

**ძირითადი პრინციპები
ქართული პოლიტიკური ზეპირმეტყველების
სენტიმენტანალიზისათვის¹
SENTIMENT-ANALYSIS
IN GEORGIAN POLITICAL DISCOURSE –
MAIN RULES FOR DETERMINING SENTIMENT²**

ელენე კადაგიშვილი
ივანე ჯავახიშვილის სახელობის
თბილისის სახელმწიფო უნივერსიტეტი
თბილისი, საქართველო

Elene Kadagishvili
Ivane Javakhishvili Tbilisi State University
Tbilisi, Georgia

ABSTRACT

Sentiment-analysis is a form of natural language processing that gives us information about the sentiment (emotion) of the text - is it positive (towards something or someone), negative or neutral;

All the examples in the presentation are drawn from the political corpus of the Georgian language.

The political discourse can be defined as political communication employed by politicians in various contexts, including public speeches, planned (inaugural texts), or unplanned speeches (television programs, participation in discussions, parliamentary speeches).

Sentiment analysis research typically encompasses three interrelated yet distinct levels (Bing Liu, 2016). These are:

- o Document level
- o Sentence level
- o Aspect level.

When defining sentiment at the aspect level, we consider several aspects: whom we value (which can be a specific person often a politician), what we value (usually, a concrete action of a specific politician), and how we value it (whether negatively, positively or neutrally) (Bing Liu, 2016). In this presentation, we will illustrate, through examples, why the aspect level of sentiment research is considered the most important.

¹ კვლევა PHDF -19 -4544 განხორციელდა შოთა რუსთაველის საქართველოს ეროვნული სამეცნიერო ფონდის მხარდაჭერით

² This research PHDF -19 - 4544 has been supported by Shota Rustaveli National Science Foundation of Georgia (SRNSFG)

We will also discuss the role of parts of speech in determining sentiment – in this case we will primarily focus on adjectives and nouns.

Usually, an adjective with positive or negative content can be attached to a semantically neutral noun and, accordingly, the sentiment of the opinion will change - it will become positive or negative; However, we may have a completely different picture if the noun itself carries sentiment. In this article, we will explore noun-adjective combinations, define types of combinations, and highlight main rules derived from the analysis of these combinations.

Based on the combinations and analyzing the contexts, we found that the most crucial element in the process of sentiment determination is not individual words but phrases. Additionally, in many cases, verb phrases (VP) play a significant role in sentiment determination, contrary to the common focus on noun phrases (NP).

We highlight the rules derived from observing the empirical material.

საკვანძო სიტყვები: ქართული პოლიტიკური ზეპირმეტყველება, პოლიტიკური დისკურსი, სენტიმენტანალიზი, კორპუსზე დაფუძნებული კვლევა.

Keywords: Georgian natural political speech, Political discourse; sentiment analysis, Corpus-based research.

ჩვენი კვლევის მიზანს წარმოადგენს ბუნებრივი ენის დამუშავება (Natural Language Processing – NLP) სენტიმენტის თვალსაზრისით, ანუ არა მხოლოდ შინაარსის, არამედ ემოციის მიხედვით. ეს საკმაოდ რთული და შრომატევადი პროცესია, რადგანაც გულისხმობს მანქანისათვის (კომპიუტერისთვის) დიდი რაოდენობის ინფორმაციის მიწოდებას და საგანგებო პროგრამების მეშვეობით მის „გაწვრთნას“, რათა შემდეგ კომპიუტერულმა პროგრამამ შეძლოს მიწოდებული წინადადების ავტომატური გაანალიზება სენტიმენტის მიხედვით.

სენტიმენტანალიზთან დაკავშირებული კვლევები გვჩვენებს იმიტომ, რომ რაც შეიძლება გავაუმჯობესოთ და დავხვეწოთ სენტიმენტანალიზატორის მუშაობა.

