

TRENDS IN COMPUTATIONAL LINGUISTICS IN DISCOURSE SEQUENCE RESEARCH

კომპიუტერული ლინგვისტიკის ტენდენციები დისკურსის თანამიმდევრობის კვლევისას

Tinatin Mshvidobadze

Doctor of Technical Sciences,
Associate Professor of Gori State University,
Gori, Chavchavadze st. #53, 1400, Georgia,
+995555118379 tinikomshvidobadze@gmail.com
<https://orcid.org/0000-0003-3721-9252>

Abstract. Today, as a result of the rapid development of network technologies, data-oriented methods are widely used in natural language processing.

The aim of the paper is to demonstrate that, using appropriate algorithms and methodologies, researchers and practitioners of computational linguistics can tackle various language processing challenges and continuously make innovative solutions to improve human language cognition.

The paper shows the development of the field of computational linguistics, theoretical and applied components, its connection with the understanding of natural language. Some of the most important applications are discussed with reference to the problems they face in teaching and learning.

From the perspective of computational linguistics, natural language processing refers to the process of using computers related to speech and semantics. On the one hand, probability theory, on the other hand, network technologies contribute to the rapid development of natural language.

The concept of "chain of events" is presented in the paper. The chain of events is a necessary condition for the analysis of discourse sequence and partial dependence.

Based on the chain of events, the paper proposes a model for representing the semantic structure of the text, on the basis of which it is possible to use the sequential resources of the text for semantic analysis.

In order to alleviate the difficulty of computer-based event recognition and to reduce the obstacles in computing text sequences, a manual recognition method is presented in this paper.

A study of discourse sequencing from a computational linguistics perspective is presented.

Key words: computational linguistics, lexical chain, chain of events, discourse.

თინათინ მშვიდობაძე

ტექნიკურ მეცნიერებათა დოქტორი,
გორის სახელმწიფო უნივერსიტეტის ასოცირებული პროფესორი,
ქ. გორი, ჭავჭავაძის ქ. #53, 1400, საქართველო,
+995555118379 tinikomshvidobadze@gmail.com
<https://orcid.org/0000-0003-3721-9252>

აბსტრაქტი. დღეისათვის ქსელური ტექნოლოგიების სწრაფი განვითარების შედეგად მონაცემებზე ორიენტირებული მეთოდები ფართოდ გამოიყენება ბუნებრივი ენის დამუშავების პროცესში.

ნაშრომის მიზანია აჩვენოს, რომ შესაბამისი ალგორითმებისა და მეთოდოლოგიების გამოყენებით გამოთვლითი ლინგვისტიკის მკვლევარებსა და პრაქტიკოსებს შეუძლიათ გაუმკლავდნენ ენის დამუშავების სხვადასხვა გამოწვევებს და განუწყვეტლივ მიიღონ ინოვაციური გადაწყვეტილებები ადამიანური ენების შემეცნების გასაუმჯობესებლად.

ნაშრომში ნაჩვენებია გამოთვლითი ლინგვისტიკის სფეროს განვითარება, თეორიული და გამოყენებითი კომპონენტები, მისი კავშირი ბუნებრივი ენის გაგებასთან. განხილულია ზოგიერთი, ყველაზე მნიშვნელოვანი აპლიკაციები, სწავლებისა და სწავლის პროცესში მათ წინაშე არსებული პრობლემების მითითებით.

გამოთვლითი ლინგვისტიკის პერსპექტივიდან ბუნებრივი ენის დამუშავება ეხება მეტყველებასა და სემანტიკასთან დაკავშირებულ კომპიუტერების გამოყენების პროცესს. ერთის მხრივ, ალბათობის თეორია, მეორეს მხრივ, ქსელური ტექნოლოგიები ხელს უწყობს ბუნებრივი ენის სწრაფ განვითარებას.

ნაშრომში წარმოდგენილია „მოვლენათა ჯაჭვის“ კონცეფცია. მოვლენათა ჯაჭვი აუცილებელი პირობაა დისკურსის თანამიმდევრობისა და ნაწილობრივი დამოკიდებულების ანალიზისათვის.

მოვლენათა ჯაჭვზე დაყრდნობით, ნაშრომი გვთავაზობს ტექსტის სემანტიკური სტრუქტურის წარმოდგენის მოდელს, რომლის საფუძველზეც შესაძლებელია სემანტიკური ანალიზისათვის ტექსტის თანამიმდევრული რესურსების გამოყენება.

