების გამოსწორების მცდელობა, რაც შეიქმნა უცხოური ენების სწავლების ახალი მეთოდების შედეგად, რომლებიც გავრცელდა საქართველოს სკოლებსა და უმაღლეს სასწავლებლებში ბოლო 10-15 წლის განმავლობაში. ამ მეთოდებმა უარყვეს მშობლიური ენის გამოყენება უცხოური ენების სწავლების პროცესში და შესაბამისად შეიზღუდა ორენოვანი ლექსიკონების გამოყენება ერთენოვანი განმარტებითი ლექსიკონების სასარგებლოდ.

იტალიელი ლექსიკოგრაფი მარტინა ნიდი, თავის პლენალურ მოხსენებაში ევროპის ლექსიკოგრაფთა ასოციაციის მე-20 კონგრესზე, აღნიშნავს, რომ ორენოვანი ლექსიკონების აკრძალვამ უცხოური ენების სწავლებისას მნიშვნელოვნად შეამცირა ინტერესი ლექსიკონებისადმი, რადგან ლექსიკონებს, კერძოდ ორენოვან ლექსიკონებს ყველაზე ხშირად სწორედ უცხოური ენების სწავლებისას იყენებდნენ (Nied, 2022).

ლექსიკონის მომხმარებელთა კვლევის მიმართულებით ჩატარებულმა არაერთმა ექსპერიმენტმა გამოავლინა ლექსიკონის გამოყენების კულტურის დაკარგვის ტენდენცია ბევრ ქვეყანაში და მისი თანმდევი უარყოფითი პროცესები. ბევრმა ენობრივმა საზოგადოებამ დაიწყო ამ პრობლემის გაცნობიერება და შესაბამისი ზომების მიღება მის გამოსასწორებლად. ზემოხსენებული განათლების ხარისხის განვითარების ეროვნული ცენტრის გადაწყვეტილება ამგვარ ზომად უნდა მივიჩნიოთ, თუმცა, დარგობრივ მახასიათებელში ამ ცვლილების შეტანა პრობლემას ავტომატურად ვერ მოაგვარებს, მით უფრო რომ მან ღრმად გაიდგა ფესვები ჩვენი განათლების სისტემაში. პრობლემას ართულებს ის გარემოებაც, რაზეც მიუთითებს მარტინა ნიდი ზემოხსენებულ სტატიაში. იგი აღნიშნავს, რომ ხშირად მასწავლებლებიც არ არიან მზად რომ ლექსიკონის გამოყენების კულტურა დააბრუნონ სკოლებსა და უმაღლეს სასწავლებლებში, უფრო მეტიც, ისინი ხშირად ვერც კი ხვდებიან ლექსიკონის გამოყენების საჭიროებას (Nied, 2022; Margalitadze and Meladze, 2023). ჩემ-

თვის, როგორც ინგლისური ენის მასწავლებლისთვის, ეს პრობლემები კარგადაა ნაცნობი, ასევე, კარგადაა ნაცნობი მასწავლებლების დამოკიდებულება ამ საკითხებისადმი, რამაც მიმიყვანა იმ დასკვნამდე, რომ საჭირო იყო ამ პრობლემის უფრო ღრმა კვლევა სკოლაში.

სწორედ ამ მიზნით 2024-2025 წლების დეკემბერ-იანვარში ჩავატარე ფართომასშტაბიანი კვლევა, რომლის მიზანი იყო სკოლის მასწავლებლებისა და მოსწავლეების გამოკითხვა ლექსიკონების სასწავლო პროცესში გამოყენებასთან დაკავშირებულ საკითხებზე. ამ მიზნით გამოვიყენე უნგრელი მეცნიერის კ. მარკუსის მიერ შემუშავებული კითხვარი, რომლის საფუძველზეც მან გამოიკვლია ინგლისური და გერმანული ენების სწავლების პროცესში ლექსიკონების გამოყენებასთან დაკავშირებული საკითხები უნგრულ უნივერსიტეტებში (P. Márkus et al., 2023.). კითხვარი ითარგმნა ქართულად და მოხდა მისი ადაპტაცია ქართული სკოლის რეალობასთან და საქართველოში ლექსიკონების გამოყენებასთან დაკავშირებულ პრობლემებთან.

ექსპერიმენტში მონაწილეობა მიიღეს ქ. თბილისისა და რეგიონების საჯარო სკოლების ქართული ენისა და სხვადასხვა უცხოური ენების (ინგლისური, გერმანული, ფრანგული, ესპანური, რუსული) მასწავლებლებმა. გამოკითხვა ინტერვიუს სახით ასევე ჩავატარე აღმოსავლური ენების (კორეული, იაპონური, სპარსული, ჩინური) მასწავლებლებთან. რაც შეეხება საჯარო სკოლების მოსწავლეების გამოკითხვას, გამოიკითხა ქ. თბილისისა და რეგიონების საჯარო სკოლების მოსწავლეები, ასევე აღმოსავლური ენების შემსწავლელები.

გამოკითხვაში მონაწილეობა მიიღო 68 მასწავლებელმა და 104 მოსწავლემ. მათ შორის 38 იყო ქ. თბილისის საჯარო სკოლების მასწავლებლები, 20 კი რეგიონების საჯარო სკოლებისა. ინტერვიუ ჩავატარე კორეული ენის 3 მასწავლებელთან, იაპონური ენის 2 მასწავლებელთან, სპარსული ენის 1 მასწავლებელთან და ჩინური ენის 4 მასწავლებელთან.

რაც შეეხება მოსწავლეებს, გამოიკითხა ქ. თბილისის საჯარო სკოლების 58 მოსწავლე, რეგიონების საჯარო სკოლების 14 მოსწავლე და აღმოსავლური ენების 32 შემსწავლელი.

ჩემს მოხსენებაში დეტალურად წარმოვადგენ ჩატარებულ გამოკითხვას და მიღებულ დასკვნებსა და რეკომენდაციებს. ეს გამოკითხვა ენების სწავლებისას ლექსიკონის გამოყენების კვლევის ერთ-ერთი პირველი მცდელობაა საქართველოს სკოლებში.

References:

Margalitadze, T. and Meladze, G. (2023). How to Solve Problems in Dictionary Use. *Lexikos*, 33(2). https://doi.org/10.5788/33-2-1840

Nied Curcio, M. (2022). Dictionaries, Foreign Language Learners and Teachers... In Klosa-Kückelhaus et al. (Eds.), *EURALEX 2022 Proceedings*, 71–84.

P. Márkus, K., Fajt, B. and Dringó-Horváth, I. (2023). Dictionary skills in teaching... International Journal of Lexicography, 36(2), 173–194. https://doi.org/10.1093/ijl/ecad004

Towards Encoding Sulkhan-Saba Orbeliani's Georgian Dictionary with the Text Encoding Initiative (TEI)

Kiknadze, Eka

European Master in Lexicography (EMLex) alumni
University of Minho, Portugal

ekakiknadze95@gmail.com

Encoding print dictionaries using a standard format is an important practice for enhancing readability, exchange and utilisation of lexicographic data between scholars and across devices and systems. The most widely implemented international standard for representing textual data in digital form is the Text Encoding Initiative (TEI). The TEI proposes guidelines for various text types, which is especially beneficial for print dictionaries as they demonstrate complex textual data.

To the best of our knowledge, the Georgian language does not have a print or online dictionary project using the TEI framework. Sulkhan-Saba Orbeliani's *Sitqvis Kona,* the first complete Georgian explanatory dictionary, represents one of the most important lexicographic works of the language. *Sitqvis Kona* is a monolingual encyclopaedic dictionary arranged in alphabetical order, reflecting the basic lexicon of the end of the 17th-century Georgian language. The dictionary has served as a model for other lexicographic projects and remains relevant for the linguistic and lexicographic analysis of the Georgian language to the present day.

The objective of the research was to propose a Text Encoding Initiative (TEI) markup for the *Georgian Dictionary by Sulkhan-Saba Orbeliani.* It was thought to serve as the TEI encoding model for a dictionary or any lexical resource of the Georgian language. It was significant and interesting to observe the limitations and advantages of encoding a less-resourced language with the TEI.

To conduct the research, the 1949 edition of *Sulkhan-Saba Orbeliani's Georgian Dictionary* was selected. It was the third edition of the dictionary, based on manuscript A prepared by Sulkhan-Saba Orbeliani's brother, Zosime. The edition is believed to closely reflect the original manuscript created by Sulkhan-Saba Orbeliani himself. Some of the lexicographer's preferences for presenting data are the use of graphic symbols and illustrations, abbreviations, and subentries for grouping semantically and grammatically related words.

As a theoretical background, the overview of the dictionary, some of the challenges with encoding print dictionaries and the TEI framework was provided. As scholars have different objectives for encoding print dictionaries, the TEI proposed two dictionary views: a textual view (preserving the original appearance of a dictionary as much as possible) and a lexical view (focusing on encoding a dictionary as a database). The combination of the views was taken as the approach for creating encoding solutions for the complex data of the *Georgian Dictionary*.

The practical part and the most crucial aspect of the research was to identify and categorise information classes in the dictionary. Based on the data categories, representative entries were selected from the dictionary and their respective encoding proposals were offered. The data categories and structural indicators

were identified and categorised based on the terms proposed by Gouws, R. H., & Prinsloo, D. J. The entries in the dictionary were grouped into simple, encyclopaedic and complex entries based on the data categories that they included. Every group provided a few examples of encoding proposals. Given that the textual view was taken as a dominant strategy, all data categories were encoded with elements. In contrast, additional or implicit, but important, missing data were introduced via attributes. Encoding entries were carried out with the x4ml, a teaching tool for XML-based lexicographical data modelling created by EMLex teacher from the Institute for the German Language (IDS), Peter Meyer.

The results showed the complexity of data categories in *Sulkhan-Saba Orbeliani's Georgian Dictionary* and the feasibility of encoding a print legacy dictionary of the Georgian language with the combination of the TEI's proposed dictionary strategies. The encoding solutions proposed were aimed at scholars interested in both textual and lexical views of the Orbeliani's Dictionary. Therefore, it can be adapted to function as an XML database, website, or study resource in electronic form. It is believed that the findings of the research can be a motivation for other Georgian scholars to engage in similar projects and propose refined solutions. Well-annotated lexical resources are the key to their long-term preservation, exchange and study.

Keywords: *Encoding Print Dictionaries, Georgian Dictionary, Sulkhan-Saba Orbeliani, Text Encoding Initiative, XML*

სულხან-საბა ორბელიანის სიტყვის კონის ანოტაცია TEI-ის საერთაშორისო სტანდარტით

ეკა კიკნაძე
ევროპული მაგისტრი ლექსიკოგრაფიაში
მინიოს უნივერსიტეტი, პორტუგალია
ekakiknadze95@gmail.com

ბეჭდური ლექსიკონების საერთაშორისო ფორმატით ანოტაცია ხელს უწყობს მკვლევრებს მარტივად დაამუშაონ, გააზიარონ და გამოიყენონ ლექსიკოგრაფიული მასალა სხვადასხვა ელექტრონულ მოწყობილობებსა და სისტემებში. TEI (Text Encoding Initiative) – „ტექსტის ანოტაციის ინიციატივა“ საერთაშორისოდ აღიარებული ტექსტური მონაცემების ციფრულ ფორმატში გადაყვანის სტანდარტია. ორგანიზაციამ სხვადასხვა ტიპის ტექსტებისთვის ანოტაციის ძირითადი პრინციპები შეიმუშავა, რაც განსაკუთრებით ხელსაყრელია ისეთი კომპლექსური ტიპის ტექსტებისთვის, როგორიც ლექსიკონია.

ამჟამად, ქართულ ბეჭდურ თუ ონლაინ ლექსიკონებში TEI-ის ფორმატით ანოტაციის მაგალითი არ გვხვდება. ქართული ლექსიკოგრაფიის მთავარ ნიმუშს სულხან-საბა ორბელიანის *სიტყვის კონა* წარმოადგენს, რომელიც ანბანურადაა განლაგებული და ასახავს მე-17 საუკუნეში ქართულ ენაში არსებულ ძირითად ლექსიკურ ერთეულებს. *სიტყვის კონა* დღემდე ერთ-ერთი უმთავრესი დამხმარე სახელმძღვანელოა სხვადასხვა ლექსიკოგრაფიული ნაშრომების შექმნისა თუ ქართული ენის კვლევისთვის.

მოცემული კვლევის მიზანი TEI-ის სტანდარტის სულხან-საბა ორბელიანის ლექსიკონისთვის მორგება იყო. ეს ქართული სიტყვა-სტატიების საერთაშორისო ფორმატით ანოტაციის პირველი მცდელობაა. მნიშვნელოვანი და საინტერესო იყო, გვეცადა და გვეჩვენებინა, რამდენად რთული და ასევე, საჭირო იქნებოდა ისეთი მცირერესურსიანი ენების მსგავსი სტანდარტით ანოტაცია, როგორიც ქართულია.

კვლევისთვის *სიტყვის კონის* მესამე გამოცემა შეირჩა, რომელიც 1949 წელს დაიბეჭდა და ეფუძნება სულხან-საბა ორბელიანის ძმის, ზოსიმეს ხელნაწერს. ეს გამოცემა ძალიან ახლოს დგას თავად ლექსიკოგრაფ სულხან-საბა ორბელიანის ხელნაწერთან. მნიშვნელოვანი იყო, მოგვეძია ისეთი გამოცემა, რომელიც სულხან-საბა ორბელიანის ხელნაწერის ისეთ თავისებურებებს გვაჩვენებდა, როგორებიცაა, მაგალითად, გრაფიკული სიმბოლოები, ილუსტრაციები, თუ ერთმანეთთან დაკავშირებული სიტყვების ერთ სიტყვა-სტატიაში განმარტება.

კვლევაში მიმოხილულია ლექსიკონის გარშემო არსებული კვლევები, ბეჭდური ლექსიკონის ანოტაციის სირთულეები და TEI-ის ძირითადი პრინციპები. TEI-ის ფორმატით შესაძლებელია ლექსიკონის ციფრულად იმგვარად წარმოდგენა, რომ, ერთი მხრივ, მისი პირვანდელი სახე შენარჩუნდეს და, მეორე მხრივ, ადვილი იყოს მკვლევრებისთვის

ლექსიკონში არსებული ინფორმაციის (მაგ. ნაგულისხმევი ინფორმაციის აღდგენა) წაკითხვა და ანალიზი. *სიტყვის კონის* სიტყვა-სტატიების ანოტაციისას უპირატესობა მიენიჭა ლექსიკონის ორიგინალური სახის მაქსიმალურად შენარჩუნებას.

სიტყვა-სტატიების ანოტაციას წინ უძღოდა კვლევის უმნიშვნელოვანესი ნაწილი – ლექსიკონის სიტყვა-სტატიების შესწავლა და მათში არსებული ინფორმაციის დახარისხება. კვლევისთვის შეირჩა ისეთი სიტყვა-სტატიები, რომლებიც მაქსიმალურად სრულყოფილად წარმოადგენდა ლექსიკონში არსებულ ინფორმაციის კლასებს. ამგვარი დახარისხება შესაძლებელი იყო ლექსიკოგრაფების, რ. ჰ. გოუვსისა და დ. ჯ. პრინსლოს მიერ შემუშავებულ ტერმინოლოგიაზე დაყრდნობით. სიტყვა-სტატიები დაჯგუფდა მარტივ, ენციკლოპედიურ და კომპლექსურ სტატიებად და თითოეულ ჯგუფში წარმოდგენილი იყო ანოტაციის რამდენიმე მაგალითი. ლექსიკონის ყველა ინფორმაციული კლასი სემანტიკური ტეგებით აღინიშნა (მაგ. სიტყვის განმარტებისთვის გამოვიყენეთ ტეგი – <def>…</def>). ხოლო დამატებითი ინფორმაცია, ატრიბუტის მნიშვნელობით – ტეგის შიგნით ელემენტის თვისებად ანოტირდა (მაგ. ლექსიკონში არსებული შემოკლებული წყაროები სრული სახით აღიწერა, როგორც შესაბამისი ელემენტის თვისება – <title expand=”გამოსვლა”> 20.10 გამოს. </title>), რითიც, ერთი მხრივ, მაქსიმალურად შენარჩუნდა ლექსიკონის სახე და ასევე, მკვლევრებისთვის მნიშვნელოვანი ნაგულისხმევი ინფორმაციის მოძიებაც გახდა შესაძლებელი. სიტყვა-სტატიის ანოტაციისთვის გამოვიყენეთ XML-ზე (TEI-ის სტანდარტი მარკირების ამ ენას იყენებს მონაცემთა სტრუქტურირებისა და ანოტაციისთვის) დაფუძნებული სასწავლო პლატფორმა x4ml, რომელიც EMLex-ის (European Master in Lexicography) პროფესორმა და გერმანული ენის ინსტიტუტის მკვლევარმა პიტერ მაიერმა შექმნა.

კვლევის შედეგებმა აჩვენა, რომ შესაძლებელია ქართული ლექსიკონის, კერძოდ კი, ისეთი კომპლექსური ლექსიკოგრაფიული ნაშრომის TEI-ის საერთაშორისო ფორმატით ანოტაცია, როგორიც *სიტყვის კონაა*. სიტყვა-სტატიები ისეა ანოტირებული, რომ ლექსიკონის სტრუქტურა და გარეგნული სახეც შენარჩუნდა და ასევე, შესაძლებელია დამატებითი, ნაგულისხმევი ინფორმაციის მოძიებაც. ეს მოდელები შეიძლება გამოვიყენოთ XML-ზე დაფუძნებული მონაცემთა ბაზის, ონლაინლექსიკონისა თუ, უბრალოდ, ელექტრონული სასწავლო მასალის შესაქმნელად. იმედია, კვლევა წაახალისებს მსგავს პროექტებს საქართველოში და სხვა მკვლევრებსაც ექნებათ სურვილი, შეიმუშაონ უკეთესი მოდელები, რაც თავისთავად, ხელს შეუწყობს ქართული ენის ნიმუშების შენარჩუნებას, გაზიარებასა და კვლევას.

საკვანძო სიტყვები: *ბეჭდური ლექსიკონების ანოტაცია, ქართული ლექსიკონი, სულხან-საბა ორბელიანი, Text Encoding Initiative, XML*

References:

Abuladze, I. (1991). Foreword. In Metreveli, E. & Qurtsikidze, Ts. (Eds.), *Sulkhan-Saba Orbeliani Georgian Dictionary I* (pp. 5-8). Merani. [In Georgian].

Budin, G., Majewski, S. and K. Mörth (2012). Creating lexical resources in tei p5. a schema for multi-purpose digital dictionaries. *Journal of the Text Encoding Initiative*, (3).
https://journals.openedition.org/jtei/522 (12.05.2025)

Burnard, L. (2020). What is TEI Conformance, and Why Should You Care?. *Journal of the Text Encoding Initiative*, (12).
https://journals.openedition.org/jtei/1777 (12.05.2025)

Burnard, L. (2014). The TEI and XML. In *What is the Text Encoding Initiative?* (ch. 2). OpenEdition Press.
https://books.openedition.org/oep/680 (12.05.2025)

Burnard, L. (2013). The evolution of the text encoding initiative: from research project to research infrastructure. *Journal of the Text Encoding Initiative*, (5).
https://journals.openedition.org/jtei/811 (12.05.2025)

Ghlonti, A. (1983). *Questions of Georgian Lexicography*. Tbilisi: Sabtchota Sakartvelo [In Georgian].

Gouws, R. H. and D.J. Prinsloo (2010). Microstructural aspects. In *Principles and practice of South African lexicography* (115-143). African Sun Media.
https://core.ac.uk/download/pdf/188222169.pdf (12.05.2025)

Ide, N. And C. M. Sperberg-McQueen (1995). The Text Encoding Initiative: History, Goals, and Future Development.
https://www.researchgate.net/publication/228930958_The_Text_Encoding_Initiative_Its_History_Goals_and_Future_Development (12.05.2025)

Ide, N. And J. Véronis (1995). Encoding dictionaries. *Text Encoding Initiative: Background and Context*, 167-179.
https://link.springer.com/book/10.1007/978-94-011-0325-1 (12.05.2025)

Margalitadze, T. and G. Meladze (2022). Lexicography in Georgia. In T. Makharoblidze (Ed.), *Issues in Kartvelian Studies* (pp. 29-54).
https://vernonpress.com/book/1552 (12.05.2025)

Orbeliani, Sulkhan-Saba. (1949). *Sulkhan-Saba Orbeliani's Georgian Dictionary* (pp. III-XXXI). Georgian SSR Publishing. [In Georgian].
https://dspace.nplg.gov.ge/handle/1234/362602 (12.05.2025)

TEI Consortium. (2023). *A Gentle Introduction to XML* (Version 4.7.0.). Text Encoding Initiative: Chapter v.
https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html (12.05.2025)

TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.7.0.
Last updated on 16th November 2023. TEI Consortium.
http://www.tei-c.org/Guidelines/P5/ (12.05.2025).

Text Encoding Initiative. (2023). *Dictionaries* (Version 4.7.0.). Text Encoding Initiative: Chapter 9.
https://tei-c.org/release/doc/tei-p5-doc/en/html/DI.html (12.05.2025).

Peculiarities of Linguistic Modelling of English-Georgian Scientific Terminology

Kvitsiani, Tamari

Ivane Javakhishvili Tbilisi State University, Georgia

tamunakviciani@yahoo.com

In the modern era, when science and technology are constantly evolving, it is especially important to study the problems of terminology. The development of different fields gives rise to many new scientific concepts and it is necessary to designate the concept, i.e. to give a name to it that should comply with international standards and the peculiarities and linguistic norms of a given language (ISO 704:2022: 83-104). Without such activities, the development of various fields is also hindered, since terminology is a necessary tool for communication between field experts.

The development of science and technology causes the influx of new scientific concepts in the Georgian language as well, which is a big challenge for modern Georgian terminology. Georgian lexicographers and field-experts emphasize that terminological work in Georgia is somewhat hampered, that many terms in various fields are mainly borrowed, which cannot be considered a positive trend (Gogia, 2023; Margalitadze, 2019; Karosanidze, 2012).

To test this tendency, in 2022-2023, we conducted a quantitative study of 1,600 terms from various subfields of biology. Our goal was to investigate the percentage of term borrowings in modern fields, for which the fields of genetics, immunology and biotechnology were selected. We analysed 800 terms of the specified fields and it was revealed that 75-80% of them are borrowed terms. We compared this result with the terminology of more traditional fields, for which we selected 800 botanical, zoological and anatomical terms. The study of the terms of these domains demonstrated a completely different tendency, in particular, almost 80% of the analysed terms turned out to be Georgian, formed on the basis of the Georgian language resources (Kvitsiani, 2023).

The results of this study determined our interest in an in-depth investigation of the term formation methods of traditional fields (botany, zoology and anatomy) and of the fields developed relatively later (immunology, biotechnology and genetics) in the Georgian language and compare these methods to the term-formation methods of the equivalent English and Russian terms. The English language was included in the study, as modern Georgian terms are mostly borrowed from English and the Russian language was added to it because Georgian terminology was heavily influenced by the Russian language in the 20th century.

For the analysis of the traditional methods of Georgian term-formation we relied on the methods and principles described in the monograph of Georgian terminologist R. Ghambashidze *Georgian Scientific Terminology and the Main Principles of its Compilation* (1986).

For the analysis of the main peculiarities of modern terminology we relied on international terminological standards, namely – the ISO standard 704-2022 "*Terminology Work – Principles and Methods*.

As mentioned above, we applied the method of contrastive study of three languages: Georgian, Russian and English. We applied the etymological method of analysis to gain a deeper understanding of the methods of formation of English terms. To determine the etymology of English terms, we used the *Oxford English Dictionary of Historical Principles* (OED) and the Etymological Online Dictionary of the English Language, published on the website etymonline.com.

Our research data was based on 1600 terms from the above-mentioned research (Kvitsiani, 2023) from which we randomly selected 200 terms covering 6 subfields of biology.

In my talk, I will present in detail the results of the research and the main principles of modeling Georgian and English terms. We believe that taking these findings into consideration will be useful for the development of Georgian terminology policy. Linguistic and terminological policy is regulated by the law in many countries. It is necessary to work out terminological policy in Georgia as well. Terminology formation process needs close collaboration of domain expertise, linguistic expertise and information management expertise in order to function properly, develop adequate terminology for the Georgian language and control terminological workflow that is somewhat chaotic at present.

Keywords: *Terminology, Georgian term-formation tendencies, English term-formation tendencies.*

ინგლისურ-ქართული სამეცნიერო ტერმინოლოგიის ლინგვისტური მოდელირების თავისებურებები

თამარ კვიციანი

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი, საქართველო

tamunakviciani@yahoo.com

თანამედროვე ეპოქაში მეცნიერებისა და ტექნიკის განვითარების ფონზე განსაკუთრებული მნიშვნელობა ენიჭება ტერმინოლოგიური პრობლემების შესწავლას. სხვადასხვა დარგის განვითარებასთან ერთად ჩნდება მრავალი ახალი სამეცნიერო ცნება, რასაც აუცილებელია დაერქვას სახელი ანუ შეიქმნას ახალი ტერმინი ისე, რომ ის ზუსტად ასახავდეს შესაბამის ცნებას. ეს პროცესი კი საერთაშორისო სტანდარტების დაცვითა და მოცემული ენის თავისებურებებისა და ენობრივი ნორმების გათვალისწინებით უნდა წარიმართოს (ISO 704: 2022: 83-104). ასეთი საქმიანობის გარეშე ფერხდება დარგის განვითარება, ვინაიდან ტერმინოლოგია დარგის წარმომადგენლებს შორის ურთიერთობის აუცილებელი პირობაა.

მეცნიერებისა და ტექნიკის განვითარებამ ქართულ ენაშიც გამოიწვია ახალი სამეცნიერო ცნებების მოზღვავება, რამაც თანამედროვე ქართული ტერმინოლოგია დიდი გამოწვევების წინაშე დააყენა. ქართველი ლექსიკოგრაფები, დარგის სპეციალისტები წერენ იმის შესახებ, რომ ტერმინოლოგიური მუშაობა საქართველოში გარკვეულად შეფერხებულია, რომ ბევრი ტერმინი სხვადასხვა დარგში, ძირითადად, სესხების გზით შემოდის და მკვიდრდება, რასაც ნამდვილად ვერ მივიჩნევთ დადებით ტენდენციად (Gogia, 2023; Margalitadze, 2019; Karosanidze, 2012).

ამ ტენდენციის შესამოწმებლად 2022-2023 წლებში ჩავატარეთ რაოდენობრივი კვლევა ბიოლოგიის სხვადასხვა დარგის 1,600 ტერმინის მასალაზე. ჩვენი მიზანი იყო გამოგვეკვლია ტერმინთა სესხების პროცენტულობა თანამედროვე დარგებში, რისთვისაც შეირჩა გენეტიკის, იმუნოლოგიისა და ბიოტექნოლოგიის დარგები. გაანალიზდა აღნიშნული დარგების 800 ტერმინი და გამოვლინდა, რომ მათი 75-80% ნასესხები ტერმინია. ეს შედეგი შევადარეთ უფრო ტრადიციული დარგების ტერმინოლოგიას, რისთვისაც შევარჩიეთ 800 ტერმინი ბოტანიკის, ზოოლოგიისა და ანატომიის დარგებიდან. ამ დარგების ტერმინების ანალიზმა სრულიად სხვა ვითარება წარმოაჩინა, კერძოდ, გაანალიზებული ტერმინების თითქმის 80% ქართული ენის რესურსებით აღმოჩნდა ნაწარმოები (Kvitsiani, 2023).

სწორედ აღნიშნული კვლევის შედეგებმა განაპირობა ჩვენი ინტერესი უფრო ჩაღრმავებულად შეგვესწავლა ტრადიციული დარგებისა (ბოტანიკა, ზოოლოგია, ანატომია) და შედარებით მოგვიანებით გა-

ნვითარებული დარგების (იმუნოლოგია, ბიოტექნოლოგია, გენეტიკა) ტერმინების წარმოების მეთოდები ქართულში და ეს მეთოდები შეგვედარებინა ეკვივალენტური ტერმინების წარმოების მეთოდებისათვის ინგლისურ და რუსულ ენებში. ინგლისური ენა იმიტომ, რომ დღეს ქართული ტერმინოლოგია, უპირატესად, ინგლისური ენიდან ნასესხები ტერმინებით ივსება, ხოლო რუსული ენა ჩავრთეთ კვლევაში იმის გამო, რომ ქართული ტერმინოლოგია რუსული ენის დიდ გავლენას განიცდიდა მე-20 საუკუნეში.

ტრადიციული ქართული ტერმინშემოქმედების მეთოდების ანალიზისათვის დავეყრდენით როგნედა ღამბაშიძის მონოგრაფიას *„ქართული სამეცნიერო ტერმინოლოგია და მისი შედგენის ძირითადი პრინციპები“* (Ghambashidze, 1986).

თანამედროვე ტერმინოლოგიის ძირითადი მახასიათებლების ანალიზისათვის გამოვიყენეთ საერთაშორისო ტერმინოლოგიური სტანდარტი ისო 704: 2022, „ტერმინოლოგიური მუშაობა – პრინციპები და მეთოდები“.

როგორც აღინიშნა, კვლევისთვის მივმართეთ სამი ენის შეპირისპირებითი ანალიზის მეთოდს. გამოვიყენეთ ეტიმოლოგიური კვლევის მეთოდი ინგლისური ტერმინების მოდელირების სიღრმისეულად წარმოჩენის მიზნით. ინგლისური ტერმინების ეტიმოლოგიის დასადგენად დავეყრდენით *ინგლისური ენის ლექსიკონს ისტორიულ პრინციპებზე* (OED) და ინგლისური ენის ეტიმოლოგიურ ონლაინლექსიკონს, რომელიც განთავსებულია შემდეგ მისამართზე: etymonline.com.

საკვლევ მასალად გამოვიყენეთ ზემოხსენებული კვლევისათვის შერჩეული 1600 ტერმინიდან (Kvitsiani, 2023) შემთხვევითი შერჩევის პრინციპით ამოკრებილი 200 ტერმინი ბიოლოგიის 6 ქვედარგიდან.

მოხსენებაში დეტალურად წარმოვადგენთ კვლევის შედეგებს, ქართული და ინგლისური ტერმინების მოდელირების ძირითად პრინციპებს, რომელთა გათვალისწინება, ვფიქრობთ, მნიშვნელოვანია ქართული ტერმინოლოგიური პოლიტიკის შემუშავებისათვის. ენობრივი და ტერმინოლოგიური პოლიტიკა ბევრ ქვეყანაში კანონით რეგულირდება. ამ კუთხით აუცილებელია საქართველოშიც წარიმართოს მუშაობა და ამ პროცესებში ჩაერთონ შესაბამისი სახელმწიფო უწყებები, დარგის სპეციალისტები, ენათმეცნიერები, რათა გაკონტროლდეს ტერმინოლოგიური მუშაობა, რომელსაც საკმაოდ ქაოსური ხასიათი აქვს დღევანდელ საქართველოში.