დროთა განმავლობაში, ინტერნეტის ფართო გავრცელებამ და მოხმარებამ, ასევე სოციალური მედია-პლატფორმებისა თუ ონლაინ-წყაროების მომრავლებამ, გამოიწვია დიდი რაოდენობით მონაცემების დაგროვება, რამაც დღის წესრიგში დააყენა ამ მონაცემების გადამუშავების, გაანალიზების საჭიროება. ბიზნესს გაუჩნდა ბუნებრივი ინტერესი და მოთხოვნილება, გაეგო საკუთარი მომხმარებლის აზრი, დამოკიდებულება პროდუქტის შესახებ, რათა, ერთი მხრივ, გაეუმჯობესებინა თავისი მარკეტინგული სტრატეგია და, მეორე მხრივ, უფრო მეტი ძალისხმევა დაეხარჯა პროდუქტის განვითარებაში. ამიტომაც სენტიმენტანალიზატორის მიმართ ინტერესი მზარდია და იგი გასცდა ბიზნეს სექტორს.

ბოლო პერიოდში საკმაოდ გაიზარდა სენტიმენტალიზის მნიშვნელობა პოლიტიკურ კვლევებში. ეს ბუნებრივიცაა, რადგან არა მხოლოდ სხვადასხვა ორგანიზაციებს ან ბიზნესის მფლობელებს სურთ გაიგონ მომხმარებლების დამოკიდებულება პროდუქტის ან სერვისის მიმართ, არამედ მთავრობასაც / სახელმწიფო სტრუქტურებსაც სურთ იცოდნენ ადამიანების აზრები თავიანთი მმართველობის მიმართ, ქვეყანაში არსებული სოციალური, ეკონომიკური თუ პოლიტიკური კლიმატის მიმართ. სენტიმენტალიზი ხშირად გამოიყენება ასევე წინასაარჩევნოდ, როდესაც სურთ, გაიგონ ამომრჩევლის დამოკიდებულება პოლიტიკური კანდიდატების მიმართ.

კვლევა მთლიანად დაეყრდნობა ქართული ენის პოლიტიკურ კორპუსს (<http://gnc.gov.ge/gnc/concordance>), რომელიც ქართული ენის ეროვნული კორპუსის ქვეკორპუსია.

სენტიმენტალიზის სამ ძირითად ტიპს გამოყოფენ, ესენია:

- პოზიტიური
- ნეგატიური
- ნეიტრალური

ეს არის ყველაზე ბაზისური დაყოფა, თუმცა, იქიდან გამომდინარე, თუ რა მონაცემებს ვაანალიზებთ, რა არის ჩვენი ინტერესის ობიექტი, სენტიმენტალიზის მიზნებიც შესაძლოა განსხვავებული იყოს და, შესაბამისად, შეიძლება განსხვავებული იყოს სენტიმენტალიზატორის ფუნქცია და საჭიროებები.

სენტიმენტალიზის კვლევას შეიძლება ჰქონდეს სამი ერთმანეთთან დაკავშირებული, თუმცა, ამავედროულად, ერთმანეთისგან განსხვავებული დონეები. ესენია (Bing Liu, Sentiment Analysis, pg. 5):

- დოკუმენტის დონე
- წინადადების დონე
- ასპექტის დონე

დოკუმენტის დონე. დოკუმენტის დონეზე ამოცანა არის ის, რომ განისაზღვროს, თუ რას გამოხატავს მთლიანი აზრი – ნეგატიურ თუ პოზიტიურ სენტიმენტს. ანალიზის დოკუმენტურ დონეზე ჩვენ ვუშვებთ, რომ საანალიზო მასალის შეფასება ეფუძნება მხოლოდ ერთ სამიზნეს (ანუ, მაგალითად, ერთ პროდუქტს ან ერთ სერვისს, ერთ პოლიტიკოსს და ა.შ.). კვლევის ამ დონეზე არ აფასებენ და ადარებენ ერთმანეთს რამდენიმე სამიზნეს, ამისათვის უფრო დეტალური ანალიზი გამოიყენება.