კომპიუტერზე დაფუძნებული მოვლენის ამოცნობის სირთულის შესამსუბუქებლად და ტექსტის თანამიმდევრობის გამოთვლისას არსებული დაბრკოლებების შესამცირებლად ნაშრომში წარმოდგენილია ხელით ამოცნობის მეთოდი.

ნაჩვენებია დისკურსის თანამიმდევრობის კვლევა გამოთვლითი ლინგვისტიკის პერსპექტივიდან.

საკვანძო სიტყვები: კომპიუტერული ლინგვისტიკა, ლექსიკური ჯაჭვი, მოვლენათა ჯაჭვი, დისკურსი.

შესავალი. კომპიუტერული ლინგვისტიკა არის ინტერდისციპლინარული სფერო, რომელიც აერთიანებს ბუნებრივი ენის შესწავლას კომპიუტერული ტექნოლოგიების გამოყენებით, ალგორითმებისა და სისტემების მეშვეობით, რომლებსაც შეუძლიათ ენის მონაცემების დამუშავება, ანალიზი და გენერირება. იგი მოიცავს როგორც ლინგვისტური თეორიის გაგებას, ასევე პრაქტიკული აპლიკაციების განვითარებას.

კომპიუტერული ლინგვისტიკა შეიძლება ჩაითვალოს ბუნებრივი ენის ავტომატური დამუშავების სინონიმადაც, რომელიც ეხება კომპიუტერული პროგრამების შექმნას

სიტყვებისა და ტექსტების ბუნებრივ ენებზე დასამუშავებლად. (Bolshakov & Gelbukh, 2004: 523).

კომპიუტერული ლინგვისტიკა განვითარდა მეცნიერების ადრეულ ხანაში. მკვლევარ ქეი მარტინის აზრით, ფრაზა "გამოთვლითი ლინგვისტიკა" წარმოიშვა 1965 წელს, როდესაც გამოჩნდა ჟურნალი "მექანიკური თარგმანი და გამოთვლითი ლინგვისტიკა" (Martin, 2003: 423).

ამ დარგის პიონერები ალან ტურინგი და ნოამ ჩომსკი ცდილობდნენ შეესწავლათ კავშირი ენას, შემეცნებასა და ხელოვნურ ინტელექტს შორის. არსებობს რამდენიმე მნიშვნელოვანი ეტაპი ამ დარგის ევოლუციაში:

1. ტურინგის ტესტი (1950) - ალან ტურინგმა შემოგვთავაზა ტესტი, მანქანის ინტელექტუალური ქცევის უნარის დასადგენად, რომელიც ექვივალენტურია ან არ განსხვავდება ადამიანის ინტელექტისაგან, როდესაც მანქანას შეუძლია ჩართვა ბუნებრივ ენობრივ საუბარში ადამიანთან. (Turing, 1950: 433).

2. ჩომსკის იერარქია (1956) - ნოამ ჩომსკიმ შემოგვთავაზა ფორმალური ენის იერარქია, რომელშიც კლასიფიცირებულია სხვადასხვა გრამატიკის ტიპები, რამაც საფუძველი ჩაუყარა ბუნებრივი ენის დამუშავების ფორმალიზებას. (Chomsky, 1956: 113)

3. პირველი მანქანური თარგმანის პროექტი (1954) - IBM-მა და ჯორჯთაუნის უნივერსიტეტმა დანერგეს სისტემა, რომელშიც ითარგმნა ინგლისურად 60 რუსული წინადადება, რაც მანქანური თარგმანის კვლევის დასაწყისს აღნიშნავს.

4. მთვარის სისტემის შექმნა (1969) - როჯერ შენკმა სტენფორდის უნივერსიტეტში შეიმუშავა სისტემა LUNAR. ესაა ბუნებრივი ენის ინტერფეისის სისტემა რომელიც საშუალებას აძლევდა მომხმარებლებს მონაცემთა ბაზების შექმნისა, იგი ინახავდა ინფორმაციას მთვარის კლდის ნიმუშების შესახებ ინგლისური წინადადებების გამოყენებით.