საკვანძო სიტყვები: *ტერმინოლოგია, ქართული ტერმინშემოქმედების ტენდენციები, ინგლისური ტერმინშემოქმედების ტენდენციები*

References:

Gambashidze, R. (1986). *Georgian Scientific Terminology and the Main Principles of its Compilation*. Tbilisi: Metsniereba.

Gogia, N. (2023). Terminological Inconsistences and Translation Problems in Legislative and Other Institutional Documents. *Proceedings of the International Conference LEXICOGRAPHY IN THE 21*ST *CENTURY* . Tbilisi

ISO 704: 2022. *Terminology work – Principles and Methods*

Karosanidze L. (2012). The Problems of Georgian Terminology (History and Modernity). *II International Symposium in Lexicography.* Batumi.

Kvitsiani, T. (2023). Georgian Terminology: Tradition and Present State. *Proceedings of the I International Conference LEXICOGRAPHY IN THE 21*ST *CENTURY.* Tbilisi: Ilia University Press.

Margalitadze, T. (2019). Towards Issues of Terminological Policy in Georgia. *Proceedings of International Conference "Humanities in the Information Society – III"*. Publishing House of Batumi State University. Batumi. ISSN 1987-7625 ISBN 978-9941-462-86-3.

Dictionaries:

English-Georgian Online Biology Dictionary. http://bio.dict.ge/ka/.

OED (2009). *Oxford English Dictionary on historical principles*. Oxford University Press.

Online Etymology Dictionary of the English Language. Available at: https://www.etymonline.com/.

Semantic and Formal Classification of Modern Georgian Neologisms

Laluashvili, Tamari
Ilia State University, Georgia
tamar.laluashvili.1@iliauni.edu.ge

New words or neologisms, that appear in a language, are crucial for its evolution. These new lexical units emerge across various fields and typically reflect the social, cultural, political, and other processes occurring within a language community. The appearance of neologisms indicates that fields, in which these new words are emerging, are evolving and developing. It is exactly these changes that keep a language alive and dynamic. In recent years, a range of new words has emerged in the Georgian language due to various developments in the country and around the world.

Based on the methodology for the detection of neologisms in Georgian, worked out by us, 1700 new words were collected from the Corpus of Georgian Neologisms (Laluashvili, Margalitadze, 2025). In my talk, I will present the results of the semantic and formal classification of these neologisms.

These linguistic innovations span multiple fields. Based on research conducted, several productive areas were identified where these innovations have taken hold. The following fields were distinguished in the Georgian language: politics, technology and the Internet, media, medicine, education, society and social life, tourism, agriculture, business, finance, sports, art, nutrition, and miscellaneous, In which we have combined those neologisms that do not belong to any specific field.

The analysis of Georgian political neologisms is very interesting, as they quite accurately reflect the processes taking place in the country from the independence to today's democratic backslide. These last developments have given rise to many derogatory words with negative connotations in contemporary Georgian.

Neologisms that have emerged in the field of education reflect the results of Georgia's accession to the Bologna Process in 2005 and the implementation of educational reforms which integrated Georgia into the Western educational and scientific space.

The COVID-19 pandemic also significantly contributed to the creation of new terms. As the global pandemic affected the world, the medical field became highly productive in generating new vocabulary. In the field of medicine, we have identified the subfield of aesthetic medicine, which indicates the popularity of this direction and active introduction of new products and spa procedures in Georgia.

Following the Russia-Ukraine war that began in 2022, a number of new terms, including military terms, emerged in international politics which were borrowed into Georgian.

Beyond these developments, linguistic innovations are also spreading due to advancements in technology and digital communications.

Concerning the formal classification of modern Georgian neologisms, we have identified six main classes. The most prevalent category is borrowing, fol-

lowed by word formation, shortening, semantic shifting, blending, and an additional category labeled "miscellaneous". In my talk, I will focus solely on loanwords and patterns of word formation, as these two classes have proven to be the most productive and numerous.

As mentioned above, borrowing is one of the most productive way of formation of neologisms in modern Georgian. With globalization, technological advancements, and the rise of Internet communication, the number of lexical units borrowed from other languages — particularly English — has increased significantly.

In this study, three types of borrowings have been identified: direct borrowings, morphologically adapted borrowings, and calques (loan translations). In case of direct borrowing, the Georgian language borrows words unchanged or with minimal changes. Examples include: *blogi* ('blog'), *vebsait'i* ('website'), *smart'poni* ('smartphone'), *brauzeri* ('browser'), *vebinari* ('webinar').

The flexible grammatical system of the Georgian language facilitates the integration of loanwords into the language. Its rich morphology allows for the adaptation and «Georgianization» of foreign words. When such words are transferred into Georgian, they acquire local characteristics. Georgian prefixes, suffixes, inflections, and other grammatical features can be added to foreign roots. For example: *digit'aluri* ('digital'), *datagva* ('to tag'), *dablok'va* ('to block').

Loan translations, or calques, represent one of the productive sources of lexical innovation in modern Georgian. A calque is a literal translation of foreign lexical units. This approach enriches the language with new lexical units created using its internal resources, eliminating the need for direct borrowing. Examples include: *emotsiuri int'elekt'i* ('emotional intelligence'); *sotsialuri kseli* ('*social network*'); *k'limat'uri tsvlileba* ('climate change').

Word formation is another productive method for creating new words. Two main directions are identified in this process: compounding and derivation. Compound words are formed by combining two or more independent roots, each with its own meaning. For example: *media* + *garemo* ('media environment'), *video* + *tvali* ('video surveillance'), *int'ernet'* + *bank'i* ('internet banking'). Some lexical units have been transformed into components that form compound words. These are: *video-, internet-, media-, online-*, which are commonly used in creating terms related to the internet and the digital world.

Derived words can be categorized into three types based on word formation processes: prefixation, suffixation, and a combination of both (prefixation and suffixation). The productive suffixes are*: -adi, -oba, -uri, -uli -iani, -eri.* From the analysis of examples of prefixation and suffixation, the following combinations were noted: ant'i- + -uri/-uli, e.g. *ant'igenderuli* ('anti-gender'), *ant'ievroat'lant'ik'uri* ('anti-Euro-Atlantic'), ara- + -uli/adi, e.g. *arasanktsirebuli* ('unsanctioned'), *arap'rovotsirebuli* ('non-provoked'), de- + -(iz)atsia, e.g. *dek'riminalizatsia* ('decriminalization'), *delegit'imatsia* ('delegitimization'), *deoligarkizatsia* ('deoligarchization'), da- + -eba/va. e.g. *daguglva* ('*to google*'), *dabust'va* ('to boost').

Through semantic and formal classification, it has become clear that modern Georgian readily absorbs new lexical units, integrates them into the language, and adapts them to fit the grammatical structure of Georgian when needed.

თანამედროვე ქართული ნეოლოგიზმების სემანტიკური და ფორმალური კლასიფიკაცია

თამარ ლალუაშვილი
ილიას სახელმწიფო უნივერსიტეტი, საქართველო
tamar.laluashvili.1@iliauni.edu.ge

ენაში წარმოქმნილი ახალი სიტყვები – ნეოლოგიზმები, მისი ევოლუციის ერთ-ერთი მნიშვნელოვანი კომპონენტია. ახალი ლექსიკური ერთეულები სხვადასხვა სფეროში ჩნდება და, როგორც წესი, ასახავს ამ ენობრივ საზოგადოებაში მიმდინარე სოციალურ, კულტურულ, პოლიტიკურ თუ სხვა პროცესებს. ნეოლოგიზმების გაჩენა ენაში მიანიშნებს, რომ სფერო, რომელშიც მრავლად შეინიშნება ახალი სიტყვები, ვითარდება, განიცდის პროგრესს. სწორედ ამგვარი ცვლილებები და სიახლეები ხდის ენას ცოცხალსა და დინამიკურს. ბოლო წლებში, ქვეყანაში და მსოფლიოში მიმდინარე პროცესების ფონზე ქართულ ენაშიც არაერთი ახალი სიტყვა გაჩნდა.

ქართულ ენაში ნეოლოგიზმების აღმოჩენის ჩვენ მიერ შემუშავებული მეთოდოლოგიის საფუძველზე, ნეოლოგიზმების კორპუსიდან შეიკრიბა 1,700 ახალი სიტყვა (Laluashvili, Margalitadze, 2025). წინამდებარე მოხსენებაში წარმოვადგენთ ამ სიტყვების შესწავლის შედეგებს სემანტიკური და ფორმალური თვალსაზრისით.

ქართული ნეოლოგიზმების შესწავლამ გამოავლინა რამდენიმე პროდუქტიული სფერო, რომელშიც ენობრივმა სიახლეებმა მეტად მოიკიდა ფეხი. ეს დარგებია: პოლიტიკა, ტექნოლოგიები და ინტერნეტი, მედია, მედიცინა, განათლება, საზოგადოება და სოციალური ცხოვრება, ტურიზმი, სოფლის მეურნეობა, ბიზნესი, ფინანსები, სპორტი, ხელოვნება, კვება და კატეგორია *სხვა*, რომელშიც გავაერთიანეთ ის ნეოლოგიზმები, რომლებიც არცერთ კონკრეტულ დარგს არ განეკუთვნება.

ძალიან საინტერესოა ქართული პოლიტიკური ნეოლოგიზმების ანალიზი, რომლებიც საკმაოდ ზუსტად ასახავენ ქვეყანაში მიმდინარე პროცესებს დამოუკიდებლობის მოპოვებიდან დღევანდელ დემოკრატიულ უკუსვლამდე. ამ უკანასკნელმა მოვლენებმა ენაში ბევრი უარყოფითი კონოტაციის დამცრობითი სიტყვა გააჩინა.

საინტერესოა განათლების სფეროში აღმოცენებული ნეოლოგიზმები, რომლებიც ასახავენ 2005 წელს საქართველოს ბოლონიის პროცესთან შეერთებისა და საგანამათლებლო რეფორმის გატარების შედეგებს, რამაც მოახდინა საქართველოს ინტეგრაცია დასავლთის საგანმანათლებლო და სამეცნიერო სივრცეში.

ახალი სიტყვების წარმოქმნაში დიდი წვლილი შეიტანა კოვიდ-19-მაც. მას შემდეგ რაც მსოფლიო პანდემიამ მოიცვა, სამედიცინო სფერო გახდა მეტად პროდუქტიული ახალი სიტყვების წარმოების პროცესში და ამ მხრივ არც ქართული ენა იყო გამონაკლისი. მედიცინის დარგში ცალკე გამოვყავით ესთეტიკური მედიცის სფერო, რაც მეტყვე-

ლებს ამ მიმართულების პოპულარობაზე და თავის მოვლის ახალი საშუალებების აქტიურ დანერგვაზე საქართველოში.

2022 წელს დაწყებული რუსეთ-უკრაინის ომის შემდეგ საერთაშორისო პოლიტიკაში გაჩნდა არაერთი ახალი ტერმინი, მათ შორის სამხედრო ტერმინები, რაც ქართულ ენაშიც შემოვიდა ნასესხობის სახით.

გარდა ამ პროცესებისა, ენობრივი სიახლეები ვრცელდება ტექნოლოგიებისა და ციფრული კომუნიკაციების განვითარების შედეგადაც, რამაც ქართულ ენაში არაერთი სიახლე შემოიტანა.

რაც შეეხება თანამედროვე ქართული ნეოლოგიზმების ფორმალურ კლასიფიკაციას, ამ მიმართულებით გამოიყო ექვსი ძირითადი კატეგორია. ყველაზე გავრცელებული კატეგორიაა სესხება, შემდეგ მოდის სიტყვათწარმოება, სიტყვათა შემოკლება, მნიშვნელობის გადაწევა, შერწყმული სიტყვები და კატეგორია *სხვა*. მოხსენებაში განვიხილავთ მხოლოდ სესხებისა და სიტყვათწარმოების ნიმუშებს, რადგან აღნიშნულთაგან ყველაზე უფრო ნაყოფიერი და მრავალრიცხოვანი სწორედ ეს ორი კატეგორია აღმოჩნდა.

როგორც აღინიშნა, სესხება არის ნეოლოგიზმების წარმოშობის ერთ-ერთი პროდუქტიული მეთოდი. გლობალიზაციის, ტექნოლოგიური პროგრესისა და ინტერნეტკომუნიკაციის პირობებში მნიშვნელოვნად გაიზარდა სხვა ენებიდან — განსაკუთრებით ინგლისური ენიდან — შემოსული ლექსიკური ერთეულები.

კვლევის ფარგლებში გამოიყო ნასესხობის სამი ტიპი: პირდაპირი ნასესხობა, მორფოლოგიურად ადაპტირებული ნასესხობა და კალკი, ანუ თარგმნითი ნასესხობა. **პირდაპირი ნასესხობის** დროს ქართული ენა სიტყვებს სესხულობს ფორმაუცვლელად ან მინიმალური ცვლილებით (მაგ.: *ბლოგი* (blog), *ვებსაიტი* (website), *სმარტფონი* (smartphone), *ბრაუზერი* (browser), *სელფი* (selfie), *ჰეშტეგი* (hashtag), *ვებინარი* (webinar)).

ქართული ენის მოქნილი გრამატიკული სისტემის წყალობით, ნასესხობების ენაში ინტეგრაცია მეტად მარტივდება. ქართული ენის მდიდარი მორფოლოგიის წყალობით ხდება უცხოური წარმომავლობის სიტყვების ენაში ადაპტაცია, მორგება და „გაქართულება“: ასეთი სიტყვები ქართულში გადმოსვლისას იღებენ ქართული ენის მახასიათებლებს და უცხოურ ფუძეს ემატება ქართული პრეფიქსები, სუფიქსები, ბრუნვის ნიშნები და სხვა (მაგ.: *ჰიბრიდ**ული**, დიგიტალ**ური**, და**თაგვა**, და**ბლოკვა**, და**გუგლვა***).

თანამედროვე ქართულ ენაში ლექსიკური ინოვაციების ერთ-ერთი ნაყოფიერი წყაროა **თარგმნითი ნასესხობა** ანუ კალკი. კალკები უცხოური ლექსიკური ერთეულების სიტყვასიტყვითი თარგმანია. ამ გზით ენა ივსება ისეთი ახალი ლექსიკური ერთეულებით, რომლებიც იქმნება ენის შიდა რესურსებით და არ ხდება საჭირო უცხოენოვანი სიტყვების პირდაპირი გადმოტანა (მაგ: *ემოციური ინტელექტი* (‘emotional intelligence’), *სოციალური ქსელი* (‘*social network*’), *კლიმატური ცვლილება* (‘climate change’)).

ქართულში სიტყვათწარმოება კიდევ ერთი პროდუქტიული მეთოდია ახალი სიტყვების შექმნის პროცესში. სიტყვათწარმოებისას

გამოიყოფა ორი ძირითადი მიმართულება: რთული სიტყვები და ნაწარმოები სიტყვები. რთული სიტყვების წარმოქმნა ხდება ორი ან მეტი დამოუკიდებელი სიტყვის შერწყმის შედეგად (*მედია+გარემო, ვიდეო+თვალი, ინტერნეტ+ბანკი*). ზოგიერთი ლექსიკური ერთეული თავად იქცა რთული სიტყვების მაწარმოებელ კომპონენტებად, ესენია: *აუდიო-, ბიზნეს-, ვიდეო-, ინტერნეტ-, მედია-, ონლაინ-*. აღნიშნული ელემენტები, გარდა ბიზნესისა, გამოიყენება ინტერნეტისა და ციფრული სამყაროს ტერმინების შექმნის პროცესში.

ნაწარმოებ სიტყვებში გამოიყო სამი ტიპის წარმოება: პრეფიქსული წარმოება, სუფიქსული წარმოება და პრეფიქს-სუფიქსული წარმოება. **სუფიქსული წარმოების** ელემენტები, რომლებიც მოპოვებული მასალის საფუძვლზე გამოიკვეთა არის: *-ადი, -ობა, -ური, -ული -იანი, -ერი, -იზაცია*. **პრეფიქს-სუფიქსული წარმოების** მაგალითების ანალიზის შედეგად გამოვლინდა შემდეგი ელემენტები: ანტი- + -ური/-ული (*ანტიგენდერული, ანტიევროატლანტიკური*), არა- + -ული/ადი (*არასანქცირებული, არაოპერირებადი, არაპროვოცირებული*), დე- + -(იზ)აცია (*დეკრიმინალიზაცია, დელეგიტიმაცია, დეოლიგარქიზაცია*), და- + -ება/ვა (*დალაიქება, დაგუგლვა, დაბუსტვა*).

თანამედროვე ქართულ ენაში გაჩენილი ნეოლოგიზმების სემანტიკური და ფორმალური კლასიფიკაციის შედეგად გამოვლინდა, რომ ენა მარტივად ითვისებს ახალ ლექსიკურ ერთეულებს, ახდენს მათ ინტეგრაციას ენაში და საჭიროების შემთხვევაში, არგებს ქართული ენის გარამატიკულ სტრუქტურასაც.

Reference

Laluashvili, T. and T. Margalitadze (2025). Semi-Automatic Detection of New Words in Modern Georgian. *Lexikos, 35*(1), 399-413. doi:https://doi.org/10.5788/35-1-2044

AI in Lexicography for Low-Resource Languages: The Case of the Georgian Language

Lomidze, Ketevani
Ilia State University, Georgia
ketevani.lomidze.1@iliauni.edu.ge

The emergence of AI has driven the process of integrating chatbots into various fields and decreasing human contribution in many instances. Lexicography has not been an exception, and a significant number of studies have been conducted since 2022 to the present day, aiming to identify the role of AI in dictionary creation. Some researchers prophesy the end of lexicography and the redundancy of lexicographers, while some still question the accuracy and consistency of chatbot-generated lexicon entries. The role of AI in lexicography for low-resource languages such as the Georgian language has not been part of this debate, as the conducted studies primarily focused on the lingua franca, high-resource language, English, which is the "mother tongue" of chatbots. Therefore, the performance of chatbots in low-resource languages remains debatable and underrepresented, and it needs to be discovered if AI can outperform lexicographers globally.

The aim of this research is to examine the performance of ChatGPT in the case of low-resource languages like Georgian and analyse its abilities in terms of lexicography. Georgian belongs to the South Caucasian (Kartvelian) language family, and it is very different from Indo-European languages. Hereby, it is important to note that chatbots like ChatGPT can generate text and provide answers based on the user's queries and the data on which they have been trained. Therefore, ChatGPT as an AI needs to process a large amount of data from various sources like Wikipedia, Newspapers or literature to comprehend input, make connections and generate meaningful responses. Interestingly, while the current version of ChatGPT can process more than 90 languages, the training data it receives in each language varies. Languages in which it has more training and obtains more data are called high-resource languages, for example, English. That is why earlier studies suggest that ChatGPT performs well in English and easily generates dictionary entries or texts. On the other hand, languages in which chatbots lack data are referred to as low-resource languages such as the Georgian language. As Georgian lacks digital presence, ChatGPT's potential regarding this language is yet to be analysed, and it is debatable to what extent ChatGPT can produce accurate, meaningful, and creative texts.

This paper is qualitative research that observes to what extent ChatGPT can provide:

- headwords,
- polysemous meanings of words,
- definitions of meanings,
- illustrative sentences
- grammar information
- labels – for a Georgian explanatory Dictionary.

The chosen methodology partly follows the similar study of Trap-Jensen (2024) regarding the English language. All entries generated by ChatGPT were compared to the entries of Georgian online monolingual dictionary (Tsotsanidze, Loladze, Datukishvili 2014) to depict correct results and errors.

The research showed that ChatGPT can easily suggest new headwords which are not yet recorded in existing dictionaries. It can also propose some interesting neologisms, mostly related to technology and social media. Some examples include: *გამომწერი (gamomtseri)* – "subscriber", *გიგაბაიტი (gigabaiti)* – "gigabyte", *სმარტფონი (smartponi)* – "smartphone", *ვებინარი (vebinari)* – "webinar", *გეოლოკაცია (geolokacia)* – "geolocation", *and ინფლუენსერი (influenseri)* – "influencer". Interestingly, ChatGPT performs well when suggesting borrowed words from English, but it struggles to provide accurate results when asked to list new Georgian words or to identify words that have been overlooked by dictionaries.

The results observed that ChatGPT performs better in creating dictionary entries when it is provided with examples of usage. For research purposes, it was instructed to use the entries of the *Georgian Dictionary* (Tsotsanidze, Loladze, Datukishvili 2014) as a model. While ChatGPT successfully copied the general style of the entries, it made numerous errors in sentence structure, spelling, and the grammar. It is evident that ChatGPT «thinks» in English and then translates content into Georgian. The fewer corrections it receives from you, the more unclear its output becomes. Therefore, although its entries can be innovative and useful, they require careful editing by a professional lexicographer.

Another frequent issue is hallucination. ChatGPT tends to hallucinate more in Georgian than in other languages by generating nonexistent words or meanings, such as *თეთრხალათიანობა (tetrkhalatianoba)* – "the wearing of white coats", *ზაფხულწათვისობა (zapxultsatvisoba)* – for "summer-ness", and *ვაშლიანობა (vashlianoba)* – for "bitterness". Interestingly, but not surprisingly, ChatGPT relies on English sources and often transfers (uses semantic borrowing) English word meanings onto Georgian equivalents. For instance, it incorrectly suggested *მგრძნობიარე თემა (sensitive topic)*, which does not exist in Georgian. The appropriate expression in this case would be *"დელიკატური თემა" (delicate topic)*.

When it comes to identifying grammatical class and labels, ChatGPT performs better and makes fewer mistakes. However, it often struggles with verbal nouns, sometimes assigning two possible grammatical categories or inventing nonexistent parts of speech, such as *ზმნიზედა სახელი (adverb-noun)* or *ზმნებითი სახელი (verb-noun)*. On the other hand, ChatGPT accurately identifies terms and their relevant domains. It can also determine whether a word is formal, informal, neutral, or derogatory, provided the word is part of its training data. Interestingly, ChatGPT is also eager to provide etymology, and they are mostly correct when it comes to English loanwords; for instance, *"The Georgian word ინფლუენსერი (inpluenceri) is a loanword derived from the English word «influencer.» – «influence» + suffix -er."*

In the presentation we will discuss both the strengths and limitations of using ChatGPT in the creation of Georgian dictionaries. It will highlight to what extent it requires lexicographers' intervention when it comes to low-resource languages such as the Georgian language.

ხელოვნური ინტელექტი მცირერესურსიან ენათა ლექსიკოგრაფიაში: ქართული ენის მაგალითი

ქეთევან ლომიძე
ილიას სახელმწიფო უნივერსიტეტი, საქართველო
ketevani.lomidze.1@iliauni.edu.ge

ხელოვნური ინტელექტის გამოჩენამ ბიძგი მისცა ჩატბოტების სხვადასხვა სფეროში ინტეგრაციას და თითქოს შეამცირა ადამიანის როლი. ამ თვალსაზრისით არც ლექსიკოგრაფიაა გამონაკლისი. 2022 წლიდან დღემდე, ბევრი კვლევა ჩატარდა იმის დასადგენად, თუ როგორია ხელოვნური ინტელექტის როლი ლექსიკონების შექმნის პროცესში. შედეგად, ზოგიერთი მკვლევარი შიშობს, რომ მომავალში ლექსიკოგრაფები აღარ იქნებიან საჭირო; სხვები კი ეჭვის თვალით უყურებენ ჩატბოტების მიერ შექმნილი სიტყვა-სტატიების სიზუსტესა და საიმედოობას. ლექსიკოგრაფიაში ხელოვნური ინტელექტის როლის ანალიზისას, მკვლევრები ძირითადად სწავლობენ დიდი რესურსების მქონე ინგლისურ ენას, რომელიც ჩატბოტების „მშობლიური ენაა." ნაკლებად არის შესწავლილი ხელოვნური ინტელექტის ეფექტურობა მცირე რესურსების მქონე ენებისთვის, მათ შორის ქართული ენისათვის. ჩატბოტების პრაქტიკული როლი მცირერესურსიან ენებში ჯერ კიდევ შესწავლის საგანია და არ არის სათანადოდ გამოკვლეული.

ამ მოხსენების მიზანია გამოიკვლიოს ChatGPT-ის შესაძლებლობები ლექსიკოგრაფიაში მცირერესურსიანი ენებისათვის ქართული ენის მაგალითზე. ქართული ენა სამხრეთ კავკასიურ (ქართველურ) ენათა ოჯახს მიეკუთვნება და საკმაოდ განსხვავდება სხვა ენებისგან. აუცილებელია აღვნიშნოთ, რომ ChatGPT, როგორც სხვა ბოტები, მომხმარებლის მოთხოვნის მიხედვით ქმნის ტექსტს და სთავაზობს პასუხს იმ ინფორმაციაზე დაყრდნობით, რომელზეც იგი გაწვრთნეს. მაშასადამე, ChatGPT-ის როგორც ხელოვნურ ინტელექტს, სჭირდება დიდი რაოდენობის მონაცემების დამუშავება, როგორიც შეიძლება იყოს ვიკიპედია, გაზეთები ან ლიტერატურა. ასე შეუძლია მას გაანალიზოს და გაიგოს ტექსტი, დააკავშიროს მასალა ერთმანეთთან და მოგვცეს აზრიანი პასუხები. საინტერესოა ის, რომ ChatGPT-ის დღევანდელ მოდელს შეუძლია 90-ზე მეტი ენის დამუშავება, თუმცა ყოველ ენაში იგი განსხვავებული მონაცემებით იწვრთნება. იმ ენას, რომელშიც მას ყველაზე მეტი მასალა და წვრთნა აქვს გავლილი, ეწოდება დიდრესურსიანი ენა. ასეთია, მაგალითად ინგლისური. სწორედ ამიტომაც, აქამდე ჩატარებულმა კვლევებმა აჩვენა კიდევაც, რომ ChatGPT საკმაოდ ადვილად ქმნის სიტყვა-სტატიებსა და ტექსტებს ამ ენაზე. საპირისპიროდ, იმ ენას, რომელზეც ჩატბოტს ნაკლები მონაცემი აქვს, მცირერესურსიან ენას უწოდებენ და ამ ჯგუფს მიეკუთვნება ქართული ენა. რადგანაც ქართულ ენას ციფრულ სამყაროში შედარებით მოკრძალებული ადგილი უკავია, ChatGPT-ის პოტენციალი ამ ენისთვის ჯერ კიდევ შესასწავლია და არ ვიცით,

რამდენად შეუძლია მას ზუსტი, აზრიანი და შემოქმედებითი ტექსტების შექმნა.

მოხსენება ეფუძნება თვისებრივ კვლევას, რომელიც აანალიზებს რამდენად შეუძლია ChatGPT-ის ქართული ენის განმარტებითი ლექსიკონისათვის დაადგინოს:

- ლექსიკონის სიტყვანი,
- სიტყვის პოლისემიური მნიშვნელობები,
- მნიშვნელობათა განმარტებები,
- საილუსტრაციო მაგალითები,
- გრამატიკული ინფორმაცია,
- კვალიფიკაციები.

ასევე გავაანალიზეთ ხელოვნური ინტელექტის უნარი ქართულ ენაში ნეოლოგიზმების აღმოჩენის თვალსაზრისით.

არჩეული მეთოდოლოგია ნაწილობრივ მიჰყვება ტრაპ-იენსენის 2024 წლის კვლევას ინგლისური ენის მაგალითზე (Trap-Jensen 2024). ChatGPT-ის მიერ შექმნილი სიტყვა-სტატიები შევადარეთ ქართული ენის განმარტებითი ლექსიკონის სიტყვა-სტატიებს მიღებული შედეგების შესაფასებლად.

კვლევამ აჩვენა, რომ ChatGPT-ის ადვილად შეუძლია ახალი სიტყვების შემოთავაზება, რომლებიც ჯერ ლექსიკონებში არ არის შეტანილი. მას ასევე შეუძლია საინტერესო ნეოლოგიზმების აღმოჩენა, რომლებიც უპირატესად დაკავშირებულია ტექნოლოგიასთან და სოციალურ ქსელებთან; ამის რამდენიმე ნათელი მაგალითია: *გამომწერი, გიგაბაიტი, სმარტფონი, ვებინარი, გეოლოკაცია და ინფლუენსერი.* აღსანიშნავია, რომ ChatGPT უკეთესად ართმევს თავს ინგლისურიდან ნასესხები სიტყვების ანალიზს, თუმცა სწორი პასუხების გაცემა უჭირს, როდესაც სთხოვ ახალი ქართული სიტყვების ჩამოწერას ან უკვე არსებული სიტყვების აღმოჩენას, რომელიც ლექსიკონებში არაა ასახული.

კვლევამ აჩვენა, რომ ChatGPT უკეთესად ქმნის სიტყვა-სტატიებს, თუ მას ვაძლევთ სასათაურო სიტყვის გამოყენების მაგალითებს. სწორედ ამიტომაც, ამ კვლევისთვის მას მოდელად მიეცა *ქართული ლექსიკონის* (Tsotsanidze, Loladze, Datukishvili 2014) სიტყვა-სტატიები. ChatGPT-მ წარმატებით შეძლო სიტყვა-სტატიის ზოგადი სტილის გადაღება, თუმცა უამრავი შეცდომა დაუშვა ქართული წინადადებების აგებისას, მართლწერასა და გრამატიკაში. აშკარაა, რომ ChatGPT ჯერ ინგლისურად „ფიქრობს," შემდგომ კი შინაარსს ქართულად თარგმნის. რაც უფრო ნაკლებად ჩაუსწორებ შეცდომებს, მით უფრო გაუგებარი ხდება მისი პასუხებიც. შესაბამისად, მიუხედავად იმისა, რომ მისი სიტყვა-სტატია შეიძლება იყოს ინოვაციური და სასარგებლო, მას პროფესიონალი ლექსიკოგრაფის ჩასწორება მაინც სჭირდება.

კიდევ ერთი ხშირი პრობლემაა ჰალუცინაცია. ChatGPT-ის ქართულ ენაში უფრო მეტი ჰალუცინაცია აქვს ვიდრე სხვა ენებში. იგი ქმნის არარსებულ სიტყვებსა და მნიშვნელობებს, როგორიცაა *თეთრხალათიანობა*

– “თეთრი ხალათის ტარება”, *ზაფხულწათვისობა* – „ზაფხულის თვისება“ და *ვაშლიანობა* – „სიმწარე“. საინტერესოა, მაგრამ არა გასაკვირი, რომ ChatGPT ინგლისური ენის წყაროებს ეყრდნობა და ხშირად ინგლისური სიტყვების მნიშვნელობები სემანტიკური კალკით გადმოაქვს ქართულ ენაში. მაგალითად, შემოთავაზებული *მგრძნობიარე თემა (sensitive topic)* არ არსებობს ქართულში და მისი სწორი შესატყვისი იქნება *„დელიკატური თემა“ (delicate topic)*.