კვლევის ამ დონეზე ხარვეზი შეიძლება იყოს ის, რომ რეალურად, როდესაც დიდ ტექსტს ვაანალიზებთ, იქ მხოლოდ ერთი კი არა, რამდენიმე სამიზნე შეიძლება შეგვხვდეს და ძნელი განსასაზღვრია, ცალსახად ვის მიეკუთვნება ესა თუ ის სენტიმენტი. ამასთან, სენტიმენტი განისაზღვრება იმის მიხედვით, თუ პროცენტულად რომელი სენტიმენტი ჭარბობს ტექსტში. ამგვარად, გამოდის, რომ კვლევის პროცესში ჩვენ რეალურ მონაცემებს ვკარგავთ და ეს შეიძლება არასწორი დასკვნის გამოტანის მიზეზი გახდეს.

წინადადების დონე. შემდეგი დონე არის იმის განსაზღვრა, თუ რას გამოხატავს ყოველი წინადადება – პოზიტიურ, ნეგატიურ თუ ნეიტრალურ აზრს. აქვე აღვნიშნოთ, რომ „ნეიტრალური აზრი“ ჩვეულებრივ ნიშნავს პოზიტიური ან ნეგატიური სენტიმენტისაგან დაცლილ აზრს. არც ანალიზის ეს დონე გვაძლევს სრულყოფილ ინფორმაციას საანალიზო ტექსტსა და სენტიმენტზე (Bing Liu, Sentiment Analysis, 2015, pg.8).

ასპექტის დონე. ანალიზის არც დოკუმენტური დონე და არც წინადადების დონე არ გვაძლევს სრულ ინფორმაციას საანალიზო ტექსტის სამიზნეებისა და სენტიმენტის შესახებ.

წინადადებას შესაძლოა ჰქონდეს რამდენიმე სენტიმენტი. მაგალითისთვის, ჩემი ბიზნესის საქმე ძალიან კარგად მიდის ამ საწყალ ეკონომიკაში – ცოტა რთულია ამ წინადადების კვალიფიკაცია, როგორც ცალსახად პოზიტიურის ან ცალსახად ნეგატიურის, რადგანაც ეს კონკრეტული წინადადება პოზიტიურია „ჩემი ბიზნესის“ მიმართ, ხოლო ნეგატიურია „ეკონომიკის“ მიმართ.

იმისათვის, რომ მოვახერხოთ საფუძვლიანი დამუშავება ტექსტისა, ჩვენ უნდა ჩავიდეთ ასპექტის დონემდე. ანალიზის ამ დონეს კვლევის უფრო ადრეულ ეტაპზე ეძახდნენ მახასიათებლების დონეს (feature level). ანალიზის ასპექტის დონეზე ვითვალისწინებთ, ასე ვთქვათ, აზრის სამიზნეებს.

ამგვარად, მიზანი უნდა იყოს არა მხოლოდ სენტიმენტის განსაზღვრა ტექსტში, არამედ იმის განსაზღვრაც, თუ რას მიემართება ესა თუ ის სენტიმენტი, რა არის მისი სამიზნე; ამ შემთხვევაში სამიზნე ორია – „ჩემი ბიზნესი“ და „ეკონომიკა“.

სენტიმენტანალიზი, როგორც ბუნებრივი ენის ანალიზის ერთ-ერთი სახე, არ არის მარტივი პროცესი და, უპირველეს ყოვლისა, გულისხმობს იმ მეტყველების ნაწილების გამოყოფას, რომლებიც სენტიმენტის განსაზღვრაში მონაწილეობენ. მაგალითად, ასეთი შეიძლება იყოს:

• ზედსართავი სახელები: კარგი, ცუდი, საუკეთესო, საშიში და ა.შ.;

• ზმნები: აუშოვს, გააუარესა და მრავალი სხვა;

• ზმნიზედები, რომლებიც უმეტესწილად, გამოიყენება, როგორც სენტიმენტის გაძლიერების მექანიზმი და მათ შეიძლება გააძლიერონ მომდევნო სიტყვის დადებითი, ან უარყოფითი შინაარსი: ძალიან, მართლაც, საკმაოდ და ბევრი სხვა;

• იდიომატური გამონათქვამები: „ფეხები არავის მოსჭამოს“, „ყურებზე ხახვი არავის დააჭრას“, „ფულებში ბანაობს“ და ა.შ.

