

BASIC PRINCIPLES OF CORPUS LINGUISTIC PROCESSING OF LEGISLATIVE TEXTS

იურიდიული ტექსტების კორპუსლინგვისტური დამუშავების ძირითადი პრინციპები¹

Tinatin Tenterashvili

PhD Student in Philology,

Ivane Javakhishvili Tbilisi State University,

Ilia Tchavchavadze Avenue 1, Tbilisi, 0179, Georgia,

+995555428714, ttenterashvili@gmail.com

<https://orcid.org/0009-0002-8113-7115>

Abstract. One of the important steps in the modern stage of the development of humanitarian sciences is the creation of big databases and the development of modern research methods. The existence of digital corpora provides an opportunity to carry out efficient research, which contributes to both the solution of a separate linguistic issue and the implementation of complex linguistic research. One such resource for Georgian language is the Georgian National Corpus, which has 202 mln tokens and including texts of different genres, is extended with two thematic corpora: a corpus of political and legal texts, although neither of them is completely filled. In particular, the sub-corpus of legal texts includes only historical legal documents. From this point of view, the genre balancing of the Georgian National Corpus is an important challenge, which is necessary to overcome for its final perfection and full functioning. One of the ways to solve the mentioned problem is to create a technological framework for digital processing of legislative texts using modern research methods.

For digital processing of the texts, it is necessary them to be available in electronic format. Nowadays, the legislative texts have already been digitized. Their electronic versions are available on the website matsne.gov.ge, however, they are not linguistically processed and therefore, lack the possibility of such research. We will present to you one of the ways to solve this issue in our report, in which we will describe the basic principles of corpus linguistic processing of legal texts on the example of specific legal texts. The obtained resource was an empirical base for linguistic research, based on which it was analyzed linguistic means with legal semantics.

The content of legal texts is transferred in different linguistic forms. From the point of view of studying sectoral vocabulary in legal texts, we should focus on cases when the vocabulary, transferring sectoral content, consists of more than one component and acquires a specific meaning for the sector only with this form. If we consider them independently of each other, in most cases, they will not be related to the legal field. In this sense, two-component verbs are important, which are given by a combination of the main verb and a noun or an infinitive closely related with it, For example, *imprisonment*, *sentencing*, etc. Such combinations are often found in legal texts to describe legal processes.

Identifying and taking into account characteristic features of legal texts is important for the corpus linguistic processing of legal texts on the one hand, and for filling the bank of legal terms on the other hand, since two-component verbs should be processed corpus-linguistically not at the level of individual lexical units, but in the form in which they are related to the field of law.

Keywords: GNC, Corpus Linguistics, Legislative, Term, Two-Component.

¹ კვლევა [PHDF-23-2595] განხორციელდა შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის მხარდაჭერით.

თინათინ ტენტერაშვილი

ფილოლოგიის სადოქტორო პროგრამის დოქტორანტი,
ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი,
ქ. თბილისი, ილია ჭავჭავაძის გამზ. 1, 0179, საქართველო
+995555428714, ttenterashvili@gmail.com
<https://orcid.org/0009-0002-8113-7115>

აბსტრაქტი. ჰუმანიტარული მეცნიერებების განვითარების თანამედროვე ეტაპზე ერთ-ერთი მნიშვნელოვანი ნაბიჯია დიდ მონაცემთა ბაზების (big data) შექმნა და კვლევის თანამედროვე მეთოდების შემუშავება. ციფრული კორპუსების არსებობა ეფიციენტური კვლევების განხორციელების შესაძლებლობას იძლევა, რაც ხელს უწყობს როგორც ცალკეული ლინგვისტური საკითხის გადაწყვეტას, ისე კომპლექსური ენათმეცნიერული კვლევების განხორციელებას. ქართულისათვის ერთ-ერთ ასეთ რესურსს წარმოადგენს ქართული ენის ეროვნული კორპუსი, რომელიც 202 მლნ. ტოკენს მოიცავს და სხვადასხვა ჟანრის ტექსტთან ერთად განვრცობილია ორი თემატური კორპუსით: პოლიტიკური და იურიდიული ტექსტების კორპუსით, თუმცა არც ერთი მათგანი არ არის ბოლომდე შევსებული. კერძოდ, იურიდიული ტექსტების ქვეკორპუსი მოიცავს მხოლოდ ისტორიულ იურიდიულ დოკუმენტებს. ამ თვალსაზრისით ქართული ენის ეროვნული კორპუსის ჟანრობრივი დაბალანსება მნიშვნელოვანი გამოწვევაა, რომლის დაძლევა აუცილებელია მისი საბოლოო სრულყოფისა და სრულფასოვანი ფუნქციონირებისათვის. აღნიშნული პრობლემის გადაჭრის ერთ-ერთ გზად გვესახება საკანონმდებლო ტექსტების ციფრული დამუშავების ტექნოლოგიური ჩარჩოს შექმნა კვლევის თანამედროვე მეთოდების გამოყენებით.