როდესაც საქმე ეხება სიტყვის მეტყველების ნაწილისა და კვალიფიკაციის შერჩევას, ChatGPT უკეთესად მუშაობს და ნაკლებ შეცდომებს უშვებს. თუმცა მას ხშირად უჭირს სახელზმნების გაანალიზება და ზოგჯერ ორ მეტყველების ნაწილს წერს ან იგონებს არარსებულ მეტყველების ნაწილს, როგორიცაა *ზმნიზედა სახელი* და *ზმნებითი სახელი*. მეორე მხრივ, ChatGPT-ის ზუსტად შეუძლია ტერმინების აღმოჩენა და მათი სფეროების დადგენა. მას ასევე შეუძლია განსაზღვროს სიტყვა წიგნურია, სასაუბროა, ნეიტრალური თუ დამცრობითი, თუმცა ეს სიტყვა მისი მონაცემების ბაზაში უნდა არსებობდეს. აღსანიშნავია, რომ ChatGPT ასევე დიდი მონდომებით ცდილობს ეტიმოლოგიის აღწერას, რომელიც უპირატესად სწორია ინგლისური ნასესხობების შემთხვევაში, მაგ: „ქართული სიტყვა *ინფლუენსერი (inpluenceri)* ნასესხებია ინგლისური სიტყვიდან «influencer» — «influence» + სუფიქსი -er.”

მოხსენებაში განვიხილავთ ChatGPT-ის ძლიერ და სუსტ მხარეებს ქართული ლექსიკოგრაფიისათვის და წარმოვაჩენთ, თუ რამდენად საჭიროებს ის ლექსიკოგრაფების ჩარევას, როცა ეს ისეთ მცირერესურსიან ენებს ეხება, როგორიცაა ქართული ენა.

Trap-Jensen, L. (2024). The Best of Two Worlds: Exploring the Synergy between Human Expertise and AI in Lexicography. T. Margalitadze (ed.). *Proceedings of the I International Conference “Lexicography in the XXI Century”*. Tbilisi: Ilia State University Press.

Tsotsanidze, G., Loladze, N., Datukishvili, K. (2014). *Georgian Dictionary.* Tbilisi: Sulakauri Publishing. https://www.ganmarteba.ge/

Semantics of Eating in Afrikaans, Northern Sotho and Georgian: Cross-linguistic Variation in Metaphor

Lomidze, Ketevani; Nadiradze, Gvantsa; Sadghobelashvili, Natia

Ilia State University, Georgia

ketevani.lomidze.1@iliauni.edu.ge, gvantsa.nadiradze.3@iliauni.edu.ge, natia.sadghobelashvili.1@iliauni.edu.ge

The study of conceptual metaphor is of particular importance in the theory of cognitive semantics (Geeraerts 2010: 203-222). It allows one to see a thing from a completely different perspective, 'to see one thing in terms of another' (Geeraerts 2010: 203). According to Lakoff and Johnson's theory of conceptual metaphor, metaphor is a cognitive phenomenon, rather than a purely lexical one; It should be regarded and analysed as a mapping between two domains, and it is grounded in experience (Lakoff and Johnson 1980). From the point of view of cognitive linguistics, metaphor also expresses cultural experience.

E. Tallard and N. Bosman's study on the semantics of *eating* in Afrikaans and Northern Sotho aims to determine whether metaphors of eating are related to similar experiences in the mentioned languages, also to determine how universal these metaphors are, taking into account the cultural differences between the language communities (Taljard and Bosman 2014).

The Northern Sotho and Afrikaans languages belong to completely different language families. Afrikaans is an Indo-European language and belongs to the West Germanic group of languages, along with Dutch, German, Frisian and English. As for Northern Sotho, it is a southeastern Bantu language and is distinguished by rich morphology.

Following the study of Tallard and Bosman, in our research we decided to analyse metaphors related to the semantics of eating in Georgian and collate the results with those of Northern Sotho and Afrikaans.

This paper is corpus-based research and follows Taljard and Bosman's (2024) methodology. Keywords associated with the semantics of eating have been selected to identify metaphors related to eating in Georgian. Chosen words include the verb *ch'ama* 'eat' itself, *qlap'va* 'swallow', *moneleba* 'digest', *ghech'va* 'chew', *khramuni* 'munch', *sansvla* 'devour' and many more. All keywords have been selected according to the semantic field of "eating," including synonyms, hypernyms and others. In the case of the Georgian language, more than 20 related lexical units were found, which were then filtered to include only the most relevant ones.

For their study, Taljard and Bosman used the Afrikaans and Sepedi corpora of the University of Pretoria to find the instances of metaphors. As for Georgian examples, we examined the Georgian National Corpus (GNC)[18], English-Georgian Parallel Corpus[19] and the webcorpus of Georgian, Araneum Georgianum[20]

18 http://gnc.gov.ge/gnc/page

19 https://enkacorpus.iliauni.edu.ge/

20 http://unesco.uniba.sk/aranea/

created by Vladimir Benco at the Slovak Academy of Sciences. The metaphors were chosen not according to frequency, but for their figurative meanings that do not literally represent food consumption.

The collected materials were organised into three different metaphor groups: agent-oriented, patient-oriented and a mixture of both. Agent-oriented metaphors refer to instances where the focus is on the consumer's internal and external state. On the other hand, patient-oriented examples highlight the state of food in the process of eating. The last group focuses on the metaphors where both the subject and the object undergo certain influence. Newly discovered exclusive Georgian metaphors were also categorised in a similar manner.

Taljard and Bosman identified the following metaphors in Northern Sotho and Afrikaans: "EMOTIONAL SATISFACTION IS EATING" – emotional contentment is conceptualised as pleasure derived from physical eating. "INTELLECTUAL SATISFACTION IS EATING" – intellectual achievement or the acquisition of knowledge is seen as satisfying mental hunger – essentially, the consumption of knowledge. "UNDERSTANDING IS EATING" – comprehension is conceptualised as a form of ingestion. "INTELLECTUAL ACTIVITY IS EATING" – intellectual engagement is portrayed as an act of feeding – the processes of thinking, analyzing, or learning are similar to "mental consumption." "SEX IS EATING" – the sexual act is metaphorically described using the language of eating, emphasizing the physical connection. "DISAPPEARANCE/ABSENCE IS EATING" – disappearance, loss, or absence is described as a process in which an agent "devours" or consumes person or object. "TORMENT (PHYSICAL AND PSYCHOLOGICAL) IS EATING" – both physical and psychological pain are represented as an internal devouring force, as if the individual is being "eaten" from within. "DEFEATING IS EATING" – defeating opponents is conceptualised as swallowing or devouring them. "CONCEDING IS EATING" – conceding or admitting defeat is associated with "swallowing" one's words, accusations, or lies. "ENDEARMENT IS EATING" – affectionate behaviour is sometimes metaphorically linked to eating (e.g., "You're so cute, I could eat you up"). "HUMILIATION IS EATING" – although eating is generally associated with pleasure, here it reflects forced acceptance of something unpleasant, such as shame.

E. Taljard and N. Bosman's study has revealed both similarities and differences between Afrikaans and Northern Sotho. However, the final results obtained showed that metaphors related to eating in these languages are universal and rely less on cultural differences.

The data of the Georgian language, obtained within the framework of our study was collated with the data of Afrikaans and Northern Sotho. Our analysis has revealed that out of these sixteen metaphorical classes identified by Taljard and Bosman, Georgian had parallels with fourteen of them. For instance, the metaphor "TORMENT (PHYSICAL AND PSYCHOLOGICAL) IS EATING" is attested not only in Northern Sotho and Afrikaans but also in Georgian, where eating is used metaphorically to represent both spiritual suffering, e.g., „ტკივილმა გამოუხრა გული“ (literally – pain hollowed out her heart) and physical torment, e.g., „ავადმყოფობამ შეჭამა“ (literally – illness has eaten him up).

Similarly, "HUMILIATION IS EATING" is found in both Georgian and Afrikaans, where humiliation or shame is expressed through verbs with the

meaning “to swallow” or “to eat,” such as „სირცხვილი ჭამა“ (literally – he/she has eaten shame), „სირცხვილი გადაყლაპა“ (literally – he/she has swallowed shame).

Moreover, Georgian exhibits additional metaphorical categories that do not appear in Northern Sotho or Afrikaans. Examples include: Depletion or exhaustion, e.g., „არსებულმა სტრუქტურამ თავისი დრო მოჭამა,“ (literally – the existing structure has eaten up its time); Annoyance or irritation, e.g., „ტვინი შემიჭამა“ (literally – he/she ate my brain); „ნუ შემჭამე, ქალო!“ (literally – don’t eat me alive, woman!).

This analysis supports the claim that while metaphors can be culturally specific, they also tend to display universality. As George Lakoff’s third postulate asserts, metaphors are based on experience—language emerges from human experience. Consequently, identical metaphors can be created across three linguistically and culturally distinct languages: Northern Sotho, Afrikaans, and Georgian. Our study also confirms the cross-linguistic convergence of conceptual metaphors between languages.

ჭამის სემანტიკა აფრიკაანსში, ჩრდილოეთ სოთოსა და ქართულ ენებში: მეტაფორის კროს-ლინგვისტური ვარიაცია

ქეთევან ლომიძე, გვანცა ნადირაძე, ნათია სადღობელაშვილი

ილიას სახელმწიფო უნივერსიტეტი, საქართველო

ketevani.lomidze.1@iliauni.edu.ge, gvantsa.nadiradze.3@iliauni.edu.ge, natia.sadghobelashvili.1@iliauni.edu.ge

კონცეპტუალური მეტაფორის კვლევას განსაკუთრებული მნიშვნელობა აქვს კოგნიტიური სემანტიკის თეორიაში (Geeraerts 2010: 203-222). მისი საშუალებით შესაძლებელია ერთი საგნის სრულიად სხვა კუთხით დანახვა. ლაკოფისა და ჯონსონის კონცეპტუალური მეტაფორის თეორიის თანახმად მეტაფორა კოგნიტიური ფენომენია, ის ორ ნიშან-თვისებას აკავშირებს ერთმანეთთან და გამოცდილებით იქმნება (Lakoff and Johnson 1980). მეტაფორა კოგნიტიური ლინგვისტიკის თვალსაზრისით, კულტურულ გამოცდილებასაც გამოხატავს.

ე. ტალიარდისა და ნ. ბოსმანის კვლევა ჭამის სემანტიკის სიტყვებზე აფრიკაანსა და ჩრდილოეთ სოთოში მიზნად ისახავს დაადგინოს, განპირობებულია თუ არა ჭამის სემანტიკასთან დაკავშირებული მეტაფორები მათ მიერ შერჩეულ ენებში მსგავი გამოცდილებით და რამდენად უნივერსალურია ისინი, იმის გათვალისწინებით, რომ ამ ენებზე მოსაუბრე ენობრივ საზოგადოებებს შორის მნიშვნელოვანი კულტურული განსხვავებები არსებობს (Taljard and Bosman 2014).

კვლევაში წარმოდგენილი ჩრდილოეთ სოთოსა და აფრიკაანსის ენები სრულიად განსხვავებულ ენათა ოჯახებს განეკუთვნება. აფრიკაანსი ინდოევროპულ ენათა ოჯახში შედის. ის ჰოლანდიურ, გერმანულ, ფრიზულ და ინგლისურ ენებთან ერთად დასავლურგერმანიკულ ენათა ჯგუფშია წარმოდგენილი. ხოლო, ჩრდილოეთ სოთოს ენა სამხრეთ-აღმოსავლეთ ბანტუს ენებს ეკუთვნის და რთული მორფოლოგიით გამოირჩევა.

ე. ტალიარდისა და ნ. ბოსმანის კროს-ლინგვისტური ანალიზის კვალდაკვალ, წარმოდგენილი კვლევით მიზნად დავისახეთ ქართულ ენაში ჭამის სემანტიკასთან დაკავშირებული მეტაფორების გაანალიზება და მათი შედარება ჩრდილოეთ სოთოსა და აფრიკაანსის ენების შედეგებთან.

წარმოდგენილი ანალიზი არის კორპუსზე დაფუძნებული კვლევა და ეყრდნობა ტალიარდისა და ბოსმანის (2014) მეთოდოლოგიას. ჭამის მეტაფორების დასადგენად შეირჩა ჭამის სემანტიკის მქონე საძიებო სიტყვები, ესენია: თავად *ჭამა* (*ჭამა* ქართულში, *eet* აფრიკაანსში, *ja* ჩრდ. სოთოში), *ყლაპვა*, *მონელება*, *ღეჭვა*, *ხრამუნი*, *სანსვლა* და სხვა. ყველა ეს საძიებო სიტყვა შეირჩა *ჭამის* სინონიმების, ჰიპერონიმებისა და სხვა სემანტიკური კავშირების მქონე სიტყვების დახმარებით. ქართული

ენაში ჭამის სემანტიკასთან დაკავშირებული სიტყვები აღმოჩნდა 20-ზე მეტი. შემდეგ ეტაპზე ეს სიტყვები გაიფილტრა და დარჩა მხოლოდ კვლევისთვის საინტერესო ვარიანტები.

შერჩეული სიტყვების გადატანითი მნიშვნელობის საძიებლად ტალიარდმა და ბოსმანმა გამოიყენეს პრეტორიის უნივერსიტეტის აფრიკაანსისა და სეპედის ენების კორპუსები. ქართული ენის მაგალითების კვლევისთვის გამოვიყენეთ ქართული ენის ეროვნული კორპუსი,[21] ინგლისურ-ქართული პარალელური კორპუსი[22] და სლოვაკეთის მეცნიერებათა აკადემიაში ვლადიმირ ბენკოს მიერ შექმნილი ქართული ენის ვებკორპუსი Araneum Georgianum[23]. ჭამის სემანტიკის მეტაფორული მაგალითების ძიებისას ყურადღება არ ექცეოდა სიხშირეს. შეირჩა მხოლოდ ის მაგალითები, რომლებშიც ჭამა გადატანითი მნიშვნელობით იყო გამოყენებული და საკვების რეალურ მიღებას არ გულისხმობდა.

კორპუსებიდან შეკრებილი მასალა დაიყო მეტაფორების სამ ჯგუფად: სუბიექტზე ორიენტირებულ, ობიექტზე ორიენტირებულ და შერეულ მეტაფორებად. სუბიექტზე ორიენტირებული მეტაფორა გულისხმობს ისეთ შემთხვევებს, სადაც ყურადღება გამახვილებულია ჭამის მოქმედების შემსრულებელი სუბიექტის შინაგან და გარეგან მდგომარეობაზე. ობიექტზე ორიენტირებული მაგალითები განიხილავს იმას, თუ რა მოსდის თავად საჭმელს. შერეულ ჯგუფში გაერთიანებულია ისეთი მეტაფორები, სადაც სუბიექტიც და ობიექტიც განიცდის ზემოქმედებას. ქართული ენის კვლევისას დამატებით აღმოჩენილი მეტაფორული მნიშვნელობებიც მსგავსად განაწილდა ამ ჯგუფებში.

ჩრდილოეთ სოთოსა და აფრიკაანსის ენებში ტალიარდმა და ბოსმანმა გამოავლინეს შემდეგი მეტაფორები: "EMOTIONAL SATISFACTION IS EATING" – ემოციური კმაყოფილების განცდა წარმოდგენილია როგორც ფიზიკური ჭამით გამოწვეული სიამოვნება; "INTELLECTUAL SATISFACTION IS EATING" – ინტელექტუალური მიღწევა ან ცოდნის დაუფლება აღიქმება როგორც გონებრივი შიმშილის დაკმაყოფილება — ანუ ცოდნის მიღება, „შეჭმა"; UNDERSTANDING IS EATING – რაღაცის გაგება, ჩაწვდომა, „მონელება"; "INTELLECTUAL ACTIVITY IS EATING"- ინტელექტუალური საქმიანობა აღწერილია როგორც კვებითი აქტი — ფიქრი, გაანალიზება ან სწავლის პროცესი მსგავსია „გონებრივი ჭამისა." "SEX IS EATING" – სექსუალური აქტის წარმოდგენა ჭამის მეტაფორით; "DISAPPEARANCE / ABSENCE IS EATING" – გაქრობა, დაკარგვა ან არყოფნა; "TORMENT (PHYSICAL AND PSYCHOLOGICAL) IS EATING" – ტკივილი, როგორც ფიზიკური ისე ფსიქოლოგიური, განიხილება როგორც შიდა მტანჯველი ძალა, რომელიც ადამიანს შიგნიდან „ჭამს". "DEFEATING IS EATING" – მოწინააღმდეგის დამარცხება აღიქმება, როგორც მისი გადაყლაპვა / შთანთქმა; "CONCEDING IS EATING" – დათმობა ან მარცხის აღიარება; "ENDEARMENT IS EATING"

21 http://gnc.gov.ge/gnc/page

22 https://enkacorpus.iliauni.edu.ge/

23 http://unesco.uniba.sk/aranea/

– სიყვარულით ან ალერსით მოპყრობა ხანდახან მეტაფორიზდება ჭამით (მაგ., „შენ ისეთი საყვარელი ხარ, რომ შეგჭამ“); “HUMILIATION IS EATING” – დამცირება განხილულია, როგორც რაიმე არასასიამოვნოს ჭამის აქტი (მაგ., „სირცხვილის ჭამა“).

ე. ტალიარდისა და ნ. ბოსმანის კვლევამ გამოავლინა როგორც მსგავსება, ისე განსხვავება ჩამოთვლილ მეტაფორებს შორის. თუმცა, საბოლოოდ კვლევამ აჩვენა, რომ აფრიკაანსსა და ჩრდილოეთ სოთოს ენებში ჭამასთან დაკავშირებული მეტაფორები უნივერსალურია, და ნაკლებად ეყრდნობა კულტურულ სხვაობას.

ჩვენი კვლევის ფარგლებში მოპოვებული ქართული ენის მონაცემები დეტალურად შევადარეთ აფრიკაანსისა და ჩრდილოეთ სოთოს შედეგებს. აღმოჩნდა, რომ ტალიარდისა და ბოსმანის მიერ გამოყოფილი თექვსმეტი მეტაფორული კლასიდან ქართული ენის მაგალითები დაემთხვა თოთხმეტ მეტაფორას, მაგალითად, “TORMENT (PHYSICAL AND PSYCHOLOGICAL) IS EATING” გვხვდება ქართულშიც – ჭამის აქტი როგორც სულიერი ტკივილი („ტკივილმა გამოუხრა გული”) და როგორც ფიზიკური ტანჯვა („ავადმყოფობამ შეჭამა“). ასევე, HUMILIATION IS EATING – დამცირება, შერცხვენა დასტურდება ქართულსა და აფრიკაანსში, ორივე ენაში დამცირება, შერცხვენა რეალიზებულია *ყლაპვა* და *ჭამა* ზმნებით, მაგალითად, „სირცხვილი ჭამა“, „სირცხვილი გადაყლაპა“.

ქართულ ენაში გამოვლინდა ჭამის სემანტიკასთან დაკავშირებული ისეთი მეტაფორები, რომლებიც არ დასტურდება ჩრდილოეთ სოთოსა და აფრიკაანსის ენებში. ასეთია, მაგალითად, „რაიმეს ამოწურვა, დასრულება და განლევა“ („არსებულმა სტრუქტურამ თავისი დრო მოჭამა“). ასეთია ასევე „ვინმეს შეწუხება, ვინმესთვის თავის მობეზრება” („ტვინი შემიჭამა“, „ნუ შემჭამე, ქალო!“)

ჩატარებული ანალიზი ასაბუთებს, რომ შესაძლოა, მეტაფორები კულტურულად სპეციფიკური იყოს, მაგრამ, რეალურად, ისინი მეტწილად უნივერსალურია, რადგან, როგორც ლაკოფის თეორიის მესამე პოსტულატი მიუთითებს, მეტაფორა იქმნება გამოცდილებით, ენა ყალიბდება ადამიანის გამოცდილებით, შედეგად, ერთი და იგივე მეტაფორა რეალიზდება სამ ერთმანეთისგან სრულიად განსხვავებულ ენაში (ჩრდილოეთ სოთო, აფრიკაანსი, ქართული). ჩვენ მიერ ჩატარებულმა კვლევამაც აჩვენა ენებს შორის კონცეპტუალური მეტაფორის კროს-ლინგვისტური თანხვედრა.

References :

Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford : Oxford University Press.

Lakoff, G. and M. Johnson. (1980). The metaphorical structure of the human conceptual system. *Cognitive science* 4, no. 2 : 195-208.

Taljard, E. and N. Bosman (2014). The Semantics of Eating in Afrikaans and Northern Sotho: Cross linguistic Variation in Metaphor. *Metaphor and Symbol,* 29 : 3, 224-245. DOI :10.1080/10926488.2014.924306.

Strategies for the Construction of a Learner's Dictionary of Verbal Locutions in Electronic Support: a New Lexicographic Solution for the Teaching of Spanish as a Foreign and Second Language

Maroto Bueno, Ana
University of Valladolid, Spain
ana.maroto@uva.es

This study is framed in the line of lexicography and the Spanish language and has been carried out following a qualitative methodology. The main objective of the study is to establish the theoretical and methodological bases for the construction of an effective tool for the teaching-learning of Spanish as a foreign or second language. The proposal is a learning dictionary of verbal locutions oriented to the production of the Spanish language. Due to the scarcity of resources of this type and the problem posed by verbal locutions, a new theoretical-practical study is presented which aims to offer a new lexicographical tool in electronic format for non-native speakers of Spanish who are in the process of learning. For this reason, an attempt is made to delimit the notion of locution, given the lack of fixation and conceptual rigour of this concept. From a diachronic point of view, although also synchronic, it can be seen how the lexicographical tradition has to a certain extent neglected this grammatical subcategory, perpetuating asystematicity in successive lexicographical repertoires. In addition, as part of the lexical wealth of Spanish, verbal locutions are of great importance in the common language and, at the same time, in the teaching-learning of Spanish as an LE/L2. For this reason, a study with a new perspective and lexicographic prospective on lexicographic technique, now motivated by the inclusion of computers and new technologies, and on the treatment of verbal locutions in dictionaries, has been formed.

Methodology: Based on a qualitative methodology, the most frequent verbal locutions in the Spanish language have been extracted and analysed in different lexicographical repertoires in order to observe the treatment they have received in lexicographical terms. Likewise, the definitional typology and the conceptual and grammatical rigour they present have been analysed, as well as the marking, if any. After the analysis which forms the core of the research, the parameters have been established for an adequate treatment of the locution and, more specifically, of the verbal locution. In order to solve the problem of how to fix this notion, an attempt at delimitation and a set of methodological bases for the lexicographical treatment of this grammatical subcategory have been offered. Subsequently, a proposal has been created for the elaboration of a learning dictionary in electronic support (with the innovation that the inclusion of computers and the application of the office automation possibilities that this offers) for the teaching-learning of verbal locutions.

Results and discussion: After a systematic review of different types of dictionaries, such as the normative academic dictionary *DLE* (*Diccionario de la Lengua Española*) (23rd ed.) of the RAE, the descriptive dictionary *DEA* (*Diccionario*

del Español Actual) by Manuel Seco *et al*, *DUE* (*Diccionario de uso del español*) by María Moliner and several learner's dictionaries, such as *DIPELE* (*Diccionario para la Enseñanza de la Lengua Española*), *CLAVE* (*Diccionario de Uso del Español Actual*), *SALAMANCA* (*Diccionario Salamanca de la lengua española*) or *DELE* (*Diccionario para la enseñanza de la lengua española*), it has been observed not only a certain asystematicity in the same dictionary, but also a lack of consensus in the treatment of verbal locutions both in didactic and lexicographic terms. The consideration of what is or is not a locution marks the beginning of this problem and the inadequacy of certain definitions for some of these pluriverbal units shows the lack of optimal repertoires for the teaching and learning of verbal locutions. Nowadays, computer technology means that space is no longer a problem in lexicographic work, and this makes it possible for definitions to be more precise, detailed and complete. Making use of hypertext links and hyperlinked redirects improves the presentation of the lexicographical article and makes the difference between printed dictionaries and the new ones in electronic format. Last but not least, it has been shown that the frequency of use of verbal locutions makes them indispensable in the Spanish language and, of course, legitimises their consideration as a cornerstone of the lexical wealth of any non-native user who wishes to learn Spanish as a second or foreign language. Thus, with this new lexicographical proposal, future learners of Spanish will go from "drawing a blank" ('*quedarse en blanco*') to "hitting the nail on the head" ('*dar en el clavo*').

ზმნური გამოთქმების ელექტრონული სასწავლო ლექსიკონის შედგენის სტრატეგიები: ახალი ლექსიკოგრაფიული გადაწყვეტა ესპანურის, როგორც უცხო და მეორე ენის სწავლებისთვის

ანა მაროტო ბუენო

ვალადოლიდის უნივერსიტეტი, ესპანეთი

ana.maroto@uva.es

ეს კვლევა განხორციელდა ლექსიკოგრაფიისა და ესპანური ენის სწავლების მიმართულებით თვისებრივი კვლევის მეთოდოლოგიის გამოყენებით. კვლევის ძირითადი მიზანია თეორიული და მეთოდოლოგიური საფუძვლების განსაზღვრა ესპანური ენის, როგორც მეორე ან უცხო ენის, სწავლა-სწავლების ეფექტური ინსტრუმენტის შესაქმნელად. მოხსენებაში წარმოდგენილია ზმნური გამოთქმების სასწავლო ლექსიკონი, რომელიც ორიენტირებულია ესპანურ ენაზე ტექსტების სინთეზზე. არსებული რესურსების სიმწირისა და ზმნური გამოთქმების მიერ შექმნილი სირთულეების გათვალისწინებით, შემოთავაზებულია ახალი თეორიულ-პრაქტიკული ჩარჩო, რომლის მიზანია შექმნას ეფექტური ელექტრონული ლექსიკოგრაფიული ინსტრუმენტი ესპანური ენის არაესპანელი შემსწავლელებისათვის. უპირველეს ყოვლისა, დაზუსტებულია გამოთქმის, ზმნური გამოთქმის განმარტება, რადგან არსებულ დეფინიციებს აკლია კონცეპტუალური სიზუსტე. დიაქრონიული, თუმცა სინქრონიული თვალსაზრისითაც, იკვეთება, თუ როგორ უგულებელყო ლექსიკოგრაფიულმა ტრადიციამ ეს გრამატიკული ქვეკატეგორია, რამაც განაპირობა მისი არათანამიმდევრული წარმოდგენა ლექსიკოგრაფიულ წყაროებში. ამასთან ერთად აღსანიშნავია, რომ როგორც ესპანური ენის ლექსიკური სიმდიდრის ნაწილს, ზმნურ გამოთქმებს დიდი მნიშვნელობა აქვს როგორც საერთო ენისათვის, ისე ესპანურის, როგორც მეორე და უცხო ენის, სწავლა-სწავლების პროცესისათვის. ამიტომ შეიქმნა ახალი ლექსიკოგრაფიული ხედვის მქონე კვლევა, რომელიც ეფუძნება როგორც კომპიუტერულ და თანამედროვე ტექნოლოგიებს, ისე ზმნური გამოთქმების დამუშავების ახლებურ ლექსიკოგრაფიულ პრინციპებს.

მეთოდოლოგია: კვლევა დაეფუძნა თვისებრივ მეთოდებს და მოიცავდა ესპანური ენის ყველაზე ხშირად გამოყენებული ზმნური გამოთქმების მოძიებასა და ანალიზს სხვადასხვა ლექსიკოგრაფიულ წყაროში, რათა შეფასებულიყო, თუ როგორ არის ისინი წარმოდგენილი ლექსიკოგრაფიულ პრაქტიკაში. ასევე გაანალიზდა და დაზუსტდა ზმნური გამოთქმების განმარტება, დადგინდა პარამეტრები გამოთქმის, უფრო კონკრეტულად კი, ზმნური გამოთქმის ადეკვატური დამუშავებისთვის, დადგინდა მისი საზღვრები და დამუშავდა ლექსიკოგრაფიული მეთოდოლოგიური საფუძვლები ამ გრამატიკული ქვეკატეგორიისთვის. მოგვიანებით, მომზადდა კონცეფცია ზმნური გამოთქმების ელექტრონული

სასწავლო ლექსიკონის შესაქმნელად, რაც მიზნად ისახავს ზმნური გამოთქმების სწავლა-სწავლების პროცესის გაუმჯობესებას.

შედეგები და დისკუსია: დამუშავდა შემდეგი ლექსიკონები : RAE-ს *ნორმატიული აკადემიური ლექსიკონი DLE* (*Diccionario de la Lengua Española*, 23-ე გამოცემა); მანუელ სეკოსა და მისი კოლეგების მიერ შედგენილი *აღწერითი ლექსიკონი DEA* (*Diccionario del Español Actual*); მარია მოლინერის *DUE* (*Diccionario de uso del español*) და რამდენიმე სასწავლო ლექსიკონი, როგორებიცაა : *DIPELE* (*Diccionario para la Enseñanza de la Lengua Española*), *CLAVE* (*Diccionario de Uso del Español Actual*), *SALAMANCA* (*Diccionario Salamanca de la lengua española*) და *DELE* (*Diccionario para la enseñanza de la lengua española*). ლექსიკონების შესწავლამ გამოავლინა შეუთანხმებლობა და არათანამიმდევრულობა ზმნური გამოთქმების წარმოდგენის პრინციპებში არა მხოლოდ სხვადასხვა ლექსიკონებში, არამედ ერთი ლექსიკონის ფარგლებშიც. ზმნური გამოთქმების რაობის განსაზღვრა და დაუზუსტებელი განმარტება, თავის როლს თამაშობს ამ ვითარებაში. შექმნილი მდგომარეობა მეტყველებს იმაზე, რომ არ არსებობს ოპტიმალური რესურსი ზმნური გამოთქმების სწავლა-სწავლებისათვის. თანამედროვე კომპიუტერული ტექნოლოგიების გამოყენება ლექსიკოგრაფიულ მუშაობაში ხსნის სივრცის პრობლემას, რის შედეგადაც შესაძლებელია განმარტებების უფრო ზუსტი, დეტალური და სრულყოფილი ფორმულირება. ჰიპერტექსტური ბმულებისა და გადამისამართებების გამოყენება აუმჯობესებს ლექსიკოგრაფიული სტატიების წარმოდგენას და ქმნის განსხვავებას ბეჭდურ და ელექტრონულ ფორმატებს შორის. და ბოლოს, დადასტურდა, რომ ზმნური გამოთქმების გამოყენების სიხშირე ესპანურ ენაში მათ შეუცვლელს ხდის და ამართლებს აღქმას, რომ ზმნური გამოთქმები ესპანური ენის ლექსიკური სიმდიდრის საფუძველია. ამიტომ ენიჭება მას ასეთი მნიშვნელობა ესპანურის, როგორც მეორე ან უცხო ენის, შესწავლის პროცესში. ამ ახალი ლექსიკოგრაფიული პროექტით, ესპანურის მომავალი შემსწავლელები გაივლიან გზას „წარუმატებლობის განცდიდან“ (*quedarse en blanco*) „ზუსტად მიზანში მოხვედრამდე“ (*dar en el clavo*).

