

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი

ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი
კომპიუტერულ მეცნიერებათა განყოფილება
გიგა მაღლაკელიძე

სამაგისტრო ნაშრომი:

ტექსტური კორპუსიდან სიტყვათა შორის კავშირების მოპოვება

ნაშრომი შესრულებულია კომპიუტერული მეცნიერებათა მაგისტრის
აკადემიური ხარისხის მოსაპოვებლად.

ხელმძღვანელი: ლიანა ლორთქიფანიძე

ანოტაცია

მოცემულ ნაშრომში განხილულია ბუნებრივ ენოვან ტექსტში სიტყვათა შორის კავშირების მოპოვების ამოცანა. მოყვანილია ამოცანის გადაწყვეტის მეთოდი. რეალიზებულია სისტემა სიტყვათა შორის კავშირების მოსაპოვებლად ქართულენოვან ტექსტში. ნაჩვენებია ექსპერიმენტების შედეგები და სხვადასხვა ალგორითმების ეფექტურობა.

Summary

The present work discusses extracting relations from natural language text. The research provides some methods solving the problem and implements the system for extracting relations from Georgian text, discusses the effectiveness of different algorithms and shows the results of experiments.

სარჩევი

შესავალი	4
ბუნებრივი ენის დამუშავების ამოცანა.....	4
თეორიული საფუძვლები.....	6
1.1 მანქანური სწავლების მეთოდების გამოყენება ტექსტებთან სამუშაოდ.....	6
1.2 მაქსიმალური ენტროპიის მოდელი.....	6
1.2.1 მაქსიმალური ენტროპიის პრინციპი.....	6
1.2.2 მაქსიმალური ენტროპიის მოდელი.....	7
2.1 ამოცანის დასმა.....	9
2.2 ანოტირების ალგორითმი	10
შეჯამება	22
გამოყენებული ლიტერატურა.....	23

შესავალი

ბუნებრივი ენის დამუშავების ამოცანა

ბუნებრივი ენის დამუშავების ამოცანა (natural language processing) გულისხმობს ტექსტის ავტომატურ ანალიზს და საინტერესო ინფორმაციის მოპოვებას.

ფართო გაგებით, ამოცანა გულისხმობს კომპიუტერისთვის ბუნებრივი ენის სწავლებას და ისეთი სისტემების შექმნას, რომელიც დიალოგურ რეჟიმში უპასუხებს მომხმარებელს.

ქვეამოცანები რომელიც ამ თემას ეხება:

- ტექსტის ანალიზი
- მანქანური თარგმანი
- ინფორმაციის მოპოვება და სხვა

შედარებით უფრო დაბალი დონის ქვეამოცანები:

- მეტყველების ნაწილების განსაზღვრა
- საკუთარი სახელების განსაზღვრა
- წინადადების სინტაქსური ანალიზი

ბუნებრივი ენის დამუშავების ამოცანა პირველად დაისვა მე-20 ე საუკუნის 50იან წლებში. ერთ-ერთ საბაზისო ნაშრომს წარმოადგენდა ალან ტურინგის კვლევა, სადაც მან წარმოადგინა ინტელექტუალური სისტემის ტესტი (ტურინგის ტესტი). სისტემა ითვლებოდა ინტელექტუალურად, თუ ადამიანი მასთან ურთიერთობისას ვერ მიხვდებოდა რომ ის მანქანას ელაპარაკებოდა (ვერ არჩევდა ადამიანისგან).

შემდგომში ასეთი სისტემების შესაქმნელად ადამიანთან ურთიერთობის ტესტი გახდა აუცილებელი.

ბევრი წარმატებული სისტემაც შეიქმნა, ერთერთი ასეთი იყო 1966 ELIZA. რომელსაც შეეძლო ადამიანთან დიალოგი.

80 იან წლებამდე, ტექსტის დამამუშავებელი სისტემები იქმნებოდა წინასწარი დაწერილი წესებით. თანამედროვე ასეთი სისტემები კი ეყრდნობიან მანქანური დაწვალის მეთოდებს.

კომპიუტერის გამოთვითი შესაძლებლობის გაძლიერებამ საფუძველი დაუდო მანქანური დაწვალის მეთოდების გამოყენებას.

მოცემულ ნაშრომში განიხილება ბუნებრივი ტექსტიდან სიტყვათა შორის კავშირების მოპოვების ამოცანა. ამოცანა გულისხმობს წინადადებიდან აზრობრივად დაკავშირებული წევრების მოპოვებას.

ნაშრომში მოყვანილია ამოცანის კვლევა და ნაჩვენებია სიტყვათა შორის კავშირების მომპოვებელი სისტემის შედეგები. განხილულია ცნობილი მიდგომები ამოცანის გადასაწყვეტად. განიხილება მანქანური სწავლების მაქსიმალური ენტროპიის მოდელი.

თეორიული საფუძვლები

1.1 მანქანური სწავლების მეთოდების გამოყენება ტექსტებთან სამუშაოდ

პირველი სისტემები ბუნებრივი ენის დასამუშავებლად იყენებდნენ წინსაწარ განსაზღვრულ წესებს.

თანამედროვე სისტემები იყენებ მანქანური დასწავლის მეთოდებს. ამ მიდგომით, წესების უმრავლესობა იწყობა ავტომატური გზით, არჩეული მოდელის დასწავლისას. როგორც წესი სწავლება ხდება დიდი მოცულობის ტექსტებიდან, რომელსაც *ლინგვისტურ კორპუსს* ეძახიან.

ნაშრომში განიხილება მაქსიმალური ენტროპიის და გრაფიკული მოდელი CRF.

1.2 მაქსიმალური ენტროპიის მოდელი

1.2.1 მაქსიმალური ენტროპიის პრინციპი

ტექსტის დამუშავების ამოცანა შეიძლება წარმოვადგინოთ როგორც კლასიფიკაციის ამოცანა, სადაც მოითხოვება განვსაზღვროთ ალბათობა **a** პასუხისა, მოცემულ კონტექსტზე **b**, აღვნიშნოთ $p(a | b)$ თი. მაგალითად წინადადებიდან საკუთარი სახელების განსაზღვრის ამოცანაში გვექნება ორი კლასი: საკუთარი და საზოგადო, კონტექსტში კი შეიძლება იყოს სიტყვები, მეტყველების ნაწილები, კავშირები და ბევრი სხვა რამ. ასე, რომ კონტექსტი შეიძლება შედგებოდეს როგორც ერთი სიტყვით, ასევე ბევრი დამატებითი ინფორმაციებით, როგორცაა მეტყველების ნაწილები. როგორც წესი, კლასი-კონტექსტის წყვილები გაცილებით მეტია ვიდრე კორპუსის ზომა და ამიტომ $p(a | b)$ შეიძლება ვერ მოიძებნოს პირდაპირ კორპუსიდან. ამიტომ აუცილებელია კარგად შევაფასოთ $p(a | b)$ ყველა შესაძლო კონტექსტ-კლას წყვილებისთვის.