ამ ისტორიულმა განვითარებამ გზა გაუხსნა თანამედროვე ენათმეცნიერებს, რომლებიც ახლა მუშაობენ აპლიკაციების ისეთ ფართო სპექტრზე, როგორცაა: მანქანური თარგმანი (MT), ტექსტის ავტომატური გადაყვანა ერთი ენიდან მეორეზე, ბუნებრივი ენის დამუშავება (NLP), ალგორითმების შემუშავება ადამიანური ენის მანიპულირებისა და გაგებისათვის, მეტყველების ამოცნობა - ტექნოლოგია, რომელიც საშუალებას აძლევს კომპიუტერს სალაპარაკო ენა ტექსტად გარდაქმნას, ტექსტის მეტყველების გენერაცია (სალაპარაკო ენის გენერირების პროცესი). ინფორმაციის მოპოვება - ტექნიკა დიდი მოცულობის ტექსტური მონაცემებიდან მნიშვნელოვანი ინფორმაციის მოსაძიებლად.

კომპიუტერული ლინგვისტიკის დინამიური ბუნება უზრუნველყოფს, რომ ეს ინტერდისციპლინარული სფერო მუდმივად იკვლევს ახალ მეთოდებს, აპლიკაციებსა და თეორიებს და აძლიერებს ჩვენს შემეცნებას მათთან ურთიერთქმედებისას.

მეთოდები მოცემულ სტატიაში გამოყენებულია აღწერითი, ანალიზის და განმარტების მეთოდები, რის საფუძველზეც გამოკვეთილია აღნიშნული კვლევის მნიშვნელოვანი საკითხები. კონცეფციის ჩამოსაყალიბებლად დავიმოწმეთ სხვადასხვა მკვლევარების შეხედულებები, მათ საფუძველზე მოვახდინეთ მსჯელობისა და დასკვნების ილუსტრირება.

შედეგები და მსჯელობა. კომპიუტერული ლინგვისტიკა ეყრდნობა მდიდარ თეორიულ საფუძველს და იყენებს სხვადასხვა მიდგომებს ენის დამუშავების ამოცანების მიერ წარმოქმნილი გამოწვევების დასაძლევად. არსებობს გამოთვლითი ლინგვისტიკის ზოგიერთი ცნობილი თეორია და მიდგომა, როგორცაა:

ტრანსფორმაციული გრამატიკა: ნოამ ჩომსკის მიერ შემოთავაზებულ იქნა თეორია, რომელიც ამტკიცებს, რომ ენის სტრუქტურა შეიძლება წარმოდგენილი იქნას წესების გამოყენებით, წინადადებების საწყისი მდგომარეობიდან, ხილულ სინტაქსურ ფორმად. ამ თეორიამ გავლენა მოახდინა იმაზე, თუ როგორ აყალიბებს კომპიუტერული ლინგვისტიკა ბუნებრივი ენის გრამატიკას აპლიკაციებში.

შეზღუდვებზე დაფუძნებული გრამატიკა: ენის დამუშავების ეს მიდგომა ხაზს უსვამს სხვადასხვა ენობრივი შეზღუდვების დაკმაყოფილებას, როგორცაა სინტაქსთან, სემანტიკასა და კონტექსტთან დაკავშირებული შეზღუდვები და წინადადების სწორი ინტერპრეტაციის დადგენა).

სასრული მდგომარეობის აპარატი (FSA): FSA არის მარტივი გამოთვლითი მოდელები, რომლებიც გამოიყენება შესატყვისი შაბლონების ამოცნობის მიზნით. მისი აპლიკაციები გამოიყენება გამოთვლით მორფოლოგიაში.

სტატისტიკური მანქანური თარგმანი (SMT): SMT არის მანქანური თარგმანის მიდგომა, რომელიც ეყრდნობა პარალელური კორპუსებიდან შესწავლილ სტატისტიკურ მოდელებს და რომლებიც შეიცავს წყაროსა და სამიზნე ენის წინადადებებს. ის იყენებს ორენოვან კორპუსებს თარგმანის ალბათობის შესასწავლად. SMT ტექნიკის მაგალითები მოიცავს ფრაზებსა და სინტაქსზე დაფუძნებულ იერარქიულ მოდელებს.

ნეირონული მანქანური თარგმანი (NMT): NMT არის თარგმანის მიდგომა, რომელიც იყენებს სიღრმისეული შესწავლის ტექნიკას, განსაკუთრებით განმეორებად ნეირონულ ქსელებს, რათა მოდელირდეს წყაროს წინადადება სამიზნე წინადადებაში თარგმანის პირობითი ალბათობით. NMT მოდელები განსაკუთრებით წარმატებულია მანქანური თარგმანის ხარისხის გაუმჯობესებაში ტრადიციულ სტატისტიკურ მეთოდებთან შედარებით.