ამა თუ იმ ტექსტის ციფრული დამუშავებისათვის აუცილებელია, რომ ისინი ელექტრონულ ფორმატში იყოს ხელმისაწვდომი. დღეისათვის საკანონმდებლო ტექსტები უკვე გაციფრებულია. მათი ელექტრონული ვერსიები თავმოყრილია ვებ-გვერდზე matsne.gov.ge, თუმცა ისინი არ არის ლინგვისტურად დამუშავებული და, შესაბამისად, მოკლებულია ამგვარი კვლევის შესაძლებლობასაც. ამ საკითხის გადაჭრის ერთ-ერთი გზაა მოცემული ქვემოთ წარმოდგენილ სტატიაში, რომელშიც აღწერილია იურიდიული ტექსტების კორპუსლინგვისტური დამუშავების ძირითადი პრინციპები კონკრეტული სამართლებრივი ტექსტების მაგალითზე. მიღებული რესურსი იყო ემპირიული ბაზა ლინგვისტური კვლევისათვის, რომელზე დაყრდნობითაც გაანალიზდა იურიდიული სემანტიკის მქონე ენობრივი საშუალებები.

სამართლებრივ ტექსტებში დარგობრივი ლექსიკის შესწავლის თვალსაზრისით საყურადღებოა ისეთი შემთხვევები, როდესაც დარგობრივი შინაარსის გადმომცემი ლექსიკა ერთზე მეტი კომპონენტისგან შედგება და მხოლოდ ამ ფორმით იძენს დარგისათვის სპეციფიკურ მნიშვნელობას. თუ მათ ერთმანეთისგან დამოუკიდებლად განვიხილავთ, ხშირ შემთხვევაში, ისინი დაკავშირებული არ იქნება იურიდიულ სფეროსთან. ამ თვალსაზრისით მნიშვნელოვანია ორკომპონენტური ზმნები, რომლებიც ძირითადი ზმნისა და მასთან მყარად დაკავშირებული არსებითი სახელის ან საწყისის კომბინაციით არის გადმოცემული, მაგ., *თავისუფლების აღკვეთა*, *გადაწყვეტილების გამოტანა* და ა.შ. ამგვარი კომბინაციები იურიდიულ ტექსტებში ხშირად გვხვდება სამართლებრივი პროცესების აღსაწერად.

იურიდიული ტექსტებისათვის ნიშანდობლივი ამ მახასიათებლების გამოვლენა და გათვალისწინება მნიშვნელოვანია, ერთი მხრივ, იურიდიული ტექსტების კორპუსლინგვისტური დამუშავებისათვის, მეორე მხრივ კი, იურიდიულ ტერმინთა ბანკის შესავსებადაც, ვინაიდან ორკომპონენტური ზმნები კორპუსლინგვისტურად უნდა

დამუშავდეს არა ცალკეული ლექსიკური ერთეულების დონეზე, არამედ იმ ფორმით, რომლითაც დაკავშირებულია სამართლის სფეროსთან.

საკვანძო სიტყვები: ქვევ, კორპუსლინგვისტური, საკანონმდებლო ტერმინი, ორკომპონენტიანი.

შესავალი. ჰუმანიტარული მეცნიერებების განვითარების თანამედროვე ეტაპზე ერთ-ერთი მნიშვნელოვანი ნაბიჯია დიდ მონაცემთა ბაზების (big data) შექმნა და კვლევის თანამედროვე მეთოდების შემუშავება. როგორც რამაზ ხალვაში აღნიშნავს მონოგრაფიაში „დიგიტალური ჰუმანიტარიის შესავალი“, დიდი მონაცემები გულისხმობს დიდი რაოდენობის რთულად სტრუქტურირებულ მონაცემებს, რომელთა დამუშავება, ძიება, გავრცელება და ანალიზი ტრადიციული მეთოდებით შეუძლებელია და საჭიროებს სპეციალური კომპიუტერული პროგრამების გამოყენებას (ხალვაში, 2018: 80).

დიდ მონაცემთა ანალიზისათვის სხვადასხვა მეთოდი გამოიყენება. მათ შორისაა **მონაცემთა მოპოვება (Data Mining)**, რომელსაც „განმარტავენ, როგორც მონაცემთა ინტელექტუალურ ანალიზს, ინფორმაციის მოპოვებას და ცოდნის მიღებას“ (ხალვაში, 2018: 82). როგორც სამეცნიერო ლიტერატურაშია აღნიშნული, მონაცემთა მოპოვება „მიზნად ისახავს არა წინასწარ მოცემული ჯგუფების კლასიფიკაციას, არამედ ახალი ცოდნის, ჯგუფებისა და კანონზომიერებების გამოვლენას დიდ მონაცემთა ბაზებში“ (ხალვაში, 2018: 81). მონაცემთა მოპოვება და დამუშავება ენობრივი კორპუსების გარეშე შეუძლებელია. მათი არსებობა კვლევის შედეგების სისტემურობასა და სიზუსტეს განაპირობებს, ვინაიდან ციფრული რესურსების დამუშავება მანქანურად ხორციელდება წინასწარ შემუშავებული წესების გამოყენებით. დღეისათვის ციფრული კორპუსები აქტიურად გამოიყენება ცალკეული ლინგვისტური საკითხისა და კომპლექსური ენათმეცნიერული კვლევების განსახორციელებლად.

