ქართული ენის ფლობის სახელური მახასიათებლების განსაზღვრა ქართული ენის კორპუსის (ქეკ) ანალიზის საფუძველზე

ნორა შირინიანი

სამაგისტრო ნაშრომი წარდგენილია ილიას სახელმწიფო უნივერსიტეტის მეცნიერებათა და ხელოვნების ფაკულტეტის გამოყენებითი ენათმეცნიერების მაგისტრის აკადემიური ხარისხის მინიჭების მოთხოვნების შესაბამისად

გამოყენებითი ენათმეცნიერება

სამეცნიერო ხელმძღვანელი: ირინა ლობჟანიძე, პროფესორი

ილიას სახელმწიფო უნივერსიტეტი

თბილისი

2023

განაცხადი

როგორც წარდგენილი ნაშრომის ავტორი, ვაცხადებ, რომ ნაშრომი არ შეიცავს სხვა ავტორების მიერ აქამდე გამოქვეყნებულ, გამოსაქვეყნებლად მიღებულ ან დასაცავად წარდგენილ მასალებს, რომლებიც ნაშრომში არ არის მოხსენიებული ან ციტირებული სათანადო წესების შესაბამისად.

ნორა შირინიანი

2023 წელი

აბსტრაქტი

ნაშრომში განხილულია ქართული ენის ათვისების დონეების მიმართება ლექსიკური მარაგის სტატისტიკურ დამუშვებასთან ქართული ენის კორპუსში წარმოდგენილი მასალის ანალიზის გზით. კერძოდ, ნაშრომის კვლევის საგანს ოთხი მეტყველების ნაწილი წარმოადგენს, როგორიცაა არსებითი სახელი, ზედსართავი სახელი, რიცხვითი სახელი და ნაცვალსახელი.

საკვლევი საკითხის პრობლემატიკა მდგომარეობს იმაში, რომ მართალია CEFR-ით არის განსაზღვრული ქართული ენის ფლობის დონეები, თუმცა, რეალურად, ლექსიკონი, რომელიც ესადაგება ამ ფლობის დონეებს, რომლის დადგენასაც სჭირდება დამატებითი მეთოდოლოგია, ქართული ენისათვის შემუშავებული არ არის. დამატებით მეთოდოლოგიაში მოიაზრება კორპუსული მეთოდოლოგია, რომლის გამოყენების საფუძველზეც შევეცადეთ შეგვეთანაბრებინა ქართული ენის ფლობის დონეები ლექსიკურ მასალასთან.

სწორედ ამ მიზნით, ნაშრომში განიხილება ქართული ენის კორპუსიდან ამოღებული 5000 ყველაზე ხშირი გამოყენების მქონე ერთეული შემდეგი მეტყველების ნაწილებისათვის: არსებითი სახელი, ზედსართავი სახელი, რიცხვითი სახელი და ნაცვალსახელი.

ჩატარებული ანალიზის საფუზველზე გამოვლინდა, რომ ქართული ენის შემთხვევაში A1 დონის ასათვისებლად საჭიროა 500-ზე მეტი ლექსიკური ერთეულის ფლობა ინგლისურის და სხვა ენებისგან განსხვავებით. შესაბამისად, მომდევნო დონეების ფლობისათვისაც ბევრად უფრო მეტი ლექსიკური ერთეულია საჭირო.

საძიებო სიტყვები: ფლობის დონეები, ქართული ენის კორპუსი, სტატისტიკური ანალიზი

Abstract

The paper discusses the relationship between the levels of acquisition of the Georgian language and the statistical processing of vocabulary through the analysis of the material presented in the corpus of the Georgian language. In particular, four parts of speech, such as noun, adjective, numeral and pronoun, are the subject of research.

The problem of the research issue lies in the fact that although the CEFR defines the levels of Georgian language proficiency, in reality, a dictionary that is suitable for these proficiency levels, which requires an additional methodology, has not been developed for the Georgian language. The additional methodology includes the corpus methodology, based on the use of which we tried to equate the levels of Georgian language proficiency with the lexical material.

For this very purpose, we took a corpus of 5,000 most frequently used words and analyzed selected parts of speech (nouns, adjectives, numerals and pronouns).

On the basis of the conducted analysis, it was revealed that mastering the A1 level requires the possession of more than 500 lexical units unlike English and other languages.

Kaywords: language proficiency levels, Georgian language corpus, statistical analyzis

მადლობა

მადლობას ვუხდი ჩემს ხელმძღვანელს, ირინა ლობჟანიძეს, ახალი ცოდნის გაზიარებისათვის, ნაშრომზე მუშაობისას გაწეული ყურადღებისა და დახმარებისათვის.