

ირინა ლობჯანიძე

Irina Lobzhanidze

ქართული ენის მორფოსინტაქსური ანოტირებისა და
სასრული პოზიციის მორფოლოგიური ანალიზის პრინციპები

Morphosyntactic Annotation and Principles of Finite-State
Morphological Analysis of Georgian Language



ილიას სახელმწიფო უნივერსიტეტი

თბილისი - 2021 წელი



დაიბეჭდა ილიას სახელმწიფო უნივერსიტეტის სტრუქტურირებული სადოქტორო საგანმანათლებლო პროგრამის პროექტის (გრანტი # DP2016_23) ფარგლებში სსიპ შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის ფინანსური მხარდაჭერით. წინამდებარე პუბლიკაციაში გამოთქმული ნებისმიერი მოსაზრება ეკუთვნის ავტორს და შესაძლებელია არ ასახავდეს ფონდის შეხედულებებს.

Published under the project on the Development of Structured PhD educational programme (Grant № DP2016_23) of the Ilia State University with financial support of Shota Rustaveli National Science Foundation. All ideas expressed herewith are those of the author and not represent the opinion of the Foundation itself.

რედაქტორი: ნ. დობორჯგინიძე

Editor: N. Doborjginidze

კომპიუტერული უზრუნველყოფა: ნ. ერისთავი

Computer service: N. Eristavi

© 2021, ირინა ლობჯანიძე

სატიტულო გვერდის ილუსტრაცია შექმნილია სპეციალურად ამ გამოცემისთვის და გამოყენებისას დაცულია ყველა უფლება.

ISBN 978-9941-18-366-9

ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა

ქაქუცა ჩოლოყაშვილის 3/5, თბილისი, 0162, საქართველო

Ilia State University Press

3/5 Cholokashvili Ave, Tbilisi, 0162, Georgia

დაიბეჭდა შპს “პრინტჯეო”-ში

Printed by LTD “Printgeo”

სარჩევი

სარჩევი.....	i
ილუსტრაციების ჩაშონათვალი	
გრაფიკების სია.....	iii
გამოსახულებათა სია.....	iv
ცხრილების სია.....	vi
წინასიტყვაობა.....	1
შემოკლებათა სია.....	3
ნაწილი 1 დერივაციული და ფლექსიური მორფოლოგია.....	6
თავი 1 შესავალი.....	7
თავი 2 სიტყვანარმოება (დერივაცია).....	9
თავი 3 სიტყვაცვლილება (ფლექსია).....	15
თავი 4 დასკვნები.....	20
ნაწილი 2 კომპიუტერული მორფოლოგია.....	22
თავი 1 შესავალი.....	23
თავი 2 სასრული პოზიციის ტექნოლოგია: სასრული პოზიციის ავტომატები (სპა) და სასრული პოზიციის გარდამქმნელები (სპგ).....	25
თავი 3 რეგულარული გამოსახულებანი და სასრული პოზიციის ოპერატორები.....	29
თავი 4 ქართული ენის მორფოსინტაქსური ანალიზი და ქსეროქსის სასრული პოზიციის საშუალებანი (<i>xfst</i> და <i>lexc</i>).....	33
ქვეთავი 4.1 სახელური ფლექსია.....	34
ქვეთავი 4.2 ზმნური ფლექსია.....	38
ქვეთავი 4.3 ქსეროქსის სასრული პოზიციის საშუალებანი (<i>lexc</i> და <i>xfst</i>).....	49

თავი 5 დასკვნები	55
ნაწილი 3 ქართული ენის სასრული პოზიციის მორფოლოგიური ანალიზი	55
თავი 1 შესავალი	59
თავი 2 მორფოლოგიური ანალიზატორი და გენერატორი	61
ქვეთავი 2.1 ლემატიზაცია	61
ქვეთავი 2.2 დიაკრიტული ნიშნულები	65
ქვეთავი 2.3 ანკერები.....	69
ქვეთავი 2.4 სისტემის საერთო დიზაინი	74
თავი 3 ტოკენიზატორი.....	85
თავი 4 ტესტირება	89
თავი 5 დასკვნები	100
დანართი 1: მორფოლოგიური ტეგსეტი	101
დანართი 2: მეტა-ანოტირების ტეგსეტი	126
ბიბლიოგრაფია	131
საძიებელი.....	148

ილუსტრაციების ჩამონათვალი

გრაფიკების სია

გრაფიკი 1 სასრული პოზიციის მარტივი ენის მაგალითი	28
გრაფიკი 2 საზოგადო არსებითი სახელის სასრული პოზიციის მარტივი ქსელის მაგალითი	51
გრაფიკი 3 მორფოლოგიური ანალიზიტაროს გამარტივებული სტრუქტურა	84
გრაფიკი 4 ამონარიდი PDF ფაილიდან	94
გრაფიკი 5 მონიშვნის პროცედურული თანმიმდევრობა	96
გრაფიკი 6 მორფოლოგიური ანალიზის პროცედურული თანმიმდევრობა	96
გრაფიკი 7 თანამედროვე ქართული ენის ამოცნობის პროცენტული მაჩვენებელი	97
გრაფიკი 8 ძველი და საშუალო ქართული ენის ამოცნობის პროცენტული მაჩვენებელი	99

გამოსახულებათა სია

გამოსახულება 1 რეგულარული გამოსახულების მაგალითი	27
გამოსახულება 2 გრამატიკული და ზედაპირული დონეები, <i>lexc</i>	50
გამოსახულება 3 მრავალნიშნის სიბმლოების ბლოკის მაგალითი .	51
გამოსახულება 4 დიაკრიტული ნიშნულების ბლოკის მაგალითი.....	52
გამოსახულება 5 ანკერების ბლოკის მაგალითი.....	52
გამოსახულება 6 მინდორი (მინდვრის) შენაცვლების წესის მაგალითი	54
გამოსახულება 7 ანკერების მოხსნის შენაცვლების წესის მაგალითი....	54
გამოსახულება 8 ზმნური ლექსიკონის შემადგენელი ნაწილების შემოკლებული მაგალითი	69
გამოსახულება 9 ზმნიზედის წარმოების დერივაციული მოდელის მაგალითი	75
გამოსახულება 10 უცხო სიტყვების ამოცნობის გარდამქმნელის მაგალითი	79
გამოსახულება 11 თანხმოვანფუძიანი არსებითი სახელის თანდებულის ფორმის გარჩევის მაგალითი	80
გამოსახულება 12 ხმოვანფუძიანი არსებითი სახელის თანდებულის ფორმის გარჩევის მაგალითი.....	80
გამოსახულება 13 ვითარებითი ბრუნვის ნიშნით გამოწვეული გრამატიკული ომონიმიის მაგალითი	80
გამოსახულება 14 ზოგიერთი რიცხვითი სახელის გარჩევის მაგალითი	81
გამოსახულება 15 ნაცვალსახელის გარჩევის მაგალითი	81
გამოსახულება 16 ერთპირიანი, ორპირიანი და სამპირიანი ზოგიერთი ზმნის გარჩევის მაგალითი ომონიმიის შემთხვევებით	81

გამოსახულება 17 ზმნიზედის გარჩევის მაგალითი.....	82
გამოსახულება 18 კავშირის გარჩევის მაგალითი გრამატიკული ოზონიზიის ჩვენებით.....	82
გამოსახულება 19 ზოგიერთი ნანილაკის გარჩევის მაგალითი.....	82
გამოსახულება 20 შორისდებულის გარჩევის მაგალითი	82
გამოსახულება 21 აბრევიატურის გარჩევის მაგალითი.....	83
გამოსახულება 22 უცხო სიტყვის გარჩევის მაგალითი	83
გამოსახულება 23 პუნქტუაციის გარჩევის მაგალითი	83
გამოსახულება 24 ზმნური მოდულის სასრული პოზიციის გარდამქმნელის შექმნის თანმიმდევრობა.....	85
გამოსახულება 25 ქართული ანბანის ტოკენიზატორში ნარმოდგენის ტიპი.....	86
გამოსახულება 26 ტოკენიზაცია ინიციალებისა და შემოკლებანის ამოცნობის მიზნით	87
გამოსახულება 27 ტოკენიზაცია წილადების ამოცნობის მიზნით.....	87
გამოსახულება 28 ერთმარცვლიანი სახელური ფუძის ამომცნობის სტრუქტურა.....	88
გამოსახულება 29 შემოკლებული ამონარიდი ტეგების ლექსიკური გრამატიკიდან.....	92
გამოსახულება 30 გრამატიკული გადაფარვის მაგალითი მე-19 და 26-ე კლასებს შორის.....	98
გამოსახულება 31 გრამატიკული გადაფარვის მაგალითი 28-ე, 29-ე, 30-ე და 47-ე კლასებს შორის.....	98

ცხრილების სია

ცხრილი 1 მეტყველების ნაწილების ანოტირება	8
ცხრილი 2 ფუძის ტიპი, პროცესები და თვისებები.....	21
ცხრილი 3 სასრული პოზიციის ავტომატების ხუთი პარამეტრი	27
ცხრილი 4 რეგულარული გამოსახულების ზოგიერთი ოპერატორისა და ქმედების განმარტება	32
ცხრილი 5 არსებითი სახელების ტიპების ტეგირება	34
ცხრილი 6 რიცხვითი სახელების ტიპების ტეგირება	35
ცხრილი 7 ნაცვალსახელთა ტიპების ტეგირება.....	36
ცხრილი 8 მზნის ტიპების ტეგირება	39
ცხრილი 9 ასპექტის ტეგირება	40
ცხრილი 10 გარდამავლობის ტეგირება	40
ცხრილი 11 გვარის ტეგირება	41
ცხრილი 12 სუბიექტის და ობიექტების კორელაციის, პირისა და რიცხვის ტეგირება.....	42
ცხრილი 13 დროისა და კილოს ტეგირება	43
ცხრილი 14 სუბიექტის, პირდაპირი და ირიბი ობიექტების ბრუნვების ტეგირება	44
ცხრილი 15 სახელის ქვედა და ზედა დონის შემოკლებული მაგალითი შუალედური დონის ჩვენებით	73

წინასიტყვაობა

წარმოდგენილი კვლევის ძირითადი ამოცანაა ქართული ენის მორფოსინტაქსური ანოტირების პრინციპებისა და ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის სტრუქტურის ზოგადად აღწერა. ბუნებრივი ენის დამუშავების აღნიშნული საშუალება შეიქმნა ქართული ენის კორპუსის ანოტირების მიზნით შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის ფინანსური მხარდაჭერით (გრანტები № DP2016_23 და № AR/320/4-105/11) და მნიშვნელოვანია როგორც ძველი, ასევე თანამედროვე ქართული ენის კომპიუტერული დამუშავების თვალსაზრისით.

მონოგრაფიაში აღწერილია ქართული ენის მდიდარი მორფოლოგიური სტრუქტურის ავტომატური ანოტირების შესაძლებლობა სასრული პოზიციის ტექნოლოგიის გამოყენებით. მონოგრაფია სამი ნაწილისგან შედგება.

პირველი ნაწილის მთავარ დანიშნულებას ქართული ენის სიტყვა-წარმოებითი და სიტყვაგვლილებითი მორფოლოგიური პროცესების აღწერა წარმოადგენს ბუნებრივი ენის დამუშავებისთვის მნიშვნელოვანი ასპექტების გათვალისწინებით.

მეორე ნაწილი მოიცავს სასრული პოზიციის მორფოლოგიური ანალიზის მიდგომებისა და ქართული ენის მორფოსინტაქსური ანოტირების პრინციპების აღწერას.

მესამე ნაწილი წარმოადგენს ინფორმაციას ქართული ტექსტის ტოკენიზატორის, ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის სტრუქტურის შესახებ შესაბამისი შემონმუშებით, გამავალი სიგნალის ანალიზითა და ქართული ენის კორპუსის (GLC) შედგენის ზოგადი პრინციპების აღწერით.

დანართები ქართული ენის სასრული პოზიციის მორფოსინტაქსური ტეგებისა და სტრუქტურული დონის ანოტირების სამაგალითო ჩამონათვალს ეთმობა.

საკვანძო სიტყვები: სასრული პოზიციის მორფოლოგიური ანალიზი, სასრული პოზიციის გარდამქმნელი, ქართული ენა

მადლობის გადახდა

აღნიშნული კვლევა წარმოადგენს სხვადასხვა დროს ილიას სახელმწიფო უნივერსიტეტისა და შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის დაფინანსებით განხორციელებული პროექტების (ციფრული ჰუმანიტარია (ციფრული ეპიგრაფიკა, კომპიუტერული ენათმეცნიერება, ციფრული პროსოპოგრაფია), პროექტი № DP2016_23, თანამედროვე ქართული ენის საანალიზო და საანოტაციო პროგრამების შექმნა, პროექტი № AR/320/4-105/11) შედეგებს.

აღნიშნული პროექტების განხორციელებაში სხვადასხვა დროს მონაწილეობას იღებდა ილიას სახელმწიფო უნივერსიტეტის ლინგვისტურ კვლევათა ინსტიტუტის მრავალი თანამშრომელი. განსაკუთრებული მადლობა მინდა მოვახსენო ნინო დობორჯგინიძეს მისი მუდმივი მხარდაჭერისათვის, პროექტზე სხვადასხვა დროს მომუშავე პროგრამისტებს: ირაკლი გუნიას, თამუნა კაპანაძესა და გიორგი მირიანაშვილს, კოორდინატორებს: სვეტლანა ბერიკაშვილს, ნატო ბილანიშვილს, ცირა ხახვიაშვილსა და გიორგი თადუშაძეს და, კვლევით ასისტენტებს: მ. აბესაძეს, ს. აბრალავას, თ. გავაშელიშვილს, თ. გეთიაშვილს, ქ. გილაშვილს, ნ. გოგიჩაიშვილს, მ. დადუნაშვილს, ნ. დათაშვილს, თ. დამენიას, ნ. დონაძეს, ლ. ვაშაყმაძეს, ლ. თეთრაშვილს, შ. თუმანიშვილს, მ. იშხანოვას, თ. კალხიტაშვილს, მ. კვასხვაძეს, მ. კვინიკაძეს, ე. კვირკველიას, ა. კიკილაშვილს, ლ. კინმარიშვილს, ლ. კუპრეიშვილს, ნ. ლათიბაშვილს, ივლ. ლობჯანიძეს, ს. მარგველაშვილს, ქ. მარიდაშვილს, ზ. მაჭავარიანს, ე. მდინარაძეს, შ. ნოზაძეს, მ. რუხაძეს, ნ. სადღობელაშვილს, ლ. სერგიას, ა. სტურუას, ა. ტყეშელაშვილს, თ. ქაროსანიძეს, თ. ქიტოშვილს, ნ. ჭიჭიკოშვილსა და თ. ხუბაშვილს.

შემოკლებათა სია

FSA სასრული პოზიციის ავტომატი, სპა
FST სასრული პოზიციის გარდამქმნელი, სპგ
GLC ქართული ენის კორპუსი, ქეგ
NLP ბუნებრივი ენის დამუშავება, ბედ
LFG ლექსიკურ-ფუნქციონალური გრამატიკა, ლფგ
PARC პალო ალტოს კვლევითი ცენტრი, პარკ
KIMMO ორდონიანი მორფოლოგია, კიმმო
TEI ტექსტის კოდირების საინიციატივო ჩარჩო, ტეი
XFST ქსეროქსის სასრული პოზიციის საშუალებანი, ქსპს
XRCE ქსეროქსის კვლევითი ცენტრი ევროპაში, ქკცე

გამოყენებული გლოსების სია¹

1, 2, 3 1, 2, 3 პირი
ADV ვითარებითი
AOR წყვეტილი
AUX დამხმარე ზმნა
CAUS კაუზატივი
CMPR შედარებითი ხარისხი
COND ხოლმეობითი
DAT მიცემითი
DObj პირდაპირი ობიექტი
EM სავრცობი

¹ გლოსირება, ძირითადად, მიჰყვება EUROTYP-ის სახელმძღვანელოს (ბუკერი და სხვ. 1993).

EMPH ემფატიკური ხმოვანი
ERG მოთხრობითი
FUT მყოფადი
GEN ნათესაობითი
IMP ბრძანებითი
IMPF უწყვეტელი
IND თხრობითი
INST მოქმედებითი
IObj ირიბი ობიექტი
IPFV უსრული ასპექტი
NOM სახელობითი
Obj ობიექტი
PASS ვნებითი
PF პირველი თურმეობითი
PFV სრული ასპექტი
PL მრავლობითი
PLUP მეორე თურმეობითი
POSS კუთვნილებითი
Posp თანდებული
pref პრეფიქსი
PRS აწმყო
PRV კორელაციის (ქცევის) ნიშანი
Ptl ნაწილაკი

prev ზმნისწინი

QUOT თემის ნიშანი

RFL უკუქცევითი

Sbj სუბიექტი

SG მხოლოდითი

SUBJ კავშირებითი

suff სუფიქსი

SUP უფროობითი

TS თემის ნიშანი

Vers ქცევა

VOC წოდებითი

ნაწილი 1

დერივაციული და ფლექსიური მორფოლოგია

მსოფლიო ენები ერთმანეთისაგან მორფოლოგიური სისტემების თვალსაზრისით განსხვავდება, თუმცა აღნიშნული სხვაობა უსასრულო არ არის და როგორც ერთეულების, ასევე პროცესების განზოგადების საშუალებას იძლევა. ბუნებრივი ენის დამუშავების (ბედ) ყველაზე მნიშვნელოვან საფეხურს, ერთი მხრივ, მორფოლოგიური შაბლონის შემუშავება, ხოლო, მეორე მხრივ, უნივერსალური თვისებების გამოვლენა წარმოადგენს. კონკრეტული ენის მორფოლოგიური შაბლონის გამოვლენა იძლევა ენის სტრუქტურირებისა და შემდგომი დამუშავების საშუალებას, ხოლო უნივერსალური თვისებების აღწერა – არსებულ სისტემებში დადასტურებული შედეგების გადააზრებისა და გაზიარების ან/და უგულებელყოფის შესაძლებლობას.