ქართულისათვის ერთ-ერთ ასეთ რესურსს წარმოადგენს **ქართული ენის ეროვნული კორპუსი**, რომელიც 202 მლნ. ტოკენს მოიცავს. სხვადასხვა ჟანრის ტექსტთან ერთად მასში ვხვდებით ორ თემატურ კორპუსს - პოლიტიკური და იურიდიული ტექსტების კორპუსს, თუმცა არც ერთი მათგანი არ არის ბოლომდე განვრცობილი. კერძოდ, „იურიდიული ტექსტების კორპუსი საშუალო ქართულის თემატურ კორპუსს წარმოადგენს“ (თანდაშვილი & ყამარაული, 2021: 142). ამ თვალსაზრისით ქართული ენის ეროვნული კორპუსის ჟანრობრივი დაბალანსება მნიშვნელოვანი გამოწვევაა, რომლის დაძლევა აუცილებელია მისი საბოლოო სრულყოფისა და სრულფასოვანი ფუნქციონირებისათვის.

„დაბალანსება კორპუსის ერთ-ერთი თვისებრივი მახასიათებელია და გულისხმობს კორპუსში ჩართული რესურსების ჟანრული მრავალფეროვნების გაწონასწორებას“ (თანდაშვილი & ყამარაული, 2021: 141). აღნიშნული მიიღწევა არა მარტო საშუალო ქართულის მეტყველებით, არამედ თანამედროვე საკანონმდებლო ტექსტების დამუშავებითაც. ამგვარად დაბალანსებული იურიდიული ქვეკორპუსი საშუალებას მოგვცემს, შევისწავლოთ იურიდიული მეტყენა როგორც სინქრონიული, ისე დიაქრონიული თვალსაზრისით. წინამდებარე კვლევის მიზანია საკანონმდებლო ტექსტების ციფრული დამუშავების ტექნოლოგიური ჩარჩოს შექმნა და მისი გამოყენება იურიდიული სემანტიკის მქონე ენობრივი საშუალებების შესასწავლად.

საკითხის აქტუალობა განპირობებულია იურიდიული ტექსტების ქვეკორპუსის განვრცობის საჭიროებით. როგორც ზემოთაც აღინიშნა, იურიდიული ტექსტების ქვეკორპუსი ამ ეტაპზე მხოლოდ საშუალო ქართულის იურიდიული დოკუმენტებით შემოიფარგლება, იგი არ მოიცავს თანამედროვე საკანონმდებლო ტექსტებს, შესაბამისად,

კორპუსლინგვისტური კვლევისას ყურადღების მიღმა რჩება რიგი ლინგვისტური საკითხები. თანამედროვე საკანონმდებლო ტექსტებით იურიდიული ქვეკორპუსის განვრცობა ხელს შეუწყობს იურიდიული მეტაენის შესწავლას როგორც სინქრონიული, ისე დიაქრონიული თვალსაზრისით.

საკითხის **მეცნიერული სიახლე** განპირობებულია იურიდიული ენის ინტერდისციპლინურ ჭრილში კვლევის შესაძლებლობით: იურიდიული ტექსტების ციფრული დამუშავების ტექნოლოგიური ჩარჩოს შექმნით შესაძლებელი გახდება თანამედროვე საკანონმდებლო რესურსების პროტოტიპული კორპუსის შექმნა და იურიდიული ტერმინების ბანკის შევსება, ტექსტების ენობრივად დამუშავება კი შესაძლებელს გახდის ლინგვისტური საკითხების კომპლექსურად კვლევას.

მეთოდები. იურიდიული ტექსტების კორპუსლინგვისტურად დამუშავება, უპირველესად, მოიაზრებს ისეთი ტექნოლოგიური პროდუქტის განვითარებას, რომელიც საშუალებას მოგვცემს, საკანონმდებლო ტექსტების პროტოტიპული კორპუსი შევქმნათ და გვექმნდეს იურიდიული ტექსტების ინტერდისციპლინურ ჭრილში კვლევის შესაძლებლობა. ამ ამოცანის განსახორციელებლად უმთავრესია **კორპუსლინგვისტური მეთოდების გამოყენება**. კორპუსლინგვისტური კვლევის უპირატესობა იმაში მდგომარეობს, რომ თანამედროვე ტექნოლოგიებით დამუშავების შედეგად მიღებული თეორიული დასკვნები, კვლევის მეთოდებიდან გამომდინარე, სიზუსტით, ადეკვატურობითა და სისტემურობით გამოირჩევა. მოცემულ კვლევაში გამოყენებული იყო კორპუსზე დამყარებული კვლევის მეთოდი, „რომელიც საშუალებას იძლევა შევამოწმოთ ლინგვისტური ჰიპოთეზის ვალიდურობა კორპუსზე დაყრდნობით“ (თანდაშვილი & ყამარაული, 2021: 181).