Harnessing AI for Creating Definitions of Georgian Military Terminological Neologisms in the context of the Russia–Ukraine War

Mchedlishvili, Ketevani
Ilia State University, Georgia
ketevani.mchedlishvili.3@iliauni.edu.ge

Each new armed conflict is followed by rapid technological advancements, thus leading to the coinage of new terms or the resurgence of previously unrecorded terminological units. The outbreak of the Russian-Ukrainian war, which has already been assessed as new generation warfare by military experts (Sorge 2023), caused a rapid influx of drone-related terminology into the Georgian military language. The preliminary analysis of the Georgian military term დრონი/ *droni* in existing Georgian military dictionaries (2009, 2017) and Georgian general language corpora revealed its absence in lexicographic resources and its recent emergence in the Georgian general language corpora.[24] This fact makes drone-related terminology candidates of Georgian Neoterms (Cabre 1999: 205; ISO 1087:2019).

Given the absence of drone-related terms in Georgian military terminology and the importance of the ongoing conflict for Georgia in the framework of NATO partnership, it is crucial to analyse, define and document these terms in the updated versions of military dictionaries. One of the vital components for compiling terminological entries is creating appropriate definitions. Growing interest towards Large Language Models and the usage of Generative AI for various lexicographic and terminological tasks (Lew 2023; Wild 2024; Sabanés and Cunha 2025), facilitated our research interest in exploring the potential of AI in creating draft definitions in the Georgian language, which is considered to be an under-resourced language.

For further analysis, we selected 10 drone-related Georgian military multiword terms alongside their English counterparts. These terms were extracted from English and Georgian military comparable corpora (MiliCorpEng, MiliCorpKa) previously compiled within the framework of PhD research. The main criteria for selecting terms were their absence from existing Georgian military dictionaries and their recent emergence (since 2014). For this purpose, we used two Georgian General Language corpora: KaWak 2013[25], which includes texts up to 2013, and Aranea, a Georgian comparable General language corpus that consists of texts crawled from 2014 onwards, including the latest 2023-2024 crawl covering the Russian-Ukraine war. The absence from the dictionaries, coupled with the presence of terminological units in Aranea, was considered a strong indicator of terminological neologisms.

To evaluate the effectiveness of AI-assisted terminological definitions, we conducted a comparative analysis produced by several large language models

24 https://corpora.uni-leipzig.de/en/res?corpusId=kat_news_2020&word=დრონი, http://aranea.juls.savba.sk/aranea/run.cgi/first?corpname=AranGeor&reload=1

25 https://cqpweb.lancs.ac.uk/kawac200mw

that support the Georgian language: ChatGPT[26] (OpenAI), Gemini[27] (Google DeepMind), Claude[28] (Anthropic), and DeepSeek[29]. In this paper, we present our findings from experiments with various prompts and templates, and explore the outputs of different LLM models, as well as the role of human editors in refining and reviewing given draft definitions.

This study contributes to the growing field of electronic lexicography and enhances terminologists' work by utilising Generative AI for under-resourced languages, such as Georgian, and specialised terminology. We anticipate that the research findings will contribute to the use of large language models in assisting with various terminological tasks, particularly in creating draft definitions in the Georgian language.

Keywords: *LLM, AI, Georgian military neoterms, creating draft definitions, Russia-Ukraine war*

26 https://chatgpt.com

27 https://gemini.google.com/app

28 https://claude.ai

29 https://www.deepseek.com

ხელოვნური ინტელექტის გამოყენება ქართული სამხედრო ტერმინოლოგიური ნეოლოგიზმების დეფინიციათა შემუშავებისთვის (რუსეთ-უკრაინის ომის კონტექსტში)

მჭედლიშვილი ქეთევანი

ილიას სახელმწიფო უნივერსიტეტი, საქართველო

ketevani.mchedlishvili.3@iliauni.edu.ge

ყოველ ახალ შეიარაღებულ კონფლიქტს თან ახლავს ახალი ტექნოლოგიების განვითარება, რაც თავის მხრივ, იწვევს ენაში ახალი ტერმინოლოგიური ერთეულების შექმნას ან მანამდე არსებული ტერმინების გააქტიურებას. რუსეთ-უკრაინის ომმა, რომელიც სამხედრო ექსპერტებმა უკვე შეაფასეს, როგორც ახალი თაობის საომარი მოქმედებები (Sorge 2023), გამოიწვია დრონებთან დაკავშირებული ტერმინების სწრაფი შემოდინება ქართულ სამხედრო დარგობრივ ენაში. ქართული ტერმინის *დრონი* პირველადმა ანალიზმა გამოავლინა, რომ აღნიშნული ტერმინი არ დასტურდება არცერთ არსებულ ქართულ სამხედრო ლექსიკონებში (2010, 2017). გარდა ამისა, ქართული საერთო ენის კორპუსების ანალიზმა გვიჩვენა, რომ აღნიშნული ტერმინი ენაში ამ ბოლო წლებში გაჩნდა[30]. აღნიშნული ფაქტი კი მეტყველებს ქართული ენისათვის დრონებთან დაკავშირებული ტერმინების ნეონიმურ ე.ი ტერმინოლოგიური ნეოლოგიზმების სტატუსზე (Cabre 1999: 205; ISO 1087:2019).

იმის გათვალისწინებით, რომ დრონებთან დაკავშირებული ტერმინოლოგია არ დასტურდება ამჟამად არსებულ ქართულ სამხედრო ლექსიკონებში და აგრეთვე, მიმდინარე შეიარაღებული კონფლიქტის მნიშვნელობიდან გამომდინარე საქართველოსთვის, როგორც ნატოს სტრატეგიული პარტნიორისთვის, აუცილებელი ხდება ამ ტერმინების ანალიზი, განმარტება და ლექსიკონების განახლებულ ვერსიებში მათი დოკუმენტირება. ერთ-ერთი მნიშვნელოვანი კომპონენტი ტერმინოლოგიური სიტყვა-სტატიების შედგენის პროცესში სწორი დეფინიციების შემუშავებაა. აღსანიშნავია, რომ ამ მიმართულებით მზარდია ინტერესი დიდი ენობრივი მოდელებისა და ხელოვნური ინტელექტის გამოყენებისადმი სხვადასხვა ლექსიკოგრაფიული თუ ტერმინოლოგიური ამოცანების გადასაჭრელად (Lew 2023; Wild 2024; Sabanés and Cunha 2025). წინამდებარე ფაქტმა გამოიწვია ჩვენი კვლევითი ინტერესი ხელოვნური ინტელექტის გამოყენებისადმი ქართული სამხედრო ტერმინების დეფინიციათა შესადგენად.

შემდგომი ანალიზისათვის კორპუსიდან შევარჩიეთ 10 დრონთან დაკავშირებული ქართული სამხედრო მრავალსიტყვიანი ტერმინი ინგლისურ ეკვივალენტებთან ერთად. წინამდებარე ტერმინები ამოვიღეთ რუსეთ-უკრაინის ომის კონტექსტში შექმნილი სამხედრო ანა-

30 https://corpora.uni-leipzig.de/en/res?corpusId=kat_news_2020&word=დრონი, http://aranea.juls.savba.sk/aranea/run.cgi/first?corpname=AranGeor&reload=1

ლიტიკური და საინფორმაციო ბლოგების საფუძველზე შედგენილი ინგლისური და ქართული დარგობრივი შედარებითი კორპუსებიდან (MiliCorpEng, MiliCorpKa), რომელიც სადოქტორო კვლევის ფარგლებში შეიქმნა. ტერმინების შერჩევის ძირითად კრიტერიუმებს წარმოადგენდა მათი არარსებობა არსებულ სამხედრო ლექსიკონებში და ბოლოდროინდელი გამოყენება კორპუსებში (2014 წლიდან). ამ მიზნით გამოვიყენეთ ქართული საერთო ენის ორი ვებკორპუსი: KaWak 2013,[31] რომელიც მოიცავს ინტერნეტიდან შეგროვებულ ტექსტებს 2013 წლამდე და „არანეას“ ქართული საერთო ენის შედარებითი კორპუსი, რომელიც შედგება 2014 წლიდან შეგროვებული ინტერნეტტექსტებისაგან და მათ შორის, მოიცავს უახლეს 2023-2024 წლების ტექსტებს რუსეთ-უკრაინის ომის თემატიკაზე. შესაბამისად, ტერმინოლოგიური ნეოლოგიზმების შერჩევის მთავარ კრიტერიუმებად განვსაზღვრეთ მათი არარსებობა ლექსიკონში და ბოლოდროინდელი გამოყენება „არანეას“ ქართული ენის კორპუსში.

იმისათვის, რომ შეგვეფასებინა ხელოვნური ინტელექტის დახმარებით შექმნილი დეფინიციების ეფექტიანობა, ჩავატარეთ შედარებითი ანალიზი რამდენიმე ცნობილი დიდი ენობრივი მოდელის გამოყენებით, როგორიცაა ChatGPT[32] (OpenAI), Gemini[33] (Google DeepMind), Claude[34] (Anthropic), and DeepSeek[35]. წინამდებარე მოხსენებაში, ჩვენ წარმოვადგენთ ხელოვნური ინტელექტისთვის მიცემული სხვადასხვა დავალების „პრომპტის“ ექსპერიმენტის შედეგებს და შევაფასებთ სხვადასხვა მოდელის მიერ შემოთავაზებულ პასუხებს. აგრეთვე, გავაანალიზებთ ტერმინოლოგის მიერ მიღებული პასუხების რედაქტირებისა და მიმოხილვის პროცესის მასშტაბებსა და მნიშვნელობას.

წინამდებარე კვლევა აქტუალურია ელეტრონულ ლექსიკოგრაფიაში ტერმინოლოგების მუშაობისათვის ხელოვნური ინტელექტის გამოყენების პოტენციალის კვლევის კუთხით, განსაკუთრებით ისეთი მწირი რესურსების მქონე ენისათვის, როგორიც ქართულია. მოველით, რომ კვლევის შედეგები სამომავლოდ შესაძლებელია გამოყენებულ იქნეს ქართულ დარგობრივ ლექსიკოგრაფიულ მუშაობაში სხვადასხვა ტერმინოლოგიური ამოცანის გადასაჭრელად, მათ შორის, ქართულ ენაზე ტერმინოლოგიური დეფინიციების შემუშავებისათვის.

საკვანძო სიტყვები: *დიდი ენობრივი მოდელები, ხელოვნური ინტელექტი, ქართული სამხედრო ტერმინოლოგიური ნეოლოგიზმები, დეფინიციების შემუშავება, რუსეთ-უკრაინის ომი*

31 https://cqpweb.lancs.ac.uk/kawac200mw

32 https://chatgpt.com

33 https://gemini.google.com/app

34 https://claude.ai

35 https://www.deepseek.com

References:

Cabré, M.T. (1999). *Terminology: Theory, methods and applications*. Amsterdam/ Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/tlrp.1.

English-Georgian Military Dictionary. https://mil.dict.ge/en/ [11 April 2025]

ISO 1087: 2019. (2019). *Terminology Work and Terminology Science – Vocabulary*. Geneva: ISO

ISO 704:2022. *Terminology work — Principles and methods*. Geneva: ISO

Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities and Social Science Communications*, 10(1), 704. https://doi.org/10.1057/s41599-023-02119-6

Medzmariashvili, E. (Ed.). 2017. *Georgian Military Encyclopedic Dictionary*. Tbilisi. (In Georgian)

Military-Explanatory Dictionary and Graphic Symbols. FM (Field Manual) 1-02. (2017). Tbilisi: The Ministry of Defense of Georgia, Command of Training and Military Education, Doctrine Development Center. (In Georgian)

Sabanés, L.V. and C. da Iria (2025). AI as a resource for the clarification of medical terminology. An analysis of its advantages and limitations. *Terminology, Computational terminology*, 31(1), 37-71. https://doi.org/10.1075/term.00083.vid

Sorge, H. (2023). War in Ukraine: Unleashing Innovation and Unseen Frontiers. *Policy Center of the New South*. https://www.policycenter.ma/publications/war-ukraine-unleashing-innovation-and-unseen-frontiers (Accessed: 01.07.2025)

Wild, K. (2024). Using Large Language Models for a Large Historical Dictionary Challenges and Opportunities for the Oxford English Dictionary. *XXI EURALEX International Congress, book of abstracts*, 225-227. https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralex_boa_20.pdf

Past Voices, Present Tools: Citizen Science and the Hesse-Nassau Dictionary

Mederake, Nathalie; Engsterhold, Robert

Marburg University, Germany

mederake@uni-marburg.de
robert.engsterhold@uni-marburg.de

The creation of dictionaries has traditionally been the responsibility not only of professional lexicographers but also of individual scholars and members of the public. There is a long and productive history of involving citizens in linguistic research – especially in dialect documentation. Their contributions have often played a crucial role in documenting regional variation and local speech forms that would otherwise be difficult to obtain. When it comes to dialect dictionary projects, laypeople have frequently served as valuable local experts, providing empirical data on behalf of researchers. As a result of broader shifts in scientific practice toward participatory research models, the number of initiatives involving cooperation between professional dialectologists and citizen scientists has increased significantly in recent decades.

Using the Hesse-Nassau Dictionary (HNWb) as a case study, this presentation explores the changing role of citizen scientists in the compilation and evaluation of dialect dictionaries. The HNWb is a large-scale historical dictionary that draws on linguistic material from three major German dialect regions: Low German, High German, and Upper German. The project's archive includes over 6,500 completed questionnaires distributed between 1914 and 1926, along with more than 300,000 lemmatized paper slips containing approximately 1.5 million individual dialectal references. Numerous topics are covered in the questionnaire responses, such as family and everyday life, traditional customs, regional flora and fauna, and local knowledge. The surveys also include semasiological and onomasiological tasks, such as naming parts of illustrated objects (e.g., a loom), and listing local expressions for specific concepts.

As part of a broader preservation effort, the entirety of the archive has now been digitized, even though only a selection of the collected material was originally transcribed and used in the dictionary. However, digitization in itself does not make the data fully accessible or usable for modern linguistic analysis. To bridge this gap, two digital platforms have been developed to enable active participation in processing and enriching the data: (1) a questionnaire app designed as a citizen science tool that allows users to transcribe historical dialect data (https://apps.dsa.info/hnwb/), and (2) a document editor based on a PostgreSQL database with a FastAPI framework which makes is easier to annotate the paper slips in detail, , including metadata such as lemma, meaning, document location, and contributor identity.

These tools not only improve access to the data but also represent a conceptual shift in the role of the citizen participant – from research subject to active collaborator. Citizen scientists are no longer seen merely as informants, but as contributors who enrich the lexicographic process and co-create linguistic knowledge. This shift reflects a growing recognition of the knowledge value of

regional expertise and the potential for non-professionals to play a significant role in scientific inquiry.

In conclusion, the Hesse-Nassau Dictionary project offers valuable insights about how historical data can be recontextualized and revitalized through modern digital methods and participatory practices. It emphasizes the importance of citizen engagement in both the preservation and further development of linguistic heritage and underscores the need to further explore questions of data quality, training, and collaborative infrastructure in citizen science-driven lexicography.

წარსულის ხმები, თანამედროვე ინსტრუმენტები: სამოქალაქო მეცნიერება[36] და ჰესენ-ნასაუს[37] ლექსიკონი

ნატალი მედერაკე, რობერტ ენგსტერჰოლდი
მარბურგის უნივერსიტეტი, გერმანია
mederake@uni-marburg.de
robert.engsterhold@uni-marburg.de

ლექსიკონების შექმნა ტრადიციულად არა მხოლოდ პროფესიონალი ლექსიკოგრაფების, არამედ აგრეთვე ცალკეული მეცნიერებისა და საზოგადოების წევრების პასუხისმგებლობაც იყო. მოქალაქეთა ჩართულობას ლინგვისტურ კვლევაში, განსაკუთრებით კი დიალექტების დოკუმენტირებაში, ხანგრძლივი და ნაყოფიერი ისტორია აქვს. მათ წვლილს ხშირად უთამაშნია გადამწყვეტი როლი ისეთი რეგიონული ვარიაციებისა და მეტყველების ადგილობრივი ფორმების დოკუმენტირებაში, რომელთა მოპოვებაც სხვა შემთხვევაში რთული იქნებოდა. როდესაც საქმე დიალექტური ლექსიკონების პროექტებს ეხება ხოლმე, არასპეციალისტები ხშირად ასრულებენ ფასეული ადგილობრივი ექსპერტების როლს და მკვლევრებს ემპირიულ მონაცემებს აწვდიან. შედეგად იმისა, რომ სამეცნიერო პრაქტიკაში სულ უფრო ხშირად გამოიყენება კვლევითი მოდელები საზოგადოების ჩართულობით, ბოლო ათწლეულებში მნიშვნელოვნადაა გაზრდილი ისეთ ინიციატივათა რიცხვი, რომლებიც პროფესიონალ დიალექტოლოგთა და სამოქალაქო მეცნიერთა შორის თანამშრომლობას გულისხმობს.

წინამდებარე პრეზენტაცია, რომელიც იყენებს ჰესენ-ნასაუს ლექსიკონს (HNWb[38]) როგორც ცალკეული შემთხვევის ანალიზს, დიალექტური ლექსიკონების შექმნასა და შეფასებაში სამოქალაქო მეცნიერთა ცვალებად როლს იკვლევს. HNWb ფართომასშტაბიანი ისტორიული ლექსიკონია, რომელშიც გამოყენებულია ენობრივი მასალები გერმანიის სამი ძირითადი დიალექტური რეგიონიდან: ქვემოგერმანულიდან, ზემოგერმანულიდან და სამხრეთგერმანულიდან.[39] პროექტის არქივი მოიცავს 6500-ზე მეტ შევსებულ ანკეტას, რომლებიც მოსახლეობას 1914-1926 წლებში დაურიგდა, და აგრეთვე 300 000-ზე მეტ ლემატიზებულ ბარათს, რომლებშიც დაახლოებით 1,5 მილიონი ცალკეული დიალექტური ჩანაწერია შესული. ანკეტის პასუხები შეეხება მრავალ სხვადასხვა თემას,

36 სამოქალაქო მეცნიერება არის კვლევა, რომელსაც აწარმოებენ ადამიანები, რომლებიც არ არიან პროფესიონალი მეცნიერები." დღეისათვის ხშირად იხმარება მოხალისეების, არაპროფესიონალების და ა. შ. სამეცნიერო საქმიანობის აღსანიშნად.

37 ჰესენ-ნასაუ (Provinz Hessen-Nassau) ადრე პრუსიის (Preussen) ნაწილი იყო; ახლა ჰესენის ფედერალურ მხარეში (Land Hessen) შედის.

38 **H**essen-**N**assauisches **W**örter**b**uch.

39 Low German (Niederdeutsch) ქვემოგერმანული; High German (Hochdeutsch) ზემოგერმანული;
Upper German (Oberdeutsch) სამხრეთგერმანული (*ზემოგერმანულის ნაწილია*).

როგორებიცაა ოჯახი და ყოველდღიური ცხოვრება, ტრადიციული ადათ-წესები, რეგიონული ფლორა და ფაუნა და ადგილობრივი ცოდნა. გამოკითხვებში ასევე შედის სემასიოლოგიური და ონომასიოლოგიური დავალებები, როგორიცაა ილუსტრირებული საგნების (მაგ. საქსოვი დაზგის) ნაწილების დასახელება და კონკრეტულ ცნებათა აღმნიშვნელი ადგილობრივი გამოთქმების ჩამოთვლა.

საარქივო მასალის შენარჩუნებისაკენ მიმართული უფრო ფართო ძალისხმევის ფარგლებში, ამჟამად უკვე მთელი არქივია გაციფრებული, და ეს მიუხედავად იმისა, რომ თავდაპირველად შეგროვებული მასალის მხოლოდ ნაწილი იყო ტრანსკრიბირებული და ლექსიკონში გამოყენებული. თუმცა, მარტოოდენ გაციფრების წყალობით მონაცემები თანამედროვე ლინგვისტური ანალიზისთვის სრულად ხელმისაწვდომი ან გამოყენებადი ვერ იქნება. ამ ხარვეზის შესავსებად შემუშავდა ორი ციფრული პლატფორმა, რათა მათი დახმარებით შესაძლებელი გახდეს საზოგადოების აქტიური მონაწილეობა მონაცემთა დამუშავებასა და გამდიდრებაში: (1) ონლაინ-ანკეტის აპლიკაცია, შექმნილი როგორც სამოქალაქო მეცნიერების ინსტრუმენტი, რომელიც მომხმარებლებს ისტორიული დიალექტური მონაცემების ტრანსკრიბირების საშუალებას აძლევს (https://apps.dsa.info/hnwb/) და (2) PostgreSQL მონაცემთა ბაზაზე დაფუძნებული დოკუმენტის რედაქტორი, აღჭურვილი FastAPI ჩარჩოთი, რაც აადვილებს ბარათების დეტალურ ანოტირებას, ისეთი მეტამონაცემების ჩათვლით, როგორიცაა ლემა, მნიშვნელობა, დოკუმენტის შედგენის ლოკაცია და ჩანაწერის ავტორის ვინაობა.

ეს ინსტრუმენტები არა მხოლოდ აუმჯობესებს მონაცემებზე წვდომას, არამედ ასახავს სამოქალაქო მონაწილის როლის კონცეპტუალურ ცვლილებას – კვლევის საგნიდან აქტიურ თანაავტორამდე. სამოქალაქო მეცნიერები უკვე აღარ აღიქმებიან მარტოოდენ ინფორმანტებად, არამედ მოიაზრებიან აქტიური წვლილის შემტან პირებად, რომლებიც ლექსიკოგრაფიულ პროცესს ამდიდრებენ და ენობრივი ცოდნის თანაშემოქმედნი ხდებიან. აღნიშნული ცვლილება ასახავს ცოდნის გამდიდრების საქმეში რეგიონული გამოცდილების წვლილის სულ უფრო მზარდ აღიარებას და არაპროფესიონალთა პოტენციალს, მნიშვნელოვანი როლი ითამაშონ სამეცნიერო კვლევაში.

დასასრულ, ჰესენ-ნასაუს ლექსიკონის პროექტი ღირებულ ინფორმაციას გვაწვდის იმის შესახებ, თუ როგორ შეიძლება ისტორიულ მონაცემთა რეკონტექსტუალიზაცია და რევიტალიზაცია თანამედროვე ციფრული მეთოდებისა და სამონაწილეო პრაქტიკების გამოყენებით. იგი განსაკუთრებით გამოყოფს მოქალაქეთა ჩართულობის მნიშვნელობას ენობრივი მემკვიდრეობის როგორც შენარჩუნებაში, ასევე შემდგომ განვითარებაში და ხაზს უსვამს აუცილებლობას, კიდევ უფრო ღრმად იქნეს შესწავლილი ისეთი საკითხები, როგორიცაა მონაცემთა ხარისხი, წვრთნა და თანამშრომლობის ინფრასტრუქტურა სამოქალაქო მეცნიერებაზე დაფუძნებულ ლექსიკოგრაფიაში.

თეზისი ქართულად თარგმნა გიორგი მელაძემ.

References:

Dorn A., Stocker R. and P. Stöckle (2024). Old dialect words through the ages – the ABCs of dialect project. ARPHA Proceedings 6: 213-216. https://doi.org/10.3897/ap.e126110.

Fischer, H. (2024). Dialektologie als Citizen Science. Perspektiven und Tools für eine bürgernahe Wissenschaft. In: Ferstl, Christian (et al.): Dialekt · unterwegs. Varietäten im Zeichen von Globalisierung und Migration. Beiträge zum 8. Dialektologischen Symposium im Bayerischen. Tirschenreuth.

Heinisch, B. (2021). Reaching the limits of co-creation in citizen science — exemplified by the linguistic citizen humanities project 'On everyone's mind and lips — German in Austria' *JCOM* 20(06), A05. https://doi.org/10.22323/2.20060205.

Formation of Analyticity in Language. Degradation or Development?
(On the example of certain trends in the modern Georgian language)

Meladze, Giorgi
Ilia State University, Georgia
giorgi.meladze.4@iliauni.edu.ge

The issue of the development of languages and their transformation from the synthetic, morphologically rich and multiform type to the analytic one, which may appear morphologically "simplified", has always been extremely controversial in linguistic circles.

The representatives of the romantic trend in linguistics, such as Jacob and Wilhelm Grimms, Johann G. Herder and others held that the diachronic development of the Old Indo-European and Germanic languages and the process of their transformation into modern languages constituted in fact their decay and degradation. They saw the loss of the variety and richness of the grammatical forms, declension, and conjugation paradigms by these languages as a clear example of such degradation. If we compare the Anglo-Saxon (Old English) language with the modern English, or Latin with Italian, Old High German with the modern German, or Gothic (almost identical to Proto-Germanic) with, say, the modern Swedish language, we will clearly see what this approach of the Romanticists was based upon.

Otto Jespersen, a well-known Danish linguist on the other hand believed that the diachronic processes and changes occurring in languages represent in fact their improvement and progress and that the said processes cause the constant increase in the efficiency of language. It is believed that his views on the subject were also influenced by Charles Darwin's evolutionary theories. Chapter XVII of Jespersen's masterpiece "Language: Its Nature, Development, and Origin" is titled "Progress or Decay?". The title of our proposed paper was inspired by this very chapter of Jespersen's book and its general spirit.

The purpose of our paper itself is to discuss analytic constructions which have become quite abundant in the modern Georgian language, although the traditional Georgian linguistics either ignores them, or regards them as the violation of the set norms of the Georgian language, as *calques* or mistakes, trying hard to eradicate them. In our paper we will try to analyse, what linguistic or communicative circumstances necessitate the appearance and spread of such analytic constructions in languages generally and in the Georgian language in particular; we will also express our humble opinion as to whether such analytic trends in diachronic linguistic change constitute indeed the decay and degradation of languages, or, on the contrary, such development only make language more efficient and expressive and are caused by the necessity to clearly and understandably express one's thoughts and ideas verbally.

ენაში ანალიტიზმის წარმოქმნის მიზეზები. დეგრადაცია თუ განვითარება?
(თანამედროვე ქართული ენისთვის დამახასიათებელი ზოგიერთი ტენდენციის მაგალითზე)

გიორგი მელაძე
ილიას სახელმწიფო უნივერსიტეტი, საქართველო
giorgi.meladze.4@iliauni.edu.ge

ენების განვითარებაზე და სინთეზური, მორფოლოგიურად მდიდარი და მრავალფეროვანი ტიპიდან ანალიზურ, თითქოსდა მორფოლოგიურად „გამარტივებულ“ ტიპზე მათი გადასვლის შესახებ ენათმეცნიერებაში ოდითგანვე ურთიერთგამომრიცხავი შეხედულებები გამოითქმებოდა.

რომანტიკული მიმდინარეობის წარმომადგენლები: იაკობ და ვილჰელმ გრიმები, იოჰან გ. ჰერდერი და სხვანი, მიიჩნევდნენ, რომ ძველი ინდოევროპული და გერმანიკული ენების დიაქრონიული განვითარება და თანამედროვე ენებად მათი ჩამოყალიბების პროცესი არსებითად მათს გადაგვარებას და დეგრადაციას წარმოადგენდა. ამგვარი დეგრადაციის მკაფიო გამოხატულებად მათ ამ ენების გრამატიკული ფორმების, ბრუნებისა და უღლების მრავალფეროვნებისა და სიმდიდრის დაკარგვა ესახებოდათ. თუ ერთმანეთს შევადარებთ ანგლოსაქსურ (ძველინგლისურ) და თანამედროვე ინგლისურ, ლათინურსა და იტალიურ, ძველ ზემოგერმანულსა და თანამედროვე გერმანულ, ანდა პროტოგერმანიკულთან ძალზე ახლოს მყოფ გოთურსა და, ვთქვათ, თანამედროვე შვედურ ენებს, თვალნათლივ დავინახავთ, რას ეფუძნებოდა ლინგვისტიკის „რომანტიკოსთა“ ამგვარი მიდგომა.

მეორე მხრივ, ცნობილი დანიელი ენათმეცნიერი ოტო იესპერსენი მიიჩნევდა, რომ ენაში მიმდინარე დიაქრონიული პროცესები და ცვლილებები არსებითად მის დახვეწასა და პროგრესში მდგომარეობს და რომ ეს პროცესები ენის ეფექტიანობის უცილობელ ზრდას იწვევს. ითვლება, რომ მის ამგვარ შეხედულებებზე გარკვეული ზეგავლენა ჩარლზ დარვინისეულმა ევოლუციის თეორიამაც მოახდინა. იესპერსენის ფუნდამენტური ნაშრომის: „ენა, მისი ბუნება, განვითარება და წარმოშობა“ XVII თავს ასევე ეწოდება: „პროგრესი თუ გადაგვარება?“ (Progress or Decay?). სწორედ ამ თავმა და მისმა საერთო სულისკვეთებამ შთაგვაგონა, წინამდებარე ნაშრომი ამგვარად დაგვესათაურებინა.

თავად ჩვენი ნაშრომის მიზანია, განვიხილოთ ანალიზური კონსტრუქციები, რომლებიც საკმაოდაა მომრავლებული თანამედროვე ქართულში, თუმცა ტრადიციული ქართული ენათმეცნიერება მათ ან ვერ ამჩნევს, ან მათ ქართული ენის ნორმების დარღვევად, კალკებად და შეცდომებად მიიჩნევს და შეძლებისდაგვარად ებრძვის. სტატიაში შევეცდებით გავაანალიზოთ, რა ენობრივი თუ საკომუნიკაციო აუცილებლობა განაპირობებს ამგვარი ანალიზური კონსტრუქციების გაჩენასა და

გამრავლებას კონკრეტულად ქართულში და ზოგადად ენებში, როგორც ასეთში, და გამოვთქვამთ ჩვენს მოკრძალებულ მოსაზრებასაც იმის თაობაზე, ენაში ამგვარი პროცესების მიმდინარეობა და საზოგადოდ ასეთი (ანალიზური) მიმართულებით ენის განვითარება მის გადაგვარებას და დეგრადაციას იწვევს, თუ პირიქით, უფრო ქმედითსა და გამომსახველს ხდის მას და აზრის მკაფიოდ და გასაგებად გამოთქმის აუცილებლობითაა განპირობებული.

References

Rastorguyeva, T. A. (1983). A History of English. Moscow «Vysšaya Škola» 1983

Britannica. Otto Jespersen https://www.britannica.com/biography/Otto-Jespersen (07.10.2022)

Jespersen, O. *Language: its Nature, Development, and Origin* (Chapter XVII) https://www.gutenberg.org/ebooks/53038 (07.10.2022)

Jespersen, O. (1894). Progress in Language. London: S. Sonnenschein.

A Corpus as a Pedagogical Tool for Language Learning – Lemmatisation and Dictionary Lookup in the GNC and the AbNC

Meurer, Paul
University of Bergen Library, Norway
paul.meurer@uib.no

In my talk, I will present pedagogical tools for language learning available in the Georgian National Corpus (GNC) and in the Abkhaz National Corpus (AbNC). Both corpora contain fictional and non-fictional texts that can be searched, and, depending on licensing, also read on-screen. The texts are neatly paginated and their structural elements – such as chapter headings, paragraphs and line breaks in poetry – are kept. Registered users can also set bookmarks for later reference, making for a smooth and user-friendly reading experience.