სენტიმენტის განმსაზღვრელად შეიძლება გვევლინებოდეს აგრეთვე უარყოფის ელემენტები და ნაწილაკები, როგორიცაა, მაგალითად, უარყოფის ნაწილაკები: არ, ვერ, ნუ; უარყოფის გამომხატველი პრეფიქსები: არა-, უ-, უ-ო, არაფერი არ, ვერაფერი ვერ, ნურაფერი ნუ და ა.შ. მაგალითისთვის: „მე არ მომწონს მათი მომსახურება“ – ამ მაგალითში სენტიმენტი ნეგატიურია

და ეს გამოხატულია ნეგაციის, ანუ უარყოფითი არ ნაწილაკის გამოყენებით. შდრ.: „მომწონს“ და „არ მომწონს“.

სენტიმენტის გადმოცემის ერთ-ერთი საშუალებაა შეიძლება იყოს სარკაზმი: „ისეთი მთავრობაა, კი უნდა იყო ამათ იმედზე“ (ავტორს სურს, გამოხატოს ირონიული, სკეპტიკური დამოკიდებულება არსებული მთავრობის მიმართ).

ზოგადად, მიჩნეულია, რომ სენტიმენტის განსაზღვრაში ზედსართავი სახელების როლი გადამწყვეტია (თუმცა სიღრმისეულმა კვლევამ გვიჩვენა, რომ ზედსართავ სახელებზე უფრო მნიშვნელოვან როლს სენტიმენტის განსაზღვრაში თამაშობენ არსებითი სახელები და ზმნები). რადგან ზედსართავი სახელების ჯგუფი საკმაოდ დიდი და მრავალფეროვანია, სენტიმენტის განსაზღვრაში მონაწილე მეტყველების ნაწილებზე მსჯელობას დავიწყებთ მსაზღვრელებით (უმჯობესია, თუ „ზედსართავი სახელის“ ნაცვლად გამოვიყენებთ ამ ტერმინს) და ჩამოვაცალიებთ სენტიმენტანალიზთან დაკავშირებულ ზოგად წესებს, რომლებიც მსაზღვრელების მონაწილეობით შეიძლება გვქონდეს.

მსაზღვრელ-საზღვრული უნდა განვიხილოთ ერთად როგორც სინტაგმა, რომელიც მთლიანობაში შესაძლოა განსაზღვრავდეს სენტიმენტს წინადადების დონეზე. თუ მსაზღვრელის (A) სენტიმენტი პოზიტიურია (Pos), არსებით სახელთან (N) კომბინაციაში სენტიმენტი იქნება პოზიტიური, თუ მსაზღვრელი ნეგატიურია (Neg), არსებით სახელთან კომბინაციაში სენტიმენტი იქნება ნეგატიური, ხოლო თუ მსაზღვრელის სენტიმენტი ნეიტრალურია (Neu), არსებით სახელთან კომბინაციაში წინადადების სენტიმენტი იქნება ნეიტრალური.