მორფოლოგიური შაბლონის შემუშავების ერთ-ერთ საფეხურს სხვადასხვა ენაში არსებული საერთო მორფოლოგიური მოვლენების აღწერა და კონკრეტული ენისათვის დამახასიათებელი მორფოლოგიური თვისებების წარმოჩენა წარმოადგენს. აღნიშნული თავი ემსახურება სიტყვანარმოებისა (დერივაციისა) და სიტყვაცვლილების (ფლექსიის) განხილვას ენობრივი შაბლონის შემუშავების თვალსაზრისით და იმ მოვლენების აღწერას, რომლებიც გავლენას ახდენს ქართული ენის დამუშავებაზე კომპიუტერული თვალსაზრისით.

პირველი ნაწილი იყოფა ოთხ ძირითად თავად: 1. შესავალში წარმოდგენილია სიტყვის მორფოლოგიური სტრუქტურის ის ასპექტები, რომლებიც ზეგავლენას ახდენს ენის გენერაციაზე; 2. მეორე თავი მოიცავს ინფორმაციას სიტყვის სიტყვანარმოებითი სტრუქტურის შესახებ; 3. მესამე თავი ეხება სიტყვის სიტყვაცვლილებით მოდელებს, და; 4. დასკვნებით კი ხდება ზემოხსენებული თავების შეჯამება.

თავი 1 შესავალი

მორფოლოგია ენათმეცნიერების ნაწილს წარმოადგენს. მისი მთავარი ამოცანაა იმ კავშირის კვლევა, რომელიც არსებობს სიტყვის ფორმასა და მნიშვნელობას შორის. მორფოლოგიური კვლევა, ერთი მხრივ, შეიძლება იყოს მორფოლოგიური და გულისხმობდეს სიტყვის ფორმის კავშირს ფონოლოგიასთან, ხოლო, მეორე მხრივ, მორფოსემანტიკური და გულისხმობდეს სიტყვის მნიშვნელობის კავშირს სემანტიკასთან. ორივე შემთხვევაში სიტყვის მორფოლოგიური სტრუქტურა ზეგავლენას ახდენს სიტყვის ფუნქციონირებაზე, რაც საბოლოო ჯამში ენის სინტაქსშიც ვლინდება. შესაბამისად, ენის მორფოლოგიური სტრუქტურის აღწერა, ერთი მხრივ, ფონოლოგიური და სინტაქსური თეორიების გადაკვეთის წერტილს წარმოადგენს, ხოლო, მეორე მხრივ, მოიაზრებს განსაზღვრულ ინფორმაციას კონკრეტული ენის ლექსიკური მარაგის (ლექსიკონის) შესახებ (ბაუერი 2003, კოატესი 1999 და სხვ.).

გრამატიკული სტრუქტურის დონეზე აღნიშნული გადაკვეთის წერტილი ლექსიკური მარაგის მეტყველების ნაწილებად წარმოდგენაში ვლინდება. მეტყველების ნაწილებად ენის სტრუქტურის დაყოფა კონკრეტული სიტყვების გრამატიკულად მსგავს თვისებებსა და წინადადების ფარგლებში მსგავს სინტაქსურ ქმედებას ეფუძნება. ბუნებრივი ენის დამუშავებისას ენის მეტყველების ნაწილების ტეგირებას საფუძვლად უდევს მეტყველების ნაწილებისთვის სპეციალური სიმბოლოების (ტეგების) მინიჭება:

მეტყველების ნაწილები	სასრული პოზიციის (სპ) ტეგები	ატრიბუტების რაოდენობა
არსებითი სახელი	+Noun	5
გმნა	+Verb	11
ზედსართავი სახელი	+Adj	5
რიცხვითი სახელი	+Num	4
ნაცვალსახელი	+Pron	5
კავშირი	+Conj	1
ნაწილაკი	+Part	1

ზმნიზედა	+Adv	2
თანდებული	+Post	1
შორისდებული	+Itj	-
შემოკლებანი	+Abr	4
პუნქტუაციის ნიშნები	+F	-

ცხრილი 1 მეტყველების ნაწილების ანოტირება

აღნიშნული მონიშვნის შედეგად ხდება ენაში არსებული სიტყვიერი მარაგის სხვადასხვა კატეგორიის კოდირება შემდგომი ქვეერთეულებისა და კონკრეტული კატეგორიებისათვის დამახასიათებელი მორფოსინტაქსური თვისებების განსაზღვრის მიზნით, ხოლო მეტყველების ნაწილების შესაბამისი ტეგები ქვეერთეულებისა და კონკრეტული კატეგორიების ტეგებთან ერთად ქმნის ენისათვის დამახასიათებელ ტეგსეტს².

ქართული ენა განეკუთვნება მორფოლოგიურად მდიდარ ენებს, რაც ნიშნავს იმას, რომ სიტყვის დონეზე სხვადასხვა დონის მორფოსინტაქსური თვისებების გამოვლენა ხდება და, რომ სიტყვათა თანმიმდევრობა წინადადების ფარგლებში ფიქსირებული არ არის. შესაბამისად, ქართული ენის თავისებურების გათვალისწინებით ბევრად უფრო მნიშვნელოვანია მისი სიტყვის მორფოსინტაქსური სტრუქტურის განხილვა. მორფოსინტაქსური სტრუქტურა ორ ნაწილად იყოფა: მორფოლოგიად და სინტაქსად, რაც ენის მოდელირების საშუალებას იძლევა სიტყვების სტრუქტურისა და ფორმირების დონეზე. აღნიშნული კი ვლინდება, არა მხოლოდ სიტყვანარმოებითი და სიტყვაცვლილებითი მორფოლოგიის ფარგლებში, არამედ ზეგავლენას ახდენს წინადადების ფორმირებაზეც. ქართული ენის მორფოსინტაქსური სტრუქტურის განხილვა უკავშირდება ზმნური კონსტრუქციის ორმაგ როგორც სიტყვანარმოებით, ასევე წინადადებაწარმოებით დანიშნულებას.

ძირითადი ამოცანიდან გამომდინარე, მორფოსინტაქსური კონსტრუქციის განხილვა გულისხმობს როგორც სიტყვის მანარმოებელი მორფემების, ასევე მისი მაფორმირებელი ერთეულების განხილვას. ამიტომაც, მნიშვნელოვანია მორფოლოგიური ანალიზის ორი ძირითადი

² ტეგსეტის სრული სიისთვის, იხ. დანართი 1.

ასპექტის განხილვა, როგორცაა სიტყვანარმოება (დერივაცია) და სიტყვაცვლილება (ფლექსია).

თავი 2

სიტყვანარმოება (დერივაცია)

მორფოლოგიური პროცესები, რომლებიც ზეგავლენას ახდენს სიტყვის წარმოებაზე სხვადასხვა ენაში სხვადასხვანაირად ვლინდება. ზოგიერთ ენაში (ინგლისური, დანიური და სხვ.) კავშირი სიტყვასა და მნიშვნელობას შორის თანხვედრას აჩვენებს და დამატებით ერთეულებს კონკრეტული სიტყვის მნიშვნელობის შესაცვლელად არ ითხოვს, ენათა ნაწილში (თურქული, აზერბაიჯანული და სხვ.) კი აღნიშნული კავშირი მორფემების შერწყმის შედეგს წარმოადგენს და მორფემების შეჯამებით შესაძლებელი ხდება კონკრეტული სიტყვის მნიშვნელობის ამოცნობა. აღნიშნული თვალსაზრისით ენები იყოფა ფლექსიურ ენებად, რომლებიც სხვადასხვა თვისების საჩვენებლად ერთ ფლექსიურ მორფემას იყენებენ, აგლუტინაციურ ენებად, რომლებიც კონკრეტული გრამატიკული კატეგორიის საჩვენებლად კონკრეტულ მორფემას საჭიროებენ და იზოლირებულ ენებად, რომლებშიც მორფოლოგია, როგორც ასეთი, არ არსებობს (იხ. პარისი 2006, კომრი 1989 და სხვ.).

ქართული ენა განეკუთვნება კომბინირებული ტიპის ენებს და სრულყოფილად არც ერთ ზემოხსენებულ ტიპს არ შეესაბამება, რადგანაც მასში დასტურდება როგორც აგლუტინაციური (1), ასევე ფლექსიური (2) ტიპის კონსტრუქციები. კომბინირებული ტიპი შეიძლება განიხილებოდეს როგორც კვაზი-პოლისინთეტური, რომლითაც დიდი რაოდენობის სიტყვის მანარმოებელი სტრუქტურების მქონე ენები აღინიშნება (იხ. ვიერი 2011). ფლექსიური და აგლუტინაციური ტიპის ენებისთვის დამახასიათებელი სტრუქტურების (იაკობინი 2006) ქართულ ენასთან შედარებისას ვლინდება, რომ ქართულ ენას არა მხოლოდ აგლუტინაციური, არამედ ფლექსიური თვისებებიც ახასიათებს. ქართულ ენაში აგლუტინაციური თვისებები ვლინდება აფიქსაციის ტიპში, მორფემულ შესაბამისობაში, მონოსილაბიზმისადმი მიდრეკილებაში, სახელებში რიცხვისა და ბრუნვის

ნიშნების მარკირებაში, სქესის გრამატიკული კატეგორიის უქონლობაში, ზედსართავი სახელების სინთეზური ხარისხის ფორმებში და ა.შ., ხოლო ფლექსიური - აფიქსაციის ტიპში, მეტყველების ნაწილების მკაფიო გამიჯვნაში, ზმნის პარადიგმულ თავისებურებაში და ა.შ.

(1) სახლ-ებ-მა

სახლ-PL-ERG

(2) სახლ-ღ=ში

სახლ-SG=in.DAT

სიტყვის სტრუქტურა შედგება თავისუფალი და ბმული მორფემებისაგან, სხვაგვარად რომ ვთქვათ, ძირისა და მასზე დართული აფიქსებისაგან. ზოგი აფიქსი სახელურ ძირს ერთვის და მონაწილეობას იღებს სახელური პარადიგმის შექმნაში, ზოგი კი - ზმნურს და, შესაბამისად, ზმნური პარადიგმის შექმნაში მონაწილეობს.

აფიქსაცია (3) გამოიყენება გრამატიკული ფუნქციისა და ახალი გრამატიკული კლასების წარმოების მიზნით, ფუძისეული ელემენტების მონაცვლეობა (4) - გრამატიკული ფუნქციის უზრუნველყოფის მიზნით, ხოლო კომპოზიტურობა (5) - კომპოზიტური ლექსემების შესაქმნელად.

(3) ქალ-ღ-მა, ქალ-ურ-ღ-ი და სხვ.

ქალ-SG-ERG

ქალურ-SG-NOM

(4) ხე-ღ -> ხ-ღ-ის და სხვ.

ხე-SG.NOM

ხე-SG-GEN

(5) ენათმეცნიერება-ღ, ენა-თ~ მეცნიერება-ღ და სხვ.

ენათმეცნიერება-SG.NOM

ენა-PL~მეცნიერება-SG.NOM

არსებობს აფიქსების შემდეგი ტიპები: 1) პრეფიქსები, რომლებიც ძირს ან სხვა პრეფიქსებს წინ ერთვის (6-7); 2) სუფიქსები, რომლებიც ძირს ან სხვა სუფიქსებს (7) მოსდევს; 3) ინფიქსები, რომლებიც ძირს ორ

ნაწილად ხლეს (8) და; 4) ცირკუმფიქსები, რომლებიც ორი აფიქსისგან შედგება და რომლებიც სიტყვას წინ და ბოლოს ერთვის (9).

- (6) ფორმანარმოებითი (ფლექსიური) პრეფიქსები: *წერ-ს* -> *უ-წერ-ს* -> *მი-ს-წერ-ს* და სხვ.;

წერ-3SGSbj:PRS.IND -> *Vers.3IObj-წერ-3SGSbj:PRS.IND* -> *prev. PFV-3SGIObj-წერ-3SGSbj:FUT.IND*

სიტყვანარმოებითი (დერივაციული) პრეფიქსები: *წერ-ს* -> *ნა-წერ-ღ-ი* და სხვ.;

წერ-3SGSbj:PRS.IND -> *ნაწერი-SG-NOM*

- (7) ფორმანარმოებითი სუფიქსები: *ა-ცხად-ებ-ს* -> *გამო-ა-ცხად-ა* და სხვ.

Vers-ცხად-TS-3SGSbj:PRS.IND -> *prev.PFV-Vers-ცხად-3SGSbj:AOR.IND*

სიტყვანარმოებითი სუფიქსები: *კაც-ღ-ი* > *კაც-ურ-ღ-ი* და სხვ.

კაც-SG-NOM -> *კაცურ-SG-NOM*

- (8) სიტყვანარმოებითი ინფიქსები: *ხნავ-ს* -> *ხ<ვ>ნა-ღ* და სხვ.

ხნავ-3SGSbj:PRS.IND -> *ხვნა-SG.NOM*

- (9) ფორმანარმოებითი ცირკუმფიქსები: *დიდ-ღ-ი* -> *უ>დიდ<ეს-ღ-ი*
დიდ-SG-NOM -> *CMPR>დიდ<CMPR-SG-NOM*

სიტყვანარმოებითი ცირკუმფიქსები: *ქალაქ-ღ-ი* > *მო>ქალაქ<ე-ღ*
ქალაქ-SG-NOM -> *მოქალაქე-SG.NOM*

აფიქსების ზემოხსენებული ტიპები საკმაოდ კარგად აღწერილია სამეცნიერო ლიტერატურაში (არონსონი 1989, ფოჩხუა 1974, ღლონტი 1964 და სხვ.) და გამოიყენება სხვადასხვა ფუნქციის საჩვენებლად, ე.ი. სხვადასხვა ფორმისა ან სხვადასხვა ლექსემის საწარმოებლად. და თუ ფორმანარმოებითი აფიქსები მორფემების დახურულ კლასს წარმოადგენს, ხოლო მათი დამუშავება სასრული პოზიციის კალკულუსში საკმაოდ მარტივია, სიტყვანარმოებითი აფიქსები როგორც რაოდენობის, ასევე წარმოებითი თავისებურების თვალსაზრისით არ იძლევა მათი

დანვრილებით აღწერის შესაძლებლობას და იზღუდება მათი ფორმითა და შინაარსით (10).

(10) მანდილ-ოსან-ი, გეც-იერ-ი და სხვ.

მანდილოსან-POSS-SG.NOM

გეც-იერ-POSS-SG.NOM

აღნიშნული სხვაობის ამოსავალი დებულებანი კარგად აღწერილია სამეცნიერო ლიტერატურაში (ბოოჭი 2006, არონოფი 1994, ანდერსონი 1992 და სხვ.) და უკავშირდება შემდეგს: 1) სიტყვანარმოება სავალდებულო არ არის, მაშინ როცა სიტყვაცვლილება სავალდებულოა; 2) სიტყვანარმოება ამოსავალი ფორმის მეტყველების ნაწილისადმი მიკუთვნებას ცვლის, მაშინ როცა სიტყვაცვლილება - არა; 3) სიტყვაცვლილება ყოველთვის უკავშირდება გრამატიკული პარადიგმის არსებობას.

ქართული ენის სიტყვათა კლასებად დაყოფა სიტყვის მნიშვნელობასა და სინტაქსურ ფუნქციას ემყარება და ათი დამოუკიდებელი ელემენტისგან შედგება, როგორიცაა არსებითი, ზედსართავი, ნაცვალსახელი, რიცხვითი სახელი, ზმნა, ზმნიზედა, კავშირი, ნაწილაკი, შორისდებული და თანდებული. სიტყვანარმოებითი მოდელი ქართული ენის ლექსიკონის გამდიდრებაზეა ორიენტირებული, სიტყვაცვლილებითი კი კონკრეტული აფიქსების გამოყენებით კონკრეტული პარადიგმების წარმოებასა და კონკრეტული ფონოლოგიური ცვლილებების ამოქმედებას განაპირობებს.

შესაბამისი აფიქსაციის გზით სიტყვანარმოებითი მოდელი სიტყვათა ერთი კლასიდან მეორე კლასის სიტყვების გენერირებას იწვევს, მაგ. მიმღეობას - ზმნისგან (11) ანდა ზედსართავ სახელს - არსებითი სახელისაგან (12) და სხვ.:

(11) ა-კეთ-ებ-ს, მ-კეთ-ებ-ელ-ი, გა-კეთ-ებ-ულ-ი და სხვ.

(12) შაქარ- ღ-ი, შაქრ-იან- ღ-ი და სხვ.

არსებული მეტყველების ნაწილები, თავის მხრივ, ღია და დახურულ კლასებად იყოფა. ახალი სიტყვებით ღია კლასის შევსება მუდმივ და პროდუქტიულ პროცესს წარმოადგენს, დახურული კლასისა კი უფრო იშვიათად ხდება. არსებითი სახელები, ზედსართავი სახელები და ზმნები

ღია კლასს განეკუთვნება და, შესაბამისად, აღნიშნული მეტყველების ნაწილები დერივაციული პროცესებისა და უცხო ენებიდან ნასესხობის გზით მიღებული სიტყვა-ფორმებით ადვილად ივსება, ხოლო ნაცვალსახელი, კავშირი, შორისდებული, კლიტიკები (თანდებული, დამხმარე ზმნა, ემფატიკური ხმოვანი, ნაწილაკი) დახურულ კლასს განეკუთვნება და, შესაბამისად, უფრო იშვიათად ივსება დერივაციული პროცესებისა ან ნასესხობის გზით. სხვაობა დახურულ და ღია კლასებს შორის უკავშირდება სხვაობას ლექსიკურ და ფუნქციონალურ კატეგორიებს შორის, რაც გულისხმობს იმას, რომ ღია კლასში მოხვედრილი სიტყვების უმეტესობას სემანტიკური მნიშვნელობა მეტი ოდენობით ახასიათებთ, ხოლო დახურულ კლასში მოხვედრილ სიტყვებს - ფუნქციური მნიშვნელობა.

ენაში მიმდინარე სიტყვანარმოებითი მოდელის განხილვისას ყურადღების გამახვილება შეიძლება შემდეგ ერთეულებზე:

- უცხო სიტყვების სესხებაზე, რომელიც ვლინდება უცხოენოვანი სიტყვებისა და შემოკლებების ენაში ადაპტირებასა და ინტეგრირებაში (ამირიძე და სხვ. 2019, ამირიძე 2018, დანელია 1975 და ა.შ.). პროცესის პროდუქტიულობა დამოკიდებულია იმაზე, თუ რამდენად ახერხებს ფლექსიური პარადიგმა უცხო ენიდან შემოსული ძირის ადაპტირებას თავისი მიზნებიდან გამომდინარე (13-14) და, შესაბამისად, ენაში განასხვავებენ ადაპტირებულ, ნახევრად ადაპტირებულ და არაადაპტირებულ ლექსემებს (გამყრელიძე და სხვ. 2008).