მაქსიმალური ენტროპიის პრინციპით სწორი იქნება ის ალბათური განაწილება რომელსაც აქვს მაქსიმალური ენტროპია.

$$H(p) = - \sum_{x \in \text{dom } p} p(x) \log p(x) dx ,$$

სადაც $\text{dom } p$ – არის ყველა შესაძლო კლასი-კონტექსტის წყვილთა სიმრავლე.

1.2.2 მაქსიმალური ენტროპიის მოდელი

მაქსიმალური ენტროპიის მოდელში გამოიყენება დიდი ლინგვისტური კორპუსი კლასებსა და კონტექსტებზე ინფორმაციის მოსაპოვებლად. მარკირებულ კორპუსში მოიპოვება სხვადასხვა მაგალითები კლასი-კონტექსტის შესახებ რის მიხედვითაც შეგვიძლია ავაგოთ ალბათური განაწილება $p^*(a,b)$.

მაქსიმალური ენტროპიის მოდელის პრინციპის გამოსაყენებლად საჭიროა განვსაზღვროთ, თუ რა იგულისხმება განაწილების შეზღუდვებში. განაწილების შეზღუდვების განსაზღვრისთვის შემოდის ფუნქცია-თვისებები (feature functions) - რომლებიც განსაზღვრულია კლასი-კონტექსტის წყვილთა სიმრავლიდან $\{0,1\}$ სიმრავლეზე.

მოცემული k ცალითვისებისთვის, შეზღუდვას აქვს შემდეგი სახე:

$$E_p f_j = E_{p^*} f_j ,$$

სადაც $1 \leq j \leq k$, f_j არის ფუნქციათვისება, $E_p f_j$ მათემატიკური მოლოდინი f_j ფუნქციის

$$E_p f_j = \sum p(a,b) f_j(a,b) ,$$

რომელიც უნდა უდრიდეს უკვე გამოკვლეული მოლოდინს $E_{p^*} f_j$,

$$E_{p^*} f_j = p^*(a,b) f_j(a,b)$$

სადაც $p^*(a,b)$ ლინგვისტური კორპუსიდან დაკვირვების შედეგად აგებული ალბათობებია.

მაქსიმალური ენტროპიის მოდელი ეფუძნება ისეთი $p(a | b)$ ალბათური განაწილების პოვნას რომელსაც ექნება მაქსიმალური ენტროპია და ასევე დააკმაყოფილებს ყველა ფუნქცია-თვისებებისთვის ზემოთ მოყვანილ შეზღუდვებს.

ფუნქცია თვისებები როგორც წესი გამოხატავს შეზღუდვას შესაძლო კლასებისა მოცემულ კონტექსტზე. მაგალითად ამოცანაში რომელიც გულისხმობს მეტყველების ნაწილების განსაზღვრას, სადაც კლასები არის მეტყველების ნაწილები და კონტექსტი კი თვითონ სიტყვები, ერთერთი ასეთი ფუნქცია-თვისება შეიძლება იყოს შემდეგი ფუნქცია

$$f(a,b) = \begin{cases} 1, & \text{თუ } a = \text{კავშირი და } b \in \{\text{და,ან,თუ,მაგრამ,ხოლო...}\} \\ 0, & \text{წინააღდეგ შემთხვევაში} \end{cases}$$

საძიებო მოდელი

$$p' = \operatorname{argmax}_{p \in P} H(p),$$

$$P = \{p \mid E_p f_k = E_{p^*} f_k \text{ ყველა } k \text{ სთვის}\}$$

$p(a,b)$ შეიძლება წარმოვადგინოთ შემდეგი სახით :

$$p(a | b) = \frac{1}{Z(b)} \exp\left(\sum_i \lambda_i f_i(a, b)\right)$$

$$Z(b) = \sum_b \exp\left(\sum_i \lambda_i f_i(a, b)\right)$$

სადაც λ_i არის წონები რომელიც გამოითვლება $E_p f_j = E_{p^*} f_j$ ტოლობიდან.

წონების გამოთვლის მრავალი ალგორითმი არსებობს, ერთერთი ასეთი არის generalized Iterative Scaling (GIS).

ამ მოდელით, კლასიფიკაცია მოხდება შემდგენიარად

$$a' = \operatorname{argmax} P(a_i | b)$$

კლასი აირჩევა იმის მიხედვით თუ რომელს უფრო მეტი ალბათობა აქვს მოდელში.

2.1 ამოცანის დასმა

ტექსტიდან სიტყვათა შორის კავშირების მოპოვების ამოცანა, რომელიც განიხილება ამ ნაშრომში გულისხმობს შემდეგ რამეს: მოცემული გვაქვს ხელით დამუშავებული სიტყვათა წყვილები რომლებსაც მიწერილი აქვს გრამატიკული და სინტაქსური მახასიათებლები. ამ მონაცემებზე დაყრდნობით უნდა დავამუშავოთ 3 სიტყვიანი წინადადებები და ვიპოოთ მათში სიტყვათა წყვილები.

სიტყვათა წყვილები ჩაწერილია ფაილში შემდეგი სახით :

```

1 გრამ. მახას. 1 სინტ. მახას.      2 გრამ. მახას. 2 სინტ. მახას.
A Pl Nom      I_VP      V Aor Pl 3   S_V
A Pl Nom      I_VP      V Impf Pl 3 S_V
A PrPz Sg Dat  I_VP      N Sg Dat Shi  S_VP
A Sg Dat Vit   I_VP      V Aor Sg 3   S_V
A Sg Nom      N      A Sg Nom Aux O  VP
A Sg Nom      I_NP      N Pl Nom      N
A Sg Nom      I_NP      N Pr Sg Nom C  N
A Sg Nom      I_S N Sg Nom      S_S
A Sg Nom      I_VP      N Sg Nom Aux  VP
A Sg Nom      I_VP      N Sg Nom Aux  VP
A Sg Nom      I_VP      N Sg Nom Aux  VP
A Sg Nom      N      V Aor Sg 3   S_V
A Sg Nom      I_VP      V Prs Sg 2   S_V
Adv I_VP      Adv S_VP
Adv S_S Cj    I_S
Adv S_S Cj    I_S
    
```