შედეგები და მსჯელობა. როგორც ზემოთაც აღინიშნა, მოცემული კვლევა მიზნად ისახავდა იურიდიული ტექსტების კორპუსლინგვისტური დამუშავების პრინციპების გამოკვეთასა და მიღებული რესურსის გამოყენებას დარგობრივი ლექსიკის შესწავლის დროს. ამ უკანასკნელის საჭიროება ტექსტების დამუშავებისას გამოიკვეთა, როცა მათში აღმოჩნდა ერთზე მეტი კომპონენტისგან შემდგარი ლექსიკური ერთეულები, რომლებიც ტერმინების დონეზე გამოიყენება საკანონმდებლო ტექსტებში.

იურიდიული ტექსტების კორპუსლინგვისტური დამუშავების დროს თავდაპირველად მოხდა მისი ტექნოლოგიური ჩარჩოს კონცეპტუალიზაცია. უპირველესად, განისაზღვრა კორპუსის შექმნის მიზანი. ვინაიდან აღნიშნული რესურსი განკუთვნილია ლინგვისტური კვლევებისათვის, მისი ანოტირება აუცილებელია, შესაბამისად, აღნიშნული ფაქტორიც გათვალისწინებული იყო ტექნოლოგიური ჩარჩოს კონცეპტუალიზაციის დროს. გარდა ამისა, ამ ეტაპის მნიშვნელოვანი ამოცანა იყო იურიდიული ტერმინების ბანკის გავრცობა, რამაც აუცილებელი გახდა ქვეკორპუსის სემანტიკური თვალსაზრისით დამუშავებაც.

კორპუსის შექმნის პირველ ეტაპზე აუცილებელია სამიზნე ტექსტების შერჩევა. ვინაიდან იურიდიული ტექსტები მრავალრიცხოვანია, მოცემული კვლევა კი - დროში შეზღუდული, ამ ეტაპზე კორპუსისათვის შეირჩა ისეთი ტექსტები, რომლებიც მოცულობითა და თემატური თვალსაზრისით თანხვდება და ავსებს ერთმანეთს. ესენია: საქართველოს კანონები - **პატიმრობის კოდექსი**, **პენიტენციური კოდექსი**² და **არასრულწლოვანთა მართლმსაჯულების კოდექსი**.

² პატიმრობის კოდექსი ძალადაკარგულად გამოცხადდა 2023 წლის 15 დეკემბერს. იგი პენიტენციური კოდექსის ადრეულ ვერსიას წარმოადგენს.

კორპუსლინგვისტური დამუშავებისას ტექსტი აუცილებლად ელექტრონულ ფორმატში უნდა იყოს მოცემული. როგორც წესი, ამ ეტაპზე ხდება დედანი ტექსტების გაციფრება და მათი ელექტრონულ ფორმატში გადატანა, თუმცა საკანონმდებლო ტექსტების შემთხვევაში ეს ამოცანა უკვე შესრულებულია: დღეისათვის არსებობს **საქართველოს საკანონმდებლო მაცნე (matsne.gov.ge)**, „რომელიც არეგისტრირებს, აქვეყნებს და სისტემატიზაციას უკეთებს ძალაში შესულ ნორმატიულ აქტებს მაცნეს ვებსაიტზე გამოქვეყნების შემდეგ. ეს საკანონმდებლო აქტები დაცულია იუსტიციის სამინისტროს მიერ. იგი იძლევა ძალაში შესული კანონების, მოქმედი კანონმდებლობის კრებული და კანონპროექტების გამოქვეყნების შესაძლებლობას, რომლებიც საზოგადოებისთვის უფასოდ არის ხელმისაწვდომი“ (მოლი & დრინოჩი, 2015: 40). მიუხედავად ამ შესაძლებლობისა, თავისი მიზნებიდან გამომდინარე, საკანონმდებლო მაცნეში ვერ შევხვდებით ენობრივად დამუშავებულ ტექსტებს, თუმცა ელექტრონული ვერსიების არსებობა აადვილებს მათი კორპუსლინგვისტურად დამუშავების პროცესს.

იურიდიული ტექსტების კორპუსლინგვისტურად დამუშავების შემდეგ ეტაპზე ისინი დავშალებთ წინადადებებად, პროგრამული უზრუნველყოფა KWIC Concordance-ის საშუალებით კი შევადგინეთ სიხშირული ლექსიკონები კორპუსში შემავალი თითოეული ტექსტისათვის. ყოველ ტოკენს მიეთითა მარტივი ლემა.

აღნიშნული კორპუსის შექმნისას განხორციელდა ტექსტების POS-ანოტაცია, რაც გულისხმობს სიტყვაფორმისათვის მეტყველების ნაწილის თეგის მინიჭებას. ანოტაციის დროს გათვალისწინებული იყო მორფოლოგიური დონის ერთეულები, რაც მორფემათა რანგობრივ განაწილებას შეესაბამება თითოეული ფორმისათვის:

1. არსებითი სახელი - ბრუნვა, რიცხვი, თანდებული, ენკლიტიკა, სავრცობი;
2. ზედსართავი სახელი - ბრუნვა, რიცხვი, ხარისხი, თანდებული, ენკლიტიკა, სავრცობი;
3. რიცხვითი სახელი - ბრუნვა, ჯგუფი;
4. ნაცვალსახელი - ბრუნვა, პირი, რიცხვი, ჯგუფი, თანდებული, ენკლიტიკა, სავრცობი;
5. ზმნა - პირი, რიცხვი, პირიანობა, ვალენტობა, მწკრივი, გვარი, ზმნის ტიპი;
6. ზმნიზედა - ჯგუფი;
7. თანდებული - ბრუნვა³;
8. კავშირი - ჯგუფი.