(13) ვორქშოპ-ი და სხვ.

ვორქშოპ-SG-NOM

(14) ბ-ი-შეარ-ებ-ს და სხვ.

1SGIObj-Vers.RFL-შეარ-TS-3SGSbj:PRS.IND

- აფიქსურ დერივაციაზე, რომელიც ვლინდება განსაზღვრული ფუნქციონალური აფიქსების სიტყვის ძირზე დართვაში ახალი სიტყვის შექმნის მიზნით. აღნიშნული პროცესი განსაკუთრებული პროდუქტიულობით ხასიათდება სახელებში ნათესაობით ბრუნვის ფორმაზე აფიქსების დართვის შემთხვევებში (15), ხოლო ზმნებში

საწყისისა და მიმღეობის წარმოების დროს (16).

(15) კაც-ის-ნაირ-ღ-ი, მზ-ის-მიერ-ღ-ი და სხვ.

კაცისნაირ-SG-NOM

მზისმიერ-SG-NOM

(16) წერ-ს -> და-წერ-ილ-ღ-ი, და-მ-წერ-ღ-ი და სხვ.

IPFV-წერ-3SGSbj:PRS.IND

დაწერილ-SG-NOM

დამწერ-SG-NOM

- კომპოზიტურობაზე, რომელიც ენაში ვლინდება ფუძეთა შეერთების გზით და მიმართავს სხვადასხვა სტრატეგიას, მათ შორის:

- o ფუძის გაორკეცებას (რედუბლიკაციას), რომელიც ონომატოპოეტური ლექსემების შექმნის წინაპირობას წარმოადგენს (17)

(17) კივ-ი-ს, კივ~კივ-ღ-ი, კივ~კივ-ი-ს და სხვ.

კივ-PRS.IND-3SGSbj

კივკივ-SG-NOM

კივკივ-PRS.IND-3SGSbj

- o ფუძეების შეერთებას დეფისის გამოყენებით (18)

(18) დღე-ღამე-ღ, მიწერ-მოწერა-ღ და სხვ.

დღე-ღამე-SG.NOM

მიწერ-მოწერა-SG.NOM

- o ფუძეების შეერთებას დეფისის გამოყენების გარეშე (19)

(19) დედა~ქალაქ-ღ-ი, მამა~ჩემ-ღი და სხვ.

დედაქალაქ-SG-NOM

მამაჩემ-SG-NOM

სიტყვანარმოებითი მორფოლოგიის მთავარი დანიშნულებაა ენის ლექსიკური მარაგის გამდიდრება. სიტყვაცვლილებით მოდელთან შედარებით როგორც ძველი, ასევე თანამედროვე ქართული ენის

სიტყვანარმოებითი კომპიუტერული მოდელირება ნაკლებად შესაძლებელია მისი ნაკლებად სტრუქტურირებული ხასიათის გამო. სიტყვაცვლილებითი მოდელის პროდუქტიულობას კი განაპირობებს ენის კომბინირებული ტიპი და სიტყვის ფარგლებში კავშირი კონკრეტულ ერთეულსა და მორფოსინტაქსურ თვისებას შორის.

თავი 3

სიტყვაცვლილება (ფლექსია)

სიტყვაცვლილებითი მორფოლოგია ყოველთვის აფიქსაციის და-ხურული კლასებითა და პარადიგმის მანარმოებელი წესებით იზღუდება. პარადიგმის წარმოების წესები, ერთი მხრივ, სიტყვანარმოებას (სპროატი 1992) უკავშირდება, რომელიც ზღუდავს აფიქსების დართვას ფონოლოგიური თვალსაზრისით და შემდგომში განაპირობებს მათ ქმედებას სიტყვის სინტაქსური ქმედებისა და სემანტიკური მნიშვნელობის განსაზღვრის მიზნით, ხოლო, მეორე მხრივ, განაპირობებს პარადიგმის ფუნქციონირების მორფოლოგიურ წესებს (სტამპი 2001-2002, ზვიკი 1977-1990 და ა.შ.), რომელიც გულისხმობს კავშირების დამყარებას მორფოლოგიასა და სინტაქსს შორის. უკანასკნელი განსაკუთრებით მნიშვნელოვანია ქართული ენის თვალსაზრისით, რადგანაც ქართული ენისათვის დამახასიათებელი პარადიგმების აღწერისას მხოლოდ მორფოლოგიური ანალიზი საკმარისი არ არის. ქართული ენის მოდელირებას პირდაპირ კავშირშია მორფოსინტაქსური ანალიზის შესაძლებლობასთან, რომელიც განაპირობებს არა მხოლოდ სიტყვის მორფოლოგიურ სტრუქტურას, არამედ მის სინტაქსურ ფუნქციასაც.

ქართული ენისათვის დამახასიათებელი სიტყვაცვლილებითი აფიქსაციის დახასიათებისას მნიშვნელოვანია შემდეგი ტიპების განხილვა:

1. **‘ნულოვანი’ მორფი**, როცა აფიქსის არსებობა იგულისხმება, მაგრამ აფიქსისთვის განკუთვნილი სლოტი ცარიელი რჩება. მხოლოდითი რიცხვის ნულოვანი მორფი გამოიყენება რიცხვის ფორმების პარალელიზმის უზრუნველყოფის მიზნით (20)

(20) კარ-შ-ი, კარ-ებ-ი და სხვ.

კარ-SG-NOM

კარ-PL-NOM

2. **‘კომბინირებული’ მორფი**, როცა აფიქსი რამდენიმე გრამატიკულ თვისებას ითავსებს და არ არსებობს ერთმნიშვნელოვანი თანხვედრა აფიქსის გრამატიკულ თვისებებსა და მის მნიშვნელობას შორის (21)

(21) გვ-ხატ-ავ-ს და სხვ.

1PLOBj-ხატ-TS-3SGSbj:PRS.IND

3. **‘ცარიელი’ მორფი**, როცა აფიქსი გრამატიკულ თვისებას არ გამოხატავს, მაგრამ არსებობს. მსგავსი მორფების არსებობა უკავშირდება იმას, რომ ძველ ქართულში არსებობდა ფუნქციონალური დატვირთვის მქონე აფიქსი, რომელსაც შემდგომში ცარიელი მორფი ჩაენაცვლა (22)

(22) ს-ძინ-ავ-ს და სხვ.

pref-ძინ-TS-3SGSbj:PRS.IND (თანამედროვე ქართულში)

3SGOby-ძინ-TS-3SGSbj:PRS.IND (ძველ ქართულში)

აღნიშნული მორფები დიდ ზეგავლენას ახდენს მორფოსინტაქსურ პარადიგმაზე და ზრდის გენერირებულ ფორმათა სიმრავლეს. საშუალოდ ერთი სახელური ძირიდან სიტყვანარმოებითი აფიქსების გარეშე მხოლოდ სიტყვაცვლილებითი აფიქსების გამოყენების გზით ხდება 3750 ერთეულის გენერირება, ხოლო ზმნური ძირიდან – 33,260 ერთეულისა. სიტყვანარმოებითი აფიქსების გამოყენება კი ზრდის ზემოხსენებული ფორმების რაოდენობას და ერთი ზმნური ძირიდან საშუალოდ მილიონ ნახევრამდე ერთეულს ქმნის. ქართული ენის აღნიშნული თვისება ზრდის კომპიუტერული ლექსიკონის ზომას და იწვევს სხვადასხვა ტიპის სირთულეს ელექტრონული რესურსების შექმნის დროს.

სიტყვის სტრუქტურის შექმნაზე ზეგავლენას ახდენს ზოგიერთი ფონეტიკური პროცესიც, როგორიცაა:

- აბლაუტი (აპოფონია), რომელიც გულისხმობს ფუძისეული ხმოვნის ცვლას და ვლინდება ზმნური პარადიგმის აორისტის წარმოებისას

-*ე*- > -*ი*- ხმოვნების მონაცვლეობის გზით. სხვა პროცესებთან შედარებით აბლაუტის შემთხვევები ქართულ ენაში იშვიათ მოვლენას წარმოადგენს (23)

(23) *დრეკ*-ს, *დრიკ*-ა და სხვ.

დრეკ-3SGSbj:PRS.IND

დრეკ-3SGSbj:AOR.IND

- კუმშვა (სინკოპე), რომელიც გულისხმობს ფუძისეული -*ა*-, -*ე*- და -*ო*- ხმოვნების დაკარგვას თანხმოვანფუძიან სახელებთან -*ლ*-, -*რ*-, -*შ*-, -*ნ*- სონანტების წინ მხოლოდითი რიცხვის ნათესაობით, მოქმედებით და ვითარებით ბრუნვებში და „ეზიანი“ მრავლობითი რიცხვის ყველა ბრუნვაში (24-26). თუმცა აღნიშნული პროცესი უნიფიცირებული არ არის (27)

(24) *ტაძარ*-*თ*-ი, *ტაძრ*-*თ*-ის, *ტაძრ*-*ებ*-ი და სხვ.

ტაძარ-SG-NOM

ტაძარ-SG-GEN

ტაძარ-PL-NOM

(25) *სოფელ*-*თ*-ი, *სოფლ*-*თ*-ის, *სოფლ*-*ებ*-ი და სხვ.

სოფელ-SG-NOM

სოფელ-SG-GEN

სოფელ-PL-NOM

(26) *ობოლ*-*თ*-ი, *ობლ*-*თ*-ის, *ობლ*-*ებ*-ი და სხვ.

ობოლ-SG-NOM

ობოლ-SG-GEN

ობოლ-PL-NOM

(27) *წელ*-*თ*-ის, *ბალ*-*თ*-ის და სხვ.

წელ-SG-GEN

ბალ-SG-GEN

- კვეცა (აპოკოპე), რომელიც გულისხმობს -*ა*- და -*ე*- ხმოვანზე

დამთავრებული ფუძეების მოკვეცას მხოლოდითი რიცხვის მოქმედებით და ვითარებით ბრუნვებში და -ა- ხმოვანზე დამთავრებული ფუძეების მოკვეცას „ებიანი“ მრავლობითი რიცხვის ყველა ბრუნვაში (28-29). იმ შემთხვევაში თუ სიტყვის ბოლო მარცვალი სონანტზე მთავრდება, კუმშვისა და კვეცის პროცესი პარალელურად მიმდინარეობს (30).

(28) ხე-~~თ~~ , ხ-~~თ~~-ის, ხე-~~ებ~~-ი და სხვ.

ხე-SG.NOM

ხე-SG-GEN

ხე-PL-NOM

(29) მამა-~~თ~~, მამ-~~თ~~-ის, მამ-~~ებ~~-ი და სხვ.

მამა-SG.NOM

მამა-SG-GEN

მამა-PL-NOM

(30) ტომარა-~~თ~~, ტომრ-~~თ~~-ის, ტომრ-~~ებ~~-ი და სხვ.

ტომარა-SG.NOM

ტომარა-SG-GEN

ტომარა-PL-NOM

- სუპლექტივიზმი, რომელიც გულისხმობს სიტყვის ძირის მონაცვლეობას პარადიგმის ფარგლებში გრამატიკულად მსგავსი, მაგრამ ფონოლოგიურად განსხვავებული ფორმების გამოყენებით. ქართულ ენაში ფუძისეული სუპლექტივიზმი ზმნური პარადიგმის წარმოებისას ვლინდება და არარეგულარულ ხასიათს ატარებს (31)

(31) ე-~~უბნ~~-~~ებ~~-ა, ე-~~ტყვ~~-~~ი~~-ს, უ-~~თხრ~~-ა, უ-~~თქვ~~-~~ამ~~-ს და სხვ.

Vers-უბნ-TS-3SGSbj:PRS.IND

Vers-უბნ-TS-3SGSbj:FUT.IND

Vers-უბნ-3SGSbj:AOR.IND

Vers-უბნ-TS-3SGSbj:PF.IND

ძველი, საშუალო და თანამედროვე ქართული ენის სიტყვა-ცვლილებითი

მორფოლოგია შემდეგ მორფოლოგიურ კატეგორიებს მოიცავს:

- რიცხვი: მხოლობითი, მრავლობითი;
- ბრუნვა: სახელობითი, მოთხრობითი, მიცემითი, ნათესაობითი, მოქმედებითი, ვითარებითი, წოდებითი. აღნიშნულ სიას ძველ ქართულში წრფელობითი ბრუნვა ემატება;
- ხარისხი: დადებითი, ოდნაობითი, უფროობითი;
- პირი: პირველი, მეორე, მესამე;
- სერიების შემთხვევაში ხდება შემდეგი სამი კატეგორიის ასახვა:
 - o დრო: ანმყო, ნამყო უსრული, ანმყოს კავშირებითი, მყოფადი, ხოლმეობით-პირობითი, მყოფადის კავშირებითი, წყვეტილი, მეორე კავშირებითი, ბრძანებითი, პირველი თურმეობითი, მეორე თურმეობითი და მესამე კავშირებითი;
 - o კილო: თხრობითი, კავშირებითი, ბრძანებითი;
 - o ასპექტი: სრული, უსრული;
- გვარი (დიათეზა)³: აქტივი, ავტოაქტივი (მედიოაქტივი), ინაქტივი, პასივი, ავტოპასივი (მედიოპასივი);
- შეთანხმება⁴: სუბიექტი, პირდაპირი ობიექტი, ირიბი ობიექტი

სხვადასხვა მეტყველების ნაწილს სხვადასხვა მორფოსინტაქსური კატეგორია ახასიათებს, თუმცა აღნიშნული კატეგორიების მთავარი დანიშნულებაა სიტყვის დონეზე სხვადასხვა გრამატიკული თვისებების გამოვლენა. რაც, ერთი მხრივ, განაპირობებს სიტყვის – სლოტებად სტრუქტურირებას, ხოლო, მეორე მხრივ, კონკრეტული სიტყვის – მორფოლოგიური ინფორმაციისა და წინადადების სხვა სიტყვებთან მისი კავშირის წარმოჩენას. ასე, მაგალითად, ზმნური ფორმა შეიძლება აჩვენებ-

3 ტრადიციული მიდგომა, რომელსაც გრამატიკების უმეტესობა იზიარებს (გოგოლაშვილი და სხვ. 2011, ვიერ 2011 და სხვ.) განასხვავებს სამ გვარს: მოქმედებით, ვნებით და საშუალოს. აღნიშნული მიდგომების ფარგლებში გვარი განიხილება როგორც მონოპერსონალური მშენებისთვის დამახასიათებელი და ევროპულ ენათა უმეტესობაში დადასტურებული გრამატიკული კატეგორია.

4 სუბიექტ-ობიექტის შეთანხმება ძირითადად სინტაქსის ნაწილად განიხილება, თუმცა მორფოლოგიური ნიშნების არსებობა განაპირობებს მათ განიხილვას სიტყვაცვლილებითი მორფოლოგიის ფარგლებში. ქართულ ენაში სუბიექტ-ობიექტის შეთანხმება წარმოადგენს არა მხოლოდ სინტაქსური, არამედ მორფოლოგიური სტრუქტურის ნაწილსაც, სხვაგვარად რომ ვთქვათ, კონკრეტული მორფოლოგიური ნიშნების გამოყენებით უზრუნველყოფს ურთიერთქმედებას მორფოლოგიასა და სინტაქსს შორის.

დეს არა მხოლოდ თუ როდის ხდება მოქმედება, არამედ იმასაც თუ ვინ ახორციელებს მას (აგენსი), ვინ განიცდის ზემოქმედებას (პაციენსი) ან ვის სასარგებლოდ ხორციელდება იგი (ბენეფიციარი) და სხვა ინფორმაციას. სიტყვები, რომლებიც ჩნდება სხვადასხვა კონტექსტში, ბუნებრივია, კონკრეტული მეტყველების ნაწილის - პარადიგმის ნაწილს წარმოადგენს. კონკრეტული სიტყვა-ფორმის გამოყენება დამოკიდებულია კონტექსტზე, რომელსაც კონკრეტული ენის გრამატიკული და მორფოლოგიური წესები განაპირობებს. აღნიშნული წესები კი პარადიგმაში ვლინება, რომელიც სიტყვათა კლასის შესაბამისს ფლექსიურ ფორმათა ჯგუფს მოიცავს.

თავი 4

დასკვნები

როგორც სიტყვანარმოებითი, ასევე სიტყვაცვლილებითი მორფოლოგია ენის განუყოფელ ნაწილს წარმოადგენს და გულისხმობს განსაზღვრული კავშირის არსებობას სხვადასხვა ენობრივ შრეს შორის. მორფოლოგია თავისი ზოგადი გაგებით სხვადასხვა ლინგვისტურ დონეს უკავშირდება: ფონოლოგიასაც და სინტაქსსაც. ენაში მიმდინარე ფონეტიკური პროცესების ნაწილი ზეგავლენას ახდენს სიტყვის მორფოლოგიურ სტრუქტურაზე, ხოლო სიტყვა-ფორმის მორფოლოგიური სტრუქტურა მის კონტექსტზე დამოკიდებულებას ავლენს და სინტაქსში ვლინდება. ამიტომაც ქართული სიტყვა-ფორმა განიხილება როგორც მორფოფონოლოგიურ, ასევე მორფოსინტაქსურ დონეზე.