სურათი 2.1.1

წინადადებები შემოდის შემდეგი სახით :

```

1 1 სიტყვა["gram": "1 გრამ. მახას. "]2 სიტყვა["gram": "2 გრამ. მახას. "]3 სიტყვა["gram": "3 გრამ. მახას. "]
2 ბრძენი["gram": "A Pl Nom"]თოქსა["gram": "N Sg Dat Long"]გრუნენ["gram": "V Fut Pl 3"]
3 პატოსანსო["gram": "A PsPz Sg Dat O"]წამოიწია["gram": "V Aor Sg 3"]ქალაქელი["gram": "N Vin Sg Nom"]
4 ღარიბის["gram": "A Sg Gen"]სიმდიდრე["gram": "N Sg Nom "]ბავშვებია["gram": "N Vin Pl Nom Aux O"]
5 მკვდარი["gram": "A Sg Nom "]ფოთლები["gram": "N Pl Nom"]მკვდარს["gram": "A PsPz Sg Dat"]
6 ყოველგვარი["gram": "A Sg Nom "]მალდატანების["gram": "N Sg Gen"]გარეშე["gram": "A Sg Nom"]
7 მეტი["gram": "A Sg Nom "]რაა["gram": "Pron Sg Nom Aux"]მკვდარი["gram": "A Sg Nom"]
8 იმედია["gram": "A Sg Nom "]სიკვდილიც["gram": "N Sg Nom C"]ადვილია["gram": "A Sg Nom Aux"]
9 მარტო["gram": "A Sg Nom "]საანგარიშოს["gram": "N Sg Dat"]ჩხაკუნის["gram": "Msd Sg Nom"]
10 მკვდარი["gram": "A Sg Nom "]მკვდარს["gram": "A PrPz Sg Dat"]ფოთლებში["gram": "N Pl Dat Shi"]
11 ცარიელი["gram": "A Sg Nom "]მყავს["gram": "A Sg Nom "]წყალია["gram": "N Sg Nom Aux"]
12 ახალი["gram": "A Sg Nom "]დაბრუნებულია["gram": "Ptcp Sg Nom Aux"]გერმანიიდან["gram": "N Top Sg Inst Dan"]
13 საშიში["gram": "A Sg Nom "]კი["gram": "Cj "]არაფერია["gram": "Pron Sg Nom Aux"]
14 მთავარი["gram": "A Sg Nom "]კი["gram": "Cj "]ესაა["gram": "Pron Sg Nom Aux"]
15 დარღობის["gram": "A Sg Nom "]მიხდა["gram": "V Prs Sg 1"]ვიყო["gram": "V Conj II Sg 1"]
16 ახალი["gram": "A Sg Nom "]უნდა["gram": "Part "]მოვიყვანოთ["gram": "V Conj_II Pl 1"]
17 მარტო["gram": "A Sg Nom "]ისინი["gram": "Pron Pl Nom"]უბედავენ["gram": "V Fut Pl 3"]
18 დიდი["gram": "A Sg Nom "]მადლობა["gram": "N Sg Nom "]მომისვენებია["gram": "V Result_I Sg 1"]
19 საწყალო["gram": "A Sg Voc O"]დასძახის["gram": "V Prs Sg 3"]ბებია["gram": "N Vin Sg Nom"]
20 ისევე["gram": "Adv"]დაუმშვენდა["gram": "V Aor Sg 3"]ცხვირი["gram": "N Sg Nom"]
21 ნამდვილად["gram": "Adv "]რომ["gram": "Cj "]სასაცილოა["gram": "A Sg Nom Aux"]
22 ასე["gram": "Adv "]იქნება["gram": "Part Long"]სულ["gram": "Adv"]
    
```

სურათი 2.1.2

ჩვენი მიზანია სურათი 2.1.1 ზე გამოსახულ ინფორმაციაზე დაყრდნობით , იმ გრამატიკული მახასიათებლებით რაც ქმნის სიტყვათა კავშირს, მოვძებნოთ ასეთივე კავშირები 2.1.2 სურათზე გამოსახულ წინადადებებში და მივუწეროთ სიტყვებს შესაბამისი სინტაქსური მახასიათებლები.

მაგალითად წინადადებაში :

ცარიელი[“gram”: “A Sg Nom “]მჟავე[“gram”: “A Sg Nom “]წყალია[“gram”: “N Sg Nom Aux”]

ვიხილავთ თითოეულ სიტყვათა წყვილს

- 1) ცარიელი[“gram”: “A Sg Nom “]მჟავე [“gram”: “A Sg Nom “]
- 2) ცარიელი[“gram”: “A Sg Nom “] წყალია[“gram”: “N Sg Nom Aux”]
- 3) მჟავე[“gram”: “A Sg Nom “]წყალია[“gram”: “N Sg Nom Aux”]

ვეძებთ კავშირებს გრამატიკული მახასიათებლების მიხედვით 2.1.1 ფაილში. პირველ წყვილში განვიხილავთ გრამატიკულ მახასიათებლებს (A Sg Nom , A Sg Nom). რადგანაც ასეთი კავშირი არ გვაქვს 2.1.1 სურათზე გამოსახულ ფაილში, ამიტომ განვიხილავთ შემდეგ წყვილს. ვეძებთ ისევ (A Sg Nom, N Sg Nom Aux) წყვილებს 2.1.1 ფაილში და ასეთ კავშირს ვპოულობთ :

N Sg Nom Aux VP A Sg Nom I_VP

ვთვლით რომ სიტყვებს ერთმანეთან აქვთ კავშირი და ვუწერთ შესაბამის სინტაქსურ მახასიათებლებს

ცარიელი[I_VP] წყალია[VP]

პროგრამა შემდეგი წყვილისთვისაც მოძებნის კავშირს და გამოიტანს ეკრანზე

მჟავე[I_VP] წყალია[VP]

2.2 ანოტირების ალგორითმი

წინადადების სინტაქსური ანალიზის აღწერას შევეცადეთ ისეთი ფორმით, სადაც წარმოჩენილია წინადადებაში გამოყოფილი სინტაქსური წყვილები, წინადადების წევრთა ურთიერთკავშირი და, ამავე დროს, შესაძლებელია სინტაქსური ხის აღდგენა.

პროგრამას შესავალში მიეწოდება ტექსტური კორპუსი. მომხმარებელი გამოსავალში ღებულობს წინადადებებად დაყოფილ ტექსტს, სადაც თითოეულ სიტყვაფორმას მიწერილი აქვს მისი ამოსავალი ფორმა, მორფოლოგიური და სინტაქსური მახასიათებლები. სიტყვაფორმის სინტაქსურ მახასიათებლებს განაპირობებს ის

მიმართებები, რომლებითაც სიტყვაფორმა დაკავშირებულია წინადადების ამა თუ იმ წევრთან.