რაც შეეხება ნაწილაკსა და შორისდებულს, ისინი ჩვენ მიერ დამუშავებულ საკანონმდებლო ტექსტებში არ გვხვდება.

ჩამოთვლილი ენობრივი მონაცემების მიხედვით იურიდიული ტექსტი კორპუსლინგვისტურად დამუშავებისას ამგვარ სახეს იღებს⁴:

სიტყვაფორმა	რაოდენობა	სალექსიკონო ფორმა	ტერმინი	მეტყველების ნაწილი	გრამატიკული მახასიათებლები	ზმნის ტიპი	პირიანობა	ვალენტობა	გვარი	მწკრივი
დანამატისათვის	19	დანამაული		1 N	Gen, Sg, L, PP:თვის	0	0	0	0	0
უკიდურესი	3	უკიდურესი		0 A	Nom, Sg, Elat	0	0	0	0	0
პირველი	33	პირველი		0 Num	Nom, Ord	0	0	0	0	0
მისთვის	10	ის		0 Pron	Pron, 3rd, Gen, Sg, PP:თვის	0	0	0	0	0
გამოიყენება	13	გამოყენება		0 V	3, Sg	Main	S:3Sg	1Val	Pass	Fut
აგრეთვე	33	აგრეთვე		0 Adv	0	0	0	0	0	0
შესაბუთ	81	შესაბუთ		0 PP	Gen	0	0	0	0	0
და	424	და		0 Cj	Coord	0	0	0	0	0

ანოტირების დროს კორპუსში გამოყენებული იყო ქართული ენის ეროვნული კორპუსის თეგები.⁵

³ ბრუნვა მითითებულია ცალკე მდგომი თანდებულისათვის, სხვა შემთხვევაში მონიშნულია კონკრეტულ სიტყვაფორმაში მორფემათა რანგობრივი განაწილების შესაბამისად.

⁴ ცხრილი მოცემულია არასრულწლოვანთა მართლმსაჯულების კოდექსის მაგალითზე.

⁵ თეგების ცხრილი იხ. ბმულზე: <http://gnc.gov.ge/gnc/gnc-features?session-id=257881584423115>.

გარდა ამისა, ანოტირება მოხდა სემანტიკურ დონეზეც: ერთი მხრივ, გამოიყო ცალკეული ლექსიკური ერთეულები, რომლებიც იურიდიულ მეტაენაში აქტიურად გამოიყენება, მათ შორის, პროფესიის აღმნიშვნელი სახელები (გამომძიებელი, ადვოკატი, პროკურორი), მეორე მხრივ კი დარგობრივი ლექსიკის ის ნაწილი, რომლებიც ერთზე მეტი კომპონენტისგან შედგება და მხოლოდ ამ ფორმით იძენს დარგისათვის სპეციფიკურ მნიშვნელობას.

დარგობრივი ლექსიკის, ტერმინოლოგიის კვლევა და სხვადასხვა ლექსიკური ერთეულის ამგვარად კვალიფიცირება ერთ-ერთი რთული და საკამათო საკითხია. „*დღეს, ტერმინი განისაზღვრება, როგორც ლიტერატურული ენის ერთეული, რომელიც ფლობს ყველა მის თვისებას და ასევე სხვა თვისებებს, რომლითაც გამოყოფენ ტერმინს, როგორც განსაკუთრებულ სიტყვას ან სიტყვას განსაკუთრებული ფუნქციით*“ (რუხიაშვილი, 2019: 45). ტერმინოლოგიის დადგენა ყოველთვის იწვევს აზრთა სხვადასხვაობას, რადგან ის ამა თუ იმ დარგის ცნებებს ზუსტად უნდა გამოხატავდეს.

სამართლის ენა მკაცრად ფორმულირებული, მკაფიო და ოფიციალურია. ტერმინოლოგიაც, როგორც წესი, ცალკეული ლექსიკური ერთეულების სახით არის გამოხატული, რომლებიც მიემართება დარგთან დაკავშირებულ პროფესიებს, სამუშაო ადგილებს/დაწესებულებებს, სასამართლო პროცესში მონაწილე პირთა სტატუსს, იურიდიულ დოკუმენტაციას და ა.შ. როგორც წესი, ისინი არსებითი სახელით არის გამოხატული და მნიშვნელობის გასაგებად კონტექსტი არ არის აუცილებელი.