სიტყვანარმოებითი და სიტყვაცვლილებითი მორფოლოგიის ფონოლოგიასა და სინტაქსთან დაკავშირების გზით როგორც ძველ და საშუალო, ასევე თანამედროვე ქართული ენისათვის დამახასიათებელი პროცესებისა და მორფოსინტაქსური მახასიათებლების წარმოდგენა შესაძლებელია შემდეგნაირად:

მეტყველების ნაწილი	ფუძის ტიპი	პროცესი	მორფოსინტაქსური თვისება
არსებითი სახელი	თანხმოვანფუძიანი, ხმოვანფუძიანი	კუმშვა, კვეცა	ბრუნვა რიცხვი

მედსართავი სახელი	თანხმოვანფუძიანი, ხმოვანფუძიანი	კუმშვა, კვეცა	ხარისხი ბრუნვა რიცხვი
ზმნა	თანხმოვანფუძიანი, ხმოვანფუძიანი (იშვიათად)	აბლაუტი, სუპლეთივიზმი (იშვიათად)	დრო კილო ასპექტი გვარი შეთანხმება რიცხვი პირი
ნაცვალსახელი	თანხმოვანფუძიანი, ხმოვანფუძიანი	კუმშვა, კვეცა	ბრუნვა რიცხვი პირი
რიცხვითი სახელი	თანხმოვანფუძიანი, ხმოვანფუძიანი (იშვიათად)	კვეცა	ბრუნვა
კავშირი	თანხმოვანფუძიანი, ხმოვანფუძიანი		
ნაწილაკი	თანხმოვანფუძიანი, ხმოვანფუძიანი		
თანდებული	თანხმოვანფუძიანი, ხმოვანფუძიანი		ბრუნვა
შორისდებული	თანხმოვანფუძიანი, ხმოვანფუძიანი		

ცხრილი 2 ფუძის ტიპი, პროცესები და თვისებები

გემოხსენებულ მეტყველების ნაწილებს ემატება შემოკლებანი და პუნქტუაციის ნიშნები. შემოკლებათა წარმოდგენის სტრატეგიები, ძირითადად, მიჰყვება სიტყვაგვლილებითი მოდელის პრინციპებს. პუნქტუაციის ნიშნები კი მნიშვნელოვანია წინადადებებისა და სიტყვების საზღვრების დადგენის თვალსაზრისით კომპიუტერული მორფოლოგიის ძირითადი ამოცანებიდან გამომდინარე.

კომპიუტერული მორფოლოგია

კომპიუტერული მორფოლოგია წარმოადგენს სიტყვა-ფორმების კომპიუტერული ანალიზისა და სინთეზის კვლევას ბუნებრივი ენის კომპიუტერულად დამუშავების მიზნით. კომპიუტერული მორფოლოგია მორფოლოგიური თვისებების კვლევის სხვადასხვა კომპიუტერულ მიდგომას მოიცავს და, ძირითადად, ფოკუსირდება როგორც ლინგვისტურ თეორიებზე, ასევე აღნიშნული თეორიების კომპიუტერულ დამუშავებაზე. ბუნებრივი ენის სხვადასხვა ტიპის კომპიუტერული სისტემის აგების მიზნით კომპიუტერული ენათმეცნიერება, ატარებს რა ინტერდისციპლინარულ ხასიათს, ლინგვისტური და კომპიუტერული მიდგომებისა და მეთოდოლოგიის გადაკვეთის წერტილს წარმოადგენს. კონკრეტული ენის მორფოლოგიის კოდირება გულისხმობს მის ხელმისაწვდომობას ბუნებრივი ენის დამუშავების მოდულებისთვის.

კომპიუტერული მორფოლოგიის ძირითად ამოცანას სიტყვების როგორც გრაფიკული, ასევე ფონემური დამუშავება წარმოადგენს, სხვაგვარად რომ ვთქვათ, მისთვის მნიშვნელოვანია როგორც წერილობითი, ასევე ზეპირი ენობრივი მასალის დამუშავება (ტროსტი 2009). მთავარი ამოცანა კი უკავშირდება შემაჯალი სიგნალის სახით სიმბოლოთა რიგის მიღებას, ხოლო გამავალი სიგნალის სახით მიღებული რიგის ანალიზს. გამავალ სიგნალს შეიძლება წარმოადგენდეს სიტყვა-ფორმაში წარმოდგენილი როგორც მორფემები, ასევე სიტყვა-ფორმის სტრუქტურის მორფოსინტაქსური ინტერპრეტაცია (32).

(32) ინანწალებდა და სხვ.

ი-ნანწალ-ებ-და

Ipfv+ნანწალებ-ს+Verb+Main+IDt+#9+RelStat+Intr+AutAct
+FutCond+<NomSubj>+Subj3Sg

დღეისათვის კომპიუტერული მორფოლოგიის განუყრელ ნაწილს იმ ინფორმაციის მოპოვება წარმოადგენს, რომელიც, ერთი მხრივ, სიტყვის

შემადგენელ ნაწილებს ეხება, როგორიცაა მორფემები, ხოლო, მეორე მხრივ, მორფემების ურთიერთკავშირის წესებს, როგორიცაა კონკრეტული ენის მორფოტაქტიკა. ყველა აღნიშნული ამოცანა კომპიუტერული ენათმეცნიერების ფარგლებში სხვადასხვა მეთოდოლოგიას საჭიროებს. ამ მხრივ, ძირითადად, განიხილება სამი ტიპის მეთოდოლოგიური ჩარჩო: ა) სტატისტიკური, ბ) აფიქსებად სიტყვის დაშლისა და გ) სასრული პოზიციის მორფოლოგიური ანალიზი. აღნიშნული ნაშრომი კი მხოლოდ სასრული პოზიციის მორფოლოგიურ ანალიზს ეთმობა.

მეორე ნაწილი ხუთ თავად იყოფა: 1. შესავალში წარმოდგენილია სიტყვის სასრული პოზიციის მორფოლოგიური ანალიზის ის ასპექტები, რომლებიც მნიშვნელოვანია ქართული ენის სასრული პოზიციის მორფოლოგიური ანალიზის თვალსაზრისით; 2. მეორე თავში განიხილება სხვაობა სასრული პოზიციის ავტომატებსა და სასრული პოზიციის გარდამქმნელებს შორის; 3. მესამე თავი ეხება სასრული პოზიციის ტექნოლოგიის რეგულარულ გამოსახულებებსა და სასრული პოზიციის ოპერატორებს, რომლებიც მნიშვნელოვანია ენის გენერირების თვალსაზრისით; 4. მეოთხე თავი სამ ქვეთავად იყოფა და კონცენტრირდება ქართული ენის მორფოსინტაქსური ანალიზის იმ პარამეტრებზე, რომლებიც მნიშვნელოვანია ენის მოდელირების თვალსაზრისით და რომელიც ზეგავლენას ახდენს ქართული ენის მოდელირებაზე ქსეროქსის სასრული პოზიციის საშუალებების გამოყენებით როგორიცაა *xfst* და *lexc*; 5. დასკვნებით კი ხდება ზემოხსენებული ოთხი თავის შეჯამება.

თავი 1

შესავალი

კომპიუტერული ენათმეცნიერება დღესდღეობით მოიაზრებს ბუნებრივი ენის დამუშავების მიმართულებით სხვადასხვა მიდგომისა და მეთოდის შექმნას, დანერგვას, განვითარებასა და გაუმჯობესებას. მთავარი ამოცანები კი სხვადასხვა კვლევით პრობლემატიკაზე კონცენტრირდება, კერძოდ, თეორიულ და კორპუსულ ენათმეცნიერებაზე, ლექსიკოგრაფიაზე, ხელოვნურ ინტელექტში ლოგიკისა და ცოდნის წარმოდგენაზე, სტა-

ტისტიკურ მოდელირებაზე ნეირული ქსელებისა და ინფორმაციის ამოღების სისტემების ფარგლებში და სხვ. (მანიჩი და სხვ. 1999, 2008).

მთავარი მიდგომა კი მოიაზრებს კომპიუტერული ან მათემატიკური მიდგომების გამოყენებას ლინგვისტური კვლევების ჩატარების დროს, რაც, თავის მხრივ, გულისხმობს ბუნებრივი ენის დამუშავებას, სხვაგვარად რომ ვთქვათ, კომპიუტერის გამოყენებას ტრადიციულად ჰუმანიტარული დისციპლინების კვლევითი ამოცანების გადაჭრის მიზნით. დღეისათვის არსებობს სხვადასხვა მეთოდოლოგია ბუნებრივი ენის დამუშავების თვალსაზრისით, რომელიც მხოლოდ ნაწილობრივ რეალიზებულია ქართული ენის შემთხვევაში სხვადასხვა დროს ქართულ ენაზე მომუშავე ჯგუფების მიერ (მოერერი 2009, გურევიჩი 2006, დათუკიშვილი 2005, მარგველანი 1999-2001 და სხვ.).

ქართული ენის რთული მორფოსინტაქსური სტრუქტურა ქმნის სხვადასხვა რესურსისა და მიდგომის გამოყენების საჭიროებას ბუნებრივი ენის დამუშავების თვალსაზრისით და წარმოადგენს სხვადასხვა ამოცანის გადაჭრის წინაპირობას, როგორცაა:

ა) მართლწერის შემმონმბებელი პროგრამების შექმნა, რომელიც საჭიროებს ენის მოდელზე დაფუძნებული სასრული პოზიციის ამომცნობების განვითარებას სიტყვების უსრული რაოდენობის შემონმბების მიზნით;

ბ) ტეგსეტების არსებული დიზაინების შეჯერება, რადგანაც თითქმის შეუძლებელია ენის მორფოსინტაქსური მოდელის სრულყოფილად აღწერა მხოლოდ სასრული პოზიციის ტეგსეტით. ერთი მხრივ, სხვადასხვა ლინგვისტური თეორია სხვადასხვანაირად აღწერს ენის სტრუქტურას (შდრ. აკ. შანიძისა და დ. მელიქიშვილის ზმნის მორფოსინტაქსი), ხოლო, მეორე მხრივ, სასრული პოზიციის ტეგსეტი არ იძლევა ენის როგორც ფლექსიური, ასევე დერივაციული აფიქსაციის დაფარვის შესაძლებლობას;

გ) ენის მორფოსინტაქსური მოდელირების სრულყოფა, რადგანაც მნიშვნელოვანია არა მხოლოდ მორფოლოგიის, არამედ ენაში არსებული მორფოსინტაქსური კავშირების წარმოდგენაც სასრული პოზიციის მიდგომების გამოყენებით;

დ) ენის სინტაქსური მოდელირება, რომელიც წარმოადგენს სიტყვათა

შორის არსებული კავშირების შესწავლისა და მორფოსინტაქსური ომონიმიის მოხსნის წინაპირობას ფორმალიზმის არსებული თეორიების გათვალისწინებით.

ყველა ზემოხსენებული ამოცანის გადაწყვეტა შესაძლებელია როგორც სტატისტიკური, ასევე სასრული პოზიციის მიდგომების გამოყენებით. სტატისტიკურ მიდგომას საფუძვლად უდევს არა ლექსიკონი ან წინასწარ განსაზღვრული წესი, არამედ ტექსტში არსებული წინასწარ განსაზღვრული შაბლონი და მისი გამოყენების სიხშირე, სიტყვების სტრუქტურის ამოცნობა გულისხმობს კორპუსის ანალიზს. აღნიშნული მეთოდოლოგია კარგად გამოიყენება მეტყველების ნაწილების ავტომატური ტეგირების დროს. თუმცა მდიდარი მორფოლოგიური სტრუქტურების ენებში ლექსიკონის ზომა ქმნის მონაცემების გაფანტულობის საშიშროებას, რაც განსაკუთრებული სიმწვავით ვლინდება ე.წ. მცირე-რესურსების მქონე ენების შემთხვევაში, როგორიცაა ქართული.

სასრული პოზიციის ტექნოლოგია კი კომპიუტერული და მათემატიკური თვალსაზრისით კარგად შეჯერებულ ერთეულს წარმოადგენს (ჟურავსკი 2000 და სხვ.). სასრული პოზიციის გარდამქმნელი, როგორც წესი, ორმიმართულებიანია, შესაბამისად, კარგად მუშაობს როგორც გენერაციის, ასევე ანალიზის დონეზეც. ქართული ენის მორფოლოგიური ანალიზატორიც შეიქმნა სასრული პოზიციის ტექნოლოგიის გამოყენებით. სტატისტიკურ მეთოდოლოგიასთან შედარებით აღნიშნული მიდგომა პრომატევადია და ლექსიკონისა და წესების შემუშავების თვალსაზრისით უფრო მეტ დროს მოითხოვს.

თავი 2

სასრული პოზიციის ტექნოლოგია: სასრული პოზიციის ავტომატები (საა) და სასრული პოზიციის გარდამქმნელები (საგ)

ბუნებრივი ენის დამუშავების სისტემათა უმეტესობა ეფუძნება, ერთი მხრივ, ტურინგის აბსტრაქტულ მოდელს (ტურინგი 1936), რომელიც კომპიუტერულ მექანიზმს აღწერს მაკონტროლებელი ერთეულის სახით,

რომელიც სასრული სიიდან (ანბანიდან) ამოღებულ სიმბოლოებს ამუშავებს, ხოლო, მეორე მხრივ, ჩომსკი-შუტცენბერგის იერარქიას (ჩომსკი 1956), რომელიც ფორმალურ გრამატიკათა იერარქიას აღწერს იმ ენის კლასის, რომელსაც იგი აგენერირებს, იმ ავტომატების ტიპის, რომელიც ენის ზემოხსენებულ კლასს ცნობს და იმ ფორმის მიხედვით, რომელიც განაპირობების მათი გენერაციის შესაძლებლობას კონკრეტულ წესებზე დაყრდნობით. სისტემათა უმეტესობის ერთ-ერთ ყველაზე მნიშვნელოვან კომპონენტს ენის მორფოლოგია წარმოადგენს. ენის კომპიუტერული მორფოლოგია აერთიანებს კომპიუტერულ და ლინგვისტურ მიდგომებს სიტყვის სტრუქტურისადმი შემდგომი გენერაციისა და/ან ანალიზის უზრუნველყოფის მიზნით.

ტურინგის მიერ აღწერილი მანქანა მანამდე ახდენს ქმედებას, სანამ არ შეწყდება შემაგალი სიგნალების თანმიმდევრობა და არ მიაღწევს სასრულ (მიმღებ ან უარყოფ) პოზიციას. სიმბოლოების მიღებული თანმიმდევრობა სიტყვას წარმოადგენს და ასეთი სიტყვების ერთობლიობა ქმნის ენის მანქანას. ტურინგის მანქანა ჩომსკი-შუტცენბერგის იერარქიის რეკურსულად თვლადი გრამატიკების ნულოვან ტიპს განეკუთვნება. იერარქიის პირველ საფეხურს კონტექსტზე დამოკიდებულ ენათა ტიპი წარმოადგენს, ხოლო მეორეს – კონტექსტზე დამოუკიდებელი. მესამე დონეს წარმოადგენს რეგულარულ გრამატიკების ტიპი და, შესაბამისად, ამ ტიპს სასრული პოზიციის ავტომატების (სპა), სხვაგვარად რომ ვთქვათ, სასრული პოზიციის მანქანის სახით აღწერენ ხოლმე.

სასრული პოზიციის ავტომატი, ტურინგის მანქანისგან განსხვავებით, მხოლოდ ერთი მიმართულებით მოძრაობს (მაგ. მარცხნიდან მარჯვნივ) და ახერხებს შემაგალი სიგნალის აღქმას სიმბოლოების თანმიმდევრულად ამოკითხვის გზით. ავტომატი, რომელიც ყოველთვის პოზიციების ბოლო რიგში თავსდება, ახალ პოზიციაზე მხოლოდ ერთი სიმბოლოსა და მისი შიდა პოზიციის ამოკითხვის შემდეგ გადადის.

სასრული პოზიციის ავტომატის საშუალებით შესაძლებელი ხდება ისეთი სისტემების მოდელირება, რომლებიც პოზიციათა სასრული რაოდენობისგან შედგება და, რომლებიც რეგულარული ენის მიღებისას

განაპირობებს მის კოდირებას რეგულარული გამოსახულების სახით (კარტუნენი 2009, ჟურავსკი და სხვ. 2000 და სხვ.). უფრო ფორმალურად, სასრული პოზიციის ავტომატის განსაზღვრა შესაძლებელია შემდეგ ხუთ პარამეტრზე დაყრდნობით:

$Q = q_0 q_1 q_2 \dots q_{n-1}$	N პოზიციების ბოლო რიგი
Σ	სიმბოლოთა შემავალი ანბანი (სიმბოლოთა სასრული რიგი)
$\delta(q, i)$	გარდამავალი მატრიცა ორ არგუმენტს შორის; ერთ არგუმენტს პოზიცია წარმოადგენს, ხოლო მეორეს - შემავალი სიმბოლო, მატრიცა კი ახალ პოზიციას ქმნის
q_0	საწყისი პოზიცია, რომელიც $Q - q_0 \in Q$ -ის წევრია
F	სასრული პოზიცია, რომელიც $Q - F \subseteq Q$ -ის ქვეწყებაა

ცხრილი 3 სასრული პოზიციის ავტომატების ხუთი პარამეტრი

სპა შეიძლება წარმოადგენდეს როგორც ენის მიმღებს, ასევე გენერატორსაც. საწყისი პოზიციიდან ბოლო პოზიციამდე სპა ტოვებს თითოეული გადაადგილების კვალს, ხოლო მისი ქმედების ბუნებრივ შემზღუდველს ანბანის სიმბოლოთა სასრული რაოდენობა წარმოადგენს. სხვაგვარად რომ ვთქვათ ენას შეიძლება ჰქონდეს რიგების უსასრულო რაოდენობა, მაგრამ ანბანი აუცილებლად სასრული იქნება.