ბუნებრივი ენის გაგებასთან დაკავშირებით არსებობს კვლევის ორი მიმართულება. ერთი მიმართულია ტექსტზე დაფუძნებული აპლიკაციებისაკენ, მეორე კი დიალოგზე დაფუძნებული აპლიკაციებისაკენ. ტექსტზე დაფუძნებული აპლიკაციები გულისხმობს ყველა სახის წერილობითი ტექსტის დამუშავებას, როგორცაა წიგნები, სტატიები, შეტყობინებები, ჟურნალები და ა.შ. დიალოგზე დაფუძნებული პროგრამული აპლიკაციებია: კითხვა-პასუხის სისტემა, მომხმარებელთა ავტომატური მომსახურება საკომუნიკაციო ხაზებით, რეპეტიტორული სისტემები, რომლებშიც სტუდენტს შეუძლია მანქანასთან ურთიერთობა. (Chapelle, 2001: 227).

მნიშვნელოვანია კომპიუტერული ლინგვისტიკის როლი ენის სწავლებისა და სწავლის პროცესში. კომპიუტერის დახმარებით ენის შესწავლა არის სწავლებისა და სწავლის პროცესის საშუალება - CALL. მკვლევარმა ლიმ დაყო CALL-ის გამოყენების პერიოდი სამ ეტაპად. (Lee, 2000: 218).

ყოველი ეტაპი გვიჩვენებს ტექნოლოგიის დონეს და მისი ეპოქის პედაგოგიურ თეორიებს: ა). ბიჰევიორისტული CALL, რომლის გამოყენებაც პირველად დაიწყო 1960-იან და 1970-იან წლებში, ძირითადად დამოკიდებული იყო ენის განმეორებით ვარჯიშებზე. ბ). კომუნიკაციური CALL, რომელიც გაჩნდა 1970-იან და 1980-იან წლებში, ისინი ორიენტირებული იყვნენ ორიგინალური გამონათქვამების გენერირებაზე, განმეორებითი სავარჯიშოების გამოყენებით. გ). ინტეგრაციული CALL, კომუნიკაციური ენის სწავლების შემეცნებითი ხედვიდან სოციალურ-კოგნიტურ ხედვამდე გადასვლის ეტაპი, რომელშიც ხაზგასმულია რეალური ენის გამოყენების მნიშვნელოვანი ავთენტური კონტექსტი. ეს ეტაპი ასევე ხაზს უსვამს ენის სწავლის ოთხი უნარ-ჩვევის (მოსმენა, ლაპარაკი, წერა და კითხვა) და ტექნოლოგიების ინტეგრაციას.

დისკურსის თანამიმდევრობის ანალიზი

„დისკურსის თანამიმდევრობა“ ყოველთვის იყო ერთ-ერთი ცხელი წერტილი ლინგვისტიკაში. ტექსტის კვლევისას დისკურსის თანამიმდევრობა სულ უფრო მნიშვნელოვან როლს თამაშობს ინფორმაციის დამუშავებასა და ბუნებრივი ენის თარგმანის გენერირებისას (Rodriguez-Aguilar & Marmolejo-Saucedo, 2019: 2686)

ბოლო წლებში დისკურსის თანამიმდევრობის ანალიზის მოდელებს სულ უფრო მეტი ყურადღება ექცევა ბუნებრივი ენის დამუშავების სფეროში. ბევრი მკვლევარი ცდილობს გამოიყენოს კომპიუტერები დისკურსის თანამიმდევრობასთან დაკავშირებული შინაარსის დასამუშავებლად (Azadegan, Srinivasan, Blome & Tajeddini, 2019: 215). ამ პროცესის არსი იმაში მდგომარეობს, რომ თანამიმდევრული ტექსტების საფუძველზე მანქანამ უნდა შეძლოს დისკურსის შესაბამისი სტრუქტურის დამუშავება.

განვიხილავთ მოვლენათა ჯაჭვზე დაფუძნებული დისკურსის თანამიმდევრობის მექანიზმს, თეორიას და მოდელს, რომელიც დაფუძნებულია მოვლენათა ჯაჭვზე, რათა გააძლიეროს ერთეულებზე დაფუძნებული დისკურსის თანამიმდევრობის მოდელების ლინგვისტური დაშვებები.