Moreover, the texts in both corpora are enriched with grammatical annotation, enabling the implementation of a range of valuable features for language learners. The annotation, carried out using a disambiguating morpho-syntactic parser tailored to each language, includes information on lemma, part of speech, and morpho-syntactic features. This information is displayed when the user clicks on a word in the text, thereby supporting learners in analysing the often morphologically complex word forms found in Georgian and Abkhaz.

The availability of lemma and part-of-speech annotation enables the implementation of a particularly useful feature for language learners: automatic dictionary lookup. In the GNC, Donald Rayfield et al.'s *Georgian-English Dictionary* is integrated into both the KWIC concordance and the text display pages. When a user clicks on a word, the corresponding dictionary entry is displayed, based on its disambiguated lemma form and part of speech. In the AbNC, the integrated dictionary is Vladimir A. Kaslandzia's *Abkhaz-Russian Dictionary*.

A similar feature exists in many web browsers for major world languages, where lemmatization is performed on the fly. However, since context is typically not taken into account, all homographs of a deduced lemma form are shown. Georgian – and especially Abkhaz – are not supported in such tools, making the corpus-integrated solution particularly valuable for learners of these languages.

Implementing the dictionary lookup feature requires careful alignment between the morpho-syntactic analyser and the dictionary. In the case of Abkhaz, this compatibility is given a-priori: the lexicon of the Abkhaz analyser is primarily derived from Kaslandzia's dictionary, making the lookup process relatively straightforward – though a few finer points still require attention.

In contrast, the Georgian case presents more complexity. The analyser lexicon is based on Kita Tschenkéli's *Georgisch-Deutsches Wörterbuch*, whose verb entries follow a very different organizational principle than those in Rayfield's *Georgian-English Dictionary*. Tschenkéli's entries are centered around verbal roots and verbal nouns, while Rayfield's dictionary uses the 3rd person singular present or future form as the lookup lemma, with verbal nouns being separate entries. To bridge this structural mismatch, the (3rd person singular present/future)

verb entries have been additionally indexed by their corresponding verbal nouns (plus other necessary features), which makes them available for lookup by verbal noun. Lookup for other parts of speech poses no such issues.

Kaslandzia's dictionary uses Russian as the target language, which can be a barrier for learners without sufficient proficiency in Russian. To address this, the Russian definitions and example translations can be automatically translated into other languages such as English, Turkish, or Ukrainian. This is made possible through integration with the DeepL translation service (deepl.com), which is accessed via its API in the AbNC backend. While translations may occasionally lack full accuracy due to limited context, they are generally of high quality and very helpful for learners. DeepL currently supports translation among 33 languages.

Finally, words looked up in the dictionaries while reading a text can be flagged for later review and memorization in the stand-alone version of the dictionaries.

Taken together, the described features turn the GNC and the AbNC into powerful and user-friendly pedagogical tools for learning these grammatically highly complex languages.

კორპუსი, როგორც ენის შესწავლის პედაგოგიური ინსტრუმენტი – ლემატიზაცია და სალექსიკონო ძიება ქართული ენის ეროვნულ კორპუსსა (GNC) და აფხაზური ენის ეროვნულ კორპუსში (AbNC)

პაულ მოირერი

ბერგენის უნივერსიტეტი, ნორვეგია

paul.meurer@uib.no

ჩემს მოხსენებაში წარმოგიდგენთ ენის შესწავლის პედაგოგიურ ინსტრუმენტებს, რომლებიც ხელმისაწვდომია ქართული ენის ეროვნულ კორპუსსა (GNC) და აფხაზური ენის ეროვნულ კორპუსში (AbNC). ორივე კორპუსი შეიცავს მხატვრულ და არამხატვრულ ტექსტებს, რომელთა მოძიება და, ლიცენზირების მიხედვით, ეკრანზე წაკითხვაც შესაძლებელია. ტექსტების გვერდები მოწესრიგებულად არის დანომრილი და მათი სტრუქტურული ელემენტები – როგორიცაა თავების სათაურები, აბზაცები და სტრიქონებს შორის გადასვლები პოეზიაში – შენარჩუნებულია. რეგისტრირებულ მომხმარებლებს ასევე შეუძლიათ ისარგებლონ სანიშნეებით კითხვის შემდგომში გასაგრძელებლად, რაც უზრუნველყოფს შეუფერხებელ და მოსახერხებელ კითხვით გამოცდილებას.

გარდა ამისა, ორივე კორპუსის ტექსტები მდიდარია გრამატიკული ანოტაციებით, რაც შესაძლებელს ხდის მნიშვნელოვანი ფუნქციების დანერგვას ენის შემსწავლელთათვის. ანოტაცია, რომელიც შესრულებულია თითოეული ენაზე მორგებული, ორაზროვნების მოსახსნელი მორფო-სინტაქსური ანალიზატორის გამოყენებით, მოიცავს ინფორმაციას ლემის, მეტყველების ნაწილისა და მორფო-სინტაქსური მახასიათებლების შესახებ. ეს ინფორმაცია ნაჩვენებია ტექსტში სიტყვაზე დაწკაპუნებისას, რაც ეხმარება შემსწავლელებს ქართულ და აფხაზურ ენებში ხშირად გავრცელებული მორფოლოგიურად რთული სიტყვა-ფორმების ანალიზში.

ლემისა და მეტყველების ნაწილის ანოტაციის არსებობა ენების შემსწავლელთათვის განსაკუთრებით სასარგებლო ფუნქციის დანერგვის საშუალებას იძლევა: ავტომატური სალექსიკონო ძიება. GNC-ში დონალდ რეიფილდის და სხვათა *ქართულ-ინგლისური ლექსიკონი* ინტეგრირებულია როგორც KWIC-ის კონკორდანსში, ასევე ტექსტის საჩვენებელ გვერდებზე. როდესაც მომხმარებელი დააწკაპუნებს სიტყვაზე, გამოჩნდება შესაბამისი სალექსიკონო სიტყვა-სტატია, რომელიც დაფუძნებულია ორაზროვნებამოხსნილი ლემის ფორმასა და მეტყველების ნაწილზე. AbNC-ში ინტეგრირებული ლექსიკონი არის ვლადიმერ ა. კასლანძიას *აფხაზურ-რუსული ლექსიკონი* .

მსგავსი ფუნქცია არსებობს მრავალ ვებბრაუზერში მსოფლიოს ძირითადი ენებისთვის, სადაც ლემატიზაცია მყისიერად ხორციელდება. თუმცა, რადგან კონტექსტი, როგორც წესი, არ არის გათვალისწინებული, დადგენილი ლემის ფორმის ყველა ომოგრაფია ნაჩვენები. ქა-

რთულისთვის — და განსაკუთრებით აფხაზურისთვის — მსგავსი ინსტრუმენტები არ არის ხელმისაწვდომი, რის გამოც მისი კორპუსში ინტეგრირების გადაწყვეტა განსაკუთრებულ მნიშვნელობას იძენს ამ ენების შემსწავლელთათვის.

სალექსიკონო ძიების ფუნქციის დანერგვა მოითხოვს მორფო-სინტაქსურ ანალიზატორსა და ლექსიკონს შორის ზუსტ შესაბამისობას. აფხაზური ენის შემთხვევაში, ეს თავსებადობა წინასწარ არის უზრუნველყოფილი: აფხაზური ანალიზატორის ლექსიკა ძირითადად კასლანძიას ლექსიკონზეა დაფუძნებული, რაც ძიების პროცესს შედარებით მარტივს ხდის — თუმცა რამდენიმე დეტალი მაინც საჭიროებს ყურადღებას.

აფხაზურისგან განსხვავებით, ქართული ენის შემთხვევა უფრო მეტ სირთულეს ქმნის. ანალიზატორის ლექსიკა დაფუძნებულია კიტა ჩხენკელის *ქართულ–გერმანულ ლექსიკონზე* (*Georgisch-Deutsches Wörterbuch*), რომლის ზმნური სიტყვა-სტატიები რეიფილდის *ქართულ-ინგლისურ ლექსიკონში* არსებულისგან სრულიად განსხვავებულ ორგანიზაციულ პრინციპს ემყარება. ჩხენკელის სიტყვა-სტატიები ზმნების ძირებსა და სახელზმნებზეა ორიენტირებული, ხოლო რეიფილდის ლექსიკონი საძიებო ლემად მესამე პირის მხოლობითი რიცხვის აწმყო ან მომავალი დროის ფორმებს იყენებს, ამასთან, სახელზმნა ცალკე სიტყვა-სტატიებად არის გამოყოფილი. ამ სტრუქტურული შეუსაბამობის აღმოსაფხვრელად, (მესამე პირის მხოლობით რიცხვის აწმყო/მომავალი დრო) ზმნური სიტყვა-სტატიები დამატებით ინდექსირებულია მათი შესაბამისი სახელზმნებით (და სხვა აუცილებელი მახასიათებლებით), რაც ხელმისაწვდომს ხდის სახელზმნების მიხედვით ძიებას. სხვა მეტყველების ნაწილების შემთხვევაში მსგავსი პრობლემები არ წარმოიშობა.

კასლანძიას ლექსიკონი სამიზნე ენად რუსულს იყენებს, რაც შეიძლება ბარიერი იყოს რუსული ენის საკმარისი ცოდნის არმქონე შემსწავლელებისთვის. ამ პრობლემის გადასაჭრელად, რუსული განმარტებები და მაგალითები შეიძლება ავტომატურად ითარგმნოს სხვა ენებზე, როგორიცაა ინგლისური, თურქული ან უკრაინული. ეს შესაძლებელი ხდება DeepL-ის თარგმნით სერვისთან ინტეგრაციის მეშვეობით (deepl.com), რომელზეც წვდომა შესაძლებელია AbNC-ის ბექენდში API-ის საშუალებით. მიუხედავად იმისა, რომ თარგმანები ზოგჯერ შეიძლება სრულად ზუსტი არ იყოს შეზღუდული კონტექსტის გამო, ისინი ზოგადად მაღალი ხარისხისაა და ძალიან სასარგებლოა შემსწავლელებისთვის. DeepL ამჟამად უზრუნველყოფს თარგმანს 33 ენაზე.

საბოლოოდ, შესაძლებელია ტექსტის კითხვისას ლექსიკონებში მოძიებული სიტყვების მონიშვნა ლექსიკონების დამოუკიდებელ (standalone) ვერსიაში შემდგომში შემოწმებისა და დამახსოვრების მიზნით.

მთლიანობაში, აღწერილი მახასიათებლები GNC-სა და AbNC-ს ძლიერ და მოსახერხებელ პედაგოგიურ ინსტრუმენტებად გარდაქნიან ამ მორფოლოგიურად რთული ენების შესასწავლად.

თეზისი ქართულად თარგმნა თამარ ლალუაშვილმა.

Frame-Based Terminography in Legal Discourse: Innovations and Ontological Limits

Nazarov, Waldemar
Johannes Gutenberg University Mainz, Germany
University of Burgundy Europe, France
wanazaro@uni-mainz.de
waldemar.Nazarov@ube.fr

The legal field holds a distinctive place in the studies of specialized language and translation due to peculiarities inherent in legal terminology. In legal and linguistic research, its characteristics are often cited as justification for isolating legal language from other specialized subject matters (see Sandrini, 1999). Adapted approaches to law within LSP research have emerged to better describe and address the semantic challenges of a "language for legal purpose (LLP)" (Cao, 2007, p. 8). This has led to the development of focused subfields such as legal linguistics, established as *Rechtslinguistik* in the Germanist tradition (Vogel, 2017) and emerging from the Canadian *jurilinguistique* (Gémar, 2015), as well as Legal Translation Studies (LTS) (Prieto Ramos, 2014), which parallels the French *juritraductologie* (Monjean-Decaudin, 2022). These subdisciplines aim, for one, to address the normativity of legal language (Jeand'Heur, 1998), which arises from the binding nature of legal terminology as part of a human-made system of societal rules (Sparer, 2002). Given the continuous evolution of case law, legislation, and legal doctrine—sources that shape legal language and terminology (Constantinesco, 1974)—the dynamic nature of legal terms plays a crucial role in terminology management. Legal concepts are captured at a specific point in time, a characteristic linked to the abstract nature of law as a "metaphysical phenomenon," (Mattila, 2006, p. 105) which also necessitates a degree of semantic vagueness (Dullion, 2015).

The socio-cultural embeddedness of the legal domain presents further challenges, which become particularly evident in contrastive settings such as translation. Legal terminology exhibits an "extreme" system dependence (de Groot, 2002, p. 228), which, even from an intralingual point of view, results in the rejection of the classification of an English, German, or French legal language due to pluricentricity. This phenomenon has led some scholars to reject the possibility of establishing *equivalence* between legal concepts from different legal systems—so-called *Bezugsrechtsordnungen* (Wiesmann, 2004, p. 20)—highlighting the necessity of legal comparison in any cross-systemic legal translation scenario (see Pommer, 2006). However, the extensive body of research emphasizing the differences between legal languages has even led some legal comparatist such as David (1974) and Kischel (2009) to question the very translatability of legal texts.

This paper examines recent developments addressing the lack of traditional and overarching LSP findings for legal language, given the terminological challenges in the field of law. While some scholars argue for insulating legal terminology from the category of specialized languages altogether (see Simonnæs, 2002), Busse (2015) presents legal language as an institutional discourse created, developed, and applied specifically by legal institutions. Given the extensive for-

mal overlap between legal terminology and general lexis (Thormann, 2019), this perspective complements ontological approaches to legal language (see Orozco & Sánchez-Gijón, 2011) by emphasizing the epistemic dimension of specialized lexicography, as seen in the perspective of *ontoterminology* (Roche, 2007). The knowledge component of legal terminology is further reinforced by recent applications of frame semantics—originally developed to describe general language—to specialized domains (Faber, 2022; Lönneker-Rodman & Ziem, 2018). A frame-based approach has also proven valuable in legal translation, as Engberg (2021) has demonstrated through the Knowledge Communication Approach (Kastberg, 2019), which enables a cognitive and epistemic representation of the ontology of legal terminology (Nazarov, 2024).

This paper assesses the limitations of a frame-based approach to ontology, terminology, and terminography in the legal field, particularly regarding the representation and structuring of legal concepts for terminological databases serving as tools for legal translation. Using examples of English legal terms from the U.S., U.K., and E.U. legal systems, such as *contract*, *good faith*, and *bailiff*, the paper explores how the dynamic, vague, normative, and system-dependent nature of legal semantics challenges the ontological representation of legal concepts. Given the recursivity and endless granularity in frame semantics, it evaluates the constraints of a frame-based approach in capturing legal terminology through slots, fillers, standard values, and inter- and intra-frame relations (Varga, 2020) within culturally embedded institutional discourses.

ფრეიმებზე-დაფუძნებული ტერმინოგრაფია სამართლებრივ დისკურსში: ინოვაციები და ონტოლოგიური შეზღუდვები

ვალდემარ ნაზაროვი

იოჰანეს გუტენბერგის სახელობის უნივერსიტეტი, მაინცი, გერმანია
ბურგუნდიის ევროპული უნივერსიტეტი, საფრანგეთი

wanazaro@uni-mainz.de
waldemar.Nazarov@ube.fr

სამართლის დარგს იურიდიული ენისათვის დამახასიათებელი თავისებურებებიდან გამომდინარე გამორჩეული ადგილი უკავია დარგობრივი ენების სწავლებასა და თარგმანში. სამართლებრივ და ლინგვისტურ კვლევებში, ხშირად ეს თავისებურებები გამოყენებულია იმის მტკიცებულებად, რომ იურიდიული ენა უნდა გამოიყოს სხვა დარგობრივი ენებისაგან (იხ. Sandrini, 1999). სპეციალიზებული ენების კვლევებში სამართლისთვის ადაპტირებული მიდგომები უკეთესად აღწერენ „იურიდიული ენისათვის“ დამახასიათებელ სემანტიკურ სირთულეებს (Cao, 2007, გვ. 8). შედეგად, აღმოცენდა ისეთი ქვედარგები, როგორიცაა იურიდიული ლინგვისტიკა, რომელიც გერმანულ ტრადიციაში ცნობილია *Rechtslinguistik*-ის სახელით (Gémar, 2015) და იურიდიული თარგმანმცოდნეობა (Prieto Ramos, 2014), რომელსაც ფრანგულში შეესაბამება *juritraductologie* (Monjean-Decaudin, 2022). ამ ქვედარგების მიზანია, ერთი მხრივ, შეისწავლონ იურიდიული ენის ნორმატიული ბუნება (Jeand'Heur, 1998), რომელიც განპირობებულია სამართლის ტერმინოლოგიის ვალდებულებითი ხასიათით, რადგან ის ადამიანის მიერ შექმნილი საზოგადოებრივი წესების სისტემის ნაწილია (Sparer, 2002). სასამართლო პრაქტიკის, კანონმდებლობისა და იურიდიული დოქტრინის უწყვეტი ბუნებიდან გამომდინარე, რომლებიც წარმოადგენენ იურიდიული ენისა და ტერმინოლოგიის წყაროებს (Constantinesco, 1974), იურიდიული ტერმინოლოგიის დინამიკური ბუნება მნიშვნელოვან როლს ასრულებს ტერმინოლოგიის მართვაში. სამართლის ცნებები მოცემულია კონკრეტული დროის მონაკვეთში, რაც სამართლის, როგორც „მეტაფიზიკური მოვლენის“ (Mattila, 2006, გვ. 105) ბუნებით არის განპირობებული. თავის მხრივ, სწორედ აღნიშნული ფაქტი, იწვევს გარკვეულ სემანტიკურ ბუნდოვანებას (Dullion, 2015).

სამართლის დარგის დამოკიდებულება სოციალურ და კულტურულ ფაქტორებზე იწვევს დამატებით სირთულეებს და თვალნათლივ იჩენს თავს, შეპირისპირების დროს, მაგალითად, თარგმანის შემთხვევაში. იურიდიული ტერმინოლოგია „მწვავედაა“ დამოკიდებული სისტემაზე (de Groot, 2002, გვ. 228), რომელიც ენათაშორისი ხედვის თანახმადაც კი, იწვევს ინგლისური, გერმანული ან ფრანგული იურიდიული ენის კლასიფიკაციის უარყოფას მრავალცენტრული ხასიათის გამო. ამ ფაქტის გამო, ბევრი ლინგვისტი უარყოფს სხვადასხვა სამართლებრივი სისტე-

მისთვის დამახასიათებელი სამართლის ცნებებს შორის ეკვივალენტურობის დადგენის შესაძლებლობას ე.წ. *Bezugsrechtsordnungen* (Wiesmann, 2004, გვ. 20), რაც ხაზს უსვამს სისტემათაშორისი იურიდიული თარგმანის შემთხვევაში შედარების აუცილებლობას. მიუხედავად ამისა, იურიდიული ენის ირგვლივ არსებულმა მრავალმა კვლევამ განაპირობა ის, რომ ზოგიერთი მკვლევარი, მათ შორის, დეივიდი (1974) და კიშელი (2009) კითხვის ნიშნის ქვეშ აყენებენ იურიდიული ტექსტების გადათარგმნის შესაძლებლობას.

ეს ნაშრომი განიხილავს უახლეს ცვლილებებს, რომლებიც ეხება სამართლის ენაში ტერმინოლოგიური გამოწვევების ფონზე ტრადიციული და ყოვლისმომცველი დარგობრივი ენების კვლევების ნაკლებობას. მიუხედავად იმისა, რომ ზოგიერთი მკვლევარი საერთოდ გამორიცხავს სამართლის ენას დარგობრივი ენების კატეგორიიდან (იხ. Simonnæs, 2002), ბუსე (2015) სამართლის ენას განიხილავს ინსტიტუციური დისკურსის ჭრილში, რომელიც შექმნილია, განვითარდა და გამოიყენება სპეციალურად იურიდიული ინსტიტუციების მიერ. სამართლის ტერმინოლოგიასა და ზოგადი ენის ლექსიკას შორის არსებული ფორმისმიერი მსგავსებებიდან გამომდინარე (Thormann, 2019), ეს მიდგომა ავსებს სამართლის ენისადმი არსებულ ონტოლოგიურ მიდგომებს (იხ. Orozco & Sánchez-Gijón, 2011), რითაც ხაზი ესმება სპეციალიზებული ლექსიკოგრაფიის ეპისტემოლოგიურ მხარეს, რაც აისახება ონტოტერმინოლოგიის პერსპექტივაში (Roche, 2007). სამართლის ტერმინოლოგიის ცოდნითი კომპონენტი კიდევ უფრო მეტად გამოიკვეთა ბოლოდროინდელი ფრეიმების სემანტიკის გამოყენებით (რომელიც თავდაპირველად შემუშავდა საერთო ენის აღწერისათვის) სპეციალიზებულ დარგებში (Faber, 2022; Lönneker-Rodman & Ziem, 2018). ფრეიმზე-დაფუძნებული მიდგომა ასევე შეუფასებელი აღმოჩნდა იურიდიულ თარგმანში, როგორც ეს აღწერა ენგბერგმა (2021) ცოდნის კომუნიკაციის მიდგომის (Kastberg, 2019) გამოყენებით, რაც იძლევა სამართლის ტერმინოლოგიის ონტოლოგიის კოგნიტიური და ეპისტემური წარმოდგენის საშუალებას (Nazarov, 2024).

ეს ნაშრომი განიხილავს ფრეიმზე-დაფუძნებული მიდგომის შეზღუდვებს სამართლის სფეროში ონტოლოგიის, ტერმინოლოგიისა და ტერმინოგრაფიისათვის, კერძოდ, სამართლის ტერმინოლოგიურ ბაზებში სამართლის ცნებების წარმოდგენისა და სტრუქტურებისათვის, რომელთა გამოყენებაც ხდება იურიდიულ თარგმანში. აშშ-ის, ბრიტანეთისა და ევროკავშირის სამართლის სისტემებიდან აღებული ინგლისური სამართლებრივი ტერმინების მაგალითებზე, როგორიცაა *contract, good faith* და *bailiff,* ნაშრომი იკვლევს, თუ როგორ ართულებს სამართლის სემანტიკის დინამიკური, ბუნდოვანი, ნორმატიული და სისტემაზე დამოკიდებული ხასიათი იურიდიული ცნებების ონტოლოგიურ წარმოდგენას. ფრეიმების სემანტიკის რეკურსიულობისა და უსასრულო დეტალიზაციის ფონზე, ნაშრომი აფასებს ფრეიმზე-დაფუძნებული მიდგომის შეზღუდვებს სამართლის ტერმინოლოგიის აღბეჭდვაში ცარიელი ადგილების, შემავსებლების, სტანდარტული მნიშვნელობებისა და ფრეიმებს შორის არსებული გარე და შიდა დამოკიდებულებების მე-

შვეობით (Varga, 2020) კულტურულად განპირობებული ინსტიტუციური დისკურსების ფარგლებში.

თეზისი ქართულად თარგმნა ქეთევან მჭედლიშვილმა.

References

Busse, D. (2015). Juristische Semantik als Frame-Semantik. In F. Vogel (Ed.), *Zugänge zur Rechtssemantik: Interdisziplinäre Ansätze im Zeitalter der Mediatisierung* (pp. 41–68). de Gruyter.

Cao, D. (2007). *Translating Law*. Multilingual Matters.

Constantinesco, L.-J. (1974). *La méthode comparative. Traité de droit comparé: II*. Pichon & Durand-Auzias.

David, R. (1974). *Les grands systèmes de droit contemporains* (6th ed.). Dalloz.

de Groot, G.-R. (2002). Rechtsvergleichung als Kerntätigkeit bei der Übersetzung juristischer Terminologie. In U. Haß-Zumkehr (Ed.), *Sprache und Recht* (pp. 222–239). de Gruyter.

Dullion, V. (2015). Droit comparé pour traducteurs : de la théorie à la didactique de la traduction juridique. *International Journal for the Semiotics of Law*, *28*, 91–106.

Engberg, J. (2021). Legal Translation as Communication of Knowledge: On the Creation of Bridges. *Parallèles*, *33*(1), 6–17.

Faber, P. (2022). Frame-Based Terminology. In P. Faber & M.-C. L'Homme (Eds.), *Theoretical Perspectives on Terminology: Explaining Terms, Concepts and Specialized Knowledge* (pp. 353–375). Benjamins.

Gémar, J.-C. (2015). De la traduction juridique à la jurilinguistique : la quête de l'*équivalence*. *Meta*, *60*(3), 476–493.

Jeand'Heur, B. (1998). Die neuere Fachsprache der juristischen Wissenschaft seit der Mitte des 19. Jahrhunderts unter besonderer Berücksichtigung von Verfassungsrecht und Rechtsmethodik. In L. Hoffmann, H. Kalverkämper, & H. E. Wiegand (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft: Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft* (pp. 1286–1295). de Gruyter.

Kastberg, P. (2019). *Knowledge Communication: Contours of a Research Agenda*. Frank & Timme.

Kischel, U. (2009). Legal Cultures – Legal Languages. In F. Olsen, A. Lorz, & D. Stein (Eds.), *Translation Issues in Language and Law* (pp. 7–17). Palgrave Macmillan.

Lönneker-Rodman, B., and Ziem, A. (2018). Frames als Repräsentationsformat in modernen Terminologiesystemen. In A. Ziem, L. Inderelst, & D. Wulf (Eds.), *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden* (pp. 251–288). düsseldorf university press.

Mattila, H. E. S. (2006). *Comparative Legal Linguistics*. Ashgate.

Monjean-Decaudin, S. (2022). *Traité de juritraductologie: Épistémologie et méthodologie de la traduction juridique*. Presses universitaires du Septentrion.

Nazarov, W. (2024). Toward a dynamic frame-based ontology of legal terminology. *Applied Ontology*, *19*(1), 73–98.

Orozco, M., and P. Sánchez-Gijón (2011). New resources for legal translators. *Perspectives: Studies in Translatology*, *19*(1), 25–44.

Pommer, S. (2006). *Rechtsübersetzung und Rechtsvergleichung: Translatologische Fragen zur Interdisziplinarität*. Lang.

Prieto Ramos, F. (2014). Legal Translation Studies as Interdiscipline: Scope and Evolution. *Meta*, *59*(2), 260–277.

Roche, C. (2007). Le terme et le concept : fondements d'une ontoterminologie. *TOTh 2007 (Terminology & Ontology: Theories and Applications)*, 1–22.

Sandrini, P. (1999). Translation zwischen Kultur und Kommunikation: Der Sonderfall Recht. In P. Sandrini (Ed.), *Übersetzen von Rechtstexten: Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache* (pp. 9–43). Narr.

Simonnæs, I. (2002). Zur Frage der rechtskulturellen Unübersetzbarkeit anhand eines Vergleiches zwischen Norwegen und Deutschland. In L. Eriksen & K. Luttermann (Eds.), *Juristische Fachsprache: Kongressberichte des 12th European Symposium on Language for Special Purposes, Brixen/Bressanone 1999* (pp. 133–156). LIT.

Sparer, M. (2002). Peut-on faire de la traduction juridique ? Comment doit-on l'enseigner ? *Meta*, *47*(2), 265–278.

Thormann, I. (2019). „Sich einlassen", „regelmäßig", „billig", „fremd" – aus der Umgangssprache vertraute Wörter mit anderer Bedeutung in der Rechtsterminologie. In Parlament der Deutschsprachigen Gemeinschaft Belgiens (Ed.), *Nationale Variation in der deutschen Rechtsterminologie* (pp. 49–56). Parlament der Deutschsprachigen Gemeinschaft Belgiens.

Varga, S. (2020). *Frames und Argumentation: Zur diskurssemantischen Operationalisierung von Frame-Relationen*. Lang.

Vogel, F. (2017). Rechtslinguistik: Bestimmung einer Fachrichtung. In E. Felder & F. Vogel (Eds.), *Handbuch Sprache im Recht* (pp. 209–232). de Gruyter.

Wiesmann, E. (2004). *Rechtsübersetzung und Hilfsmittel zur Translation: Wissenschaftliche Grundlagen und computergestützte Umsetzung eines lexikographischen Konzepts*. Narr.

New Words in Old Chambers: Tracking Polish Parliamentary Neologisms

Ogrodniczuk, Maciej; Tomaszewska, Aleksandra
Institute of Computer Science, Polish Academy of Sciences, Poland
maciej.ogrodniczuk@ipipan.waw.pl
aleksandra.tomaszewska@ipipan.waw.pl

One might think that parliamentary debates are unlikely to be associated with the use or creation of new words: speeches are often prepared in advance, and political theatre demands precision of expression. However, both the deliberation in parliament on issues of importance to society and the temperature of the debate often make parliamentary chambers a rich environment for linguistic evolution, where the pressures of political communication generate new words or expressions that contribute to the discussion or change its perspective. Moreover, the combination of planned discourse and spontaneous exchange allows for both calculated innovation and improvised linguistic creativity. In this way, studies of parliamentary neologisms have important implications for both linguistic and political research, contributing to our understanding of how linguistic innovation shapes and is shaped by political communication.

In this talk, we will present insights from the analysis of the parliamentary debates gathered in the CLARIN-PL Polish Parliamentary Corpus (from 2020 to now) using our newly developed NeoN software. Our research material comprises all the speeches delivered in the Sejm and the Senate during that period. These speeches were compared against a set of filters: a contemporary Polish dictionary sjp.pl and Polish language corpora – the National Corpus of Polish and the Contemporary Corpus of Polish. Words that met certain criteria were automatically filtered out. Additionally, the list of candidate words was manually verified.

The list of words extracted from the current debates will help us trace how new words find their way into dictionaries by comparing the results with the contents of dictionaries of new vocabulary, both language observatories at universities and slang dictionaries maintained by amateur lexicographers and collected by crowd-sourcing.

The timeframe analyzed is particularly interesting because it covers both the period of the pandemic, when entirely new word-formative nests related to COVID were created and used in the public sphere, such as *kowidianie* (covidians) or *kowidioci* (covidiots, etc.), and the period when they were taken over by the vocabulary related to the Russian-Ukrainian war, such as *weaponizacja* (weaponization) or the famous *Bayraktar*. These findings are consistent with studies that explain that linguistic constructions often emerge at moments of conceptual transformation, when existing vocabulary proves inadequate to capture evolving political realities. But different types of linguistic innovation can also be used to trace how linguistic creativity serves substantive rhetorical purposes in advancing policy arguments and political narratives.

Another observation confirmed by our study is that many recent political neologisms derive from the names of prominent political figures and function as linguistic shorthand for complex policy approaches or ideological positions.

In the Polish parliament, the most prominent is *putinflacja* (also as *putininflacja*; refers to the rise in inflation attributed to Vladimir Putin; with the highest frequency over the whole period). Local figures are represented by such terms as *bodnaryzacja* (Bodnar-isation; excessive interference of power in the legal system, leading to a violation of the principles of law and logic; from the name of Adam Bodnar, the Minister of Justice) or *sasinada* (Sasinade; a situation or decision that ends in failure, controversy or unexpected consequences, often in an unfortunate or comical way; from the name of Jacek Sasin, associated with the chaotic organisation of the 2020 presidential election by postal vote during the COVID-19 pandemic, which was later declared unconstitutional).

Compound neologisms in particular serve as markers of discursive evolution, allowing researchers to trace conceptual developments over time. In the parliamentary corpus they are represented, for example, by the word nest with the prefix *pato-*, e.g. *patowładza*, *patobudżet*, even *patoparkowanie* (pathological government, budget, parking) or *alkotubki* (alcohol sold in small bags that look like fruit pulp).