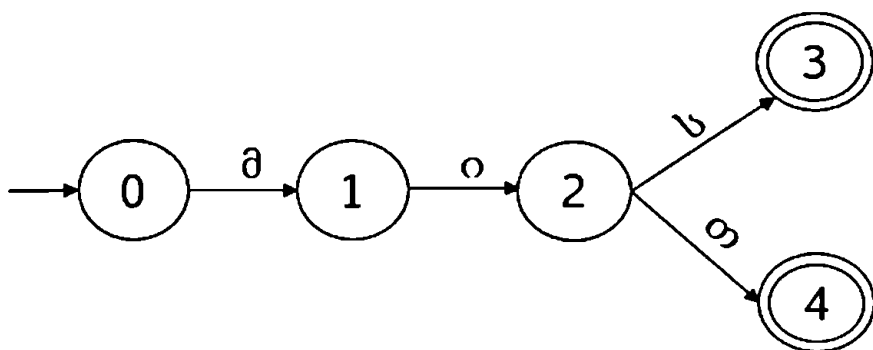
სპა-ს დანიშნულებაა რეგულარული ენების მიღება და საჭიროებიდან გამომდინარე დამუშავება. კომპიუტერული თვალსაზრისით ენა რეგულარულია თუ რომელიმე ენობრივი ერთეული რიგის განსაზღვრულ ადგილას თავსდება და იზღუდება მხოლოდ ინფორმაციით წინ მდგომი ერთეულების შესახებ. შესაბამისად, რეგულარული ენის დახასიათება რეგულარული გამოსახულებებით შეიძლება. რეგულარული გამოსახულებები კი გამოიყენება სასრული პოზიციის ავტომატების განსაზღვრის დროს. სამაგალითოდ შესაძლებელია 'მის' და 'მით' ნაცვალსახელების გამოყენება, რომლებშიც -ი-ს შეიძლება მოსდევდეს ან -ს ანდა -თ. რეგულარული გამოსახულების დონეზე აღნიშნული ერთეულები შემდეგნაირად წარმოიჩინდება:

მ ი [ს | თ]

გამოსახულება 1 რეგულარული გამოსახულების მაგალითი

კომპიუტერული თვალსაზრისით, როგორც *მის*, ასევე *მით* ნაცვალ-სახელი სიმბოლოთა რიგს წარმოადგენს. რეგულარული გამოსახულება კი აღნიშნული რიგის ამოკითხვას ახდენს. რიგი, თავის მხრივ, მოიცავს ნულს ან მეტ სიმბოლოს. თუ რიგი არც ერთ სიმბოლოს არ მოიცავს მას **ცარიელ რიგს** უწოდებენ და ენასაც, რომელიც ცარიელი რიგებისგან შედგება **ცარიელი რიგის ენას** იმ ენის საპირისპიროდ, რომელშიც არც ერთი რიგი საერთოდ წარმოდგენილი არ არის და რომელსაც **ცარიელ ან ნულოვან ენას** უწოდებენ.

სასრული პოზიციის ავტომატი განიხილება როგორც **ქსელი**, რომელიც პოზიციებისა და რკალებისაგან შედგება. ქსელს გააჩნია როგორც სანწყისი პოზიცია, ასევე ბოლო პოზიცია. პოზიციები გრაფიკულად წრეებისა და რკალების სახით წარმოგვიდგება. სანწყისი პოზიცია ყოველთვის მარცხნივ თავსდება, ხოლო ბოლო – მარჯვნივ. ნებისმიერი პოზიცია ნულოვანი ან მეტი პოზიციით იწყება და შემდგომი პოზიციისკენ მიემართება. თუ ქსელის ნებისმიერი რკალი ერთი სიმბოლოსგან შედგება, მაშინ ქსელს **მარტივ ავტომატს** უწოდებენ, ხოლო თუ წყვილისაგან, მაშინ ქსელს **გარდამქმნელს** უწოდებენ. რკალების თანმიმდევრობას სანწყისი პოზიციიდან ბოლო პოზიციამდე **ბილიკს** უწოდებენ. ბილიკების საერთო რაოდენობა ქმნის **სასრული პოზიციის ენას** ან **სასრული პოზიციის რელაციას**.



გრაფიკი 1 სასრული პოზიციის მარტივი ენის მაგალითი

სასრული პოზიციის ქსელის გამოყენება მნიშვნელოვანია სიტყვათა სიის კოდირებისას, რადგანაც მხოლოდ ამ გზით შესაძლებელი ხდება სიტყვების კონკრეტული ნაწილების გაზიარება სიტყვა-ფორმის საწყის და ბოლო პოზიციებს შორის.

სასრული პოზიციის ავტომატების ერთ-ერთ ტიპს სასრული პოზიციის გარდამქმნელი (სპგ) წარმოადგენს, მას სხვანაირად მიეღოს მანქანას უწოდებენ (მიელი 1955). სასრული პოზიციის ავტომატისაგან განსხვავებით, სპგ-ს შეუძლია ენობრივი ერთეულის არა მხოლოდ მიღება, არამედ გენერირებაც. სხვაგვარად რომ ვთქვათ, სპგ არის ორმომართულებიანი. სპა ენას რიგების წყების განსაზღვრის გზით, ხოლო სპგ ენათა შორის ურთიერთობას რიგების წყებებს შორის ურთიერთობის განსაზღვრის გზით ადგენს (დჰონჩადხა 2002, ჟურავსკი 2000 და სხვ.). აღნიშნული ურთიერთობა ვლინდება **რეგულარული რელაციის** სახით, რომელიც სიმბოლოთა წყვილებს შორის რეგულარული კავშირების გამოვლენას ზედა და ქვედა დონის დაკავშირების გზით უზრუნველყოფს.

შესაბამისად, სასრული პოზიციის გარდამქმნელი ლექსიკურ (ზედა) და ფორმობრივ (ქვედა) ენებს შორის რეგულარული კავშირის კოდირებას ახდენს. გარდამქმნელი შემავალი სიგნალის მიღების შემდეგ სხვადასხვა ტიპის გამავალ სიგნალს აგენერირებს და ამით განაპირობებს ლექსიკური ორმომართულებიანი გარდამქმნელის შექმნას.

თავი 3

რეგულარული გამოსახულებანი და სასრული პოზიციის ოპერატორები

სასრული პოზიციის ავტომატის შემავალ სიგნალს რეგულარული ენა წარმოადგენს. რეგულარული ენის დასაშუაშვებლად ანუ ტექსტის საძიებო ხაზის განსაზღვრის მიზნით კომპიუტერულ მეცნიერებაში **რეგულარული გამოსახულება** გამოიყენება.

რეგულარული გამოსახულება პირველად კლინმა (1956) შეიმუშავა და განიხილავდა მას, როგორც რიგების მარტივი კლასების განსაზღვრის

სპეციალურ ფორმულას. ფორმალურად რეგულარული გამოსახულება არის რიგთა წყების ალგებრული დახასიათების ცდა (ჟურავსკი 2000 და სხვ.) და საჭიროებს როგორც საძიებო შაბლონს, ასევე ტექსტების კორპუსს. ამ ეტაპზე რეგულარული გამოსახულება ისეთ კომპიუტერულ ენებში გამოიყენება როგორიცაა python, Java და ა.შ. და, ბუნებრივია, მისი გამოყენების საჭიროება კომპიუტერული ენათმეცნიერების ამოცანების გადასაწყვეტად. რეგულარული გამოსახულების ყველაზე მარტივი ლინგვისტური ერთეულების ჩვენება შემდეგნაირად შეიძლება:

მ ი [ს | თ] - შეესაბამება ნაცვალსახელის ფორმებს *მის* ან *მით* ;

(3) ყ ო ფ ს - შეესაბამება ფორმებს *ყოფს* ან *ყოფს* და აჩვენებს, რომ კ- პირობითია;

და ა [?+] ს - შეესაბამება ნებისმიერ სიტყვა-ფორმას, რომელიც იწყება და- თი, რომელსაც ერთი ან მეტი სიმბოლო მოსდევს და სრულდება -ს-თი;

და ა ნ ე რ [?*] - შეესაბამება ნებისმიერ სიტყვა-ფორმას, რომელიც იწყება *დანერ-* ფორმით და მთავრდება ერთ ან მეტ სიმბოლოთა თანმიმდევრობით;

[?*] ნ ე რ # - შეესაბამება ნებისმიერ სიტყვა-ფორმას, რომელიც იწყება ერთი ან მეტი სიმბოლოს თანმიმდევრობით და სრულდება -*ნერ* ფორმით.

ჩამოთვლილ ატომურ გამოსახულებებს შორის აღსანიშნავია: 1. **კითხვის ნიშანი (?)**, რომელიც ნებისმიერ სიმბოლოსა ან ნებისმიერ სიმბოლოთა თანმიმდევრობას აღნიშნავს; 2. **კლინის ვარსკვლავი (*)**, რომელიც რეგულარული გამოსახულების წინა სიმბოლოს ნულოვან ან მეტჯერ გამეორების შემთხვევებს აღნიშნავს; 3. **დიეზის ნიშანი (#)**, რომელიც აბსოლუტურ დასაწყისსა ან აბსოლუტურ დასასრულს აღნიშნავს.

ანბანის სიმბოლოთა და ენობრივი წესების ურთიერთქმედება სასრული პოზიციის ავტომატის შექმნის წინაპირობას ქმნის. სასრული პოზიციის გარდამქმნელი, სასრული პოზიციის ავტომატისგან განსხვავებით, აღნიშნავს არა რეგულარულ ენას, არამედ ენებს შორის ურთიერთობას (რელაციას). სასრული პოზიციის გარდამქმნელის მათემატიკურ ოპერაციათა სიმრავლე დანვრილებით წარმოდგენილია ბისლი და კარტენენტან (ბისლი და სხვ. 2003, ჟურავსკი 2000 და სხვ.),

თუმცა ზოგიერთი ოპერატორის აღნიშვნა მნიშვნელოვანია სისტემური თვალსაზრისით:

რეგულარული გამოსახულება	ოპერატორებისა და ქმედებების განმარტება
განმეორება (იტერაცია)	
a^+	ასახავს a -ს a -სთან ერთჯერად ან მრავალჯერად გადაკვეთას. აღნიშნულ ოპერატორს კლინის პლუსს ან სიგმა-პლუსს უწოდებენ.
a^*	ასახავს a^+ -ის კავშირს ცარიელი რიგის ენასთან. აღნიშნულ ოპერატორს კლინის ვარსკვლავს ან სიგმა-ვარსკვლავს უწოდებენ.
მატება	
$\sim a$	აღნიშნავს რიგს, რომელიც a ენაში არ დასტურდება. $\sim a$ უტოლდება $[?^* - a]$.
$\backslash a$	მარცხენი დახრილი ხაზის ($\backslash$) ოპერატორს სასაზღვრო უარყოფას უწოდებენ. უტოლდება $[? - a]$.
პირობითობა	
(a)	ასახავს ენისა ან რელაციის კავშირს ცარიელი რიგის ენასთან და, ფაქტიურად, უტოლდება $[a \mid 0]$ -ს.
კასკადი (კონკატენაცია)	
$[a \ b]$	ასახავს ორი ენისა და რელაციის კავშირს. უტოლდება $a \ b$ -ს თანმიმდევრობას, ხოლო ინტერვალური რეგულარულ გამოსახულებათა შორის მაკავშირებელი ოპერატორის ფუნქციას ასრულებს.
$\{ a \}$	ასახავს კავშირს სიმბოლოთა რიგს შორის, და, ფაქტიურად, ერთმანეთს უტოლებს $\{ გული \}$ და $[გ \ უ \ ლ \ ი]$ გამოსახულებებს.
ლოკალიზაცია	

5 ა ნებისმიერ რეგულარულ გამოსახულებას ასახავს.

ა	ასახავს ენასა ან რელაციას, რომელიც მიიღება უნივერსალური ენის აფიქსაციის რეგულარულ გამოსახულებაზე დართვის გზით. ა უტოლდება [?* ა ?*].
გაერთიანება	
[ა ბ]	ასახავს ორი ენისა ან რელაციის გაერთიანებას.
გადაკვეთა	
[ა .x. ბ]	ასახავს რელაციას, რომელიც ერთმანეთს აკავშირებს ა (ზედა) ენის ნებისმიერ რიგს ბ (ქვედა) ენის ნებისმიერ რიგთან. .x. ოპერატორი : ოპერატორს უტოლდება, თუმცა ორი წერტილი (:), როგორც წესი, უფრო პრიორიტეტულია.
შეკავშირება	
[ა .o. ბ]	ა-ს რელაციისა ბ-ს რელაციასთან შეკავშირების შესაძლებლობას ასახავს.
პროექცია	
ა.ა	ა-ს რელაციის ზედა ენას ასახავს.
ა.ი	ა-ს რელაციის ქვედა ენას ასახავს.

ცხრილი 4 რეგულარული გამოსახულების ზოგიერთი ოპერატორისა და ქმედების განმარტება

პრიორიტეტულობის თვალსაზრისით ზოგიერთი ოპერატორი წინ უსწრებს დანარჩენ ოპერატორებს, ხოლო მათ გამოყენებას კონკრეტული თანმიმდევრობა ახასიათებს, კერძოდ,

- შეკავშირების ჩვენებისას პრიორიტეტულობა ენიჭება : ოპერატორს;
- რეგულარულ გამოსახულებამდე გამოყენებული ოპერატორების თანმიმდევრობა შემდეგია: $\sim$, $\backslash$, $\$$;
- რეგულარული გამოსახულების შემდეგ გამოყენებული ოპერატორებისა კი: $+$, $*$, $^$, .ა და სხვ. და ბოლოს;
- გადაკვეთის ჩვენებისას პრიორიტეტულია .x. და მხოლოდ შემდეგ .o. აღსანიშნავია რეგულარული გამოსახულების რიგის განსაზღვრულ

ადგილზე მიბმის ორი ოპერატორი, როგორიცაა $\wedge$ და $\$$. აღნიშნულ ოპერატორებს ანკერებს უწოდებენ.

თავი 4

ქართული ენის მორფოსინტაქსური ანალიზი და ქსაროქსის სასრული კოზიციის საშუალებანი (*xfst* და *lexc*)

ბუნებრივი ენის დამუშავების საწყისებს მორფოლოგიური დამუშავება წარმოადგენს, რაც განსაკუთრებით აქტუალურია მორფოლოგიურად მდიდარი ენების შემთხვევაში როგორიცაა ქართული. ქართული ენის კომპიუტერული მოდელირების თვალსაზრისით მორფოლოგიური ანალიზის სირთულე დამოკიდებულია ლინგვისტურ სპეციფიკაზე, რომელიც ქართული ენის ანოტირებას საფუძვლად უდევს. ლინგვისტური სპეციფიკა ემსახურება მრავალფეროვანი მორფოსინტაქსური ინფორმაციის ასახვას, რომელიც წარმოაჩენს როგორც ერთი ძირიდან სხვადასხვა სიტყვა-ფორმის გენერირებისა, ასევე წარმოებული სიტყვა-ფორმების სხვა ერთეულებთან სინტაქსურ დონეზე ურთიერთობას. ქართული ენის მორფოლოგიური სტრუქტურა გულისხმობს ფლექსიური კატეგორიების სიმრავლეს, რომლებიც დასტურდება როგორც ზმნური, ასევე სახელური პარადიგმების ფარგლებში, წარმოქმნის თვალსაზრისით სხვადასხვა ელემენტების ერთმანეთზე ურთიერთდამოკიდებულებასა და დიდი რაოდენობის რეგულარული, ნაწილობრივ რეგულარული და ირეგულარული შაბლონების არსებობას. რაც ნიშნავს იმას, რომ მორფოლოგიურად მდიდარი ენის სტრუქტურა სიტყვის დონეზე სხვადასხვა ინფორმაციას ავლენს. აღნიშნული ინფორმაცია კი ზეგავლენას ახდენს, ერთი მხრივ, სალექსიკონო მარაგზე და, შესაბამისად, ლექსიკოგრაფიაზეც, რაც კომპიუტერული თვალსაზრისით ლემის დადგენის პრობლემას უკავშირდება, ხოლო, მეორე მხრივ, წარმოების წესებზე, რაც კომპიუტერული თვალსაზრისით სწორი ფორმების გენერირების აუცილებლობაში ვლინდება. როგორც სალექსიკონო, ასევე ფორმის წარმოების წესების დონე განაპირობებს სრულმამულებიანი კომპიუტერული სისტემის შექმნას და, თავისთავში მოიაზრებს როგორც დახურული,

ასევე ღია კლასების წარმოებას. ღია კლასის ელემენტების წარმოებას სჭირდება გამართული წესის შემუშავება, ხოლო დახურული კლასის ელემენტებს - მხოლოდ ლექსიკონში ასახვა. ამიტომაც ქართული ენის მორფოსინტაქსური გენერირება ორ მძლავრ წარმოებით მოდელს უკავშირდება: სახელურსა და ზმნურს.

ქვეთავი 4.1 სახელური ფლექსია

ქართული ენისათვის დამახასიათებელ სახელურ პარადიგმას მსგავსი მორფოსინტაქსური კატეგორიები ახასიათებს, თუმცა წარმოებითი მოდელები ერთმანეთისგან განსხვავდება (გოგოლაშვილი და სხვ. 2011, პიუიტი 2005-1995, ტუიტი 1998-1984, შანიძე 1973 და ა.შ.). სახელურ პარადიგმას განეკუთვნება შემდეგი მეტყველების ნაწილები: არსებითი სახელი, ზედსართავი სახელი, რიცხვითი სახელი და ნაცვალსახელი. ბრუნების ტიპი დამოკიდებულია სახელური ფუძის ტიპზე, ხოლო ელემენტთა უმეტესობა სუფიქსაციის გზით ინარმოება, გამონაკლისს მხოლოდ ზედსართავი სახელის ხარისხის ფორმების ცირკუმფიქსული წარმოებითი მოდელი აჩვენებს.

სახელური პარადიგმის მოდელირებისას მნიშვნელოვანია მეტყველების ნაწილის შესაბამისი კლასების წარმოდგენა. არსებითი სახელების შემთხვევაში კლასი საზოგადო და საკუთარი სახელების ტიპებად იყოფა, ხოლო საკუთარი სახელები - პირადი და გეოგრაფიული სახელების ქვეტიპებად.

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
ტიპი		საზოგადო	+Prop
	ქვეტიპი	საკუთარი	+Com
		პირადი სახელი	+Name
		გეოგრაფიული სახელი	+Geog

ცხრილი 5 არსებითი სახელების ტიპების ტეგირება

რიცხვითი სახელების ტიპებად დაყოფას წინ უსწრებს შათი ფორმებად დაყოფის პრინციპი, რომელიც რიცხვითი სახელის სიტყვა-ფორმისა ან

სპეციალური სიმბოლოების სახით წარმოდგენას ეყრდნობა. შესაბამისად, რიცხვითი სახელი შემდეგ ფორმებს აჩვენებს:

1. სიტყვა-ფორმის სახით წარმოდგენილი რიცხვითი სახელი (33)

(33) *ორი, თორმეტი* და სხვ.

2. არაბული ციფრებით წარმოდგენილი რიცხვითი სახელი (34)

(34) *2, 12* და სხვ.