მოვლენათა კვლევა ჩვეულებრივ იყოფა სამ დონედ: მოვლენის მონიშვნა; მოვლენების ურთიერთობები, იდენტიფიკაცია და ამოღება; მოვლენების გამოყენების ტექნიკა. ზოგიერთი მკვლევარი ხშირად აერთიანებს მოვლენებსა და მოვლენათა სტრუქტურებს სინტაქსისა და ლექსიკური სემანტიკის შესწავლასთან, რათა გამოიკვლიონ, როგორ ასახავს მოვლენა სინტაქსს.

ამ კვლევებს უკვე გააჩნია შედარებით სრულყოფილი ტექნოლოგიები და ალგორითმები, რის გამოც მოვლენებს და მოვლენათა სტრუქტურებს აქვს გარკვეული ფუნქციონირება დისკურსის დამუშავების პროცესში. ეს არის მთავარი მიზეზი, რის გამოც ჩვენ ვიყენებთ მოვლენებს და მოვლენათა სტრუქტურებს დისკურსის თანამიმდევრობის გასაანალიზებლად.

მოვლენათა ჯაჭვზე დაფუძნებული ტექსტის თანამიმდევრობის შესწავლა აისახება და განისაზღვრება ტექსტში მოცემული მოვლენების თანამიმდევრობით. ყველა ძირითადი მოვლენა დაკავშირებულია ერთმანეთთან, რათა ჩამოყალიბდეს ერთი ან მეტი მოვლენათა ჯაჭვი გარკვეული ფორმით. ტექსტის თანამიმდევრობის პირობა არის ის, რომ ტექსტს შეუძლია შექმნას თანამიმდევრობა მხოლოდ იმ შემთხვევაში, თუ იგი შეიცავს მოვლენათა მინიმუმ ერთ ჯაჭვს. სხვა სიტყვებით, თუ ტექსტში ბევრია ლექსიკური ჯაჭვი მოვლენის ჯაჭვის გარეშე, ტექსტი, სავარაუდოდ არათანამიმდევრული იქნება.

მოვლენათა ჯაჭვი განისაზღვრება ერთეულის მდებარეობით დისკურსში. მოვლენათა ჯაჭვისათვის, დისკურსში თითოეულ წინადადებას შეიძლება ჰქონდეს მრავალი მოვლენა, შემდეგ შესაძლებელია მოვლენათა ჯაჭვის აგება და თითოეული ჯაჭვი შეიძლება იყოს უწყვეტი. თუ როგორ უნდა დადგინდეს დისკურსის მოვლენათა ჯაჭვის მოდელი, მოითხოვს კვლევას. ერთის მხრივ, ეს მოიცავს მოვლენათა ჯაჭვის აღწერას და, მეორე მხრივ, ამ ჯაჭვის გამოთვლის სირთულესაც.

ლექსიკური და მოვლენათა ჯაჭვის აგება

ნორმალურ ვითარებაში ტექსტი აყალიბებს დისკურსის სემანტიკას ძირითადი მოვლენების (მოვლენა - როგორც წესი, მოიცავს ცვლილებებს მოქმედებებში, მდგომარეობებსა ან თვისებებში) ამოღებით, რომლებიც დაკავშირებულია ერთმანეთთან, რათა შექმნას მოვლენათა ძირითადი ნაწილი. მოვლენათა ჯაჭვის ძირითადი კონცეფცია არის ლექსიკური ჯაჭვი. ამიტომ, ლექსიკური ჯაჭვი გამოიყენება, როგორც მინიშნება მოვლენისა და მოვლენათა ჯაჭვის დასადგენად.

ლექსიკური ჯაჭვი შედგება ურთიერთდაკავშირებული სიტყვების თანამიმდევრობისაგან. ეს სიტყვები ტექსტში განაწილებულია და ქმნიან გრამატიკული სტრუქტურისგან დამოუკიდებელ სემანტიკურ ჯაჭვს, მაგრამ მიუთითებენ საერთო

თემაზე. ლექსიკური ჯაჭვი, როგორც წესი, არის დაკავშირებული არსებითი სახელები. თანამიმდევრულ ტექსტში მოვლენათა ჯაჭვი მჭიდროდ არის დაკავშირებული ლექსიკურ ჯაჭვთან (Lynn, Choi & Kim, 2018).