In conclusion, some new words act as linguistic tools to conceptualise new ideas, raise awareness of certain issues and introduce alternative perspectives into political discourse. In their ability to name previously unnamed concepts or to reframe existing ones, neologisms have the potential to influence the direction and content of parliamentary debates and the way in which issues are understood by both politicians and the public.

The work was financed as part of the investment: CLARIN ERIC – European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure (period: 2024-2026) funded by the Polish Ministry of Science and Higher Education (Programme: "Support for the participation of Polish scientific teams in international research infrastructure projects"), agreement number 2024/WK/01, and by CLARIN-PL, the European Regional Development Fund, FENG programme (FENG.02.04-IP.040004/24).

ახალი სიტყვები ძველ დარბაზებში: პოლონური საპარლამენტო ნეოლოგიზმების კვალდაკვალ

მაცეი ოგროდნიჩუკი
ალექსანდრა ტომაშევსკა
პოლონეთის მეცნიერებათა აკადემია, პოლონეთი
aleksandra.tomaszewska@ipipan.waw.pl
maciej.ogrodniczuk@ipipan.waw.pl

შესაძლოა ვინმემ იფიქროს, რომ საპარლამენტო დებატები არ არის დაკავშირებული ახალი სიტყვების შექმნასთან ან მათ გამოყენებასთან: პოლიტიკოსების სიტყვით გამოსვლები წინასწარ არის მომზადებული და პოლიტიკური არენაც მოითხოვს აზრის მკაფიოდ ჩამოყალიბებას. თუმცა, ერთი მხრივ, საზოგადოებისათვის მნიშვნელოვანი საკითხების განხილვა და, მეორე მხრივ, დებატების სიმწვავე ხშირად აქცევს პარლამენტის დარბაზებს ნოყიერ გარემოდ ენის განვითარებისათვის. ეს არის ადგილი, სადაც პოლიტიკური ურთიერთობით გამოწვეული ზეწოლის შედეგად იქმნება ახალი სიტყვები თუ გამოთქმები, რომელიც ხელს უწყობს დისკუსიის წარმართვას და ცვლის მის სამომავლო პერსპექტივას. გარდა ამისა, წინასწარ მოფიქრებული გამონათქვამებისა და სპონტანური პასუხების ერთობლიობა ხელს უწყობს სიახლეების გაჩენას და ენის შემოქმედებითი უნარის განვითარებას.

ამგვარად, საპარლამენტო ნეოლოგიზმების შესწავლა მნიშვნელოვანია ენობრივი თუ პოლიტიკური მიმართულების კვლევებისათვის, რაც შესაძლებლობას გვაძლევს შევისწავლოთ როგორ ახდენს გავლენას ენობრივი სიახლეები პოლიტიკურ ურთიერთობებზე და პირიქით, პოლიტიკური ურთიერთობები ენაში სიახლეების გაჩენაზე.

მოცემულ მოხსენებაში წარმოვადგენთ პოლონეთის საპარლამენტო დებატების ანალიზის შედეგებს, რისთვისაც გამოვიყენეთ პოლონეთის საპარლამენტო კორპუსი CLARIN-PL (2020 წლიდან დღემდე) და ჩვენ მიერ ახლად შემუშავებული პროგრამული ინსტრუმენტი NeoN. საკვლევ მასალად აღებულია ამ დროის მონაკვეთში წარმოთქმული ყველა სიტყვა სეიმსა თუ სენატში. სიტყვით გამოსვლები შედარებულია სხვადასხვა ფილტრთან, როგორიცაა: თანამედროვე პოლონური ენის ლექსიკონი sjp.pl და პოლონური ენის კორპუსები – პოლონური ენის ეროვნული კორპუსი (the National Corpus of Polish) და პოლონური ენის თანამედროვე კორპუსი (the Contemporary Corpus of Polish). კვლევის პროცესში სიტყვები, რომლებიც აკმაყოფილებდნენ გარკვეულ კრიტერიუმებს ავტომატურად ამოიკრიბა. ამასთანავე, კანდიდატი სიტყვები ხელით გადამოწმდა.

მოცემული დებატებიდან ამოკრებილი სიტყვები დაგვეხმარება თვალი მივადევნოთ იმას, თუ როგორ ხვდება ახალი სიტყვა ლექსიკონში, რისთვისაც ერთმანეთს შევადარებთ ახალი ლექსიკის შემცველ სხვადასხვა სახის ლექსიკონებს, როგორიცაა, მაგალითად, უნივერსიტე-

ტებში არსებული ენაზე დაკვირვების ჯგუფების მიერ შექმნილი ლექსიკონები, ან მოყვარული ლექსიკოგრაფების მიერ შედგენილი და სახალხო რესურსით შეგროვებული სლენგის ლექსიკონები.

დროის მონაკვეთი, რომელიც შესასწავლად შეირჩა, განსაკუთრებულ ინტერესს იწვევს, ვინაიდან მოიცავს ორ მნიშვნელოვან პერიოდს. ერთი მხრივ, ეს არის პანდემია, როცა ახალი სიტყვების შექმნა სრულად უკავშირდებოდა კოვიდ-პანდემიას და ფართოდ გამოიყენებოდა ისეთი სიტყვები, როგორიცაა *kowidianie* (covidians) და *kowidioci* (covidiots და ა.შ.). მეორე პერიოდი კი მოიცავს რუსეთ-უკრაინის ომის დროინდელ ლექსიკას, როცა გაჩნდა ისეთი სიტყვები, როგორიცაა *weaponizacja* (weaponization) და ცნობილი *Bayraktar*. ჩვენ მიერ ჩატარებული კვლევის შედეგები კიდევ ერთხელ ადასტურებს, რომ ენობრივი კონსტრუქციები ხშირად თავს იჩენს ცნების ცვლილების დროს, როდესაც არსებული ლექსიკა აღმოჩნდება შეუსაბამო ახალი პოლიტიკური რეალობისათვის. მაგრამ სხვადასხვა სახის ლინგვისტური სიახლეებით შესაძლოა ასევე თვალი მივადევნოთ იმას, თუ რამდენად შესაძლებელია ენობრივი შემოქმედებითობის უნარის გამოყენება პოლიტიკური კამათებისა და ნარატივების დასახვეწადაც.

ჩვენმა კვლევამ ასევე დაადასტურა, რომ ბევრი პოლიტიკური ნეოლოგიზმი ცნობილი პოლიტიკური ფიგურის სახელს უკავშირდება და ფუნქციონირებს როგორც ლინგვისტური სტენოგრაფია რთული პოლიტიკური მიდგომებისა და იდეოლოგიური შეხედულებების გამოსახატავად. პოლონეთის პარლამენტში ყველაზე ცნობილია სიტყვა *putinflacja* (აგრეთვე *putininflacja*; გამოიყენება ვლადიმერ პუტინის პოლიტიკის მიერ გამოწვეული ინფლაციის აღსანიშნად, რაც საკმაოდ ხშირია მთელი ამ პერიოდის განმავლობაში). ადგილობრივი პოლიტიკური ფიგურები წარმოდგენილი არიან ისეთი ტერმინებით, როგორიცაა *bodnaryzacja* (Bodnar-isation; აღნიშნავს ძალაუფლების გამოყენებას სასამართლო სისტემაში, რასაც მივყავართ სამართლისა და ლოგიკის კანონების დარღვევამდე; უკავშირდება იუსტიციის მინისტრის სახელს Adam Bodnar) ან *sasinada* (Sasinade; მდგომარეობა ან გადაწყვეტილება, რომელიც სრულდება კრახით, საკამათო თუ მოულოდნელი შედეგით, კომიკურად თუ უიღბლოდ. დაკავშირებულია Jacek Sasin-ის სახელთან, რომელიც ასოცირდება კოვიდ 19-ის პანდემიის დროს, 2020 წლის საპრეზიდენტო არჩევნებში ფოსტით ხმის მიცემის ქაოსურ ორგანიზებასთან. მოცემული არჩევნების შედეგები მოგვიანებით არაკონსტიტუციურად გამოცხადდა).

შედგენილი ნეოლოგიზმები წარმოადგენს დისკურსული ევოლუციის მარკერებს, რაც მკვლევარს შესაძლებლობას აძლევს თვალყური ადევნოს ცნების განვითარებას დროთა განმავლობაში. მაგალითად, საპარლამენტო კორპუსში ისინი წარმოდგენილია სიტყვათა ჯგუფით, რომლებიც იწყება თავსართით *pato-*, მაგ: *patowładza*, *patobudżet* და *patoparkowanie* (პათოლოგიური მთავრობა, ბიუჯეტი, პარკინგი) ან *alkotubki* (აღნიშნავს ალოკოჰოლს, რომელიც იყიდება ხილის რბილობის მსგავს პატარა პაკეტებში).

საბოლოო ჯამში, ზოგიერთი ახალი სიტყვა არის ერთგვარი ენობრივი ინსტრუმენტი ახალი იდეის გასააზრებლად, გარკვეულ საკითხებზე მეტი აქცენტის გასაკეთებლად და პოლიტიკურ დისკურსში ალტერნატიული პერსპექტივების წარმოსაჩენად. ნეოლოგიზმებს შეუძლია გავლენა იქონიოს საპარლამენტო დებატების შინაარსზე და მიმართულებაზე და იმაზე თუ როგორ აღიქვამენ გარკვეულ საკითხებს პოლიტიკოსები თუ საზოგადოება, ვინაიდან მათ აქვთ უნარი ახალ ცნებას დაარქვან სახელი ან გადაიაზრონ არსებული ცნება.

თეზისი ქართულად თარგმნა თამარ კვიციანმა.

Developments in the Structure of Dictionaries of Foreign Words in Slovak

Panocová, Renáta
Pavol Jozef Šafárik University in Košice, Slovakia
renata.panocova@upjs.sk

Language contact often leads to borrowing of lexical items. However, the status of loanwords in the recipient language has always been critically observed by various categories of speakers. The starting point for lexicographers is the observation of the use of loanwords in texts. This is followed by a decision-making process, determining whether the loanwords qualify for inclusion in the dictionary they are compiling. The criteria for selection and inclusion depend on the type of dictionary. A type of dictionary where loanwords play a central role is the dictionary of foreign words. This type includes specialized dictionaries covering a relatively narrowly delimited part of the vocabulary of a language. Swanepoel (2003: 64) briefly mentions dictionaries of neologisms and foreign words in his overview of typology of specialized or restricted dictionaries.

In the description of German, dictionaries of foreign words have a well-established tradition of more than two hundred years. Hausmann (1985: 391) gives two main functions of dictionaries of foreign words in German, on one hand the purist attempts to replace these words by words of native origin and the other the explanation of difficult words. The latter corresponds to the situation in English at the time of its first monolingual dictionaries, which focused on words that were hard to understand (Osselton 1983: 15). The main functions of the earliest English dictionaries up to 1750 were cultural and educational. Béjoint (2000: 94) explains that especially many children, women and foreigners did not have access to education and dictionaries of hard words were often used as self-teaching tools.

In Slovak, there is also a long tradition of dictionaries of foreign words, inspired in part by German and Slavic examples. Since 1921, when the first dictionary of foreign words was published, more than forty dictionaries of this type have appeared. Such a quantity is impressive as on average, a new dictionary or revised edition was published nearly every two years. Given the fact that there are around 5,115,000 total users of Slovak (Eberhard, Simons, and Fenning, 2019), the degree of popularity of this type of dictionary is striking. The popularity of dictionaries of foreign words among Slovak users is closely connected to a general interest in so-called language cultivation, which includes a good knowledge of frequently used words of non-native origin, the ability to explain their meaning and to provide their native equivalents and mastering their use in stylistically adequate contexts.

Slovak dictionaries developed significantly over the course of the past century. The main aim of this paper is to investigate the nature of this development. In order to do so, I compared five dictionaries of foreign words in Slovak SCSVS (1939), SCS (1953), SCS (1966), VSCS (1997) and SCSA (2005) with a central focus on their properties such as macrostructure, microstructure, intended users, and medium. A key selection criterion was that each of the five dictionaries rep-

resents a different historical period, which can be assumed to have an influence on these properties. The results of the analysis are compared with the situation in German and English in the respective periods. This comparison determines the position of the Slovak tradition of dictionaries of foreign words in the broader European context.

უცხო სიტყვათა ლექსიკონების სტრუქტურის განვითარება სლოვაკურ ენაში

რენატა პანოცოვა

პავოლ იოზეფ შაფარიკის სახელობის კოშიცეს უნივერსიტეტი, სლოვაკეთი

renata.panocova@upjs.sk

ენების კონტაქტი ხშირად იწვევს ლექსიკური ერთეულების სესხებას. თუმცა, მიმღებ ენაში ნასესხები სიტყვების სტატუსს ყოველთვის კრიტიკულად უდგებიან ამ ენაზე მოსაუბრე სხვადასხვა პროფესიის წარმომადგენლები. ლექსიკოგრაფებისათვის უმნიშვნელოვანესია ნასესხები სიტყვების ტექსტებში გამოყენების შესწავლა. ამას მოსდევს გადაწყვეტილების მიღების პროცესი, რაც გულისხმობს იმის განსაზღვრას, თუ რამდენად ღირს ამა თუ იმ ნასესხები სიტყვის ლექსიკონში შეტანა. სიტყვების შერჩევის კრიტერიუმები კი დამოკიდებულია იმაზე, თუ რა ტიპის ლექსიკონთან გვაქვს საქმე. ლექსიკონს, რომელშიც ნასესხები სიტყვები მთავარ როლს ასრულებს, უცხო სიტყვათა ლექსიკონი ეწოდება. უცხო სიტყვები ასევე მნიშვნელოვან როლს თამაშობს დარგობრივ ლექსიკონებშიც, სადაც წარმოდგენილია ენის საერთო ლექსიკისაგან გამიჯნული სპეციალიზებული ნაწილი. სვანეფული (Swanepoel 2003: 64), მიმოიხილავს რა დარგობრივი ლექსიკონების ტიპოლოგიას, ასევე ახსენებს ნეოლოგიზმებისა და უცხო სიტყვათა ლექსიკონებსაც.

გერმანულ ენაში უცხო სიტყვათა ლექსიკონებს აქვს ორას წელზე მეტი ხნის მყარად დამკვიდრებული ტრადიცია. ჰაუსმანი (1985: 391) ხაზს უსვამს გერმანულში უცხო სიტყვათა ლექსიკონის ორ მთავარ ფუნქციას. ერთი მხრივ, ესაა უცხო სიტყვების მშობლიური წარმოშობის სიტყვებით ჩანაცვლების პურისტული მცდელობა და მეორე მთავარი ფუნქცია არის რთული სიტყვების განმარტება. ეს უკანასკნელი შეესატყვისება ინგლისურში არსებულ მდგომარეობას იმ დროისათვის, როდესაც იქმნებოდა პირველი განმარტებითი ლექსიკონები, რომლებიც ორიენტირებული იყო რთული სიტყვების განმარტებაზე (Osselton 1983: 15). 1750 წლამდე გამოცემული უძველესი ინგლისური ლექსიკონების ფუნქცია ძირითადად იყო კულტურული და საგანმანათლებლო. ბეჟუანი (2000: 94) განმარტავს, რომ იმ დროისათვის ბევრ ბავშვს, ქალსა თუ უცხოელს არ ჰქონდა განათლების მიღების შესაძლებლობა და რთული სიტყვების ლექსიკონი იყო თვითგანსწავლის საშუალება.

სლოვაკურ ენაშიც არის უცხო სიტყვათა ლექსიკონების არსებობის ხანგრძლივი ტრადიცია შთაგონებული გერმანული და სლავური მაგალითებით. 1921 წლიდან მოყოლებული, მას შემდეგ რაც პირველი უცხო სიტყვათა ლექსიკონი გამოიცა, ორმოცზე მეტი ასეთი ლექსიკონი შეიქმნა. ეს რაოდენობა ნამდვილად შთამბეჭდავია, ვინაიდან ახალი ლექსიკონი, თუ ყოველი განახლებული გამოცემა საშუალოდ ყოველ ორ წელიწადში ერთხელ იბეჭდებოდა. იმ ფაქტის გათვალისწინებით, რომ სლოვაკურ ენას დაახლოებით ჰყავს 5 115 000 მომხმარებელი (Eberhard,

Simons და Fenning, 2019), ასეთი სახის ლექსიკონების პოპულარობა განსაცვიფრებელია. უცხო სიტყვათა ლექსიკონების ასეთი პოპულარობა მჭიდროდ უკავშირდება სლოვაკი ხალხის ინტერესს ენის განვითარებისადმი, რაც გულისხმობს ხშირად გამოყენებული უცხო წარმოშობის სიტყვების ცოდნას, მათი მნიშვნელობების განმარტების უნარს, მშობლიურ ენაზე მათი შესატყვისების მოძიებასა და ამ სიტყვების სტილისტურად შესაბამის კონტექსტში გამოყენებაში დახელოვნებას.

გასულ საუკუნეში სლოვაკური ლექსიკონები მნიშვნელოვნად განვითარდა. მოცემული ნაშრომის მთავარი მიზანია ამ განვითარების ბუნების შესწავლა. ამ მიზნით ერთმანეთს შევადარე სლოვაკურ ენაში არსებული უცხო სიტყვათა ხუთი ლექსიკონი SCSVS (1939), SCS (1953), SCS (1966), VSCS (1997) და SCSA (2005), რომლის დროსაც აქცენტი გავაკეთე მათ ისეთ მახასიათებლებზე, როგორიცაა მაკროსტრუქტურა, მიკროსტრუქტურა, სავარაუდო მომხმარებელი და შედგენის პრინციპები. ლექსიკონების შერჩევის მთავარი კრიტერიუმი იყო ის, რომ თითოეული შედგენილია ისტორიულად სხვადასხვა პერიოდში, რამაც შესაძლოა გავლენა იქონია სწორედ ამ მახასიათებლებზე. კვლევის შედეგები შედარდა გერმანულსა და ინგლისურში არსებულ მდგომარეობას შესაბამის პერიოდში. ეს შედარება განსაზღვრავს უცხო სიტყვათა ლექსიკონების სლოვაკური ტრადიციის ადგილს ფართო ევროპულ კონტექსტში.

თეზისი ქართულად თარგმნა თამარ კვიციანმა.

References

Béjoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.

Eberhard, D., Simons, G. and C. Fennig (eds). (2019). *Ethnologue: Languages of the World.* Twenty-second edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

Hausmann, F. J. (1985). Lexicographie. In Schwarze, Ch. and D., Wunderlich (eds), *Handbuch der Lexikologie*. Königstein/Ts.: Athenäum, 367-411.

Osselton, N. E. (1983). 'On the history of dictionaries: the history of English-language dictionaries.' In Hartmann, R. R. K. (ed), *Lexicography: Principles and practice*. London: Academic Press, 13-21.

Swanepoel, P. (2003). 'Dictionary typologies: A pragmatic approach.' In Van Sterkenburg, P. (ed), *A Practical Guide to Lexicography*. Amsterdam: John Benjamins, 44-69.

SCSVS. (1939). *Slovník cudzích slov a výrazov v slovenčine*. [Dictionary of Foreign Words and Expressions in Slovak]. Prídavok, J. M. (ed). Praha: Československá grafická Unie úč. Spol.

SCS. (1953). *Slovník cudzích slov* (1. vyd.) [Dictionary of Foreign Words, 1st edition]. Hollý, A. and Ľ. Zúbek (eds). Bratislava: Tatran.

SCS. (1966). *Slovník cudzích slov* (2. zrev. vyd.) [Dictionary of Foreign Words, 2nd revised edition]. Šaling, S., Šalingová, M., and O. Peter (eds). Bratislava: Slovenské pedagogické nakladateľstvo.

VSCS. (1997). *Veľký slovník cudzích slov*. (1. vyd.). [Large Dictionary of Foreign Words, 1st edition]. Šaling, S., Ivanová-Šalingová, M., and Z. Maníková (eds). Veľký Šariš: Vydavateľstvo SAMO.

SCSA. (2005). *Slovník cudzích slov (akademický)*. (2. doplnené a upravené vydanie). [Dictionary of Foreign Words (Academic), 2nd revised edition]. Balážová, Ľ., and J. Bosák (eds). Bratislava: Slovenské pedagogické nakladateľstvo – Mladé letá.

Dictionary Without Borders – a Hungarian Case Study

P. Márkus, Katalin
M. Pintér, Tibor
Károli Gáspár University of the Reformed Church in Hungary
p.markus.kata@kre.hu
m.pinter.tibor@kre.hu

Following the Treaty of Trianon in 1920, the borders of Hungary were reorganised, and with minor exceptions, these newly created borders have remained largely unchanged to the present day. The establishment of these new political borders also resulted in the displacement of over three million Hungarians, who consequently found themselves residing outside the borders of their mother country. These communities continue to reside in the border regions of Hungary's neighbouring states, and the vast majority have maintained their Hungarian identity and native language dominance. Prior to the political transition of 1989, the language use of Hungarians residing beyond the borders received comparatively less attention than it has in the last three decades.

Hungarian is a pluricentric language, meaning it is spoken in a variety of forms in different areas, including the spheres of state administration, education, mass communication, religious ceremonies, cultural life, and scientific research in several countries of the Carpathian Basin. In 2001, the *Termini Hungarian Language Research Network* was established with the aim of uniting Hungarian linguistic research institutions operating in neighbouring countries. The network's linguists initiated a comprehensive lexical documentation project (viz. *Termini Online Hungarian–Hungarian Dictionary and Database*), termed the 'debordering' dictionary project, with the objective of cataloguing the distinctive lexical elements of Hungarian dialects used beyond the borders of Hungary. These elements were to be incorporated into Hungarian dictionaries and reference books. This initiative sought to replace linguistic divergence with a policy of lexical convergence. The linguists of the Termini Hungarian Language Research Network advocated for the inclusion of lexical elements frequently used by the members of Hungarian-speaking communities outside Hungary (which differ from the standard variety) alongside standard Hungarian lexemes in newly developed lexicographical works and dictionaries (e.g. the *Concise Dictionary of Hungarian; Concise Dictionary of Hungarian Orthography; Dictionary of Foreign Words and Phrases*).

The *Termini Online Hungarian–Hungarian Dictionary and Database* describes the lexicon of the Hungarian language as spoken in the countries surrounding Hungary. It is considered to be a general dictionary of Present-day Hungarian and stands as a resource for several linguistic research conducted on contact varieties of Hungarian. Each dictionary entry contains authentic example sentences to illustrate the use of the headword, making it possible to examine the special use of a word or construction in a grammatical and pragmatic context. The lexicographical database is edited online in eight countries. The dictionary was created as an online source for linguistic research and as a material to gain knowledge on the lexical system of contact varieties of the Hungarian language.

The presentation demonstrates how to eliminate bottlenecks in the editing process and outlines the methodological aspects (best practices) that allow the dictionary to be integrated into the teaching process – since the National Core Curriculum includes the study of Hungarian language varieties spoken in the neighbouring countries. Online editing makes it possible for the dictionary to expand – even simultaneously – as a result of activity in eight countries, making it an up-to-date and inexhaustible resource for researchers and students alike. In the presentation, the novelties and peculiarities of the dictionary will be highlighted, touching on the following topics: dictionary structure, IT support, database character, multimedia elements, and finally, effective strategies and techniques to support the teaching of dictionary use.

ლექსიკონი საზღვრების გარეშე – უნგრული მაგალითი

კატალინ პ. მარკუში
ტიბორ მ. პინტერი
რეფორმატული ეკლესიის კაროლი გაშპარის სახელობის უნივერსიტეტი, უნგრეთი
p.markus.kata@kre.hu
m.pinter.tibor@kre.hu

1920 წელს დადებული ტრიანონის ხელშეკრულების შედეგად შეცვლილი უნგრეთის საზღვრები, მცირე გამონაკლისების გარდა, დღემდე თითქმის უცვლელად შემორჩა. საზღვრების ცვლილებამ და ახალი პოლიტიკური საზღვრების დაწესებამ სამ მილიონზე მეტი უნგრელი ლტოლვილად აქცია, რადგან ისინი აღმოჩნდნენ საკუთარი სამშობლოს ფარგლებს მიღმა. ეს უნგრული თემები ამჟამად ცხოვრობენ უნგრეთის მეზობელი სახელმწიფოების სასაზღვრო რეგიონებში და მათი უმრავლესობა ინარჩუნებს უნგრულ იდენტობასა და მშობლიურ ენას. 1989 წლის პოლიტიკურ ცვლილებამდე, საზღვრებს გარეთ მცხოვრები უნგრელების ენის გამოყენების საკითხს შედარებით ნაკლები ყურადღება ექცეოდა ბოლო სამ ათწლეულთან შედარებით.

უნგრული არის პლურიცენტრული ენა, რაც ნიშნავს იმას, რომ ის სხვადასხვა ფორმით გამოიყენება სხვადასხვა რეგიონში, მათ შორის: სახელმწიფოს ადმინისტრაციაში, განათლებაში, მასობრივი ინფორმაციის საშუალებებში, რელიგიურ ცერემონიებში, კულტურულ ცხოვრებასა და სამეცნიერო კვლევებში, კარპატების (პანონიის) აუზის რამდენიმე ქვეყანაში. 2001 წელს დაარსდა **Termini Hungarian Language Research Network** (ტერმინის უნგრული ენის კვლევის ქსელი), რომელიც მიზნად ისახავდა მეზობელ ქვეყნებში მოქმედი უნგრული ლინგვისტური კვლევითი ინსტიტუტების გაერთიანებას. ამ ქსელში ჩართულმა ლინგვისტებმა დაიწყეს მასშტაბური ლექსიკური დოკუმენტაციის პროექტი (ტერმინის უნგრულ-უნგრული ონლაინლექსიკონი და მონაცემთა ბაზა), რომელსაც "საზღვრების გაუქმების" ლექსიკონის პროექტი ეწოდა. მისი მიზანი იყო უნგრული დიალექტების იმ განსაკუთრებული ლექსიკური ელემენტების აღწერა, რომლებიც გამოიყენება უნგრეთის საზღვრებს მიღმა. ეს ელემენტები უნდა ჩართულიყო უნგრულ ლექსიკონებსა და ცნობარებში. ეს ინიციატივა მიზნად ისახავდა ენობრივი განსხვავებების ჩანაცვლებას ლექსიკური დაახლოების პოლიტიკით. Termini-ს ქსელის ლინგვისტები მხარს უჭერდნენ სტანდარტული უნგრული ლექსემების გვერდით, საზღვრებს გარეთ მცხოვრები უნგრულენოვანი თემების მიერ ფართოდ გამოყენებული (სტანდარტული ვარიანტისგან განსხვავებული) ლექსიკური ერთეულების შეტანას ახალ ლექსიკოგრაფიულ ნაშრომებსა და ლექსიკონებში (მაგ., *Concise Dictionary of Hungarian*, *Concise Dictionary of Hungarian Orthography*, *Dictionary of Foreign Words and Phrases*).

Termini-ს უნგრულ-უნგრული ონლაინლექსიკონი და მონაცემთა ბაზა აღწერს უნგრული ენის ლექსიკას, როგორც ის გამოიყენება უნ-

გრეთის მიმდებარე ქვეყნებში. იგი ითვლება თანამედროვე უნგრული ენის ზოგად ლექსიკონად და წარმოადგენს რესურსს უნგრული ენის ვარიანტებზე ჩატარებული რამდენიმე ლინგვისტური კვლევისათვის. ლექსიკონის თითოეული სიტყვა-სტატია შეიცავს ავთენტიკურ საილუსტრაციო წინადადებებს, რომლებიც ასახავს სასათაურო სიტყვის გამოყენებას, რაც შესაძლებელს ხდის სიტყვის ან კონსტრუქციის განსაკუთრებული გამოყენების შესწავლას გრამატიკულ და პრაგმატულ კონტექსტში. ლექსიკოგრაფიული მონაცემთა ბაზა რედაქტირებულია ონლაინრეჟიმში რვა ქვეყანაში. ლექსიკონი შეიქმნა, როგორც ონლაინწყარო ლინგვისტური კვლევისთვის და როგორც მასალა უნგრული ენის ვარიანტების ლექსიკური სისტემის შესახებ ინფორმაციის მისაღებად.

პრეზენტაცია წარმოადგენს, თუ როგორ უნდა აღმოიფხვრას რედაქტირების პროცესში არსებული შეფერხებები და განიხილავს მეთოდოლოგიურ ასპექტებს (საუკეთესო პრაქტიკებს), რომლებიც ლექსიკონის ინტეგრირებას უზრუნველყოფს სასწავლო პროცესში – განსაკუთრებით იმის გათვალისწინებით, რომ ეროვნული სასწავლო გეგმა მოიცავს მეზობელ ქვეყნებში გამოყენებული უნგრული ენის ვარიანტების შესწავლას. ონლაინრედაქტირება შესაძლებელს ხდის ლექსიკონის ერთდროულ განვრცობას რვა ქვეყნის სპეციალისტების მუშაობის შედეგად, რაც მას აქცევს უახლეს და ამოუწურავ რესურსად როგორც მკვლევრებისთვის, ასევე მოსწავლეებისთვის. პრეზენტაციაში ყურადღება გამახვილდება ლექსიკონის შესახებ სიახლეებსა და თავისებურებებზე, მათ შორის: ლექსიკონის სტრუქტურაზე, ტექნოლოგიურ მხარდაჭერაზე, მონაცემთა ბაზის მახასიათებლებზე, მულტიმედიურ ელემენტებზე და ბოლოს ლექსიკონის გამოყენების სწავლების ეფექტიან სტრატეგიებსა და ტექნიკებზე.

Integrating AI into the Term Creation Process: Case Study of English and Georgian Terms of Emerging Technologies

Tchigladze, Ketevani
Ilia State University, Georgia
European Master in Lexicography (EMLex) alumni
ketitchighladze@gmail.com

As the global pace of technological innovation accelerates, new concepts are emerging — particularly in such fast-evolving areas as blockchain, cryptocurrency, artificial intelligence, finance based on modern technologies, etc. This process poses a particular challenge to languages represented in the digital world with limited resources, such as the Georgian language. Without a systematic process of creating new Georgian terms, it is obvious that direct transliteration of English terms occurs uncontrollably, which in turn makes the Georgian-speaking community excessively dependent on foreign-language terminology, and creates significant challenges in terms of perception and understanding of new concepts and their content. Therefore, developing timely and accurate terminology for such emerging concepts is not only a linguistic challenge but also a cultural and strategic necessity. The paper explores how artificial intelligence (AI), such as large language models (e.g., ChatGPT and others), can be integrated into the term creation process as a support tool for terminologists.