სისტემით დამუშავებული ორსიტყვიანი წინადადების ნიმუში:

წინადადება	I სიტყვა	I ბაზისური ფორმა	I გრამ. მახას.	I სინტ. მახას.	II სიტყვა	II ბაზისური ფორმა	II გრამ. მახას.	II სინტ. მახას.
ღმერთმა დაიფაროს.	ღმერთმა	ღმერთი	N Rel Sg Erg	N	დაიფაროს	დაიფარავს	V Conj_II Sg 3	V
უკვე დაღამებულია	უკვე	უკვე	Adv	I_VP	დაღამებულია	დაღამდება	V Result_I Sg 3	V
აქა ვარ.	აქა	აქ	Adv Long	I_VP	ვარ	იქნება	V Prs Sg 3	V
ალბათ ვერა.	ალბათ	ალბათ	Adv	I_S	ვერა	ვერ	Part Long	I_S

ქართული ტექსტის სინტაქსური ანოტირების ლინგვისტურ მოდელს საფუძვლად დაედო დაქვემდებარებული კავშირების გრამატიკა.

სინტაქსური გარჩევის ხე წარმოდგენელია წინადადების სიტყვებს შორის ბინარული ურთიერთორიენტირებული კავშირებით. თითოეულ კავშირში მოცემულია მთავარი და არამთავარი სიტყვა.

იმისათვის, რომ შედეგად მიღებული გრაფი იყოს სინტაქსური გარჩევის ხე, უნდა შესრულდეს შემდეგი წესები:

- 1) დაცული უნდა იყოს ხის ერთიანობა. ალგორითმში დაშვებულია ამოსავალი წინადადების დაყოფა ორ და მეტ ერთმანეთისაგან დამოუკიდებელ სინტაქსურ ხედ. ამიტომ, ხის ერთიანობის შესანარჩუნებლად, წინადადების წინ შემოდის ცარიელი სიტყვა *S, რომელიც მთავრის როლშია ერთმანეთზე დამოკიდებული ხეების დაქვემდებარებაში მყოფი სიტყვათა კვანძებისთვის.
- 2) წინადადების ყველა სიტყვა, ჩართული უნდა იყოს ხის სტრუქტურაში და ყოველ მათგანს უნდა ჰყავდეს თავისი მთავარი.
- 3) დაცული უნდა იყოს ე.წ. სუბორდინაციის მოთხოვნა:
 - ა) თუ ორი, გარკვეული თანმიმდევრობით მდგარი a და b სიტყვა, სინტაქსური კავშირითაა ერთმანეთთან დაკავშირებული, მაშინ ტექსტში მათ შორის მდებარე სიტყვა არ შეიძლება დაუკავშირდეს არც ერთ სიტყვას, რომელიც მოთავსებულია a-დან მარცხნივ და b-დან მარჯვნივ. მაგალითად, გვაქვს წინადადება:

შაბათ დილას სასტუმრო "თბილისი" 2 საათში გაივსო უცხოელებით.

ვთქვათ, განვიხილავთ კავშირს a სიტყვა *სასტუმრო* და b სიტყვა *გაივსოს* შორის. აქედან *გაივსო* არის **მთავარი** და *სასტუმრო* არის **არამთავარი**. სუბორდინაციის ა) მოთხოვნის შესაბამისად, სიტყვები: *“თბილისი”*, *2*, და *საათში* არ უნდა დავუკავშიროთ არც a-ს მარცხნივ მდგარ სიტყვებს: *შაბათ* და *დილას* და არც b-ს მარჯვნივ მდგარ სიტყვას: *უცხოელებით*.

ბ) **არამთავარს**, რომელიც ტექსტში **მთავრის** მარცხნივაა მოთავსებული, არ შეიძლება ჰყავდეს თავისი **არამთავარი მთავრის** მარჯვნივ, და **არამთავარს**, რომელიც **მთავრის** მარჯვნივაა მოთავსებული, არ შეიძლება ჰყავდეს თავისი **არამთავარი მთავრის** მარცხნივ.

4) დაცული უნდა იყოს პირობა: არ შეიძლება *S სიტყვასთან არაპროექტირებული სიტყვების მიერთება.

სინტაქსური გარჩევის შედეგად თითოეულ სიტყვას მიეწერება კავშირის ის სახე, რომელშიც მოცემული სიტყვა არის **არამთავარი** ან **მთავარი**. ასეთი კავშირი შეიძლება იყოს მხოლოდ ერთი.

რაიმე კავშირში მთავრის სახით სიტყვის მონაწილეობა განსაზღვრავს მის სინტაქსურ როლს წინადადებაში და, შესაბამისად, მასზე დაქვემდებარებული ხის სინტაქსურ როლს მთლიანად. ყველა კავშირს აქვს თავისი სახელი, რომელიც აღნიშნავს **მთავრის** სინტაქსურ როლს.

თუ საქმე გვაქვს ელიფსისთან, ანუ გამოტოვებულია წინადადების რომელიმე წევრი, და მის მაგივრად ტექსტში გამოყენებულია ახალი ტოკენი, რომელიც არ არის სიტყვაფორმა, მაგრამ გარკვეული შინაარსის მატარებელია, მაგ., ტირე, მას მივუწერთ ამ შინაარსის გამომხატველ სინტაქსურ მახასიათებლებს. წინადადებაში: *შეხედავ – გაგახსენდება* ტირე აღნიშნავს კავშირს და. ამიტომ მას მივუწერთ კავშირის შესაბამის ინფორმაციას.

სინტაქსური ანალიზის დროს ჩვენ ვეყრდნობით თეორიას, რომლის მიხედვით წინადადების **ცენტრალური კვანძი** არის **პრედიკატი**, ხოლო წინადადების სტრუქტურა ძირითადად განისაზღვრება მოცემული **პრედიკატის** შესაძლო **აქტანტების** (**სუბიექტის**, **ობიექტის** და მსგ.) მიხედვით. ყურადღება ექცევა ისეთი კონსტრუქციების გარჩევას, რომლებშიც **პრედიკატი** მართავს სხვა ზმნურ ფორმებს (მაგალითად მასდარს), და ასევე ერთგვაროვანი წევრების მქონე კონსტრუქციებს.

წინადადებაში ყოველი მომდევნო ერთგვაროვანი წევრი **არამთავრის** სახით წინა წევრს უერთდება და აქვს საკუთარი სახელდებული სინტაქსური როლი. ერთგვაროვან წევრებს

შორის არსებული კავშირები და პუნქტუაციის ნიშნები (ტოკენები) შესაბამისად მარკირებულია და თავისი სინტაქსური როლით მიუერთდება წინადადების იმ წევრს, რომელიც დგას მის მარცხნივ.