მაშასადამე, შეგვიძლია მივყვეთ ლექსიკური ჯაჭვის მინიშნებებს ტექსტში მოვლენათა ჯაჭვის ასაგებად. ლექსიკური და მოვლენათა ჯაჭვი ურთიერთკავშირშია ტექსტის ცენტრალურ თემასთან. ამიტომ, უნდა არსებობდეს მოვლენათა მაგისტრალური და მრავალი განშტოების ჯაჭვი, რომლებიც მიუთითებს დისკურსის ცენტრალურ თემაზე. ძირითადი მოვლენების ჯაჭვი გასდევს მთელს ტექსტს და მოიცავს დისკურსებსა და აბზაცებს, რათა უზრუნველყოს ტექსტის საერთო თანამიმდევრობა.

მოვლენათა ჯაჭვის აგების შემდეგ, გამოითვლება თანამიმდევრობის სიმძლავრე თითოეულ ორ მიმდებარე მოვლენას შორის. თანამიმდევრობის სიმძლავრის გამოთვლა გულისხმობს მოვლენებს შორის თანამიმდევრობის გამოთვლას მთელს მოვლენათა ჯაჭვზე (Lippmann, 2019: 192).

თანამიმდევრობის სიმძლავრე Δe_i და $\Delta e_i + 1$ მოვლენას შორის აღვნიშნოთ C_i . ხოლო e_j , და e_k არის ურთიერთდაკავშირებული მოვლენათა წყვილი. უნდა ვიპოვოთ ყოველი ასეთი (e_j , e_k) წყვილი ყველა მოვლენათა წყვილს შორის, რომლებიც მონიშნულია. მათ შორის, $\text{sim}(e_j, \Delta e_i)$, რომელიც წარმოადგენს მსგავსებას e_j -სა და Δe_i -ს შორის. ამის მიზეზი იმაში მდგომარეობს, რომ თუ არსებობს მსგავსება e_j -სა და Δe_i -ს მოვლენას შორის, მოვლენა e_k მსგავსია - $\Delta e_i + 1$ მოვლენის. თუ არსებობს თანამიმდევრობა e_j , e_k მოვლენათა წყვილებს შორის, გამოდის, რომ მოვლენა Δe_i . თანამიმდევრულია - $\Delta e_i + 1$ მოვლენის.

სხვა სიტყვებით რომ ვთქვათ, მოვლენებს შორის თანამიმდევრობის სიმძლავრე გამოითვლება როგორც პრედიკატს შორის მსგავსების ჯამი და არგუმენტებს შორის მსგავსება (მაქსიმალური მნიშვნელობა ყველა კომბინაციას შორის). სიტყვებს შორის სემანტიკური მსგავსების გამოსათვლელად ($\text{sim}(e_j, \Delta e_i)$), უნდა შეფასდეს ლექსიკური ჯაჭვი, მოვლენათა ჯაჭვი და თანამიმდევრობის სიმძლავრე. შეფასების კრიტერიუმად გამოიყენება სიზუსტის დონე და გაანგარიშების მაჩვენებელი.

$$\text{სიზუსტის მაჩვენებელი} = \frac{|N \cap R|}{|N|} \quad (1)$$

$$\text{გაანგარიშების მაჩვენებელი} = \frac{|N \cap R|}{R} \quad (2)$$

მათ შორის, N არის დაკავშირებული ლექსიკური წყვილი (ან წინადადების თანამიმდევრული წყვილი), რომელიც გამოითვლება შესაბამისი ალგორითმით და R არის ხელით მონიშნული დაკავშირებული ლექსიკური წყვილი. (Fan & Xu, 2020).

ინგლისური დისკურსის თანამიმდევრული რესურსების აგება

დისკურსის თანამიმდევრული რესურსების აგების პროცესი, ფაქტობრივად, წარმოადგენს კორპუსის აგების პროცესს, რომელსაც უდიდესი მნიშვნელობა აქვს ბუნებრივი ენის დამუშავებისას. დისკურსის თანამიმდევრობის კორპუსი არის მანქანით წაკითხვადი ელექტრონული რესურსების ბიბლიოთეკა, რომელიც შედგება რეალური ბუნებრივი ტექსტებისაგან და შემთხვევითი შერჩევის მეთოდით აღინიშნება გარკვეული ლინგვისტური პრინციპების მიხედვით. ინგლისურ დისკურსში თანამიმდევრული რესურსების აგება მოვლენათა ჯაჭვზე მოიცავს შემდეგ პროცედურებს: კორპუსის შერჩევა, კორპუსის წინასწარ დამუშავება, ლექსიკური ჯაჭვის მონიშვნა, და მოვლენათა ურთიერთობის მონიშვნა.