თუმცა, ცალკეული ლექსიკური ერთეულების გარდა, ჩვენ მიერ დამუშავებულ ტექსტებში გვხვდება ისეთი შემთხვევები, როდესაც დარგობრივი შინაარსის გადმოცემი ლექსიკა ერთზე მეტი კომპონენტისგან შედგება და მხოლოდ ამ ფორმით იძენს დარგისათვის სპეციფიკურ მნიშვნელობას. თუ სინტაგმის შემადგენელ ნაწილებს ერთმანეთისგან დამოუკიდებლად განვიხილავთ, შესაძლოა, მათი გამოყენების არეალი გაცილებით მრავალფეროვანი აღმოჩნდეს. ამ თვალსაზრისით მნიშვნელოვანია ორკომპონენტიანი ზმნები, რომლებიც ძირითადი ზმნისა და მასთან მყარად დაკავშირებული არსებითი სახელის, ან საწყისის კომბინაციით არის გადმოცემული.

მაგალითისათვის, განვიხილოთ სინტაგმა „თავისუფლების აღკვეთა“, რომელიც ჩვენ მიერ დამუშავებულ სამივე საკანონმდებლო ტექსტში ყველაზე ხშირად არის გამოყენებული და გამოხატავს ჩადენილი დანაშაულისათვის განსაზღვრული სასჯელის ერთ-ერთ სახეს:

- „არასრულწლოვანს ვადიანი თავისუფლების აღკვეთა შეიძლება დაენიშნოს მხოლოდ ძირითად სასჯელად“ (არასრულწლოვანთა მართლმსაჯულების კოდექსი, 2015);
- „თავისუფლების აღკვეთის დაწესებულებაში ბრალდებულთათვის/მსჯავრდებულთათვის უნდა შეიქმნას პროფესიული განათლების მიღების პირობები“ (პატიმრობის კოდექსი, 2010);
- „თავისუფლების აღკვეთის დაწესებულება არის საქართველოს სისხლის სამართლის კოდექსის მე-40 მუხლის პირველი ნაწილის „ზ“ და „თ“ ქვეპუნქტებით გათვალისწინებული სასჯელების აღსრულების დაწესებულება“ (პენიტენციური კოდექსი, 2023).

თავისუფლების აღკვეთა ამგვარი ფორმით მხოლოდ სამართლებრივ კონტექსტში გამოიყენება, რასაც ადასტურებს მისი გამოყენების სიხშირე ქართული ენის ეროვნულ კორპუსში. ქვეკ-ის მონაცემების მიხედვით, „თავისუფლების აღკვეთა“ სხვადასხვა ჟანრის ტექსტში 3081-ჯერ არის გამოყენებული და ნებისმიერ მათგანში სასჯელის ფორმას გულისხმობს:

- „საზოგადოების მოტყუებისათვის ცაწავალას მისაჯოს ერთი დღე-ღამით თავისუფლების აღკვეთა კანკემის ღეროზე მიბმით, – გამოიტანა მოსამართლემ განაჩენი“;

- „1994 წელს მას რვა წლით თავისუფლების აღკვეთა და ქონების კონფისკაცია მიესაჯა“;
- „დამნაშავედ ცნო ღალატში და მას 17 წლით თავისუფლების აღკვეთა მიუსაჯა“ და ა.შ. (ქეევ, 2011 – 2024).

სინტაგმის ერთეულების დამოუკიდებლად განხილვის შემთხვევაში ცხადი ხდება, რომ ისინი შეიძლება სამართლებრივი კონტექსტის მიღმა შეგვხვდეს. „თავისუფლება“ ნათესაობითი ბრუნვის ფორმით 32439-ჯერ დასტურდება ქართული ენის ეროვნულ კორპუსში და ყოველთვის არ არის დაკავშირებული იმ სემანტიკასთან, რასთანაც „თავისუფლების აღკვეთა“. რაც შეეხება საწყისს „აღკვეთა“, იგი 4496-ჯერ გამოიყენება და, მიუხედავად იმისა, რომ ეს უკანასკნელი უმეტეს შემთხვევაში სამართლებრივ სფეროს მიემართება, ყოველთვის „პატიმრობის“ მნიშვნელობით არ გამოიყენება - არც თუ იშვიათად „აღმოფხვრის“ სემანტიკას ატარებს:

- „სომეხთა მეფემ აჯანყება წარმატებით აღკვეთა“;
- „ტანსაცმლისა და ფეხსაცმლის ქურდობის აღკვეთა სკოლაში“;
- „ეს დროებითი დაქსაქსულობა მალევე აღკვეთა „მხედართმთავარმა“ მარიკამ“ და მისთ. (ქეევ, 2011 – 2024).

აღსანიშნავია, რომ დასახელებული სიტყვაფორმები ყველაზე ხშირად ერთმანეთთან ქმნის კოლოკაციებს, ფაქტობრივად, მყარ შესიტყვებად არის ქცეული და სამართლებრივ კონტექსტს მიემართება როგორც ზეპირ მეტყველებაში, ისე ოფიციალურ ტექსტებში. აქედან გამომდინარე, მნიშვნელოვანია, რომ იგი ორკომპონენტიან ტერმინად კვალიფიცირდეს ტექსტების კორპუსლინგვისტურად დამუშავების დროს.