3. რომაული ციფრებით წარმოდგენილი რიცხვითი სახელი (35)

(35) *II, XII* და სხვ.

4. ქართული ანბანის ნიშნებით წარმოდგენილი რიცხვითი სახელი* (36)

(36) *ბ 'ორი', იბ 'თორეტი'* და სხვ.

სიტყვა-ფორმის, ისევე როგორც სპეციალური სიმბოლოების სახით წარმოდგენილი რიცხვითი სახელი შემდეგ ტიპებად იყოფა:

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სასრული პოზიციის (სპ) ტეგი
ტიპი		რაოდენობითი	+Card
		რიგობითი	+Ord
		ნილობითი	+Fract
		მიახლოებითი	+Approx
		განმეორებითი	+Rep

ცხრილი 6 რიცხვითი სახელების ტიპების ტეგირება

ტიპებად დაყოფის თვალსაზრისით ყველაზე მრავალრიცხოვანია ნაცვალსახელი, რომელიც შემდეგ ტიპებს აერთიანებს:

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
ტიპი		პირისა	+Pers
		ჩვენებითი	+Dem
		კუთვნილებითი	+Poss

* პირველი სამი ფორმა აქტიურად გამოიყენება თანამედროვე ქართულში, ხოლო მეოთხე ფორმა გამოიყენებოდა ძველსა და საშუალო ქართულში და მე-18 და მე-19 საუკუნის ტექსტებშიც შიგადაშიგ დასტურდებოდა და, შესაბამისად, უმეტესწილად ინერებოდა ასომთავრული და ნუსხური ანბანის სიმბოლოების გამოყენებით.

განუსაზღვრელობითი	+Indf
კითხვითი	+Int
მიმართებითი	+Rel
ურთიერთობითი	+Recp
უარყოფითი	+Neg
განსაზღვრებითი	+Det

ცხრილი 7 ნაცვალსახელთა ტიპების ტეგირება

მორფოლოგიური კატეგორიების უმეტესობა, რომელიც ზეგავლენას ახდენს სახელური პარადიგმის წარმოებაზე შემდეგია:

- ბრუნვა (სახელობითი, მოთხრობითი, მიცემითი, ნათესაობითი, მოქმედებითი, ვითარებითი და წოდებითი),
- რიცხვი (მხოლობითი და მრავლობითი),
- კლიტიკები (თანდებულები, ემფატიკური ხმოვანი, ნაწილაკები).

ზედსართავი სახელის შემთხვევაში ამას ემატება ხარისხის, ხოლო ნაცვალსახელის შემთხვევაში – პირის კატეგორია. სულიერება/უსულოების გრამატიკული კატეგორია შეესაბამება არსებით სახელებსა და ნაცვალსახელებს, თუმცა მხოლოდ წინადადების დონეზე ვლინდება სუბიექტის ზმნასთან შეთანხმებისას. სულიერი არსებითი სახელი ან ნაცვალსახელი ზმნას როგორც მხოლობით, ასევე მრავლობით რიცხვში ითანხმებს, ხოლო უსულო – მხოლობითში, მრავლობითში კი ზმნისგან მხოლობითი რიცხვის ფორმას ითხოვს.

არსებითი სახელის, რიცხვითი სახელისა და ნაცვალსახელის კატეგორიათა თანმიმდევრობის სქემა შემდეგნაირია:

ტიპი -> რიცხვი -> ბრუნვა -> თანდებული -> კლიტიკები

აღნიშნულ თანმიმდევრობას ზედსართავი სახელის შემთხვევაში ხარისხის კატეგორია ემატება:

ტიპი -> ხარისხი -> რიცხვი -> ბრუნვა -> თანდებული -> კლიტიკები

სლოტების რაოდენობა დამოკიდებულია მეტყველების ნაწილზე. არსებითი სახელის, რიცხვითი სახელისა და ნაცვალსახელის სქემაში

სლოტების მაქსიმალური რაოდენობა ცხრას შეადგენს და შემდეგი ერთეულებისაგან შედგება:

1. სახელური ძირი, რომელზეც დამოკიდებულია ბრუნების ტიპი, სხვაგვარად რომ ვთქვათ ის, თუ რა პროცესს დაექვემდებარება კონკრეტული სახელი: კვეცას თუ კუმშვას;
2. რიცხვის ნიშანი;
3. ბრუნვის ნიშანი;
4. ემფატიკური ხმოვანი;
5. თანდებული;
6. ემფატიკური ხმოვანი;
7. ნაწილაკი;
8. დამხმარე ზმნა;
9. არაპირდაპირი თქმის ნიშანი (37).

(37) ფრინველ-ფ-ის-ა-თვის-ა-ცა-ა-მეთქი და სხვ.

ფრინველ-SG-GEN-EMP-POST-EMP-PTCL-AUX-1.QUOTE:PRS.IND

ზედსართავი სახელის სქემაში სლოტების საერთო რაოდენობა 11 ერთეულს შეადგენს:

1. ხარისხის ნიშანი, რომელიც წარმოდგენილია ცირკუმფიქსის სახით, რომელიც იკავებს სტრუქტურული შაბლონის პირველსა და მესამე სლოტს;
2. ძირი, რომელზეც დამოკიდებულია ხარისხის წარმოება და ის თუ რა პროცესს დაექვემდებარება ესა თუ ის ზედსართავი სახელი (კვეცასა თუ კუმშვას);
3. ხარისხის ნიშანი;
4. რიცხვის ნიშანი;
5. ბრუნვის ნიშანი;
6. ემფატიკური ხმოვანი;
7. თანდებული;

8. ემფატიკური ხმოვანი;
 9. ნაწილაკი;
 10. დამხმარე ზმნა;
 11. არაპირდაპირი თქმის ნიშანი (38).
- (38) მო-თეთრ-ო-ღ-ს-ა-ვით-ა-ცა-ა-მეთქი და სხვ.

COMP-თეთრ-CIRC-SG-GEN-EMP-POST-EMP-PTCL-AUX-1.
 QUOT:PRS.IND

სლოტების მორფოტაქტიკა და მორფოლოგიური კატეგორიების თანმიმდევრობა ზეგავლენას ახდენს მორფოლოგიური ანალიზატორის საერთო დიზაინზე, ლექსიკონის ბაზაზე და გენერაციის წესებზე. სახელური პარადიგმის საწყის წერტილს უმეტესწილად ძირის საწყისი ასო წარმოადგენს და გენერაციის თვალსაზრისით საკმაოდ მარტივ სტრუქტურირებას ექვემდებარება. შედარებით სირთულეს ზედსართავი სახელის პარადიგმა წარმოადგენს, რომელიც ხარისხის ფორმების გენერირებისას აფიქსურ წარმოებას მიმართავს და, შედეგად, პარადიგმის ამოსავალ წერტილს არა ძირი, არამედ პრეფიქსი წარმოადგენს. აღნიშნული სირთულის დაძლევა ქსეროქსის კალკულუსის ტექნოლოგიის ფარგლებში დიაკრიტული ნიშნულების გამოყენებით ხდება.

ქვეთავი 4.2 ზმნური ფლექსია

ზმნური პარადიგმის დახასიათებისას მნიშვნელოვანია არა მხოლოდ მორფოსინტაქსური თვისებების ასახვა, არამედ მათი წარმოჩენა ტეგსების სახით. ტეგსების სირთულე დამოკიდებულია მომხმარებელზე ანუ იმაზე, თუ რის გამოვლენა სჭირდება კონკრეტული რესურსის მომხმარებელს. ამიტომაც მნიშვნელოვანია, ერთი მხრივ, არსებული ტეგსების შეჯერება, ხოლო, მეორე მხრივ, ტეგსების შემუშავება კონკრეტულ ლინგვისტურ თეორიაში არსებული თვისებებისა და ფუნქციების გათვალისწინებით.

ბოლო პერიოდში, კომპიუტერული მიდგომების ფარგლებში განიხილება ე.წ. თეორიაზე დამოუკიდებელი ტეგსების უპირატესობა, რაც მიზანშეწონილია სხვადასხვა ენაში არსებული სტრუქტურების შეჯერებისა და

შესწავლის მიზნით. თუმცა კონკრეტული ენის ლინგვისტური თავისებურების გათვალისწინებით თითქმის შეუძლებელია თეორიაზე დამოუკიდებლად ენობრივი თავისებურების სრულყოფილი აღწერა (ანოტირება), ნებისმიერი ყველაზე მარტივი ტეგირების სქემაც კი თეორიულ ვარაუდს ითვალისწინებს (ატველი 2008). შესაბამისად, ტეგსეტის შერჩევისას გასათვალისწინებელია ევროპულ ენათა ტეგსეტები როგორიცაა EAGLES (1996), Multext (2004) და სხვ., მაგრამ არანაკლებ მნიშვნელოვანია ქართული ენის შესაბამისი გრამატიკული თვისებების აღწერა ტეგსეტების დონეზე და მათი გამოყენება ტექსტის მორფოსინტაქსური ანოტირების პროცესში.

ზმნის მახასიათებლების საწყისი აღწერა ზმნის ტიპისა და ქვეტიპის განსაზღვრას გულისხმობს. ფუნქციურად ზმნა შეიძლება იყოს როგორც ძირითადი, ასევე დამხმარე. ძირითადი ზმნა, თავის მხრივ, აბსოლუტურ-სტატიკურ, რელატიურ-სტატიკურ, დინამიკურ და რელატიურ-დინამიკურ ზმნებად იყოფა და შემდეგნაირად აისახება:

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
ტიპი		ძირითადი	+Main
		დამხმარე	+Aux
	ქვეტიპი	აბსოლუტურ-სტატიკური	+AbsStat
		რელატიურ-სტატიკური	+RelStat
		დინამიკური	+Dyn
		რელატიურ-დინამიკური	+RelDyn

ცხრილი 8 ზმნის ტიპების ტეგირება

თანამედროვე ქართულში ზმნისწინიანი ფორმები მყოფად, წყვეტილსა და თურმეობითში სრულ ასპექტს აღნიშნავს, მაშინ როცა უზმნისწინო – უსრულს. ზმნისწინის პოზიციის გათვალისწინებით სასრული პოზიციის გარდამქმნელის ტეგი წინ უსწრებს ძირს და არ მოსდევს მას, თუმცა იმ შემთხვევაში, როცა ზმნისწინიანი ფორმა არ არის არც სრული და არც უსრული ასპექტის თვალსაზრისით, მაშინ ტეგებით მისი აღწერა არ ხდება.

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
ასპექტი		სრული	Ipfv+
		უსრული	Pfv+

ცხრილი 9 ასპექტის ტეგირება

გარდამავლობის თვალსაზრისით ზმნური ფორმა შეიძლება იყოს ა) გარდამავალი და ითხოვდეს პირდაპირ ობიექტს, ბ) გარდაუვალი ობიექტის გარეშე და, გ) ირიბ-გარდამავალი და ითხოვდეს ირიბ ობიექტს:

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
გარდამავლობა		გარდაუვალი	+Intrans
		ირიბ-გარდამავალი	+IndTrans
		გარდამავალი	+Trans

ცხრილი 10 გარდამავლობის ტეგირება

ზმნის გარდამავლობა დამოკიდებულია გვარისა და ვალენტობის ცვლილებაზე. დიონისე თრაკიელის მიხედვითი გვარის გრამატიკულ კატეგორიას ზოგჯერ დიათეზას უწოდებენ და ქართული ზმნის ფორმები განსხვავებული დიათეზის სამ ჯგუფში ნაწილდება. თითოეული დიათეზა მორფო-სინტაქსურ ერთიანობას წარმოადგენს და, შესაბამისად, მოიაზრებს გვარის შემდეგ ტიპებს: 1) აქტივს, რომელიც გრამატიკულ სუბიექტს ასახავს; 2) ავტოაქტივს, რომელიც ავტოაქტიურ სუბიექტს ასახავს; 3) ინაქტივს, რომელიც ობიექტს არ საჭიროებს; 4) პასივს, რომელიც ძირითადი ზმნის თემასა ან პაციენსს ასახავს და, 5) ავტოპასივს, სხვაგვარად რომ ვთქვათ, მედიოპასივს, რომელიც აღნიშნავს, რომ ქმედება აგენსის გარეშე ხორციელდება.

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
გვარი		I დიათეზა	+IDt
		II დიათეზა	+IIDt
		III დიათეზა	+IIIDt
	ქვეტიბი	აქტივი	+Act

ავტოაქტივი	+AutAct
ინაქტივი	+Inact
პასივი	+Pass
მედოპასივი	+MPass

ცხრილი II გვარის ტეგირება

დ. მელიქიშვილის (2001) მიხედვით I დიათეზა მოიაზრებს ავტოაქტივისა და აქტივის ფორმებს, II დიათეზა - ავტოპასივისა და პასივის ფორმებს, ხოლო III დიათეზა - ინაქტივის ფორმებს.

ზმნური გარდამავლობა უკავშირდება ქართული ზმნის პოლიპერსონალიზმს და სუბიექტისა და ობიექტის მნიშვნელობის ჩამოყალიბებაში მონაწილე სხვადასხვა ფუნქციას. აღნიშნული ფუნქციების აღწერა ქმედების ბენეფიციარის, პაციენსის, რეციპიენსის, მიმართულების და ა.შ. ჩვენების შესაძლებლობას იძლევა. აღნიშნულ სემანტიკურ ფუნქციებს ითავსებს ქცევის ნიშანი, თუმცა დ. მელიქიშვილი მათ როგორც კონკრეტული თვისებების მქონე სპეციალურ აფიქსებს განიხილავს, სხვაგვარად რომ ვთქვათ, სუბიექტისა და ობიექტის კორელაციის დამატებით ნიშნებს. ტეგირების პრინციპიც სუბიექტის და ობიექტის აღნიშნული თვისებების ჩვენებას ეფუძნება.

ატრიბუტი	დამ. ატრ.	მნიშვნელობა	სპ ტეგი
სუბიექტისა		პირველი პირის სუბიექტი, მხ. რიცხვი	+Subj1Sg
და		მეორე პირის სუბიექტი, მხ. რიცხვი	+Subj2Sg
ობიექტების,		მესამე პირის სუბიექტი, მხ. რიცხვი	+Subj3Sg
პირისა და		პირველი პირის სუბიექტი, მრ. რიცხვი	+Subj1Pl
რიცხვის		მეორე პირის სუბიექტი, მრ. რიცხვი	+Subj2Pl
კორელაცია		მესამე პირის სუბიექტი, მრ. რიცხვი	+Subj3Pl
		პირველი პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen1Sg

მეორე პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen2Sg
მესამე პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen3Sg
პირველი პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen1Pl
მეორე პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen2Pl
მესამე პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen3Pl
პირველი პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus1Sg
მეორე პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus2Sg
მესამე პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus3Sg
დანვრილებით იხ. დანართი 1.	

ცხრილი 12 სუბიექტის და ობიექტების კორელაციის, პირისა და რიცხვის
ტეგირება

დროის, ასპექტისა და კილოს მორფოლოგიური კატეგორიები წარმო-
დგენილია სერიების სახით და მათი ასახვა მწკრივების ჩვენების გზით
ხდება:

ატრიბუტი	დამ. ატრიბუტი	მნიშვნელობა	სპ ტეგი
დრო და კილო		ანმყო	+Pres
		უწყვეტელი (ნამყო უსრული)	+Imperf
		ანმყოს კავშირებითი	+PresSbj
		მყოფადი	+Fut
		ხოლმეობით-პირობითი	+FutCond
		მყოფადის კავშირებითი	+FutSbj

წყვეტილი	+Aor
მეორე კავშირებითი	+AorSbj
ბრძანებითი	+AorImp
პირველი რეზულტატივი/ თურმეობითი	+Res1
მეორე რეზულტატივი/თურმეობითი	+Res2
მესამე კავშირებითი	+PerfSbj

ცხრილი 13 დროისა და კილოს ტეგირება

ზმნური პარადიგმის ნაწილს საწყისი (+VerbalNoun) და მიმღეობა (+VerbalAdj) წარმოადგენს. როგორც ერთი, ასევე მეორეც სახელური პარადიგმის თვისებებს იზიარებს და არსებითი და ზედსართავი სახელების წარმოებით მოდელს შეესაბამება. ზედსართავი სახელის შემთხვევაში იმ მოდელს, რომელსაც ხარისხის ფორმები არ ეწარმოება. როგორც ერთი, ასევე მეორის ზმნური წარმომავლობა დერივაციული თვალსაზრისით მანარმოებელ ზმნისწინებში ვლინდება, და ზმნის ანწყოსა და მყოფადის ფორმებზე შესაბამისი აფიქსების დართვით იქმნება. წინადადების დონეზე კი მათი ფუნქცია არაფრით განსხვავდება სახელური შესიტყვების ფუნქციისგან.

ზმნური ფორმები დამოკიდებულია არგუმენტების სტრუქტურაზე და, შესაბამისად, კონსტრუქციის მარკირება განაპირობებს ზმნის დონეზე სუბიექტისა და ობიექტების ბრუნვების მონიშვნას:

- I სერიის მწკრივებთან – ნომინატიური კონსტრუქცია;
- II სერიის მწკრივებთან – ერგატიული კონსტრუქცია;
- III სერიის მწკრივებთან – დატიური კონსტრუქცია.