კორპუსის წინასწარი დამუშავება

წინასწარი დამუშავება მოიცავს კორპუსიდან არასტანდარტული კორპუსის ამოღებას და თითოეული ახალი ამბების კორპუსის სემანტიკურებას (Partha, Gondy &

Kauchak, 2019: 2164). არასტანდარტული კორპუსების აღმოფხვრა ხდება ხელით წაკითხვის გზით. ტექსტის სემანტიკა და წინადადების სემანტიკა ხორციელდება მანქანის ავტომატური ამოცნობის საშუალებით: პუნქტუაციის ნიშნები ტექსტში გამოიყენება წინადადების ავტომატური სემანტიკისათვის, ხოლო იდენტიფიკატორი ტექსტში გამოიყენება აზრის ავტომატური სემანტიკისათვის.

ლექსიკური ჯაჭვის ანოტაცია

ლექსიკური ჯაჭვის სახელმძღვანელო ანოტაცია ლექსიკური ჯაჭვის ასაგებად ძირითადად იყენებს ესეებში არსებულ არსებით სახელებს და სახელობით ფრაზებს, შემდეგ კი სიტყვებს ათავსებს შესაბამის ჯაჭვში, მათ შორის სემანტიკური დამოკიდებულების მიხედვით.

მოვლენათა ჯაჭვის ანოტაცია

დისკურსში მოვლენათა ჯაჭვის ანოტაცია არსებითად წარმოადგენს დისკურსში მოვლენებისა და მათი ურთიერთობების ანოტაციას. მონიშვნის პროცესი შეიძლება შეჯამდეს, როგორც: მოვლენების განსაზღვრა ტექსტში და მოვლენათა ჯაჭვის აგება. მოვლენების დეფინიციის მიხედვით შეგვიძლია განვსაზღვროთ ზმნები, როგორც მოვლენების ტრიგერები და გამოვიყენოთ ეს ტრიგერები შესაბამისი მოვლენების დასადგენად, შემდეგ კი განვსაზღვროთ ურთიერთობა ამ მოვლენებს შორის.

ტექსტში ყოველ ორ მიმდებარე წინადადების სტრუქტურას შორის შეიძლება შეიქმნას მრავალი ლექსიკური ჯაჭვი, რამაც შეიძლება გამოიწვიოს შესაბამისი მოვლენები. სწორედ ამ მოვლენებზე დაყრდნობით, ჩვენ უნდა განვსაზღვროთ, ავამჩნიოთ თუ არა მოვლენათა ჯაჭვი. (Kong, Linnik & SamPo, 2018: 406).

დასკვნა. კომპიუტერული ლინგვისტიკა პირველად 1949 წელს დაიწყო საინტერესო თარგმანით. შემდეგ იგი გაფართოვდა და მოიცავს სხვადასხვა აქტივობებს, რომლებიც ემსახურება კომპიუტერის მომხმარებლებს ზოგადად და მათ, ვინც დაინტერესებულია ენის სწავლებითა თუ სწავლით. იგი ფუნქციონირებს ორი მიზნის მისაღწევად: ერთი არის ენის სწავლება კომპიუტერით, ინტერნეტით ან დისკებით. მეორე ისეთი პროგრამების შექმნაა, რომლითაც ლინგვისტებს შეუძლიათ ისარგებლონ ლექსიკონებით და თარგმანის შემოწმების ინსტრუმენტებით.

სტატია გვთავაზობს „მოვლენათა ჯაჭვის“ კონცეფციას, აყალიბებს დისკურსის თანამიმდევრობის თეორიას და იძლევა ახალ პერსპექტივას დისკურსის ანალიზისა და გაგებისათვის.

გენერირების თვალსაზრისით, ის ამდიდრებს დისკურსის თანამიმდევრული შეფასების გზას ენის სწავლების სფეროებში, რომელსაც ასევე აქვს გარკვეული სახელმძღვანელო მნიშვნელობა ინოვაციებისათვის. ასევე საფუძველს უყრის დისკურსის თანამიმდევრობის მექანიზმის ყოვლისმომცველ და მრავალმხრივ გამოკვლევებს.