განვიხილოთ კიდევ ერთი შესიტყვება, *გადაწყვეტილების გამოტანა*, რომელიც ასევე აქტიურად გამოიყენება ჩვენ მიერ დამუშავებულ საკანონმდებლო ტექსტებში:

- „არასრულწლოვანი მსჯავრდებულის რესოციალიზაცია-რეაბილიტაციის გეგმა უნდა მომზადდეს მის გათავისუფლებამდე სულ მცირე 2 თვით ადრე, გარდა იმ შემთხვევისა, როდესაც ამ ვადის დაცვა შეუძლებელია პირობით ვადამდე გათავისუფლების ან ამნისტიის/შეწყალების აქტის გამოცემის ანდა სასამართლო გადაწყვეტილების გამოტანის გამო“ (არასრულწლოვანთა მართლმსაჯულების კოდექსი, 2015);
- „დაუშვებელია დასაბუთებული გადაწყვეტილების გამოტანის სხვა დროისათვის გადადება“ (პატიმრობის კოდექსი, 2010);
- „ეს გადაწყვეტილება გამოიტანება მხოლოდ იმ შემთხვევაში, თუ პირის მიერ შესაბამისი დისციპლინური დარღვევის ჩადენა სასამართლო სხდომაზე საქმის განხილვის დროს სასამართლოს მიერ გამოკვლეული უტყუარი მტკიცებულებების საფუძველზე დამტკიცდა“ (პენიტენციური კოდექსი, 2023).

ქართული ენის ეროვნულ კორპუსში ფორმა „გადაწყვეტილების“ 18864-ჯერ გამოიყენება, ყველაზე ხშირად კი წყვილდება საწყისის „მიღება“ ფორმებთან: *გადაწყვეტილების მიღებისას*, *გადაწყვეტილების მიღება*, *გადაწყვეტილების მიღებამდე*, *გადაწყვეტილების მიღების*, *გადაწყვეტილების მიღებაში* და ა.შ. ქეევ-ის მონაცემების მიხედვით, აღნიშნული ყოველთვის არ არის დაკავშირებული სამართლებრივ კონტექსტთან. ანალოგიური შეიძლება ვთქვათ საწყისზე „გამოტანა“, რომლითაც ყველაზე ხშირად შემდეგი კოლოკაციები იქმნება: *დასკვნის გამოტანა*, *განაჩენის გამოტანა*, *სააშკარაოზე გამოტანა*, *ვერდიქტის გამოტანა*. მათგან ორი, *განაჩენის გამოტანა* და *ვერდიქტის გამოტანა* სამართლებრივი დისკურსის ნაწილია, *გადაწყვეტილების გამოტანა* კი - მათი სინონიმი.

სუბსტანტიური მსაზღვრელ-საზღვრული *გადაწყვეტილების გამოტანა* ქართული ენის ეროვნულ კორპუსში გამოიყენების მაღალი სიხშირით არ გამოირჩევა, თუმცა ყველა შემთხვევაში სამართლებრივ კონტექსტში გვხვდება, იმის მიუხედავად, თუ რა ჟანრის ტექსტში დასტურდება იგი. აქედან გამომდინარე, შეიძლება ითქვას, რომ ისიც ქმნის მყარ შესიტყვებას, რომელიც იურიდიულ დისკურსში აქტიურად გამოიყენება და

მნიშვნელოვანია, რომ იგი დარგობრივი ლექსიკის შემადგენელ კომპონენტად მოინიშნოს კორპუსში.

გარდა ზემოთქმულისა, ჩვენ მიერ დამუშავებულ ტექსტებში დასტურდება სხვა მაგალითებიც:

პირობითი ვადა, ზეპირი მოსმენა, სასჯელის დანიშვნა, მტკიცების ტვირთი, შემამსუბუქებელი გარემოება, ხანდაზმულობის ვადა, ჯგუფური დაუმორჩილებლობა, გადაუდებელი აუცილებლობა, დაუძლეველი ძალა, [არასრულწლოვანის] საუკეთესო ინტერესები. დამუშავებულ ტექსტებში აღნიშნული შესიტყვებები აქტიურად გამოიყენება, ქართული ენის ეროვნული კორპუსის ბაზაზე კი დასტურდება მათი მიმართება სამართლებრივ კონტექსტთან. აქედან გამომდინარე, ვფიქრობთ, რომ საკანონმდებლო ტექსტების კორპუსლინგვისტური დამუშავებისას აუცილებელია მათი ტერმინებად მონიშვნა.