როდესაც ვიწყებთ წინადადების სინტაქსურ ანალიზს, პირველ რიგში, წინადადების წევრებს ვყოფთ იმის მიხედვით, თუ რომელ ჯგუფს ექვემდებარებიან ისინი – ქვემდებარის თუ შემასმენლის (ჯგუფის სახელის შერჩევა ხდება ხის საბოლოო წვეროდან გამომდინარე).

შემოგვაქვს შემდეგი აღნიშვნები:

N - ქვემდებარე

V - შემასმენელი

VP - შედგენილი შემასმენელი

I_V – მოდალური ზმნა შემასმენელთან

I_NP - ქვემდებარის ჯგუფში შემავალი სინტაქსური წყვილის არამთავარი წევრი

S_NP - ქვემდებარის ჯგუფში შემავალი სინტაქსური წყვილის მთავარი წევრი

I_VP - შემასმენლის ჯგუფში შემავალი სინტაქსური წყვილის არამთავარი წევრი

S_VP - შემასმენლის ჯგუფში შემავალი სინტაქსური წყვილის მთავარი წევრი

I_S - წინადადების ისეთი წევრი, რომელიც არ უკავშირდება არც ერთ წინადადების სხვა წევრს, მაგ. ჩანართი, შორისდებული, მიმართვა და ა.შ. და თუ ასეთი სიტყვები სინტაქსურ წყვილს ქმნიან, მათში არამთავარია;

S_S - წინადადების ისეთი წევრი, რომელიც არ უკავშირდება არც ერთ წინადადების სხვა წევრს, მაგ. ჩანართი, შორისდებული, მიმართვა და ა.შ. და თუ ასეთი სიტყვები სინტაქსურ წყვილს ქმნიან, მათში მთავარია.

ზემოთ ჩამოთვლილ მარკერებში ყველგან „_“ ნიშნის მარცხენა მხარეს S აღნიშნავს წყვილის მთავარ წევრს, ხოლო I - არამთავარს.

განვიხილოთ მაგალითი:

ხატის წინ ბებიას სანთელი აუნთია.

პირველ რიგში ვეძებთ შემასმენელს – *აუნთია*, აღვნიშნავთ მას V მარკერით, შემდეგ ვეძებთ ქვემდებარეს – *ბებიას*, აღვნიშნავთ მას N-ით. პირდაპირი დამატება *სანთელი* უკავშირდება ზმნას და აღვნიშნავთ I_VP-თი. *ხატის წინ* სინტაქსურად არის ადგილის გარემოება და ის უკავშირდება შემასმენელს, ამიტომ მასაც მიეწერება მარკერი I_VP. სინტაქსური ანალიზის შემდეგ ჩვენი წინადადება ასე გამოიყურება:

ხატის (I_VP) წინ (S_VP) ბეზიას (N) სანთელი (I_VP) აუნთია (V); (a)

იგივე წინადადებას დავამატოთ ახალი წევრები. *პატარა ხატის წინ მოხუც ბეზიას თაფლის სანთელი აუნთია*. ეს გამოიწვევს სინტაქსური მახასიათებლების შეცვლას.

პატარა (I_VP) ხატის (S_VP) წინ (S_VP) მოხუც (I_NP) ბეზიას (N) თაფლის (I_VP) სანთელი (S_VP) აუნთია (V); (b)

a) წინადადებაში სიტყვა *ხატის* სინტაქსური მახასიათებელი იყო I_VP. b)-ში მას დაემატა განსაზღვრება *პატარა*, რის შემდეგაც ის გახდა S_VP, როგორც *მთავარი* შესიტყვებისა - *პატარა ხატის*. ასევე შეიცვალა *სანთელის* მახასიათებელიც. *თაფლის სანთელი* სახელური ფრაზაა, მაგრამ *სანთელი*, როგორც ობიექტი, შემასმენელს უკავშირდება და ამიტომაც ის აღინიშნება _VP-თი, და არა _NP-თი.

შემასმენლის აღწერისას განვიხილეთ რამდენიმე სხვადასხვა შემთხვევა:

1. მარტივი შემასმენელი - *დახატა* (V)
2. შედგენილი შემასმენელი - *დახატულია* (VP)
3. მოდალური ზმნა მარტივ შემასმენელთან - *უნდა* (I_V) *დახატოს* (V)
4. ბივერბალური შემასმენელი - *უბრძანა* (S_V) *დაეხატა* (S_V)

წინადადების სინტაქსური გარჩევის მოდულის შესავალს მიეწოდება მორფოლოგიური ანალიზის შედეგად ტეგირებული ტექსტი. გამოსავალზე სინტაქსურად ანოტირებული ტექსტი მიიღება.

წინადადების თითოეულ სიტყვაფორმას მარკერები შემდეგი თანმიმდევრობით მიეწერება: სიტყვის ნომერი წინადადებაში; სიტყვაფორმა; სიტყვაფორმის ნორმირებული ფორმა, ანუ ლემა; წინადადებაში *მთავარი* სიტყვაფორმის ნომერი; მოცემული სიტყვაფორმის *მთავარ* სიტყვაფორმასთან სინტაქსური კავშირის მარკერი; მოცემული სიტყვაფორმის გრამატიკული მახასიათებლების მარკერების კონა.

წინადადების სინტაქსური სტრუქტურის ასეთი სახით ჩაწერა გარკვეული წესების თანმიმდევრული დაცვით, საშუალებას გვაძლევს ჩავატაროთ უკუპროცესი, ანუ შესაძლებელია წინადადების სინტაქსური ხის აღდგენა.

გარდა ამისა, წინადადების ყოველ წევრს ვუწერთ ამოსავალ ფორმას და მის მორფოლოგიურ მახასიათებლებს. სახელებში ამოსავალ ფორმად მიღებულია მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმა.

მაგალითად განვიხილოთ ზედსართავი სახელი: წითლებისაგანაცააო, საკუთარი სახელი: ბაგრატიონისავთქო და ზმნა: დაუძახა.

სახელებში პირველ რიგში აღინიშნება შემდეგი მორფოლოგიური მახასიათებლები:

კრებითობა, საკუთარი, გეოგრაფიული, მითოლოგიური, რელიგიური, დროის აღმნიშვნელი, სულიერი, ვინ ჯგუფის, რა ჯგუფის. ზმნას – V Obj თუ ობიექტური უღლებაა და V – ყველა დანარჩენს.