ზემოთქმული წესების სახით რომ ჩამოვწეროთ, შემდეგი შემთხვევები გვექნება:

- A (Pos) + N (Neu) >>> Sentiment (Positive)
- A (Neg) + N (Neu) >>> Sentiment (Negative)
- A (Neu) + N (Neu) >>> Sentiment (Neutral)

განვიხილოთ შესაბამისი მაგალითი:

• დარჩება საზოგადო მოღვაწეობის <pos> უძლიერეს</pos> ტიპათ, <pos> მისაბამ </pos> მაგალითათ, <pos> წარმტაც </pos> იდეალათ >>> Sentiment (Positive)

ყველა მოცემულ შემთხვევაში ჩვენ ვუშვებთ, რომ არსებითი სახელის (N) სენტიმენტი არის ნეიტრალური, რადგანაც თუ არსებით სახელს ექნება მკვეთრად გამოყოფილი ორი სენტიმენტიდან ერთ-ერთი, მაშინ ჩვენ სხვა სურათი გვექნება:

- A (Neu) + N (Pos) >>> Sentiment (Positive)
- A (Neu) + N (Neg) >>> Sentiment (Negative)

• მეგობრებო, ამ ერთმა წელმა პუტჩის შემდგომ მიგვიყვანა <neu> ეროვნულ </neu> <neg> კატასტროფამდის </neg>, ამ სიტყვის სრული მნიშვნელობით >>> Sentiment (Negative).

ზოგადად, არსებითი სახელის სენტემენტს მხოლოდ მაშინ არ აქვს გადამწყვეტი მნიშვნელობა სენტემენტის განსაზღვრისას, როცა მსაზღვრელის სენტემენტი ნეიტრალურია. მას აქვს მნიშვნელობა მაშინაც, როცა მსაზღვრელსაც აქვს სენტემენტი (დადებითი ან უარყოფითი). ამის მიხედვით განსაზღვრელი გვექნება შემდეგი შემთხვევები:

A (Pos) + N (Pos) >>> Positive

A (Pos) + N (Neg) >>> Negative

A (Neg) + N (Neg) >>> Negative

A (Neg) + N (Pos) >>> Negative

უფროხილდით <pos> ქარიზმატულ </pos> <neg> მანიპულატორებს </neg>, რომლებიც ცდილობენ გაყოფას და დაპყრობას >>> negative

სენტემენტის კვლევის პროცესში მნიშვნელოვანი საკითხი არის ის, თუ რა უნდა მივიჩნიოთ სენტემენტის განმსაზღვრელ ერთეულებად – ცალკეული სიტყვები თუ ფრაზები/გამონათქვამები? კვლევამ აჩვენა, რომ ანალიზი უფრო ეფექტურია, როცა ჩვენ პირობითი მარკერებით (“pos”, “neg”, “neu”...) მოვნიშნავთ არა ცალკეულ სიტყვებს, არამედ ფრაზებს.

შეიძლება წინადადებაში მოცემული იყოს ცალკეული სენტემენტ-სიტყვები (ვგულისხმობთ ისეთ სიტყვებს, რომელთაც, ცალკე აღებულს, ცალსახად აქვს პოზიტიური ან ნეგატიური სენტემენტი), მაგრამ გადამწყვეტი მნიშვნელობა სენტემენტის განსაზღვრის დროს მაინც მთლიანად ზმნურ ფრაზებს ენიჭებათ. განვიხილოთ ამის საილუსტრაციოდ მაგალითი:

• ამითი სომხეთის მთავრობამ დაარღვია საუკუნოებით შემუშავებული კეთილმეზობლური განწყობილება – მოცემულ წინადადებაში გვაქვს სენტემენტ-სიტყვა – კეთილმეზობლური, რომელსაც ცალკე აღებულს თუ განვიხილავთ, დადებითი სენტემენტის გამოხატუელია. მაგრამ ამ წინადადების სენტემენტი მთლიანობაში განისაზღვრება, როგორც ნეგატიური და ამას განაპირობებს ზმნური ფრაზა და ზმნური ფრაზის ბირთვი – ზმნა „დაარღვია“ – რომლითაც გადმოცემულია ნეგატიური სენტემენტი. მაშასადამე, მთლიანობაში, ამ გამონათქვამის სენტემენტი იქნება ნეგატიური.