დასკვნა. ჩვენ მიერ კორპუსლინგვისტურად დამუშავებული ტექსტებისა და ქართული ენის ეროვნული კორპუსის მონაცემების შედარების საფუძველზე ვეცადეთ გვეჩვენებინა, რომ საკანონმდებლო ტექსტებში გვხვდება დარგობრივი შინაარსის გადმომცემი ლექსიკა, რომელიც ერთზე მეტი კომპონენტისგან შეიძლება შედგებოდეს და ამ ფორმით იძენდეს დარგისათვის სპეციფიკურ მნიშვნელობას. თუ მათ ერთმანეთისგან დამოუკიდებლად განვიხილავთ, შესაძლოა, იურიდიულ სფეროს არც უკავშირდებოდეს. როგორც წესი, ამგვარი სინტაგმები შედგება ძირითადი ზმნისა და მასთან მყარად დაკავშირებული არსებითი სახელის ან საწყისის კომბინაციისაგან, რომლებიც საკანონმდებლო ტექსტებში სამართლებრივი პროცესების აღსაწერად გამოიყენება.

იურიდიული ტექსტებისათვის ნიშანდობლივი ამ მახასიათებლების გამოვლენა და გათვალისწინება მნიშვნელოვანია, ერთი მხრივ, იურიდიული ტექსტების კორპუსლინგვისტური დამუშავებისათვის, მეორე მხრივ კი, იურიდიულ ტერმინთა ბანკის შესავსებადაც, ვინაიდან ორკომპონენტური ზმნები კორპუსლინგვისტურად უნდა დამუშავდეს არა ცალკეული ლექსიკური ერთეულების დონეზე, არამედ იმ ფორმით, რომლითაც დაკავშირებულია სამართლის სფეროსთან. ამ გზით უფრო სრულყოფილად წარმოჩნდება საკანონმდებლო ტექსტების ლინგვისტური მახასიათებლები, რაც სანდო შედეგების მიღების წინაპირობაა.

გამოყენებული ლიტერატურა

- არასრულწლოვანთა მართლმსაჯულების კოდექსი.* (2015). მოძიებულია ვებგვერდიდან: <https://matsne.gov.ge/>
- თანდაშვილი, მ., & ყამარაული, მ. (2021). *შესავალი დიგიტალურ ქართველოლოგიაში*. თბილისი: გამომცემლობა „ივერიონი“.
- პატიმრობის კოდექსი.* (2010). მოძიებულია ვებგვერდიდან: <https://matsne.gov.ge/>
- პენიტენციური კოდექსი.* (2023). მოძიებულია ვებგვერდიდან: <https://matsne.gov.ge/>
- რუხიაშვილი, ლ. (2019). *იურიდიული ტერმინოლოგია ენობრივი და სოციოკულტურული თვალსაზრისით*. ფილოლოგიის დოქტორის აკადემიური ხარისხის მოსაპოვებლად წარმოდგენილი დისერტაცია. თბილისი. მოძიებულია ვებგვერდიდან: http://press.tsu.ge/data/image_db_innova/რუხიაშვილი%20ლია-დისერტაცია-საბოლოო-ვერსია.pdf
- ხალვაში, რ. (2018). *დიგიტალური ჰუმანიტარიის შესავალი*. ბათუმი: ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტი.
- მოლი, კ., & დრინოჩი, თ. (2015). *საკანონმდებლო პროცესის შეფასება საქართველოში*. ეუთო. მოძიებულია ვებგვერდიდან:

https://legislationline.org/sites/default/files/documents/0e/256_Legis_Assessment_GEO_Jan_2015_geo.pdf

ქართული ენის ეროვნული კორპუსი. ვებგვერდი: <http://gnc.gov.ge>

Satoru, T. (2023). *KWIC Concordance*, ver. 5.4.

REFERENCES

- arasruts'lovanta martlmsajulebis k'odeksi [Juvenile Justice Code]*. (2015). Retrieved from: <https://matsne.gov.ge/>
- tandashvili, m., & qamarauli, m. (2021). *shesavali digit'alur kartvelologiashi [Introduction to Digital Kartvelology]*. Tbilisi: Publishing House „Iverioni“.
- p'at'imrobis k'odeksi [Imprisonment Code]*. (2010). Retrieved from: <https://matsne.gov.ge/>
- p'enit'entsiuri k'odeksi [Penitentiary Code]*. (2023). Retrieved from: <https://matsne.gov.ge/>
- rukhiashvili, l. (2019). *iuridiuli t'erminologia enobrivi da sotsiok'ult'uruli tvalsazrisit [Legal terminology from a linguistic and sociocultural perspective]*. Dissertation submitted for the academic degree of Doctor of Philology. Tbilisi. Retrieved from: http://press.tsu.ge/data/image_db_innova/რუხიაშვილი%20ლია-დისერტაცია-საბოლოო-ვერსია.pdf
- khalvashi, r. (2018). *digit'aluri humanit'ariis shesavali [Introduction to Digital Humanities]*. Batumi: Shota Rustaveli Batumi State University.
- moli, k'., & drinochi, t. (2015). *sak'anonmdeblo p'rotsesis shepaseba sakartveloshi [Assessment of the legislative process in Georgia]*. OSCE. Retrieved from: https://legislationline.org/sites/default/files/documents/0e/256_Legis_Assessment_GEO_Jan_2015_geo.pdf
- kartuli enis erovnuli k'orp'usi [Georgian National Corpus]*. Website: <http://gnc.gov.ge>
- Satoru, T. (2023). *KWIC Concordance*, ver. 5.4.