წითლებისაგანაცააოწითელი|A

ბაგრატიონსავითთქო|ბაგრატიონი|N Pr

ბაგრატიონსავითთქო|ბაგრატიონი|N Pr

დაუძახა|დაუძახებს|V

შემდეგ სახელებისთვის ამას მოსდევს: პოზიციურობა – იმ შემთხვევაში თუ გვაქვს მსაზღვრელ–საზღვრული; გვაქვს პრეპოზიციული – PrPz (მაღალი სახლი), პოსტპოზიციური - PsPz (სახლი მაღალი) და აპოზიციური - NoPz (ვანო მასწავლებელი ან ექიმი გიორგი) წყობის მაჩვენებელი მახასიათებლები. მარკერს ვუწერთ მსაზღვრელს. ამავე დროს, აპოზიციურობის შემთხვევაში, მარკერს ვუწერთ I ადგილზე მდგომ სიტყვაფორმას. თუ სახელი I ადგილზეა, ხოლო მისი მსაზღვრელი II-ზე, სახელში ბრუნვის აღნიშვნა აღარ არის საჭირო, იგი ჭარბია, რადგან ამ შემთხვევაში სახელი არ იბრუნვის და ამიტომ ბრუნვის მარკერს აღარ ვწერთ. (ვანო მასწავლებელი, ვანო მასწავლებელმა, ვანო მასწავლებელს და ა.შ.), ხოლო, თუ მსაზღვრელი წინ უსწრებს საზღვრულს, მაშინ ბრუნვის ფორმას ორივეგან ვუთითებთ. (ექიმი გიორგი, ექიმმა გიორგიმ, ექიმ გიორგის და ა.შ.).

შემდეგ სახელებს მიეწერებათ მარკერები შემდეგი თანმიმდევრობით:

რიცხვი – ჩვეულებრივ ვუწერთ მოცემული ფორმა მხოლოდითია თუ მრავლობითი. ანუ:

ბრუნვა – მარკერით აღვნიშნავთ ბრუნვას.

თანდებული – მივუწერთ თანდებულის აღმნიშვნელ შესაბამის მარკერს.

ნაწილაკი – აღვნიშნავთ შესაბამისი მარკერით.

სავრცობი – მივუწერთ აქვს თუ არა სავრცობი.

შედგენილი შემასმენელი – მივუწერთ იმ შემთხვევაში, თუ საქმე გვაქვს შედგენილ შემასმენელთან.

სხვათა სიტყვა – აღვნიშნავთ შესაბამისი მარკერებით.

მორფოლოგიური მარკერების თანმიმდევრობა ზმნებში:

1. მწკრივი;
2. რიცხვი;
3. პირი;
4. სხვათა სიტყვა.

ამასთანავე აღწერისას მკაცრად უნდა იქნას დაცული აღნიშნული მარკერების რიგიც:

წითლებისაგანაცააოწითელი|A Pl Gen Gan C Aux O

ბაგრატონისა|ბაგრატონი|N Pr Sg Gen Long|ბაგრატონი|N Pr PsPz Sg Nom

ბაგრატონსავითთქო|ბაგრატონი|N Pr Sg Dat Vit Tko

დაუძახა|დაუძახებს|V Aor Sg 3

ჩვენ ვეყრდნობით განმარტებით ლექსიკონს, თუ მოცემული სიტყვაფორმა შეტანილია განმარტებით ლექსიკონში ცალკე ერთეულად, ჩვენც მას ასევე ცალკე ერთეულად განვიხილავთ. მაგ., „აქაც“ არ განვიხილავთ როგორც „აქ“ ფორმისგან –ც ნაწილაკის დართვით ნაწარმოებ ფორმას, არამედ როგორც დამოუკიდებელ ერთეულს, რადგან ის განმარტებით ლექსიკონში ცალკე ერთეულად შედის. ასევე ვიქცევით სხვა მსგავს შემთხვევებში.

ზმნის ბაზისურ ფორმად პირობითად ვიღებთ მის მყოფადის მხოლობითი რიცხვის მესამე პირის ფორმას. იმ შემთხვევაში, თუ მოცემულ კონკრეტულ ზმნას არ გააჩნია მყოფადის ფორმა, მაშინ ასეთ ზმნების ბაზისურ ფორმაში აწმყოს მხოლობითი რიცხვის მესამე პირის ფორმას ვწერთ. მაგ., ფრაზაში: *მე რაღას მერჩოდი?* შემასმენლის მორფოლოგიური მახასიათებლების მიწერისას, მის ბაზისურ ფორმად ავიღებთ „ერჩის“ ფორმას.

ზმნური ფორმების აღწერისას გამოყენებულია შემდეგი მორფოლოგიური მახასიათებლები: ზმნის ობიექტური წყობა, მწკრივი, რიცხვი, პირი, სხვათა სიტყვა.

როგორც ვთქვით, მკაცრად უნდა იქნას დაცული გრამატიკული მარკერების თანმიმდევრობა.

ისეთ ზმნებს, რომლებიც შინაარსით *ყოფნა* ზმნას უკავშირდება, როგორცაა მაგ., ბრძანდება გახლავს და ა.შ. ბაზისურ ფორმაში ჩავუწერთ იქნება–ს, რადგა მათ არ გააჩნიათ საკუთარი მყოფადის ფორმები. მაგ., *ასეთი გახლავთ მარგალიტა*.

თუ საქმე გვაქვს ელიფსისთან ანუ გამოტოვებულია წინადადების რომელიმე წევრი, მაგრამ ადვილად ვხვდებით რა იგულისხმება და მის მაგივრად ტექსტში გამოყენებულია ახალი ტოკენი, რომელიც არ არის სიტყვაფორმა, მაგრამ გარკვეული შინაარსის

მატარებელია მაგ., ტირე, მას მივუწერთ ამ შინაარსის გამომხატველ სინტაქსურ მახასიათებლებს. მაგ., *შეხედავ – გაგახსენდება*. ამ შემთხვევაში, ტირე აღნიშნავს კავშირ და–ს. ამიტომ მას მივუწერთ კავშირის შესაბამის ინფორმაციას.