Traditionally, terminologists possess strong linguistic and lexicographic expertise, although they are not always (very rarely) specialists in the fields they may work on. On the other hand, developing accurate and valuable terms, especially for the emerging and technical concepts, requires in-depth and comprehensive research of the domain-specific resources and scholarly articles. This is one of the important parts of the work of terminologists. This step is time-consuming and resource-intensive, particularly when the concepts under research are interdisciplinary or still evolving. At this stage, the work of a terminologist often also requires the involvement of field experts to discuss the content and context of the use of new terms. AI can be considered an important alternative at this stage and can even be seen as a virtual field expert. AI platforms have access to vast databases of expert-level knowledge, explanations, case studies, and conceptual frameworks. We can say that it potentially has much more information about a given field than any human expert or group of experts in the same area. In addition, available AI models can interact with humans and clearly and intelligibly explain simple or complex ideas and concepts in a way we understand (although here it also worth considering the high probability of errors!).

Considering the approach used in our research, AI can play the role of a field expert for terminologists, constantly available and ready to answer any of our questions and discuss this or that issue with us. It can also replace or significantly reduce the lengthy process of studying domain-specific literature.

The paper explores how artificial intelligence can be used to analyze the emerging concepts and corresponding English terms in the areas of cryptocurrency and blockchain technology in order to develop new Georgian terms. In

particular, at the stage of explanation, clarification and semantic decomposition of the given concepts.

As part of an iterative dialogue with AI, we prompted it to define selected English-language terms, identify the central and peripheral semantic components of the corresponding concepts, structurally decompose the concept and compare/contrast it with related concepts. For example, in the area of blockchain technologies, for the English terms – utility token and security token – AI decomposed the conceptual structure, identified the key, nuanced distinctions in terms of its function and purpose, and discussed their usage contexts. This deconstruction and juxtaposition of concepts sparked the ideas about which component could be used at the stage of naming the corresponding Georgian terms.

AI-generated Georgian terms suggestions (due to its limited "knowledge" of the Georgian language) are linguistically imperfect and mismatched, however, it inspires creative thinking. It is known that general language lexical units are often embedded in English terms (usually in multi-word terms). In addition, one of the approaches to naming target language terms is to transfer the general language lexical units of the source term, not by transliteration, but by using translation equivalents. Accordingly, as part of our research, regarding general lexical units used in the selected terms, we asked artificial intelligence questions in two directions: 1. Why could this lexical unit have been selected for this term, what connects it to the concept, and which semantic component of the concept does it specifically reflect? 2. What are the close synonyms of the general language lexical unit used in the term in English? AI-generated answers to such questions, on the one hand, helped us to better understand on what basis the given concept could be expressed with the given lexical units and to what extent the same tactics could be used in relation to the words of the Georgian language. On the other hand, if the Georgian equivalent of a specific English word was not relevant (did not sound natural), to what extent could the Georgian equivalents of the English synonyms of the given English word be more suitable in the process of naming the Georgian term under consideration. It should be noted that using this method, artificial intelligence can inspire the human mind to use a number of Georgian equivalents that are within the same conceptual framework. Instead of using calques and literal translations, this approach can help us come up with valuable and intuitive Georgian terms that sound natural and are acceptable to the Georgian vocabulary and also express the content of the concept in a clear and precise way.

In conclusion, at present AI cannot be considered as a reliable and ultimately trustworthy source for the development of Georgian terms through direct responses (with the rapid development of AI models, their 'knowledge' of the Georgian language may also be further refined, though), although it can provide us with significant assistance in this process. In particular, we can confidently say that AI can fulfill the role of an ever-available virtual field expert (in all fields) and replace the long and laborious process of reading and exploring the vast scientific and technical literature. On the other hand, it can be of considerable help in the process of finding Georgian equivalents for emerging English terms by explaining the reasons why given lexical units are used in English terms or by providing comprehensive information on possible English synonyms. Of course, the final decision on the term equivalents requires a human terminologist's judgment and

inspiration, however, AI significantly reduces the time and laborious process of exploring and understanding new concepts, and also helps us in generating new ideas. This will potentially help accelerate the process of developing contextually appropriate and natural-sounding Georgian terminology for the new concepts in emerging fields, thereby ensuring that Georgian terminology keeps pace with global trends.

Keywords: *terminology, artificial intelligence, ChatGPT, term nomination, technologies*

ხელოვნური ინტელექტის ინტეგრირება ტერმინების შექმნის პროცესში: უახლესი ტექნოლოგიების დარგის ინგლისური და ქართული ტერმინების მაგალითზე

ქეთევან ჭიღლაძე
ილიას სახელმწიფო უნივერსიტეტი, საქართველო
ევროპული მაგისტრი ლექსიკოგრაფიაში
ketitchighladze@gmail.com

გლობალურ დონეზე ტექნოლოგიური ინოვაციების ტემპის აჩქარებასთან ერთად, ჩნდება ახალი ცნებები, განსაკუთრებით ისეთ უახლეს მიმართულებებში, როგორებიცაა ბლოკჩეინი, კრიპტოვალუტა, ხელოვნური ინტელექტი, თანამედროვე ტექნოლოგიებზე დაფუძნებული ფინანსები და ა.შ. ეს პროცესი განსაკუთრებული გამოწვევების წინაშე აყენებს ციფრულ სამყაროში შეზღუდული რესურსებით წარმოდგენილ ენებს, მაგალითად, ქართულ ენას. ახალი ქართული ტერმინების შექმნის სისტემატური პროცესის გარეშე, ცხადია, უკონტროლოდ ხდება ინგლისური ტერმინების პირდაპირი ტრანსლიტერაცია, რაც თავის მხრივ ქართულენოვან საზოგადოებას გადაჭარბებულად დამოკიდებულს ხდის უცხოენოვან ტერმინოლოგიაზე, ხოლო ახალი ცნებებისა და შინაარსის აღქმისა და გაგების თვალსაზრისით მათ მნიშვნელოვან სირთულეებს უქმნის. აქედან გამომდინარე, უახლესი ცნებებისთვის დროული და ზუსტი ტერმინოლოგიის შექმნა არა მხოლოდ ენობრივი გამოწვევაა, არამედ კულტურულ და სტრატეგიულ აუცილებლობასაც წარმოადგენს. წინამდებარე ნაშრომი იკვლევს თუ როგორ შეიძლება ხელოვნური ინტელექტის (AI), როგორიცაა დიდი ენობრივი მოდელები (მაგ., ChatGPT და სხვები), ინტეგრირება ტერმინების შექმნის პროცესში, როგორც ტერმინოლოგიის სპეციალისტების დამხმარე საშუალება.

ტრადიციულად, ტერმინოლოგიის სპეციალისტებს ცოდნა და გამოცდილება ლინგვისტური და ლექსიკოგრაფიული მიმართულებით აქვთ, თუმცა ისინი ყოველთვის არ არიან (ძალიან იშვიათად თუ არიან) იმ საგნისა და დარგის ექსპერტები, რის ტერმინოლოგიაზეც შეიძლება მოუწიოთ მუშაობა. მეორე მხრივ კი, ზუსტი და ღირებული ტერმინების შექმნა, განსაკუთრებით უახლესი და ტექნიკური ცნებებისთვის, მოითხოვს დარგობრივი ლიტერატურისა და სამეცნიერო კვლევების სიღრმისეულ და ყოვლისმომცველ კვლევას, რაც ტერმინოლოგიის სპეციალისტების სამუშაოს ერთ-ერთი მნიშვნელოვანი ნაწილია. ტერმინოლოგიის სპეციალისტების მუშაობაში ეს მნიშვნელოვანი გამოწვევაა და დიდ დროსა და რესურსებს მოითხოვს, განსაკუთრებით იმ შემთხვევაში, თუ სამუშაო ცნებები ინტერდისციპლინურია ან კვლავ დახვეწის პროცესშია. ტერმინოლოგიის სპეციალისტის მუშაობაში ამ ეტაპზე ხშირად ასევე საჭიროა დარგის ექსპერტების ჩართულობა და მათთან ერთად ახალი ცნებების შინაარსისა და გამოყენების კონტექსტების განხილვაც. AI შეგვიძლია

ამ ეტაპზე მნიშვნელოვან ალტერნატივად განვიხილოთ და ვირტუალურ დარგის ექსპერტადაც კი მოვიაზროთ. AI პლატფორმებს წვდომა აქვთ ექსპერტული დონის ცოდნის, განმარტებების, შემთხვევათა ანალიზებისა და კონცეპტუალური ჩარჩოების უზარმაზარ მონაცემთა ბაზებზე. შეიძლება ითქვას, რომ მას აქვს პოტენციურად ბევრად მეტი ინფორმაცია ამა თუ იმ დარგის შესახებ, ვიდრე ამავე დარგის რომელიმე ექსპერტს (ადამიანს) ან ექსპერტთა ჯგუფს. ამასთანავე, ხელმისაწვდომ AI მოდელებს შეუძლიათ ადამიანებთან ინტერაქცია და ჩვენთვის გასაგებ ენაზე მარტივი თუ კომპლექსური იდეებისა და კონცეფციების ნათლად და გასაგებად ახსნა (თუმცა აქვე გასათვალისწინებელია მათ მიერ შეცდომების დაშვების დიდი ალბათობაც!).

ნაშრომში წარმოდგენილი მიდგომის გათვალისწინებით, AI-მ შეიძლება შეასრულოს ტერმინოლოგიის სპეციალისტებისთვის დარგის ექსპერტის როლი, რომელიც მუდმივად ხელმისაწვდომია და მზადაა ჩვენს ნებისმიერ კითხვას უპასუხოს და ესა თუ ის საკითხი ჩვენთან ერთად განიხილოს. ასევე, ჩაანაცვლოს ან მნიშვნელოვანწილად შეამციროს დარგობრივი ლიტერატურის შესწავლის ხანგრძლივი პროცესი.

ნაშრომში გამოკვლეულია თუ როგორ შეიძლება, ახალი ქართული ტერმინების შემუშავების მიზნით, გამოვიყენოთ ხელოვნური ინტელექტი კრიპტოვალუტისა და ბლოკჩეინის ტექნოლოგიების მიმართულებით არსებული უახლესი ცნებებისა და შესაბამისი ინგლისური ტერმინების ანალიზის პროცესში. კერძოდ, აღნიშნული ცნებების ახსნა-დაზუსტებისა და სემანტიკური დეკომპოზიციის ეტაპზე. AI-თან იტერაციული დიალოგის ფარგლებში, მას ვთხოვეთ შერჩეული ინგლისურენოვანი ტერმინების განმარტება, შესაბამისი ცნების ცენტრალური და პერიფერიული სემანტიკური კომპონენტების გამოყოფა, ცნების სტრუქტურულად ჩაშლა და დაკავშირებულ ცნებებთან შედარება/შეპირისპირება. მაგალითად, ბლოკჩეინის მიმართულებით, შემდეგი ინგლისურენოვანი ტერმინებისთვის – utility token და security token – ხელოვნურმა ინტელექტმა დაშალა მათი კონცეპტუალური სტრუქტურა, გამოყო ძირითადი, ნიუანსური სხვაობები ფუნქციური და მიზნობრივი თვალსაზრისით და განიხილა გამოყენების კონტექსტები. ცნებების ამგვარად დეკონსტრუქციამ და შეპირისპირებამ საშუალება მოგვცა სხვადასხვა მიმართულებით გვეფიქრა იმ საკითხზე, თუ რომელი კომპონენტი შეიძლება გამოგვეყენებინა შესაბამისი ქართული ტერმინის სახელდების ეტაპზე. ხელოვნურ ინტელექტს (ქართული ენის მწვრი „ცოდნიდან“ გამომდინარე) არ შეუძლია თავად შემოგვთავაზოს შესაბამისი ქართულენოვანი ტერმინები, თუმცა, მას შეუძლია მნიშვნელოვანი ბიძგი მისცეს ჩვენს შთაგონებას ამ მიმართულებით. ცნობილია, რომ ინგლისურენოვან ტერმინებში (უფრო მეტად მრავალსიტყვიან ტერმინებში) ხშირად გამოიყენება ისეთი ლექსიკური ერთეულები, რომლებიც ზოგადი ინგლისური ენის ლექსიკას განეკუთვნება. ასევე ხშირად, ტერმინების სახელდების ერთ-ერთი მიდგომაა ინგლისურენოვან ტერმინში გამოყენებული ზოგადი ლექსიკის სიტყვის მიხედვით ქართულენოვანი დასახელების შექმნა, არა ტრანსლიტერაციის, არამედ მათი თარგმნის გზით. შესაბამისად, კვლევის

ფარგლებში, შერჩეულ ტერმინებში გამოყენებული ზოგადი ლექსიკის სიტყვებთან მიმართებით, ხელოვნურ ინტელექტს კითხვები დავუსვით ორი მიმართულებით: 1. რატომ შეიძლებოდა აერჩიათ ეს ლექსიკური ერთეული ამ ტერმინისთვის, რა აკავშირებს მას ცნებასთან და კონკრეტულად ცნების რომელ სემანტიკურ კომპონენტს ასახავს? 2. ინგლისურ ენაში რომლებია ინგლისურ ტერმინში გამოყენებული ზოგადი ლექსიკის სიტყვის ახლო სინონიმები. ასეთი კითხვები, ერთი მხრივ დაგვეხმარა კარგად გაგვეაზრებინა თუ რის საფუძველზე შეიძლებოდა მოცემული ცნება მოცემული ლექსიკური ერთეულებით გამოეხატათ და რამდენად შეიძლებოდა იგივე ტაქტიკა გამოგვეყენებინა ქართული ეკვივალენტის შერჩევისას. მეორე მხრივ, თუ კონკრეტული ინგლისური სიტყვის ქართული ეკვივალენტი არ გამოდგებოდა (არ ჟღერდა ბუნებრივად), მოცემული ინგლისური სიტყვის ინგლისურივე სინონიმების ქართული ეკვივალენტები რამდენად შეიძლებოდა უფრო შესაფერისი ყოფილიყო განსახილველი ტერმინის სახელდების პროცესში. უნდა აღინიშნოს, რომ ამ მეთოდის გამოყენებით შეიძლება ხელოვნურმა ინტელექტმა ადამიანის გონებას შთააგონოს არაერთი ისეთი ქართული ეკვივალენტის გამოყენება, რომელიც იგივე კონცეპტუალურ ჩარჩოშია. კალკებისა და სიტყვასიტყვითი თარგმანის გამოყენების ნაცვლად ეს მიდგომა შეიძლება დაგვეხმაროს ისეთი ღირებული და ინტუიციური ქართული ტერმინების მოფიქრებაში, რომელიც ქართული ლექსიკისთვის ბუნებრივად და მისაღებად ჟღერს და ცნების შინაარსსაც გასაგებად და ზუსტად გამოხატავს.

დასკვნის სახით, ხელოვნური ინტელექტი ქართულენოვანი ტერმინების შექმნისას ამჟამად (თუმცა, AI მოდელების სწრაფად განვითარების კვალდაკვალ, შეიძლება მისი ქართული ენის „ცოდნაც“ უფრო მეტად დაიხვეწოს) ვერ განიხილება სანდო და საბოლოოდ სარწმუნო წყაროდ პირდაპირი თარგმნის თვალსაზრისით, თუმცა მას ნამდვილად შეუძლია ამ პროცესში მნიშვნელოვანი დახმარება გაგვიწიოს. კერძოდ, დარწმუნებით შეიძლება ვთქვათ, რომ მას შეუძლია შეასრულოს ყოველთვის ხელმისაწვდომი ვირტუალური დარგის ექსპერტის როლი (ყველა დარგში) და ჩაანაცვლოს სამეცნიერო ტექნიკური ლიტერატურის კითხვის ხანგრძლივი და შრომატევადი პროცესი. მეორე მხრივ, მას შეუძლია მნიშვნელოვნად დაგვეხმაროს ტერმინებისთვის ქართული ეკვივალენტების მოფიქრების პროცესში, ინგლისურენოვან ტერმინებში გამოყენებული ლექსიკური ერთეულების გამოყენების მიზეზის ახსნით თუ სავარაუდო სინონიმური წყვილების შესახებ მრავლისმომცველი ინფორმაციის მოწოდების გზით. რა თქმა უნდა, ტერმინების საბოლოო ვერსიაზე გადაწყვეტილება კვლავ ტერმინოლოგიის სპეციალისტმა (დარგის ექსპერტებთან ერთად) უნდა მიიღოს ადამიანური განსჯისა და შთაგონების საფუძველზე, თუმცა, AI მნიშვნელოვნად ამცირებს ახალი ცნებების შესწავლისა და გააზრების შრომატევად პროცესს და დროს, ასევე გარკვეულწილად გვეხმარება ახალი იდეების გენერირებაში. ეს კი პოტენციურად ხელს შეუწყობს ახალი და განვითარებადი დარგების უახლესი ცნებებისთვის კონტექსტურად შესაფერისი და ქართული ენისთვის ბუ-

ნებრივი ტერმინების შექმნის პროცესის დაჩქარებას, რაც მცირე ლინგვისტური რესურსების მქონე ქართულ ტერმინოლოგიას და ტერმინოლოგიის სპეციალისტებს დაეხმარება არ ჩამორჩეს და ფეხი აუწყოს ამ მიმართულებით არსებულ გლობალურ ტენდენციებს.

საკვანძო სიტყვები: *ტერმინოლოგია, ხელოვნური ინტელექტი, ChatGPT, ტერმინების სახელდება, ტექნოლოგიები*

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

Tchighladze, Salome
Ilia State University, Georgia
salome.tchigladze.1@iliauni.edu.ge

In my talk I will introduce a model of a Georgian-English learner's online dictionary of collocations, developed in accordance with the principles of collocations' dictionaries and the theory of lexical functions proposed by Igor Mel'čuk.

The significance of word combinations in the language acquisition process was already emphasized in the 1920s–30s by Harold Palmer and Albert Hornby. However, the linguistic phenomenon of *collocation*, as a distinct term, was first introduced in the 1950s by the British linguist J. R. Firth. Building on Firth's ideas, John Sinclair's corpus-based research further elaborated the practical application of collocations, establishing them as a core principle of corpus linguistics and lexical analysis.

The development of collocations' dictionaries as a distinct lexicographic genre has been facilitated by advancements in computational and corpus linguistics. This type of dictionary gained particular importance once large-scale processing of linguistic data became feasible. These dictionaries aim to describe and document natural language use, emphasizing not only individual lexical units and their definitions, but also their collocability. They rely on extensive language corpora, which contributes to their reliability and accuracy. An early and influential example is the COBUILD project (Collins Birmingham University International Language Database), which produced the first corpus-based dictionary, the *COBUILD English Dictionary*. The project prioritized lexical co-occurrence over traditional grammar rules and presented real-life usage patterns drawn from a representative electronic corpus of English.

The concept of collocation also plays a central role in Mel'čuk's Explanatory Combinatorial Dictionary (ECD) and his theory of lexical functions. According to this theory, words are connected within particular semantic contexts through predictable relations, termed *lexical functions*. Approximately 70 such functions are distinguished, and they are categorized into syntagmatic and paradigmatic functions. Syntagmatic functions describe lexical co-occurrence patterns, specifying which nouns, adjectives, adverbs, or verbs are typically used with a given lexeme. These functions help identify semantically and contextually appropriate collocates. In contrast, paradigmatic functions (e.g., synonymy, antonymy etc.) describe lexical units that share a semantic category and may substitute the target lexeme, while syntagmatic functions co-occur with it in actual usage.

This study proposes a methodology for compiling collocation entries for a Georgian-English learner's dictionary, grounded in the aforementioned theoretical frameworks. Each dictionary entry classifies collocates according to parts of speech and includes relevant lexical functions. This methodology provides learn-

ers with insight into the combinatorial potential of lexical units and supports the syntactic construction of grammatically and semantically coherent expressions.

Only those lexical functions that are relevant to Learner's dictionary—both syntagmatic and paradigmatic—have been selected for inclusion. The presentation will present and analyze in detail a model of an entry for nouns developed according to this methodology. This approach supports the creation of an active, synthesis-oriented learner's dictionary that will significantly enhance the efficient use of lexicographic resources in both teaching and research contexts.

Each entry not only lists collocations but also provides authentic usage examples with English translations, enabling learners to grasp the practical application of words in real-life discourse. The dictionary's structure draws on the organizational principles of the *Oxford Collocations Dictionary* and the *Macmillan Collocations Dictionary*, where collocates are grouped by part of speech and, in some cases, include extended phrases. This enhances the practical utility of the dictionary for learning purposes.

The proposed model—based on collocational data and enriched with lexical functions—is a novel contribution to the field of Georgian-English Learner's lexicography.

Keywords*: collocations dictionary, lexical functions, a model of a dictionary entry, Georgian-English lexicography, dictionary microstructure*

ქართულ-ინგლისური შესიტყვებების სასწავლო ონლაინლექსიკონის მოდელი
(შესიტყვებების ლექსიკონების პრინციპებისა და ი. მელჩუკის ლექსიკური ფუნქციების თეორიის გამოყენებით)

სალომე ჭიღლაძე
ილიას სახელმწიფო უნივერსიტეტი
salome.tchigladze.1@iliauni.edu.ge

მოხსენებაში წარმოვადგენ ქართულ-ინგლისური შესიტყვებების სასწავლო ონლაინლექსიკონის მოდელს, რომელიც დაფუძნებულია კოლოკაციების ლექსიკონების შედგენის პრინციპებსა და იგორ მელჩუკის ლექსიკური ფუნქციების თეორიაზე.

ჯერ კიდევ 1920-30-იან წლებში ჰაროლდ პალმერი და ალბერტ ჰორნბი საუბრობენ სიტყვათა კომბინაციის მნიშვნელობაზე ენის შესწავლის პროცესში, თუმცა ეს ენობრივი მოვლენა როგორც ტერმინი, ლინგვისტიკაში პირველად ბრიტანელმა ენამეცნიერმა ჯონ რუპერტ ფერთმა შემოიტანა 1950-იან წლებში. ფერთის შემდეგ ჯონ სინკლერის კვლევებმა მეტად განავრცო და გააღრმავა კოლოკაციების პრაქტიკული გამოყენების მნიშვნელობა, რაც კორპუსული კვლევებისა და ლექსიკური ანალიზის ერთ-ერთ მთავარ პრინციპად იქცა.

კოლოკაციების ლექსიკონის, როგორც ჟანრის განვითარებას ხელი შეუწყო კომპიუტერული და კორპუსული ლინგვისტიკის განვითარებამ. კერძოდ, ამ ტიპის ლექსიკონები აქტუალური გახდა მას შემდეგ რაც შესაძლებელი გახდა ენობრივი მონაცემების კომპიუტერული დამუშავება. ლექსიკონის ეს ჟანრი მიზნად ისახავს ბუნებრივი ენის გამოყენების აღწერას და მის დოკუმენტირებას. ამ ტიპის ლექსიკონები ორიენტირებულია არა მხოლოდ ცალკეულ სიტყვებზე და მათ განმარტებებზე, არამედ ამ სიტყვათა ურთიერთკავშირზე, მათ შესიტყვებადობაზე. შესიტყვებების ლექსიკონები უმეტესად ეყრდნობა დიდი მოცულობის ენობრივ კორპუსებს, რაც მათ უფრო სანდოს და ზუსტს ხდის. ამ მხრივ აღსანიშნავია პროექტი COBUILD (Collins Birmingham University International Language Database), რომელიც იყო პირველი ელექტრონული კორპუსი. მასში თავმოყრილია ინგლისური ენის ბუნებრივი გამოყენების მაგალითები და მათზე დაყრდნობით შეიქმნა პირველი კორპუსზე დაფუძნებული ლექსიკონი – COBUILD English Dictionary. ამ პროექტით სინკლერმა გრამატიკულ წესებზე მეტად წინა პლანზე წამოსწია სიტყვათა შესიტყვებადობის საკითხი.

კოლოკაციების საკითხი აქტუალურია იგორ მელჩუკის როგორც განმარტებით კომბინატორული ლექსიკონის თეორიაში, ისე მისი ლექსიკური ფუნქციების თეორიაშიც. ლექსიკური ფუნქციების თეორიის მიხედვით, სიტყვები სხვადასხვა სემანტიკურ კონტექსტში გარკვეულ კავშირში არიან ერთმანეთთან და ამ კავშირს შეგვიძლია ლექსიკური

ფუნქცია ვუწოდოთ. ამ თეორიაში სულ აღწერილია 70-მდე ლექსიკური ფუნქცია, რომლებსაც მელჩუკი ჰყოფს სხვადასხვა კატეგორიებად, მათ შორის სინტაგმატურ და პარადიგმატულ ფუნქციებად. სინტაგმატური ლექსიკური ფუნქციები ძირითადად აღწერენ სხვადასხვა მეტყველების ნაწილებს და გვთავაზობენ იმ არსებით თუ ზედსართავ სახელებს, ზმნიზედებსა და ზმნებს, რომლებიც გვხვდება მოცემულ ლექსემასთან. შესაბამისად ფუნქციები გვაძლევენ სემანტიკურად მოსალოდნელ და შესაბამის სიტყვათა წყვილების კომბინაციებს. პარადიგმატული ლექსიკური ფუნქციები, (როგორიცაა სინონიმები, ანტონიმები, კონვერსივები და სხვა) აღწერენ სიტყვებს, რომლებიც ერთსა და იმავე სემანტიკურ კატეგორიას მიეკუთვნება და შესაძლოა ჩაანაცვლონ მოცემული ლექსემა. სინტაგმატური ლექსიკური ფუნქცია კი გამოიყენება მოცემულ სიტყვასთან ერთად.

კვლევის ფარგლებში კოლოკაციებთან დაკავშირებული თეორიის, ასევე კოლოკაციების ლექსიკონების შედგენის პრინციპებისა და მელჩუკის ლექსიკური ფუნქციების თეორიის გაცნობის შემდეგ, შემუშავდა მეთოდოლოგია. აღნიშნული მეთოდოლოგიის მიხედვით შესიტყვებები ქართულ-ინგლისური შესიტყვებების სასწავლო ლექსიკონის სიტყვა-სტატიაში კლასიფიცირდება მეტყველების ნაწილების მიხედვით და მიეთითება შესაბამისი ლექსიკური ფუნქციები. აღნიშნული მეთოდი მომხმარებელს საშუალებას აძლევს დეტალურად გაიაზროს ლექსიკური ერთეულების კომბინაციური შესაძლებლობები და საჭიროების შემთვევაში თავად ააგოს გამართული სინტაქსური კონსტრუქციები. მელჩუკის ლექსიკური ფუნქციებიდან შეირჩა მხოლოდ ის სინტაგმატური და პარადიგმატული ლექსიკური ფუნქციები, რომლებიც აქტუალურია ქართულ-ინგლისური ლექსიკოგრაფიული მიზნებისთვის შესიტყვებების სასწავლო ლექსიკონის კონტექსტში. მოხსენებაში დეტალურად განვიხილავ აღნიშნული მეთოდის მიხედვით შედგენილი არსებითი სახელის სიტყვა-სტატიის მოდელს. ვფიქრობ, შერჩეული მიდგომა ხელს შეუწყობს ტექსტის სინთეზზე ორიენტირებული, აქტიური ტიპის ლექსიკონის შექმნას, რაც მნიშვნელოვნად განაპირობებს ლექსიკოგრაფიული რესურსების ეფექტიან გამოყენებას როგორც სწავლების, ისე კვლევის კონტექსტში.

თითოეულ სიტყვა-სტატიაში განსაზღვრულია არა მხოლოდ სიტყვათა კავშირები, არამედ მათი გამოყენების მაგალითებიც ინგლისური თარგმანებით, რაც ენის შემსწავლელს აძლევს რეალურ წარმოდგენას ამა თუ იმ სიტყვის გამოყენებაზე ყოველდღიურ მეტყველებაში. სიტყვა-სტატიის სტრუქტურა ეყრდნობა ოქსფორდის შესიტყვებების ლექსიკონსა (*Oxford Collocations Dictionary*) და მაკმილანის შესიტყვებების ლექსიკონში (*Macmillan Collocations Dictionary*) კოლოკაციების წარმოდგენის პრინციპს. ამ ლექსიკონებში, თითოეულ სიტყვა-სტატიაში წარმოდეგნილი კოლოკაციები დაჯგუფებულია მეტყველების ნაწილების მიხედვით და რიგ შემთხვევებში მითითებულია ვრცელი ფრაზებიც, რაც ამ ტიპის ლექსიკონს ხდის პრაქტიკულად გამოყენებადს ენის შესწავლის მიზნისთვის.

კოლოკაციებზე დაფუძნებული, ლექსიკური ფუნქციებით გამდიდრებული ლექსიკონის მოდელი სიახლეა ქართულ-ინგლისური სასწავლო ლექსიკოგრაფიის მიმართულებით.

საკვანძო სიტყვები: *მესიტყვებების ლექსიკონი, ლექსიკური ფუნქციები, სიტყვა-სტატიის მოდელი, მელჩუკი, ქართულ-ინგლისური ლექსიკოგრაფია.*

Beyond Dictionaries: A Methodological Framework for Studying Lexical Innovation in Corpora

Tomaszewska, Aleksandra
Ogrodniczuk, Maciej
Institute of Computer Science, Polish Academy of Sciences, Poland
aleksandra.tomaszewska@ipipan.waw.pl
maciej.ogrodniczuk@ipipan.waw.pl

Traditional lexicographic methods, while essential for describing stabilized lexical units, inherently document new words with significant delays. These delays result from lexicographic processes requiring lexical stability, widespread usage, institutional recognition, and the publication timeline of dictionary entries. However, contemporary linguistic communication occurs in highly dynamic digital environments – such as online journalism, social media, blogs, discussion forums, and other interactive platforms – where lexical innovations often emerge and spread long before their formal documentation in dictionaries. Consequently, lexicographic resources rarely capture vocabulary at its earliest stages, potentially missing linguistically, socially, or culturally significant developments. Moreover, the temporal criterion commonly employed in studies of neologisms (sometimes even 20–25 years) often results in the inclusion of items that contemporary language users no longer perceive as new, and sometimes even regard as outdated.

Recent advancements in corpus linguistics, computational lexicography, and large language models offer an effective solution to these issues. These methodologies facilitate systematic, near-real-time identification of new lexical phenomena, enabling the study of lexical innovations that reflect ongoing social, cultural, and technological transformations in contemporary communication.

In this presentation, we will propose methodological criteria and a replicable analytical workflow for identifying lexical innovations adjusted to this changing reality based on corpus data, supported by computational and manual validation techniques. The proposed methodology includes corpus compilation, automated detection of candidate words, computational and manual validation, and sociolinguistically contextualized analysis. Integrating manual validation is particularly valuable because it allows human experts to interpret nuanced linguistic phenomena and contextual subtleties, complementing the efficiency and scalability of computational methods. Importantly, our approach does not seek to replace dictionary-based methodologies, but rather complements them by documenting words at the earliest stages of their use in language and by reducing the time required for identification procedures through criteria that can later be incorporated into NLP tools designed for filtering neologisms. We note that our methods do not account for semantic shifts; they focus exclusively on new lexical forms.