ემპირიულ მასალაზე დაკვირვებამ აჩვენა, რომ სენტემენტის განმსაზღვრელად ცალკე აღებული რომელიმე ზედსართავი სახელი ან არსებითი სახელი კი არ უნდა მივიჩნიოთ, არამედ უმჯობესია, თუ სენტემენტის განმსაზღვრელად მივიჩნივთ ფრაზას და ხშირ შემთხვევაში, ეს არის არა სახელური ფრაზა (Noun phrase – NP), არამედ უმეტესწილად, ეს არის ზმნური ფრაზა (Verb phrase – VP). ზმნური ფრაზა არის სიტყვათა ჯგუფი, რომელიც შეიცავს ძირითად ზმნას, ე.წ. წინადადების ბირთვს და ის აგრეთვე შეიძლება შეიცავდეს დამხმარე ზმნასაც.

მაშასადამე, ზმნური ფრაზები გვაწვდიან მეტ ინფორმაციას წინადადების სენტიმენტის შესახებ, განსაზღვრავენ რა საკუთრივ სენტიმენტს, დროს, ასპექტს, მოდალობას და სენტიმენტის ინტენსივობას წინადადებაში. სენტიმენტანალიზის ალგორითმებს შეუძლია გამოიყენოს ეს ინფორმაცია იმ ინფორმაციასთან ერთად, რომელსაც ის აგრეთვე მიიღებს სხვა ინდივიდუალური სიტყვებიდან და განსაზღვროს მოცემული ტექსტის მთლიანი სენტიმენტი ან გამოავლინოს სპეციფიკური აზრი თუ ემოცია ტექსტის შიგნით.

დამოწმებული ლიტერატურა და წყაროები

- მორფოლოგია 2011:** თანამედროვე ქართული ენის მორფოლოგია, გ. გოგოლაშვილის საერთო რედაქციით. თბილისი, 2011.
- სარჯველაძე... 2000:** სარჯველაძე ზ., ფენრიხი ჰ., ქართველურ ენათა ეტიმოლოგიური ლექსიკონი. თბილისი, 2000.
- ფეიქრიშვილი 1992:** ფეიქრიშვილი ჟ., ქართული ენის მორფოლოგია. ქუთაისი, 1992.
- ქეგლ:** ქართული ენის განმარტებითი ლექსიკონი. <http://ena.ge/explanatory-online>.
- ქეგლ 2008:** ქართული ენის განმარტებითი ლექსიკონი. ახალი რედაქცია, ავთანდილ არაბულის რედაქტორობით, ტომი I, თბილისი, 2008.
- ქეგლ 2010:** ქართული ენის განმარტებითი ლექსიკონი. ახალი რედაქცია, ავთანდილ არაბულის რედაქტორობით. ტომი II, თბილისი, 2010.
- შანიძე 1980:** შანიძე ა., ქართული ენის გრამატიკის საფუძვლები. თხზულებანი თორმეტ ტომად, ტ. III, თბილისი, 1980.
- Bing:** Bing Liu, Sentiment Analysis (Mining options, Sentiments and Emotions), University of Illinois at Chikago, p. 610.
- Cinque 2010:** Cinque G., The Syntax of Adjectives. London, 2010.
- Cinque 2015:** Cinque G., Typological Studies. New York, Routledge, 2015.
- Dixon 2005:** Dixon R. M. W., Aikhenvald A. Y., Adjective Classes. New York, 2005.
- Dixon 2010:** Dixon R. M. W., Basic Linguistic Theory, volume 1, New York, 2010.
- Dixon:** Dixon R. M. W., Where have all the adjectives gone? And other essays in semantics and syntax, Berlin, Mouton.
- Löbner 2013:** Löbner S., Understanding Semantics, New York, Routledge, 2013. <http://gnc.gov.ge/gnc/page>
- Maltseva... 2019:** Maltseva A., Shilkina N., Tiomniy I., Makhnytkina O., Lizunova I., Theory of Semantic Field for Sentiment-Analysis of the Language of Specific Users' Group in Social Media (Case of Freelancer Groups), November 2019.