ამას ემატება სუბიექტის, პირდაპირი და ირიბი ობიექტების ტეგირების შესაძლებლობა:

ატრიბუტი	დამ. ატრ.	მნიშვნელობა	სპ ტეგი
სუბიექტისა		სუბიექტი სახელობითში	+<NomSubj>

და	სუბიექტი მოთხრობითში	+<ErgSubj>
ობიექტების	სუბიექტი მიცემითში	+<DatSubj>
ბრუნვები	სუბიექტი სახელობითში, ბენეფიციარი	+<NomSubjBen>
	სუბიექტი მოთხრობითში, ბენეფიციარი	+<ErgSubjBen>
	სუბიექტი მიცემითში, ბენეფიციარი	+<DatSubjBen>
	ობიექტი სახელობითში	+<NomObj>
	ობიექტი მიცემითში	+<DatObj>
	ობიექტი მიცემითში, ბენეფიციარი	+<DatObjBen>
	ობიექტი მიცემითში, რეციპიენსი	+<DatObjRec>
	ობიექტი მიცემითში, ლოკატივი	+<DatObjLoc>
	ობიექტი მიცემითში, პაციენსი	+<DatObjPat>
	ობიექტი ნათესაობითში	+<GenObj (for) >
	სუბიექტი სახელობითში, კაუზატორი	+<NomSubjCaus>
	სუბიექტი მოთხრობითში, კაუზატორი	+<ErgSubjCaus>
	სუბიექტი მიცემითში, კაუზატორი	+<DatSubjCaus>
	ობიექტი მიცემითში, კაუზაციის ადრესატი	+<DatObjRecCaus>

ცხრილი 14 სუბიექტის, პირდაპირი და ირიბი ობიექტების ბრუნვების ტეგირება

განხილული ტეგების საფუძველზე ნათელია, რომ გმნური პარადიგმის წარმოება დამოკიდებულია შემდეგ ძირითად მორფოლოგიურ კატეგორიებზე:

ა) დრო-კილო-ასპექტის სერიებზე, რომლებიც განაპირობებენ ბრუნვის მარკირებას, კავშირის დამყარებას ქმედების მონაწილეთა შორის, როგორიცაა აგენსი და პაციენსი და გრამატიკულ ურთიერთობას სუბიექტსა და ობიექტებს შორის გმნისწინების, სუბიექტ-ობიექტის კორელაციის (ქცევის) ნიშნებისა და თემატური სუფიქსების გამოყენებით;

- ბ) დიათეზაზე (გვარზე), რომელიც შეიძლება იყოს აქტიური, ავტოაქტიური, ინაქტიური, პასიური და ავტოპასიური;
- გ) პირიანობაზე, რომელიც წარმოგვიდგება ერთპირიანი, ორპირიანი და სამპირიანი ზმნების სახით⁷ და;
- დ) რიცხვზე.

სუბიექტ-ობიექტის კორელაციის ნიშნები (ქცევის კატეგორია) შეიძლება იყოს სათავისო, საარვისო და სასხვისო (მელიქიშვილი 2001, შანიძე 1973 და სხვ.) და ტრადიციულად განიხილება ზმნის ძირითად მორფოსინტაქსურ კატეგორიებს შორის. მნიშვნელობის დონეზე აღნიშნული კატეგორია მიუთითებს გრამატიკულ ურთიერთობაზე სუბიექტსა და ობიექტებს შორის და ამიტომაც მისი განხილვა ხდება, უშუალოდ, აღნიშნული ურთიერთობიდან გამომდინარე.

ქართულ ზმნას სხვადასხვა ტიპის მორფემა ახასიათებს, რაც, ერთი მხრივ, ბუნებრივია, აგლუტინაციური სტრუქტურების წარმოების თვალსაზრისით, ხოლო, მეორე მხრივ, ასევე, ფლექსიურ სტრუქტურებს ახასიათებს. შესაბამისად, ზმნური პარადიგმაც წარუვი სახით წარმოგვიდგება (ბოედერი 2005, ჰარისი 1981-2006, ჰიუიტი 1995 და სხვ.). ზმნის სლოტების მაქსიმალური რაოდენობა კი 9-დან 12-მდე მერყეობს და მოიცავს შემდეგს:

1. ზმნისწინი;
2. ზმნისწინი, რომელიც მო- ზმნისწინის სახით წარმოდგენილია და მიმართულებას აღნიშნავს;
3. პრეფიქსული პირის ნიშანი, რომელიც ზოგჯერ რიცხვის ნიშნის მნიშვნელობასაც ითავსებს;
4. ქცევის ნიშანი, განიხილება როგორც ობიექტის კორელაციის ნიშანი;
5. ძირი, რომელიც შეიძლება ერთი (39) ან ექვსი (40) თანხმოვნისგან ხმოვნით ან ხმოვნის გარეშე შემდგარი რიგით იყოს წარმოდგენილი. ძირი არ განაპირობებს ზმნების ფლექსიური კლასების მიხედვით განაწილებას, თუმცა მისი პოზიცია სლოტების თანმიმდევრობაში

⁷ მიქმიე ოთხპირიანი ზმნის ვარიანტი უნიფიცირების თვალსაზრისით არ განიხილება.

ინვეს ლემატიზაციის პრობლემას, რომელიც განსაკუთრებული სიმწვავეით ქართულ ლექსიკოგრაფიაში ვლინდება;

6. პასივის ნიშანი;
7. თემის ნიშანი;
8. კაუზაციის ნიშანი;
9. სავრცობი, რომელსაც ზოგჯერ როგორც იმპერფექტივის ნიშანს (მახარობლიძე 2018) განიხილავენ;
10. დროის ნიშანი;
11. პირის ნიშანი, რიცხვის ნიშანი, დამხმარე ზმნა;
12. არაპირდაპირი თქმის ნიშანი.

(39) ა-გ-ებ-ს და სხვ.

Vers-გ-TS-3SGSbj:PRS.IND

(40) გვ-ბრდღვნ-ი-ს და სხვ.

IPLObj-ბრდღვნ-TS-3SGSbj:PRS.IND

ზმნათა უმეტესობას შეუძლია ლექსიკურად განპირობებული ზმნის-წინების დართვა, თუმცა ზოგიერთ ზმნურ ძირს ზმნისწინების დართვა არ ახასიათებს. ზმნისწინების კლასიფიცირება შესაძლებელია როგორც ერთეულთა დახურული კლასისა, რომელიც გამოიყენება მყოფადის, წყვეტილისა და თურმეობითის სანარმოებლად. თუმცა დ. მელიქიშვილი (2001) დამატებით აღწერს ზმნისწინიან ფორმებს, რომლებიც პირველი სერიის ანმყოს, ანმყოს უწყვეტელისა და ანმყოს კავშირებითში გვხვდება, მაგ. დახტის და სხვ. ზმნისწინების რაოდენობა თანამედროვე ქართულში ოცდაერთს არ აღემატება, კერძოდ, მი-, მო-, ა-, ამო-, აღ-, აღმო-, უკუ-, და-, ჩა-, შე-, გა-, გან-, გადა- და სხვ. გარდა ამისა, ზმნისწინები გამოიყენება ასპექტის ჩვენების მიზნით (41), თუმცა ზოგიერთი ზმნისწინი ამ დანიშნულებას არ ასრულებს (42).

(41) და-ხატ-ავ-ს და სხვ.

PFV-ხატ-TS-FUT.3SGSBJ

(42) მი-დი-ს და სხვ.

prev-დი-PRES.3SGSBJ

ნებისმიერი მორფემა მოიცავს ინფორმაციას სიტყვის გრამატიკული ფუნქციის შესახებ. ზმნა ქართულში, ძირითადად, გრამატიკული ფუნქციის საჩვენებლად ბმულ მორფემებს იყენებს. წარმოების ტიპების თვალსაზრისით მნიშვნელოვანია შემდეგი:

1. აფიქსაცია, მაგ., ხატ-ავ-დ-ა და სხვ.

ხატ-TS-PAS-IMPERF.3SGSBJ

2. ფუძისეული ხმოვნის მონაცვლეობა, მაგ., გრეხ-ს, გრიხ-ა და სხვ.

გრეხს-PRES.3SGSBJ

PV-გრეხ-AOR.3SGSBJ

3. ძირის მონაცვლეობა, მაგ., ე-ტყვ-ი-ს, უ-თხრ-ა და სხვ.

prv-უბნ-TS-FUT.3SGSBJ

Vers-უბნ-AOR.3SGSBJ

აღსანიშნავია, რომ ზმნური ფორმებიდან ძირთან ყველაზე ახლოს მდგომ ერთეულს მეორე პირის მხოლოდობითი რიცხვის ანმყოს ფორმა წარმოადგენს და იმის გათვალისწინებით, რომ აღნიშნული კვლევა სასრული პოზიციის გარდამქმნელის სახით ქართული ენის მოდელირებას ემსახურება, მისი ერთ-ერთი ყველაზე მნიშვნელოვანი ნაწილი ამოსავალ ფორმათა, სხვაგვარად რომ ვთქვათ, ლემათა ინდექსის განსაზღვრა იყო.

კომპიუტერული ლექსიკონი ისეთ ფორმებს უნდა მოიცავდეს, რომ-ლებზეც დაყრდნობით შესაძლებელი იქნება როგორც სასუბიექტო, ასევე საობიექტო პარადიგმების გენერირება და ანალიზი. იმის გათვალისწინებით, რომ ქართულ ლექსიკონთა უმრავლესობა (ცოცხანიძე და სხვ. 2014, ჩიქობავა 1950-1964 და სხვ.) მიჰყვება ზმნური ამოსავალი ფორმის წარმოდგენის ნარეკ პრინციპებს და წარადგენს მას ზოგჯერ როგორც საწყისს, ზოგჯერ როგორც მესამე პირის მხოლოდობითი რიცხვის ანმყოს და/ან მყოფადის ფორმებს, ჩვენი ამოცანების გათვალისწინებით ამოსავალ ფორმად მიჩნეულია მესამე პირის მხოლოდობითი რიცხვის ანმყოს ან

მყოფადის ფორმები როგორც წარმოდგენილია არნ. ჩიქობავას ქართული ენის განმარტებით ლექსიკონში (1950-1964) და დ. მელიქიშვილის ზმნური ფორმების ინდექსში (2001).

ქართული ლექსიკოგრაფიის ლემატიზაციის პრობლემა (ლობჯანიძე 2019) სასრული პოზიციის გარდამქმნელის ლექსიკონზეც ახდენს ზეგავლენას და უკავშირდება ზმნისწინების, პირის ნიშნებისა და პირების კორელაციის (ქცევის ნიშნების) არსებობას, რომლებიც სტრუქტურის საწყის სლოტებს იკავებენ და განაპირობებენ გრძელმანძიდილიან დამოკიდებულებებს სლოტებს შორის. თუმცა ზმნისწინებისგან განსხვავებით, პირისა და ქცევის ნიშნები ემსახურება აგენსის, პაციენსისა და ბენეფიციარის ჩვენებას და, შესაბამისად, ურთიერთობას განსაზღვრავს სუბიექტსა და პირდაპირ და ირიბ ობიექტებს შორის.

ზმნის წარმოებითი მოდელის პრობლემატიკის შეჯამება შემდეგნაირად შეიძლება:

1. ქართულ ლექსიკონებში წარმოდგენილი ზმნის ამოსავალი ფორმების სიჭრელე, რაც ვლინდება ამოსავალი ფორმის ხან მესამე პირის მხოლოდობით რიცხვის ანმყოსა და მყოფადის, ხან საწყისის ფორმებით წარმოდგენით. მსგავსი პრობლემა ინფინიტივის არმქონე თითქმის ყველა ენაში დასტურდება. ზდრ. თანამედროვე ბერძნულ ენაში სალექსიკონო ერთეულის ამოსავალი ფორმა წარმოდგენილია პირველი პირის მხოლოდობით რიცხვის ანმყო დროის სახით;
2. ზმნის პოლიპერსონალიზმი, რომელიც განაპირობებს სუბიექტისა და ობიექტის ნიშნების სხვადასხვა აფიქსის სახით წარმოდგენას. აღნიშნული აფიქსები ზმნის სუბიექტს, პირდაპირ და ირიბ ობიექტებსა და ლოკატიურ და კაუზატიურ მნიშვნელობებს ასახავენ. უფრო ზოგადად, ზმნას შესწევს ერთიდან სამ არგუმენტამდე შეთანხმების უნარი, რაც სხვადასხვა მორფოლოგიური სტრუქტურების მქონე ზმნური ფორმების ლექსიკონებში ასახვას იწვევს;
3. ქართული ლექსიკოგრაფიის თვისობრივად სხვადასხვა მიდგომა ლემისადმი განაპირობებს ლემატიზაციის პრობლემას და საგრძნობლად ართულებს ნაბეჭდი და ელექტრონული ლექსიკონების საძიებო

სისტემებს და, შედეგად, ზეგავლენას ახდენს ენის მოდელირებაზეც.

ქვეთავი 4.3 ქსეროქსის სასრული პოზიციის საშუალებანი (*lexc* და *xfst*)

ბუნებრივი ენის დამუშავების ამოცანათა უმეტესობა როგორიცაა ტოკენიზაცია, მორფოლოგიური ანალიზი და სხვ. სასრული პოზიციის ტექნოლოგიის გამოყენებით ხორციელდება (ბისლი და კარტუნენი 2003, კოსკენიემი 1983 და სხვ.). სასრული პოზიციის გარდამქმნელები იქმნება რეგულარული გამოსახულებების გამოყენების გზით და, შესაბამისად, სასრული პოზიციის მეთოდოლოგიური ჩარჩოს განუყრელ ნაწილს სპეციალური პროგრამული უზრუნველყოფა წარმოადგენს, რომელიც სხვადასხვა ენობრივი მასალის დამუშავებას ემსახურება.

რეგულარული გამოსახულებების სასრული პოზიციის ავტომატების სახით (მოჰრი 1997, მოჰრი და სპროატი 1996, ოფლაბერი 1996 და სხვ.) რეალიზება კომპიუტერული ენათმეცნიერების ჭრილში სხვადასხვა ვებ-გვერდისა თუ აპლიკაციის სახით ხდება. აღნიშნული აპლიკაციების ნაწილი არაკომერციულ მიზნებს ემსახურება (პალდენი 2009 და სხვ., შმიდი 2005), ნაწილი კი - კომერციულს (ბისლი და კარტუნენი 2003).

აღნიშნული მიმართულების განვითარება ჯერ კიდევ 1980 წლიდან ქსეროქსის პალო ალტოს ცენტრმა (პარკ) დაიწყო და, ძირითადად, ფონოლოგიური წესების სასრული პოზიციის ავტომატების სახით წარმოდგენას ემსახურებოდა, შემდგომში ორდონიანი მორფოლოგიის სახით განვითარდა, რომელსაც კიმმო-ს ეძახდნენ (კოსკენიემი 1983). ამ ეტაპზე ქსეროქსის სასრული პოზიციის ტექნოლოგია კომერციული პროექტის შედეგია, რომელსაც საფუძვლად კიმმო-ს სისტემა უდევს.

აღნიშნული ტექნოლოგია წარმოადგენს კომპიუტერულ რესურსს, რომელიც არაა ენაზე დამოკიდებული. ეს გულისხმობს იმას, რომ მისი გამოყენებით შესაძლებელია სტრუქტურულად განსხვავებული ენობრივი მასალის აღწერა. ეს მიდგომა განსაკუთრებით კარგად მუშაობს აგლუტინაციური ენების შემთხვევაში, როგორიცაა ფინური (ლინდენი 2009), თურქული (ოფლაბერი 2018) და სხვ.

ქსეროქსის სასრული პოზიციის ტექნოლოგიის ორი ძირითადი მოდული როგორიცაა *lexc* და *xfst* დანვრილებით აღწერილია ბისლისა და კარტუნენის მიერ (2003), თუმცა ყურადღება უნდა გამახვილდეს აღნიშნული რესურსის რამდენიმე ასპექტზე:

- *lexc* არის მაღალი დონის დეკლარატიული პროგრამირების ენა, რომელიც ორიენტირებულია ორმიმართულებიანი მორფოტაქტიკის წარმოდგენაზე. შემდგომი დამუშავების მიზნით მას შეუძლია მომხმარებლის მიერ შექმნილი ლექსიკონის მიღება შემდეგი ძირითადი სინტაქსური წესის დაცვით:

ამოსავალი ლემა	კლასი ;
Lemma	class ;

ამოსავალ ლემას, სხვაგვარად რომ ვთქვათ, ლექსიკურ ერთეულს წარმოადგენს ძირი, ან ლექსიკონში ანდა სიტყვათა ინდექსში დადასტურებული სიტყვა-სტატიის სასათაურო ფორმა. ქართული ენის თავისებურების გათვალისწინებით, ამოსავალ ფორმას ფუძე წარმოადგენს, რომელსაც ფლექსიური აფიქსების სახით სხვადასხვა მორფემა ერთვის. სლოტების თანმიმდევრობის დაცვით, ერთი კლასი მეორეში გადადის, სანამ სტრუქტურის ბოლომდე არ მივა. შესაბამისად, თითოეულ მეტყველების ნაწილს კლასების თავისი რაოდენობა ახასიათებს. ორდონიანი მორფოლოგიის თავისებურების გათვალისწინებით, ლექსიკური ერთეულების ნაწილი ორი დონისგან შედგება: გრამატიკული (ზედა) და ზედაპირული (ქვედა) დონეებისგან, რაც პროგრამის დონეზე შემდეგნაირად ვლინდება:

პურ-ი: პური	N1 ;
გრამატიკული დონე : ზედაპირული დონე	N1 ;

გამოსახულება 2 გრამატიკული და ზედაპირული დონეები, *lexc*

გრამატიკული დონე შესაბამისი ტეგების გამოყენებით მორფოლოგიური ანალიზის ნაწილს უზრუნველყოფს, ხოლო ზედაპირული – სიტყვა-ფორმების გენერირებას სლოტების მორფოტაქტიკის გათვალისწინებით. სხვაგვარად რომ ვთქვათ, გრამატიკული დონე შეესაბამება

სასრული პოზიციის გარდამქმნელის ზედა დონეს, ხოლო ზედაპირული – სასრული პოზიციის გარდამქმნელის ქვედა დონეს.