The proposed methodological procedure involves the following steps. First, lexical candidates are identified based on recent corpus attestations and explicit absence from authoritative dictionaries as well as historical reference corpora compiled prior to a specified cutoff date. This cutoff date varies depending on the research topic and the availability of reference corpora; however, for studies of

the most recent phenomena, we propose five years before as the cutoff. Second, to exclude isolated or idiosyncratic lexical forms, candidate neologisms must occur across more than one textual domain. Additionally, each candidate term must appear at least a few times, across multiple authors or sources. The research process also includes data filtering procedures integrating automatic and semi-manual validation techniques. Particular attention is given to common data-quality issues found in digital texts, such as typographic inconsistencies, OCR errors, or orthographic stylizations resulting from repeated letters (e.g., *yeahhh* instead of *yeah*), which often produce false positives that unnecessarily lengthen validation time and can be efficiently eliminated through predefined rule-based filters. Measures such as edit-distance metrics (e.g., Levenshtein distance) serve as additional identification criteria, as substantial differences from existing words can further indicate genuine lexical innovation rather than orthographic variants.

In our presentation, we illustrate these criteria in practice through testing our methodological framework implemented in the specialized software NeoN, which can be used by linguists with no programming skills. The presented case studies will encompass lexical innovations from contemporary linguistic domains, such as gender-inclusive language, digital communication, discourse related to large language models, and parliamentary discourse, highlighting the sociocultural significance of studying lexical innovations in authentic communicative contexts. We show the value of applying clearly defined criteria and available computational tools tailored to contemporary linguistic realities, emphasizing the dynamic nature of language evolution. We hope our methodological framework will have practical implications not only for linguists and lexicographers but also for educators, technologists, and policymakers seeking to understand and respond to rapid linguistic change.

ლექსიკონების მიღმა: მეთოდოლოგიური ჩარჩო კორპუსებში ლექსიკური სიახლეების შესწავლისათვის

ალექსანდრა ტომაშევსკა,
მაცეი ოგროდნიჩუკი

პოლონეთის მეცნიერებათა აკადემია, პოლონეთი

aleksandra.tomaszewska@ipipan.waw.pl
maciej.ogrodniczuk@ipipan.waw.pl

ტრადიციული ლექსიკოგრაფიული მეთოდები, რომლებიც მნიშვნელოვანია ენაში დამკვიდრებული ლექსიკური ერთეულების აღწერისათვის, თავისთავად ასახავენ ახალ სიტყვებსაც, თუმცა მნიშვნელოვანი დაგვიანებით. ეს შეფერხება დროის თვალსაზრისით გამოწვეულია ლექსიკოგრაფიული პროცესებით, რაც უნდა განვლოს სიტყვამ ლექსიკონში მოხვედრამდე. ეს პროცესები მოიცავს ლექსიკურ სტაბილურობას, სიტყვის ფართოდ გამოყენებას, ინსტიტუციურ აღიარებასა და ლექსიკური ერთეულის გამოქვეყნების დროს. თუმცა, თანამედროვე ეპოქაში ლინგვისტური კომუნიკაცია ძალზედ აქტიურ ციფრულ გარემოში მიმდინარეობს – როგორიცაა ონლაინჟურნალისტიკა, სოციალური ქსელები, ბლოგები, სხვადასხვა სადისკუსიო ფორუმები და სხვა საურთიერთო პლატფორმები – სადაც ხშირად თავს იჩენს ლექსიკური სიახლეები და ვრცელდება დიდი ხნით ადრე, მანამ სანამ ისინი ლექსიკონებში აისახება. მაშასადამე, ლექსიკოგრაფიული რესურსები იშვიათად აღბეჭდავს ენის ლექსიკას განვითარების ადრეულ საფეხურზე, პოტენციურად ამ დროს გამორჩება ხოლმე განვითარების ის ეტაპები, რომლებიც ლინგვისტური, სოციალური თუ კულტურული თვალსაზრისით მნიშვნელოვანია. გარდა ამისა, ნეოლოგიზმების კვლევაში ხშირად გამოყენებული დროითი კრიტერიუმი (ზოგჯერ 20–25 წელიც კი), ხშირად იწვევს ლექსიკონებში ისეთი ერთეულების შეტანას, რომლებიც უკვე არა თუ ნეოლოგიზმად, არამედ მოძველებულადაც კი შეიძლება მიიჩნიოს მომხმარებელმა.

უახლესი მიღწევები კორპუსის ლინგვისტიკაში, კომპიუტერულ ლექსიკოგრაფიაში და დიდი ენობრივი მოდელები გვთავაზობს ამ საკითხების ეფექტურად გადაჭრის გზებს. ეს მეთოდოლოგიები ამარტივებს ახალი ლექსიკური ფენომენის სისტემატურ და რეალურ დროსთან მაქსიმალურად მიახლოებულად იდენტიფიკაციას. ეს კი იძლევა შესაძლებლობას ლექსიკური სიახლეები ისე იქნეს შესწავლილი, რომ ისინი ასახავდნენ მიმდინარე სოციალურ, კულტურულ და ტექნოლოგიურ ცვლილებებს თანამედროვე კომუნიკაციის პირობებში.

ამ პრეზენტაციაში ჩვენ შემოგთავაზებთ მეთოდოლოგიურ კრიტერიუმებს და აღვწერთ ლექსიკური სიახლეების იდენტიფიცირების პროცესს, რომელიც მორგებულია კორპუსულ მონაცემებზე დაფუძნებულ მუდმივად ცვალებად რეალობას. მეთოდი გამყარებულია კომპი-

უტერული და ხელით გადამოწმების ტექნიკებით. შემოთავაზებული მეთოდოლოგია მოიცავს კორპუსის შექმნას, კანდიდატი სიტყვის ავტომატიზებულ აღმოჩენას, მის კომპიუტერითა და ხელით გადამოწმებას და სოციოლინგვისტურ კონტექსტში ანალიზს. ხელით გადამოწმება არის მნიშვნელოვანი, ვინაიდან ის შესაძლებლობას აძლევს ექსპერტს განმარტოს ლინგვისტური მოვლენისა და მისი კონტექსტში გამოყენების ფაქიზი ნიუანსები, რაც ავსებს კომპიუტერული მეთოდის ეფექტურობასა და მასშტაბურობას. მნიშვნელოვანია აღინიშნოს, რომ ჩვენ არ ვცდილობთ ლექსიკონზე დაფუძნებული მეთოდოლოგიის ჩანაცვლებას, არამედ ვამდიდრებთ მას სიტყვების დოკუმენტირებით ენაში მათი გამოყენების ადრეულ ეტაპზე. ჩვენ მიერ შემუშავებული მეთოდოლოგია ასევე ამცირებს სხვადასხვა კრიტერიუმის გათვალისწინებით ლექსიკური სიახლეების იდენტიფიცირებისათვის საჭირო დროს, რომელიც შემდგომში შეიძლება ინტეგრირდეს ნეოლოგიზმების ფილტრაციისთვის განკუთვნილ NLP ინსტრუმენტებში. ასევე ხაზს ვუსვამთ, რომ ჩვენი მეთოდები არ გულისხმობს სემანტიკურ ცვლილებებს, ისინი მხოლოდ ახალ ლექსიკურ ერთეულებზეა ორიენტირებული.

შემოთავაზებული მეთოდოლოგია შედგება რამდენიმე ეტაპისაგან. პირველ ეტაპზე ლექსიკური კანდიდატები იდენტიფიცირდებიან უახლეს კორპუსებში დამოწმების საფუძველზე. ისინი არ უნდა იყოს შეტანილი ავტორიტეტულ ლექსიკონებში და არ უნდა ფიქსირდებოდეს ისტორიულ რეფერენციალურ კორპუსებში, რომლებიც მოიცავს რომელიღაც კონკრეტულ თარიღამდე გამოცემულ ტექსტებს. ეს თარიღი განსხვავდება საკვლევი თემისა და კორპუსთან ხელმისაწვდომობის მიხედვით. თუმცა უახლესი ფენომენის შესასწავლად ჩვენ ვთავაზობთ მაქსიმუმ ხუთი წლით ადრე პერიოდს. მეორე ეტაპი გულისხმობს განცალკევებული ან იდიოსინკრაზიული ლექსიკური ფორმების გამორიცხვას, კანდიდატი ნეოლოგიზმები უნდა გვხვდებოდეს სხვადასხვა ტიპის ტექსტებში. ამასთანავე, თითოეული კანდიდატი ტერმინი უნდა გვხვდებოდეს რამდენიმეჯერ რამდენიმე ავტორთან ან რამდენიმე წყაროში. კვლევის პროცესი ასევე მოიცავს მონაცემების ფილტრაციას ავტომატურ და ნახევრად ხელით დამოწმების ტექნიკების გამოყენებით. ყურადღება ასევე ექცევა საკვლევ მონაცემთა ხარისხის საკითხსაც, რომლებიც წარმოდგენილია ციფრულ ტექსტებში, მაგალითად როგორიცაა ტიპოგრაფიული შეუსაბამობები, OCR-ის უზუსტობები თუ ორთოგრაფიული სტილიზაცია, რაც გამოწვეულია ასოთა გამეორებით (მაგალითად *yeahhh* ნაცვლად *yeah*), რომლებიც ხშირად გვაძლევს მცდარ დადებითებს, ყოველივე კი დამოწმების პროცესს დროში ახანგრძლივებს და ეს მაშინ, როდესაც ის შესაძლოა ეფექტურად იქნეს აღმოფხვრილი წინასწარ განსაზღვრული გარკვეულ წესებზე დაფუძნებული ფილტრების მეშვეობით. საზომები, როგორიცაა მაგალითად ლევენშტეინის მანძილი არის დამატებითი კრიტერიუმი იდენტიფიცირებისათვის, ვინაიდან არსებული სიტყვებისაგან მნიშვნელოვანი განსხვავებები შესაძლოა უფრო მიუთითებდეს უტყუარ ლექსიკურ ინოვაციაზე, ვიდრე ორთოგრაფიულ ვარიანტზე.

პრეზენტაციაში ჩვენ წარმოვადგენთ ამ კრიტერიუმებს პრაქტიკაში, რისთვისაც დავტესტავთ ჩვენს მეთოდოლოგიურ ჩარჩოს, რომელიც დანერგილია სპეციალურ პროგრამაში NeoN. მისი გამოყენება მარტივად შეუძლია ნებისმიერ ენათმეცნიერს ყოველგვარი პროგრამული უნარების გარეშე. წარმოდგენილი კვლევა მოიცავს ლექსიკურ ინოვაციებს თანამედროვე ლინგვისტური დარგებიდან, როგორიცაა გენდერულად ინკლუზიური ენა, ციფრული კომუნიკაცია, დიდ ენობრივ მოდელებთან დაკავშირებული დისკურსი და საპარლამენტო დისკურსი. ყოველივე ხაზს უსვამს ლექსიკური ინოვაციების კვლევის სოციოკულტურულ მნიშვნელობას ნამდვილ კომუნიკაციურ კონტექსტებში. ჩვენ ვაჩვენებთ მკაფიოდ განსაზღვრული კრიტერიუმების და თანამედროვე ლინგვისტურ რეალობებს მორგებული კომპიუტერული ინსტრუმენტების გამოყენების მნიშვნელობას ენის აქტიურად ცვლილების პირობებში. ვიმედოვნებთ, რომ ჩვენი მეთოდოლოგიური ჩარჩო დაეხმარება არა მხოლოდ ენათმეცნიერებსა და ლექსიკოგრაფებს, არამედ მასწავლებლებს, ტექნოლოგებსა და პოლიტიკოსებს, რომლებიც ცდილობენ შეისწავლონ და, შესაბამისად, ფეხი აუწყონ სწრაფ ლინგვისტურ ცვლილებებს.

თეზისი ქართულად თარგმნა თამარ კვიციანმა.

FAIRterm 2.0: A Concept-Oriented Web Application for Terminology Management

Vezzani, Federica
Di Nunzio, Giorgio Maria
University of Padova, Italy
federica.vezzani@unipd.it, giorgiomaria.dinunzio@unipd.it

The development and dissemination of high-quality terminology resources are fundamental to ensuring semantic clarity, multilingual consistency, and interdisciplinary knowledge exchange – especially in the context of open science and research infrastructures. This proposal presents FAIRterm 2.0, a newly released version of a web-based terminology management system developed at the University of Padua [1]. Building upon the conceptual and practical foundations of the FAIR terminology approach [2], this system aims to facilitate the creation, curation, and sharing of terminological data that are aligned with the FAIR principles for Open Science [3].

FAIRterm 2.0 represents a significant advancement from its predecessor, both in terms of user experience and underlying terminological modeling. While FAIRterm 1.0 [4] adopted a primarily term-based, translation-oriented layout, the second version introduces a fully concept-oriented architecture, aligned with international ISO standards for terminology management (ISO 16642, ISO 12620, ISO 30042) [5]. This new structure enables the clear separation of conceptual and linguistic layers, allowing multilingual terminologies to be accurately modeled. Each concept is uniquely identified and enriched with metadata, definitions, and domain classifications (e.g., EuroVoc categories), supporting both hierarchical and associative relationship documentation, and effective knowledge organization.

The application is not only a tool for managing terminological data but also a platform for collaborative terminology work. It has been successfully deployed in multilingual academic and institutional projects, including the European Parliament's "Terminology Without Borders" initiative [6], and is currently used by over 180 users across 30 countries. Designed with usability in mind, FAIRterm 2.0 allows both expert terminologists and students to contribute to structured terminology projects, thereby bridging pedagogical practice and professional terminology work. Moreover, the system facilitates data export in TBX (TermBase eXchange) format, ensuring compliance with recognized standards and enabling integration with external repositories and translation tools.

By presenting the design rationale, methodological principles, and real-world applications of FAIRterm 2.0, this proposal contributes to current discussions on the modernization of terminology resources. It advocates for a convergence between terminological and lexicographical approaches, and for the importance of aligning terminology management tools with emerging standards in open science, semantic web technologies, and multilingual knowledge representation. Ultimately, FAIRterm 2.0 demonstrates how a carefully designed tool can enhance the visibility, usability, and interoperability of terminological data across domains, languages, and research communities.

FAIRterm 2.0: ცნებაზე ორიენტირებული ვებაპლიკაცია ტერმინოლოგიის მართვისთვის

ფედერიკა ვეცანი
ჯორჯო მარია დი ნუნციო
პადუის უნივერსიტეტი, იტალია
federica.vezzani@unipd.it, giorgiomaria.dinunzio@unipd.it

მაღალი ხარისხის ტერმინოლოგიური რესურსების შემუშავება და გავრცელება უაღრესად მნიშვნელოვანია სემანტიკური სიცხადის, მრავალენოვანი შეთანხმებულობისა და ინტერდისციპლინური ცოდნის გაცვლის უზრუნველსაყოფად, განსაკუთრებით ღია მეცნიერებისა და კვლევითი ინფრასტრუქტურის კონტექსტში. ეს მოხსენება წარმოგიდგენთ FAIRterm 2.0-ს — ვებპლატფორმაზე დაფუძნებული ტერმინების მართვის სისტემის განახლებულ ვერსიას, რომელიც შემუშავდა პადუის უნივერსიტეტში [1]. იგი ეყრდნობა როგორც კონცეპტუალურ, ისე პრაქტიკულ საფუძვლებს, რომლებიც ჩამოყალიბდა FAIR ტერმინოლოგიის მიდგომაში [2]. ეს სისტემა მიზნად ისახავს ტერმინოლოგიური მონაცემების შექმნის, მართვისა და გაზიარების გამარტივებას, რაც შეესაბამება ღია მეცნიერებისთვის განსაზღვრულ FAIR პრინციპებს [3].

FAIRterm 2.0 მის წინამორბედთან შედარებით მნიშვნელოვნად გაუმჯობესებული ვერსიაა როგორც მომხმარებლის გამოცდილების, ასევე ტერმინოლოგიური მოდელირების თვალსაზრისით. FAIRterm 1.0-ს [4] ძირითადად ჰქონდა ტერმინებზე დაფუძნებული, თარგმანზე ორიენტირებული განლაგება, მეორე ვერსია კი წარმოგვიდგენს სრულად ცნებაზე ორიენტირებულ არქიტექტურას, რომელიც შეესაბამება ტერმინოლოგიის მართვის საერთაშორისო ISO სტანდარტებს (ISO 16642, ISO 12620, ISO 30042) [5]. ეს ახალი სტრუქტურა ცნებითი და ენობრივი დონეების მკაფიო გამიჯვნის საშუალებას იძლევა, რაც მრავალენოვანი ტერმინოლოგიების ზუსტი მოდელირების შესაძლებლობას ქმნის. თითოეულ ცნებას აქვს უნიკალური იდენტიფიკატორი და გამდიდრებულია მეტამონაცემებით, განმარტებებითა და დარგობრივი კლასიფიკაციებით (მაგ., EuroVoc კატეგორიები), რაც ქმნის როგორც იერარქიული, ისე ასოციაციური ურთიერთობების დოკუმენტირებისა და ცოდნის ეფექტური ორგანიზების საშუალებას.

აპლიკაცია არა მხოლოდ ტერმინოლოგიური მონაცემების მართვის ინსტრუმენტია, არამედ ტერმინოლოგიური თანამშრომლობის პლატფორმაც. ის წარმატებით გამოიყენეს არაერთ მრავალენოვან აკადემიურ და ინსტიტუციურ პროექტებში, მათ შორის ევროპარლამენტის „ტერმინოლოგია საზღვრების გარეშე“ ინიციატივაში [6] და ამჟამად მას 30 ქვეყანაში 180-ზე მეტი მომხმარებელი იყენებს. გამოყენებადობის გათვალისწინებით შექმნილი FAIRterm 2.0 საშუალებას აძლევს როგორც ექსპერტ ტერმინოლოგებს, ასევე სტუდენტებს, წვლილი შეიტანონ სტრუქტურირებულ ტერმინოლოგიურ პროექტებში, რითაც ერ-

თმანეთთან აკავშირებს პედაგოგიურ პრაქტიკასა და პროფესიულ ტერმინოლოგიურ მუშაობას. უფრო მეტიც, სისტემა მონაცემთა ექსპორტის შესაძლებლობასაც იძლევა TBX (TermBase eXchange) ფორმატში, რითაც უზრუნველყოფს აღიარებულ სტანდარტებთან შესაბამისობას და ინტეგრირებულია გარე საცავებთან და თარგმნის ინსტრუმენტებთან.

FAIRterm 2.0-ის დიზაინისა და მეთოდოლოგიური პრინციპების, ასევე მისი გამოყენების შესაძლებლობების წარდგენით, ეს მოხსენება თავის წვლილს შეიტანს ტერმინოლოგიური რესურსების მოდერნიზაციის შესახებ მიმდინარე დისკუსიებში. ეს პლატფორმა მხარს უჭერს ტერმინოლოგიური და ლექსიკოგრაფიული მიდგომების დაახლოებას, ასევე ხაზს უსვამს ტერმინოლოგიის მართვის ინსტრუმენტების შესაბამისობის მნიშვნელობას ღია მეცნიერების, სემანტიკური ვებტექნოლოგიებისა და მრავალენოვანი ცოდნის წარმოდგენის ახალ სტანდარტებთან. საბოლოო ჯამში, FAIRterm 2.0 აჩვენებს, თუ როგორ შეუძლია საგულდაგულოდ შემუშავებულ ინსტრუმენტს გააუმჯობესოს ტერმინოლოგიური მონაცემების ხილვადობა, გამოყენებადობა და თავსებადობა დარგებს, ენებსა და კვლევით დაწესებულებებს შორის.

References

1. **FAIRTerm.** (2025) FAIRTerm web page. Accessed: June 2, 2025. [Online]. Available: https://purl.org/fairterm
2. **Vezzani, F.** (Ed.) (2022). Terminologie numérique: conception, représentation et gestion. Peter Lang International Academic Publishers, 2022, accepted: 2022-06-02T13:51:41Z. [Online]. Available: https://library.oapen.org/handle/20.500.12657/56656
3. **Wilkinson, M.D. et al.** (2016). "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, no. 1, p. 160018, Mar. 2016, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/sdata201618
4. **Vezzani, F.** The FAIRterm resource: Between pedagogical practice and professionalization in specialized translation. *Synergies Italie*, no. 17, pp. 51–64, 2021.
5. **Vezzani, F., Di Nunzio, G.M. and R. Costa** (2023). ISO standards for terminology resources management: Are they FAIR enough? *Digital Translation*, vol. 10, no. 2, pp. 233–252, Dec. 2023, publisher: John Benjamins. [Online]. Available: https://www.jbe-platform.com/content/journals/10.1075/dt.00009.vez
6. **Terminology Without Borders.** (2025) YourTerm web page. Accessed: June 2, 2025. [Online]. Available: https://terminologynetwork.eu/terminology-without-borders/

Where Generative AI Fails: Explaining international standards through an explanatory dictionary

Williams, Geoffrey
University Grenoble-Alpes, France

Pensec, Emmanuelle
Independent Researcher, France
williamg@univ-grenoble-alpes.fr
emmanuellepensec@outlook.fr

With generative AI and Large Language Models in vogue, it is interesting to look at their limits in terms of dictionary construction when addressing precise user communities. Indeed, AI cannot understand the specificities of discourse communities and a very large mass of data cannot describe the realities that become clear in a smaller balanced and representative corpus. This can be illustrated through the building of an explanatory dictionary aimed at users who have to apply standards that they do not necessarily fully understand. In such cases, a generative model may find the appropriate terms, but cannot explain them. In this paper, we shall look at the problem of Sustainability in French, where translation from English brings in cultural variations that a simple translation hides.

For a number of years now, the concept of Sustainability has gained considerable importance and is enshrined in EU and national legislation. This used to be a subject exclusively for large companies, so that Small and Medium Sized Companies (henceforth SMEs) simply do not understand what the rules are and why it is to their advantage to adopt them. To adapt a discourse to the needs of SMEs, we propose to look at them as discourse communities, that is to say "socio-rhetorical networks that form in order to work towards sets of common goals" (Swales, 1990).

Part of the problem is that SMEs as a community have to apply international standards and these are not also clear to users. The problem arises in part from the fact that international standards are primarily written in English and that the terminological translation does not take into account cultural specificities and forgets that the spirit of the text may be interpreted in different ways. This has been clearly demonstrated in the study of the notion of 'work' (Williams & Pensec, 2025) where corpus, dictionary and sociological analyses show radically different perceptions of an apparently simple concept through the application of collocational resonance.

Evolving legislation appears first in English and whilst this may not be a problem for major corporations, it is certainly the case for SMEs, which do not have the capacity to follow changes in international legislation, and nor do they have dedicated services to allow translation. In this way, the tendency is a domination of Anglo-Saxon practices that ignores the situation in all other countries. Thus, SMEs must rely on mediation through experts who may not themselves be involved at reflection at such a level. In this way, they become victim of changes rather than active participants. Different cultural values lead to the difficulty of 'lost in translation' as the concepts originally defined in English are translated

into other languages. Another problem is that the terminology, whether it be in English or not, remains terminology and thus not directed to non-specialist readers. This is where explanatory lexicography plays a role by acting as a mediator between terms and what users within a community need to understand.

This paper illustrates the problem of introducing Sustainability to French SMEs, but focusses on the decisions in building an appropriate corpus and their exploitation using corpus linguistics methodology and the adoption of TLex as a tool for dictionary creation and management. In order to do this, we need first to set the scene by explaining the relationship between the concept of Sustainability and the UN policy initiative set out in its 17 Sustainable Development Goals. We will then concentrate on the issues concerning corpus construction and exploitation and the building of a dictionary that addresses a clearly defined community. As will be seen, the vocabulary is riddled with acronyms, and that increase the danger of misunderstanding. We shall thus pay particular attention to these in the dictionary.

სად ვერ გამოვიყენებთ გენერაციულ ხელოვნურ ინტელექტს: საერთაშორისო სტანდარტების ახსნა განმარტებითი ლექსიკონის მეშვეობით

ჯეფრი უილიამსი
გრენობლის ალპების უნივერსიტეტი, საფრანგეთი

ემანუელ პენსეკი
ინდივიდუალური მკვლევარი, საფრანგეთი
williamg@univ-grenoble-alpes.fr
emmanuellepensec@outlook.fr

გენერაციული ხელოვნური ინტელექტისა და დიდი ენობრივი მოდელების პოპულარობის ფონზე, საინტერესოა, განვიხილოთ მათი შეზღუდვები ლექსიკონის შექმნის მიმართულებით, როდესაც საქმე მომხმარებელთა კონკრეტულ ჯგუფებს ეხება. ხელოვნურ ინტელექტს არ შეუძლია დისკურსული ჯგუფების სპეციფიკის გაგება, ხოლო მონაცემთა ძალიან დიდ მასას არ შეუძლია აღწეროს ის რეალობა, რაც უფრო მცირე, დაბალანსებულ და რეპრეზენტაციულ კორპუსში ცხადად იქნება ნაჩვენები. ეს შეიძლება ვაჩვენოთ განმარტებითი ლექსიკონის შექმნის მაგალითზე, რომელიც იმ მომხმარებლებისთვისაა განკუთვნილი, რომლებსაც ისეთი სტანდარტების გამოყენება უწევთ, რომლებიც მათთვის სრულად გასაგები არ არის. ასეთ შემთხვევაში, გენერაციულმა მოდელმა შესაძლოა შესაბამისი ტერმინები იპოვოს, მაგრამ ვერ შეძლოს მათი სათანადოდ განმარტება. ამ ნაშრომში ჩვენ განვიხილავთ მდგრადობასთან დაკავშირებული ცნებების (Sustainability) პრობლემას ფრანგულ ენაში, სადაც ინგლისურიდან თარგმანს თან სდევს კულტურული ვარიაციები, რაც მარტივად შესრულებულ თარგმანში არ ჩანს.

უკვე რამდენიმე წელია, მდგრადობის ცნებამ დიდი მნიშვნელობა შეიძინა და გაერთიანებული ერების ორგანიზაციისა და ეროვნულ კანონმდებლობაშიც დაიმკვიდრა ადგილი. ეს საკითხი ადრე მხოლოდ მსხვილ საწარმოებს (კომპანიებს) ეხებოდა, შესაბამისად, მცირე და საშუალო ზომის საწარმოებს (კომპანიებს) (შემდგომში – მსს) უბრალოდ არ ესმით, რა წესები არსებობს და რატომ არის მათთვის სასარგებლო ამ წესების შემოღება. მსს-ების საჭიროებებთან ადაპტირების მიზნით, ჩვენ გთავაზობთ განვიხილოთ ისინი, როგორც დისკურსული ჯგუფები ანუ „სოციალურ-რიტორიკული ქსელები, რომლებიც ყალიბდება საერთო მიზნებზე მუშაობისთვის“ (Swales, 1990).

პრობლემის ნაწილი ისაა, რომ მსს-ებს, როგორც ჯგუფს, უწევთ იმ საერთაშორისო სტანდარტების გამოყენება, რაც მათთვის ყოველთვის ნათლად გასაგები არ არის. პრობლემა ნაწილობრივ იქიდან გამომდინარეობს, რომ საერთაშორისო სტანდარტები, ძირითადად, ინგლისურ ენაზეა დაწერილი და რომ ტერმინოლოგიური თარგმანი არ ითვალისწინებს კულტურულ სპეციფიკას. მხედველობაში არაა მიღებული ის ფაქტიც, რომ ტექსტის სულისკვეთება შეიძლება სხვადასხვაგვარად იყოს ინტე-

რპრეტირებული. ეს ნათლად იქნა ნაჩვენები „სამუშაოს“ (‘work’) ცნების კვლევაში (Williams & Pensec, 2025), სადაც კორპუსული, ლექსიკოგრაფიული (ლექსიკონების მონაცემების) და სოციოლოგიური ანალიზები რადიკალურად განსხვავებულ აღქმებს აჩვენებს ერთი შეხედვით მარტივი ცნების მიმართ, კოლოკაციური რეზონანსის გამოყენებით.

კანონმდებლობაში ცვლილებები ჯერ ინგლისურ ენაზე ჩნდება. ეს შეიძლება მსხვილ კორპორაციებს პრობლემას არ უქმნის, თუმცა ეს ნამდვილად პრობლემური საკითხია მსს-ებისთვის, რომელთაც არ აქვთ საერთაშორისო კანონმდებლობაში მიმდინარე ცვლილებებისთვის თვალის მიდევნების საშუალება და შესაბამისი თარგმნისთვის გამოყოფილი რესურსები. ამგვარად, სახეზეა ანგლოსაქსური პრაქტიკის დომინირების ტენდენცია, რომელიც ყველა სხვა ქვეყანაში არსებულ სიტუაციას უგულებელყოფს. ამრიგად, მსს-ებს უწევთ დაეყრდნონ შუამავალ ექსპერტებს, რომლებიც შესაძლოა თვითონაც არ იყვნენ ჩართული ასეთი დონის ანალიზში. ისინი ხდებიან ცვლილებების მსხვერპლი და არა მისი აქტიური მონაწილეები. განსხვავებული კულტურული ღირებულებები წარმოქმნის „თარგმანში დაკარგვის“ საფრთხეს, რადგან თავდაპირველად ინგლისურად განმარტებული ცნებები სხვა ენებზე ითარგმნება. კიდევ ერთი პრობლემა ისაა, რომ ტერმინოლოგია, იქნება ეს ინგლისურ თუ სხვა ენაზე, ტერმინოლოგიად რჩება და შესაბამისად, არ არის განკუთვნილი არასპეციალისტი მკითხველისთვის. სწორედ აქ იკვეთება განმარტებითი ლექსიკოგრაფიის მნიშვნელობა, რომელიც შუამავლის როლს ასრულებს ტერმინებსა და იმ ინფორმაციას შორის, რაც საზოგადოების წევრებმა უნდა გაიგონ.

ეს ნაშრომი ასახავს ფრანგულ მსს-ებში მდგრადობის ცნების დანერგვის პრობლემას. თუმცა, იგი ორიენტირებულია იმ გადაწყვეტილებებზე, რომელიც უკავშირდება კორპუსის შექმნას და მის გამოყენებას კორპუსული ლინგვისტიკის მეთოდოლოგიაზე დაყრდნობით, ასევე TLex-ის, როგორც ლექსიკონის შექმნისა და მართვის ინსტრუმენტის გამოყენებაზე. ამისათვის, ჩვენ ჯერ აღვწერთ მდგრადობის ცნებასა და გაეროს მიერ დადგენილ „მდგრადი განვითარების 17 მიზანს“ შორის არსებულ ურთიერთკავშირს. შემდეგ განვიხილავთ კორპუსის შექმნისა და გამოყენების, ასევე ლექსიკონის შექმნის საკითხებს, რომელიც ნათლად განსაზღვრულ საზოგადოებრივ ჯგუფს მოერგება. თქვენ ნახავთ, რომ ლექსიკონი აბრევიატურებითაა სავსე, რაც ინფორმაციის არასწორად გაგების საფრთხეს ზრდის. ამიტომ, ლექსიკონში მათ განსაკუთრებული ყურადღება ექცევა.

თეზისი ქართულად თარგმნა ქეთევან ჭიღლაძემ.

References

Swales, J. (1990). *Genre analysis*. Cambridge University Press.

Williams, G. C. and E. Pensec (2025). «All Work and No Play”: The Collocational Resonance of ‘Work’. *International Journal of Lexicography*. https://doi.org/10.1093/ijl/ecaf006

ავტორთა საძიებელი

ილიას სახელმწიფო უნივერსიტეტის
გამომცემლობა

ILIA STATE UNIVERSITY
PRESS