მოგვყავს გამოყენებული მეტყველების ნაწილის და გრამატიკული მახასიათებლების მარკერების სია:

მეტყველების ნაწილი	მარკერი
არსებითი სახელი, კონკრეტული საკუთარი	N Pr
არსებითი სახელი, კრებითი	N Coll
არსებითი სახელი, პიროვნების გვარი	N Pr Lst
არსებითი სახელი, პიროვნების სახელი	N Pr Fst
არსებითი სახელი, საკუთარი, გეოგრაფიული	N Top
არსებითი სახელი, საკუთარი, მითოლოგიური	N Mit
არსებითი სახელი, საკუთარი, რელიგიური	N Rel
არსებითი სახელი, სახელი	N
არსებითი სახელი, სულიერი, ვინ ჯგუფი	N Vin
არსებითი სახელი, სულიერი, რა ჯგუფი	N Ra
არსებითი სახელი, დროის აღმნიშვნელი (თვე, კვირის დღე)	N Temp
ზედსართავი სახელი	A
ზმნა მოდალური	V Mod
ზმნა ობიექტური წყობა	V Obj
ზმნა სუბიექტური წყობა (Subj არ აღინიშნება მარკერით)	V
ზმნიზედა	Adv
თანდებული	Pstp
კავშირი	Cj
მასდარი	Msd
მიმღობა	Ptcp
მხლებელი	Adv
ნაწილაკი	Part
რიცხვითი სახელი	Num
რიცხვითი სახელი გაურკვეველი	Num Quant
რიცხვითი სახელი მიახლოებითი რაოდენობითი	Num Crd Approx
რიცხვითი სახელი რაოდენობითი	Num Crd
რიცხვითი სახელი რაოდენობითი თანდართული –ვე	Num Crd Encl Ve
რიცხვითი სახელი რიგობითი	Num Ord
რიცხვითი სახელი წილობითი (broken numbers)	Num Br
შორისდებული	Interj

გრამატიკული მახასიათებელი	მარკერი
IIIპირი	3
IIპირი	2

პირი	1
პოზიცია აპოზიციური	NoPz
ბრუნვა ვითარებითი	Ad
ბრუნვა მიცემითი	Dat
ბრუნვა მოთხრობითი	Erg
ბრუნვა მოქმედებითი	Inst
ბრუნვა ნათესაობითი	Gen
პოზიცია პოსტპოზიციური	PsPz
პოზიცია პრეპოზიციული	PrPz
ბრუნვა სახელობითი	Nom
ბრუნვა წოდებითი	Voc
დრო IIIკავშირებითი	Conj_III
დრო IIთურმეობითი	Rezult_II
დრო IIკავშირებითი	Conj_II
დრო Iთურმეობითი	Rezult_I
დრო აწმყო	Prs
დრო აწმყოს კავშირებითი	ConjPres
დრო ბრძანებითი	Imp
დრო მყოფადი	Fut
დრო მყოფადის კავშირებითი	ConjFut
დრო უწყვეტელი	Impf
დრო წარსული	Past
დრო წყვეტილი	Aor
დრო ხოლმეობითი	Cond
თანდებული გან	Gan
თანდებული დამი	Dami
თანდებული დან	Dan
თანდებული დმი	Dmi
თანდებული ებრ	Ebr
თანდებული ვით	Vit
თანდებული ზე	Ze
თანდებული თან	Tan
თანდებული თანავე	Tanave
თანდებული თვის	Tvis
თანდებული კენ	Ken
თანდებული მდე	Mde
თანდებული ურთ	Urt
თანდებული ში	Shi
ნაცვალსახელი	Pron

ნაწილაკი ვე	Ve
ნაწილაკი ღა	Gha
ნაწილაკი ყე	Ye
ნაწილაკი ც	C
რიცხვი მრავლობითი	Pl
რიცხვი მხოლოდობითი	Sg
რიცხვი ნართანიათი მრავლობითი	OldPl
სავრცობი	Long
სხვათა სიტყვა –თქვა	Tkva
სხვათა სიტყვა –თქო	Tko
სხვათა სიტყვა –მეთქი	Metki
სხვათა სიტყვა –ო	O
შედგენილი შემასმენელი	Aux

როდესაც ვიწყებთ წინადადების სინტაქსურ ანალიზს, პირველ რიგში, მის წევრებს ვყოფთ ორ ნაწილად, იმის მიხედვით, თუ რომელ ჯგუფში შედიან ისინი – ქვემდებარის ჯგუფში (noun phrase) თუ შემასმენლის ჯგუფში (verb phrase). შემოგვაქვს შემდეგი აღნიშვნები:

N - ქვემდებარე

S_V - შემასმენელი

VP - შედგენილი შემასმენელი

I_NP - ქვემდებარის ჯგუფში შემავალი სინტაქსური წყვილის არამთავარი წევრი

S_NP - ქვემდებარის ჯგუფში შემავალი სინტაქსური წყვილის მთავარი წევრი

I_VP - შემასმენლის ჯგუფში შემავალი სინტაქსური წყვილის არამთავარი წევრი

S_VP - შემასმენლის ჯგუფში შემავალი სინტაქსური წყვილის მთავარი წევრი

S - საწყისი წინადადება ან ფრაზა

I_S - წინადადების ისეთი წევრი, რომელიც არ უკავშირდება არც ერთ წინადადების სხვა წევრს, მაგ. ჩანართი, შორისდებული, მიმართვა და ა.შ. და თუ ასეთი სიტყვები სინტაქსურ წყვილს ქმნიან, მათში არამთავარია;

S-S - წინადადების ისეთი წევრი, რომელიც არ უკავშირდება არც ერთ წინადადების სხვა წევრს, მაგ. ჩანართი, შორისდებული, მიმართვა და ა.შ. და თუ ასეთი სიტყვები სინტაქსურ წყვილს ქმნიან, მათში მთავარია;

ერთმანეთთან სინტაქსურად დაკავშირებულ წყვილებში განვიხილავთ თუ რომელია მთავარი ანუ მშობელი სიტყვა და რომელია მემკვიდრე ანუ არამთავარი სიტყვა. შესაბამისად მთავარ წევრებს ყველგან აღვნიშნავთ S –ით, ხოლო არამთავარს კი I- თი.

როგორც უკვე აღვნიშნეთ, ჩვენთვის გადამწყვეტი მნიშვნელობა აქვს განვსაზღვროთ მოცემული სიტყვა ქვემდებარის ჯგუფში შედის თუ შემასმენლის ჯგუფში. თუ სიტყვა ქვემდებარის ჯგუფში შედის, მას ავლნიშნავთ I_NT-თი ან S_NT-თი იმის მიხედვით, მათ შორის რომელია მთავარი და არამთავარი.