გამოყენებული ლიტერატურა

- Azadegan, A., Srinivasan, R., Blome, K., & Tajeddini K. (2019). Learning from near-miss events: An organizational learning perspective on supply chain disruption response. *International Journal of Production Economics*. p. 215.
- Bolshakov, I. & Gelbukh, A. (2004). *Computational Linguistics: Models, Resources, Applications*. Mexico, Center for Computing Research. p. 523.
- Fan, M., & Xu, S. (2020). Research on Discourse Coherence based on the Analysis Model of Event Chain from the Perspective of Computational Linguistics. *E3S Web of Conferences* 189, 03025.

- Kong, P., Linnik, A., & SamPo, L. (2018). Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory. *International journal of speech-language pathology*. p. 406.
- Lippmann, J. (2019). Snorkelling and breath-hold diving fatalities in Australia, 2001 to 2013. Demographics, characteristics and chain of events. *Diving and hyperbaric medicine*. p. 192.
- Lynn, H., Choi, C., & Kim, P. (2018). An improved method of automatic text summarization for web contents using lexical chain with semantic-related terms. *Soft Computing*, p. 413.
- Lee, K. (2000). English Teachers' Barriers to the Use of Computer-Assisted Language Learning. In *The Internet TESL Journal*, Vol. VI, No. 12, p. 218.
- Martin, K. (2003). Linguistic Computing. in R. Mitkov, *The Oxford Handbook of Computational Linguistics*. p. 423. U.S.A.: Oxford University Press.
- Partha, M., Gondy, L., & Kauchak, D. (2019). Using Lexical Chains to Identify Text Difficulty: A Corpus Statistics and Classification Study. *IEEE journal of biomedical and health informatics*. p. 2164.
- Rodriguez-Aguilar, R., & Marmolejo-Saucedo, J. (2019). Structural Dynamics and disruption events in Supply Chains using Fat Tail Distributions. *IFAC PapersOnLine*., p. 2686.
- Turing, A. (1950). *Computing machinery and intelligence*. p. 433.
- Chomsky N. (1956). *Three models for the description of language*. *IRE Transactions of Information Theory IT-2*. p. 113.
- Chapelle C. (2001). *Computer Applications in Second Language Acquisition: Foundations for teaching, testing and research*. Cambridge University Press, Cambridge, UK, p. 227.

REFERENCES:

- Azadegan, A., Srinivasan, R., Blome, K., & Tajeddini K. (2019). Learning from near-miss events: An organizational learning perspective on supply chain disruption response. *International Journal of Production Economics*. p. 215.
- Bolshakov, I. & Gelbukh, A. (2004). *Computational Linguistics: Models, Resources, Applications*. Mexico, Center for Computing Research. p. 523.
- Fan, M., & Xu, S. (2020). Research on Discourse Coherence based on the Analysis Model of Event Chain from the Perspective of Computational Linguistics. *E3S Web of Conferences* 189, 03025.
- Kong, P., Linnik, A., & SamPo, L. (2018). Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory. *International journal of speech-language pathology*. p. 406.
- Lippmann, J. (2019). Snorkelling and breath-hold diving fatalities in Australia, 2001 to 2013. Demographics, characteristics and chain of events. *Diving and hyperbaric medicine*. p. 192.
- Lynn, H., Choi, C., & Kim, P. (2018). An improved method of automatic text summarization for web contents using lexical chain with semantic-related terms. *Soft Computing*, p. 413.
- Lee, K. (2000). English Teachers' Barriers to the Use of Computer-Assisted Language Learning. In *The Internet TESL Journal*, Vol. VI, No. 12, p. 218.
- Martin, K. (2003). Linguistic Computing. in R. Mitkov, *The Oxford Handbook of Computational Linguistics*. p. 423. U.S.A.: Oxford University Press.
- Partha, M., Gondy, L., & Kauchak, D. (2019). Using Lexical Chains to Identify Text Difficulty: A Corpus Statistics and Classification Study. *IEEE journal of biomedical and health informatics*. p. 2164.
- Rodriguez-Aguilar, R., & Marmolejo-Saucedo, J. (2019). Structural Dynamics and disruption events in Supply Chains using Fat Tail Distributions. *IFAC PapersOnLine*., p. 2686.
- Turing, A. (1950). *Computing machinery and intelligence*. p. 433.
- Chomsky N. (1956). *Three models for the description of language*. *IRE Transactions of Information Theory IT-2*. p. 113.
- Chapelle C. (2001). *Computer Applications in Second Language Acquisition: Foundations for teaching, testing and research*. Cambridge University Press, Cambridge, UK, p. 227.