მაგ., ფრაზაში –*ძალიან ლამაზმა გოგონამ* – *ძალიან* იქნება I_NP, *ლამაზმა* იქნება S_NP, ხოლო *გოგონამ* კი – N. ხოლო თუ ეს სიტყვა შედის შემასმენლის ჯგუფში, შესაბამისად ავლნიშნავთ I_VP და S_VP-თი. მაგ.,

ჰქონდა ამის მოლოდინი = S_V I_VP S_VP, რაც ნიშნავს *ჰქონდა* = S_V, *ამის* = I_VP, *მოლოდინი* = S_VP;

S_V-ს საპირისპიროდ, I_V-თი ავლნიშნავთ მოდალურ ზმნას, მაგალითად,

უნდა წახვიდე სახლში

უნდა|უნდა|V Mod

სახლში|სახლი|N Sg Dat Shi

წახვიდე|წავა|V Imp Sg 2

უნდა წახვიდე სახლში = I_V S_V I_VP; რაც ნიშნავს შემდეგს:

უნდა = I_V, წახვიდე = S_V და სახლში = I_VP;

შემოგვაქვს გარკვეული წესები:

- თუ წინადადება მხოლოდ ქვემდებარე და შემასმენლისაგან შედგება, მათ შესაბამისად ავლნიშნავთ N –ით და S_V-თი; მაგ.,

მაგ., *აწყვეტინებს ლამარა* = S_V N;

- თუ ჩანართი ან მიმართვა ან შორისდებული და ა.შ. ერთწევრიანია, ის ავლნიშნება S_S-ით;

მაგ., *ქალბატონო!* = S_S;

- თუ ჩანართი ორწევრიანია და მისი წევრები სხვადასხვა ფრაზაში შედიან (noun phrase)–ში და (verb phrase)–ში, ორივე აღიწერება I_S-ით, წინააღმდეგ შემთხვევაში მთავარი იქნება S_S და არამთავარი კი – I_S.

მაგ., *ჩხირი კედელს!* = I_S I_S;

ლამაზო ქალბატონო! = I_S S_S;

განვიხილოთ ასეთი წინადადება:

ხატის წინ ბებიას სანთელი აუნთია.

პირველ რიგში ვეძებთ შემასმენელს – *აუნთია*, აღვნიშნავთ მას, როგორც შევთანხმდით, S_V-თი, შემდეგ ვეძებთ ქვემდებარეს – *ბეზიას*, აღვნიშნავთ მას N -ით, მოცემულ წინადადებაში noun phrase-ში მეტი არც ერთი წევრი არ შედის, ყველა დანარჩენი შედის verb phrase-ში.

აუნთია სანთელი, ამიტომ *სანთელი* იქნება I_VP, დარჩა *ხატის წინ*, სადაც მთავარია *წინ*, რასაც აღვნიშნავთ S_VP-თი და არამთავარია *ხატის*, რაც იქნება უკვე I_VP. საბოლოოდ მივიღეთ: I_VP S_VP N I_VP S_V;

განვიხილოთ პირიქითა პროცესი, ანუ, როგორ აღვადგინოთ მოცემული ფორმულებით თავდაპირველი წინადადების სტრუქტურული ხე.

ვიწყებთ პირველი წევრიდან – I_VP, მის შემდეგ მოდის S_VP, ვთვლით, რომ მათ უკვე შექმნეს კვანძი, ანუ სინტაქსური წყვილი *ხატის წინ*, რომლის თავშიც იქნება *წინ*, როგორც მთავარი წევრი. შემდეგია ქვემდებარე N – *ბეზიას*, შემდეგ მოდის I_VP, რადგან მას თავისი მეწყვილე S_VP არ ყავს, ის პირდაპირ დაუკავშირდება მომდევნო V-ს; ე.ი სტრუქტურული ხის აღდგენის დროს ვიწყებთ პირველი წევრიდან და მივდივართ მარჯვნივ, თითოეული არამთავარი წევრისთვის ვეძებთ მთავარ წევრს, თუ უშუალოდ მათ გვერდზე არ დგას, ვაგრძელებთ პროცესს მარჯვენა მიმართულებით, თუ მოხდა ისე, რომ გავედით ბოლოში და ვერ ვნახეთ მისი მთავარი წევრი, მაშინ ვბრუნდებით უკან და იქ ვუკავშირებთ S-ს ან უშუალოდ N-ს ან S_V-ს.

განვიხილოთ ასეთი მაგალითი:

სამარცხვინო საქციელია და მეტი არაფერი

აქ ქვემდებარე საერთოდ არ გვაქვს და გარდა ამისა, გვაქვს შედგენილი შემასმენელი: *საქციელია*, რომელსაც არვნიშნავთ VP-თი, დანარჩენი წევრები ყველა აღვნიშნავთ I_VP-თი, გარდა ბოლო სიტყვისა. მოცემული წინადადების სტრუქტურა ასე გამოისახება: I_VP VP I_VP I_VP S_VP.

შეჯამება

პროგრამა რომელიც დაიწერა ამ ნაშრომის ფარგლებში, სიტყვათა შორის კავშირების მოსაპოვებლად იყენებს 2.1 თავში აღწერილ ალგორითმს. რა თქმა უნდა ეს ალგორითმი არც ისე ეფექტურია, მაგრამ გამომდინარე იქედან რომ ქართულ ენას ძალიან რთული გრამატიკა აქვს ამ ეტაპზე შეიძლება ითქვას რომ პროგრამა თავის საქმეს აკეთებს და უმეტესობა კავშირებს პოულობს. შემდგომში იგეგმება პროექტის განვითარება და მაქსიმალური ენტროპიის მოდელის გამოყენებით უფრო ეფექტური სისტემის შექმნა.

2.1 თავში აღწერილ ალგორითმის ეფექტურობა ნაწილობრივ დამოკიდებულია სასწავლო მონაცემების ზომაზე 2.1.1 სურათზე გამოსახულ ფაილზე. რაც მეტია აღწერილი წყვილების რაოდენობა(სასწავლო მონაცემებში), სისტემა მით უფრო მეტ კავშირებს მოიპოვებს წინადადებებში.

გამოყენებული ლიტერატურა

- 1) C. D. Manning, H. Schuetze «Foundations of Statistical Natural Language Processing»
- 2) T. M. Mitchell «Machine Learning»
- 3) A. Ratnaparkhi «A Simple Introduction to Maximum Entropy Models for Natural Language Processing».
- 4) J. N. Darroch, D. Ratcliff «Generalized Iterative Scaling for Log-Linear Models»
- 5) Daniel Jurafsky & James H. Martin Speech And Language Processing

ლინკები ინტერნეტიდან

- 1) <http://www.cs.cmu.edu/afs/cs/user/aberger/www/html/tutorial/node3.html>
- 2) <https://class.coursera.org/nlp/lecture>
- 3) http://cseweb.ucsd.edu/~elkan/254/ari_talk.pdf
- 4) <http://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/>