ზედა დონე - პურ-ი+Noun+Com+Inanim+Sg+Nom

ქვედა დონე - პური

მორფოლოგიური თვისებების ამსახველი ტეგები, როგორიცაა +Noun (არსებითი სახელი), +Com (საზოგადო სახელი), +Inanim (უსულო), +Sg (მხოლოდითი რიცხვი) და +Nom (სახელობითი ბრუნვა) ზედა დონეზე ჩნდება და *lexc* სინტაქსის მრავალსიმბოლოიანი ელემენტების სახით ფაილის დასაწყისშივე ცხადდება:

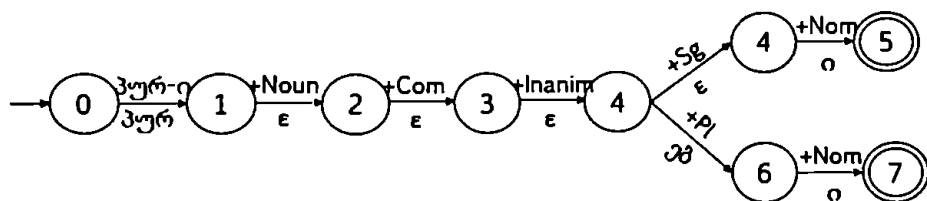
Multichar_Symbols

+Noun +Prop +Com +Anim +Inanim +Sg +Pl +Nom +Erg +Dat

+Gen +Ins +Advb +Voc +Emph +Post +Ptcl +Aux

გამოსახულება 3 მრავალნიშნაიანი სიმბოლოების ბლოკის მაგალითი

ზედა და ქვედა დონის ურთიერთქმედების უმარტივესი ქსელის წარმოდგენა შემდეგნაირად შეიძლება:



გრაფიკი 2 საზოგადო არსებითი სახელის სასრული პოზიციის მარტივი ქსელის მაგალითი

აღნიშნული ქსელის დამუშავებისას საბოლოო ჯამში შემდეგი უმარტივესი სიტყვა-ფორმები გენერირდება:

ზედა დონე: პურ-ი+Noun+Com+Anim+Sg+Nom

ქვედა დონე: პური

ზედა დონე: პურ-ი+Noun+Com+Anim+Pl+Nom

ქვედა დონე: პურები

ტეგების ჩამონათვალს დიაკრიტული ნიშნულების სია მოსდევს.

დიაკრიტული ნიშნულები გამოიყენება თვისების განსაზღვრისა და უნიფიცირების ოპერაციების განხორციელებისა და სასრული პოზიციის გარდამქმნელში სიტყვა-ფორმების ზედმეტი გენერირების თავიდან აცილების მიზნით. გარდა ამისა, მათი გამოყენება ხდება გრამატიკული ოპერაციების გამართივებისა და გრძელმანძიდილიან დამოკიდებულებათა კომპიუტერული რეალიზების დროს:

! ბრუნვები

@U.CASE.NOM@	@R.CASE.NOM@	@D.CASE.NOM@	@U.CASE.NOM1@
@D.CASE.NOM1@	@U.CASE.ERG@	@R.CASE.ERG@	@D.CASE.ERG@
@U.CASE.DAT@	@U.CASE.DAT1@	@U.CASE.DAT2@	@R.CASE.DAT@
@R.CASE.DAT1@	@R.CASE.DAT2@	@D.CASE.DAT1@	@U.CASE.GEN@
@R.CASE.GEN@	@U.CASE.GEN1@	@D.CASE.GEN1@	@U.CASE.INS@
@U.CASE.INS1@	@R.CASE.INS@	@R.CASE.INS1@	@U.CASE.ABL@
@U.CASE.ABL1@	@R.CASE.ABL1@	@U.CASE.VOC@	და ა.შ.

გამოსახულება 4 დიაკრიტული ნიშნულების ბლოკის მაგალითი

დიაკრიტული ნიშნულების ჩამონათვალის შემდეგ, *lexc*-ის სინტაქსი იმ ანკერების ჩამონათვალს ითხოვს, რომლებიც გამოიყენება შენაცვლების წესების ამოქმედების მიზნით:

! triggers ^S - მხოლოდითი რიცხვი, ^P - „ებიანი“ მრავლობითი, ^NT - „ნართანიანი“ მრავლობითი და ა.შ. ^S ^S1 ^S1 ^P ^NT და ა.შ.

გამოსახულება 5 ანკერების ბლოკის მაგალითი

ძახილის ნიშნის (!) ოპერატორი გვაძლევს შენიშვნების ჩამოთვლის შესაძლებლობას, მას შემდეგ წარმოდგენილი ერთეულები კოდის ნაწილს არ წარმოადგენს და მხოლოდ ინფორმაციულ ხასიათს ატარებს. შენაცვლების ყველა წესი ხორციელდება რეგულარული გამოსახულების სახით ანკერების *xfst*-ს მოდულიდან გამოძახებისა და ამოქმედების გზით.

- *xfst* არის რეგულარული გამოსახულების მიღებისა და სასრული პოზიციის ქსელის შექმნაზე ორიენტირებული რესურსი. აღნიშნული რესურსის გამოყენებით მომხმარებელს შეუძლია სასრული

პოზიციის ქსელების მართვა, შენაცვლების წესების ამოქმედება და კონკრეტული რეგულარული გამოსახულებების აღწერა. მას, ასევე, შეუძლია იმ ქსელების მართვა, რომლებიც იქმნება *lexc*-ში, კერძოდ, პირდაპირ *xfst*-დან შეიძლება *lexc*-ის მაკომპილირებელი პროგრამის გამოძახება. საჭიროების შემთხვევაში *xfst* აძლევს პირდაპირ წვდომას ქსეროქსის კალკულუსის ალგორითმებზე როგორიცაა გაერთიანება, გადაკვეთა და სხვ. *xfst*-ს ისევე, როგორც *lexc*-ის შესაძლებლობანი კარგად აღწერილია ბისლისა და კარტუნენტან (2003).

ქართული ენის სასრული პოზიციის მორფოლოგიური ანალიზატორის სტრუქტურული თვალსაზრისით შენაცვლების წესების რეგულარული გამოსახულების ფორმატი იცავს *xfst*-ს ძირითად სინტაქსს, როგორიცაა:

ა - > ბ ;

ა გადადის ბ-ში ;

ა - > ბ || დ _ ე ;

ა გადადის ბ-ში, როცა მას წინ უსწრებს დ- (მარცხენა კონტექსტი) და მოსდევს - ე (მარჯვენა კონტექსტი) ;

ა - > ბ || # _ ;

სიტყვის აბსოლუტურ დასაწყისში ა გადადის ბ-ში ;

ა - > ბ || _ # ;

სიტყვის აბსოლუტურ ბოლოს ა გადადის ბ-ში ;

აღნიშნული წესი ნიშნავს იმას, რომ ა შეესაბამება ქვედა ღონის რიგს ან რიგებს და ენაცვლება ბ-ს, მაგრამ მხოლოდ იმ შემთხვევაში, როცა აკმაყოფილებს მარცხენა და მარჯვენა კონტექსტის კონკრეტულ პირობებს. როგორც მარცხენა, ასევე მარჯვენა კონტექსტი პირობითია და იცვლება ენაში არსებულ კონკრეტულ წესებზე დაყდრნობით. მაგალითად, ქართულ ენაში სახელური ფუძე იკუმშება სონანტების წინ, როგორიცაა -რ-, -ლ-, -შ- და -ნ- . კუმშვისას ფუძიდან იკარგება -ა-, -ე- და -ო- ხმოვნები და ეს ხდება მხოლოდითი რიცხვის ნათესაობით, მოქმედებით და ვითარებით ბრუნვებში და -ებ-იანი მრავლობითის ყველა ბრუნვაში. თუმცა ზოგიერთ შემთხვევაში -ო- არ იკარგება, არამედ -ვ-ში გადადის. იმ ინფორმაციაზე

დაყრდნობით, რომელიც უკვე სლოტების დონეზე *lexc*-შია განერილი, შენაცვლების წესის ამოქმედება შესაძლებელია *xfst*-ს მოდულიდან შემდეგნაირად:

```
define R2 [ თ -> ვ || _ რ %^S2 ?* $["^G"] ~$["^NT"] .ო.
თ -> ვ || _ რ %^S2 ?* $["^I"] .ო. თ -> ვ || _ რ %^S2 ?*
$["^T"] .ო. თ -> ვ || _ რ %^P2 ?* ] ;
```

გამოსახულება 6 მინდორი (მინდვრის) შენაცვლების წესის მაგალითი

პროცენტის ნიშნის (%) ოპერატორი პუნქტუაციის ნებისმიერი ნიშნის წინ გამოიყენება მაშინ, როცა პუნქტუაციის ნიშანი ოპერატორის ფუნქციას არ ასრულებს. პროცენტის ნიშნის გარეშე პუნქტუაციის ნიშანი ოპერატორის ფუნქციას იძენს და გამოიყენება პროგრამული მიზნებისათვის. გასათვალისწინებელია, რომ *lexc*-ში ქვედა დონის ფორმების აღწერის დროს ^ პუნქტუაციის ნიშნის გამოყენება ხდება პროცენტის ნიშნის (%) ოპერატორის გარეშე.

გემოხსენებული გამოსახულების შენაცვლების წესი არის უფრო დიდი სკრიპტის ნაწილი, რომელშიც წინასწარ განსაზღვრულია თანხმოვნებიც, ხმოვნებიც და სონანტებიც. წესების კასკადის განხორციელების შემდეგ, ანკერების გაუქმება შემდეგი გზით ხდება:

```
define R12 [%^NT -> [] .ო. %^S -> [] .ო. %^P -> [] .ო.
%^E -> [] .ო. %^D -> [] .ო. %^G -> [] .ო. %^I -> [] .ო.
%^T -> [] და ა.შ. ] ;
```

გამოსახულება 7 ანკერების მოხსნის შენაცვლების წესის მაგალითი

სასრული პოზიციის გარდამქმნელის დაკავშირება ლექსიკონის გარდამქმნელთან ხდება ისე, რომ პირველი გარდამქმნელის (*lexc*-ში შექმნილი სალექსიკონო ფაილის) ქვედა დონე შეესაბამება მეორე გარდამქმნელის (*xfst*-ში შექმნილი შენაცვლების წესების) ზედა დონეს. დაკავშირების შედეგად მიღებული გარდამქმნელი ორმიმართულებიანია და, საბოლოო ჯამში, ქმნის როგორც ზედა, ასევე ქვედა დონეს ყოველგვარი შეცდომების გარეშე. სხვაგვარად რომ ვთქვათ, თუ ანალიზატორში შემავალი სიტყვა

არის 'მინდვრის', გარჩევისას ანალიზატორი გამავალ სიგნალს შემდეგი სახით აწარმოებს 'მინდორ-ი+Noun+Com+Inanim+Sg+Gen' და პირუკუ.

გრამატიკული ომონიმიის შემთხვევები მდიდარი მორფოლოგიის მქონე ენების შემთხვევაში ხშირია, რადგანაც სიტყვათა უმეტესობას ერთზე მეტი მორფოსინტაქსური ანალიზი ახასიათებს. ასე მაგალითად, 'და' სიტყვა-ფორმა შეიძლება განიხილებოდეს როგორც კავშირი და როგორც არსებითი სახელი. ამოსავალი ლემაც წარმოდგენილია არა ერთ, არამედ მორფოლოგიური ანალიზატორის ორ ლექსიკონში (არსებითი სახელისა და კავშირის) და, შედეგად, ქვედა დონე ერთმანეთს ემთხვევა, ხოლო ზედა დონე გარჩევის ორ ვარიანტს იძლევა.

და: და+Noun+Com+Anim+Sg+Nom

და: და+Conj+Coord

თავი 5

დასკვნები

რეგულარული გამოსახულებების სასრული პოზიციის გარდამქმნელების სახით წარმოდგენა აქტუალურია ბუნებრივი ენის დამუშავების თვალსაზრისით. ქართული ენის სტრუქტურაც იძლევა მასში მიმდინარე პროცესების სასრული პოზიციის გარდამქმნელების სახით წარმოდგენის საშუალებას და განაპირობებს ორი მოდულის შექმნას. პირველი მოდული ემსახურება ენის სალექსიკონო ერთეულების სტრუქტურირებას ენაში არსებული მორფოტაქტიკის გათვალისწინებით, ხოლო მეორე შენაცვლების წესების აღწერას კონკრეტული სალექსიკონო ერთეულების ცვლილებების ასახვის მიზნით.

ნაწილი 3

ქართული ენის სასრული პოზიციის

მორფოლოგიური ანალიზი

მორფოლოგიური ანალიზის ნებისმიერი ტიპი ეყრდნობა სიტყვის შიდა სტრუქტურას. შიდა სტრუქტურაში არსებული მორფემების ურთიერთ-

კავშირი განაპირობებს ენაში როგორც ღია, ასევე დახურული კლასების არსებობას. კავშირი, ნაწილაკი და შორისდებული მეტყველების ნაწილების დახურულ კლასს განეკუთვნება, რაც გულისხმობს იმას, რომ ლექსიკონში წარმოდგენილ ერთეულებს რაიმე ტიპის ფლექსია არ ახასიათებთ და საკმარისია შესაბამისი ტეგებით მათი მონიშვნა მორფოლოგიური ანალიზის კომპიუტერულად უზრუნველყოფის მიზნით. სახელური და ზმნური ფორმები კი ხასიათდება თავისებურებებით, რომლებსაც ენის მორფოტაქტიკური წესები და მორფოფონოლოგიური თვისებები განაპირობებს. აღნიშნული წესებისა და თვისებების აღწერა შემდგომი კომპიუტერული შესრულებით წინასწარ შერჩეულ თეორიულ ჩარჩოზეა დამოკიდებული.

ქართული ენის კომპიუტერულ მოდელირებას სხვადასხვა დროს სხვადასხვა ჯგუფი ცდილობდა, თუმცა გადანყვეტილებები, მიდგომები და რეალიზაცია განსხვავებული იყო. ქ. დათუკიშვილმა, ნ. ლოლაძემ და მ. ზაკალაშვილმა სხვადასხვა მორფემის შეკავშირების წესები აღწერეს და შაბლონური მოდელების სახით მორფოლოგიური პროცესორის კომპიუტერული მოდელის შესაქმნელად გამოიყენეს (დათუკიშვილი 2005).

ლ. მარგველანი ახდენდა რა სიტყვა-ფორმების ანალიზს და თანამედროვე ქართული ენის მართლწერის შემოწმებას, განსაკუთრებულ ყურადღებას მრავალფეროვანი აფიქსებით წარმოებულ სხვადასხვა მეტყველების ნაწილის შესაბამისს რთულ მოდელებზე ამახვილებდა. სისტემაც შედგებოდა ძირისა და ალგორითმისაგან, რომელიც აფიქსებსა და ძირებზე დაყრდნობით გრამატიკულად სწორ ფორმებს ქმნიდა (მარგველანი 1999-2001).

ო. გურევიჩმა (2006) და ო. კაპანაძემ (2009) თანამედროვე ქართული ენის მოდელირებისას სასრული პოზიციის ტექნოლოგია გამოიყენეს და, ძირითადად, მიჰყვებოდნენ ა. შანიძის (1953) თეორიას ქართული ენის გრამატიკული სტრუქტურის შესახებ. ო. კაპანაძე (2009) ჰყოფდა ქართულ ზმნურ მოდელს გარდამავალ, გარდაუვალ, საშუალ, ინვერსიულ და სტატიკურ ჯგუფებად, მაშინ როცა ო. გურევიჩი (2006) მიუხედავად იმისა, რომ აღნიშნა დ. მელიქიშვილის (2001) კლასიფიკაციის დანვრილებითი

ხასიათი, ზმნური პარადიგმის დამუშავებისას შეიქმნა ა. შანიძის თეორიაზე, თუმცა სცადა ურთიერთდამოკიდებულებათა გამოვლენა მყოფადსა და ხოლმეობითს ანდა წყვეტილსა და პირველ თურმეობითს შორის.

სრულმაშტაბიანი კომპიუტერული გრამატიკის შედგენის დროს პ. მოერერი (2007) ლექსიკურ-ფუნქციონალური გრამატიკის (ლფგ) თეორიული ჩარჩოს გათვალისწინებით, ასევე, სასრული პოზიციის კალკულუსს იყენებდა. თუმცა ზემოაღნიშნული მიდგომებისგან განსხვავებით მისი მოდელი კ. ჩხენკელის ლექსიკონს (1965) დაეყრდნო, რომელშიც ამოსავალ ლემას ზმნებისთვის და სხვა მეტყველების ნაწილებისთვის ძირი წარმოადგენს, ხოლო ლექსიკური დონის მორფოტაქტიკა წარმოდგენილია შესაბამისი აფიქსებით, რომლებიც, თავის მხრივ, კონკრეტულ ძირებთან ასოცირდება. აღნიშნული მიდგომის დადებით მხარეს ზმნის პირიანობის შესაძლებლობა წარმოადგენს, რადგანაც კ. ჩხენკელის ლექსიკონს აღნიშნული ინფორმაცია ახლავს.

ქართული ენის მოდელირებისადმი კონცეპტუალურად საპირისპირო მიდგომას ჯ. ანთიძის (2009-2010) მოდელი წარმოადგენს. აღნიშნული მოდელი, სასრული პოზიციის მიდგომისგან განსხვავებით, არა-დეტერმინისტულია. შესაბამისად, სრულყოფილი მორფოლოგიური ანალიზის უზრუნველყოფისათვის ხდება არა-დეტერმინისტური ალგორითმის გამოყენება, რომელიც მარცხნიდან-მარჯვნივ ღრმა ძიებას ბექ-ტრეკინგით ახორციელებს. ძირითად მეთოდოლოგიას, რომელსაც აღნიშნული სისტემა ეყრდნობა, თვისებათა სტრუქტურების განსაზღვრა წარმოადგენს. უკან დაბრუნების ალგორითმის გამოყენებით სისტემა უზრუნველყოფს მორფოლოგიურ ანალიზს, ხოლო კონტექსტზე დამოკიდებულ გრამატიკებზე დაყრდნობით – სინტაქსურს.

ქართული ენის ცენტრალურ პრობლემას ზმნური პარადიგმის გენერირება წარმოადგენს. ენაში მიმდინარე როგორც დერივაციული, ასევე ფლექსიური პროცესების თავისებურების გათვალისწინებით ზმნურ ძირზე დაყრდნობით ხდება ერთეულთა ძალიან დიდი რაოდენობის გენერირება. მოდელირების თვალსაზრისითაც ზემოხსენებული მიდგომების ფარგლებში სხვადასხვა ავტორი ზმნური პარადიგმის საწარმოებლად,

ძირითადად, მანიქისეულ მიდგომას ირჩევდა. მიდგომათა უმეტესობა დ. მელიქიშვილის სისტემას განიხილავდა მხოლოდ სპორადულად, მიუხედავად იმისა, რომ მორფოსინტაქსური თვალსაზრისით აღნიშნული სისტემა ზმნური სალექსიკონო ერთეულების სამ დიათეზად დაყოფის შესაძლებლობას იძლევა.

პირველი დიათეზის ზმნები შეიძლება იყოს ან აქტიური ან ავტო-აქტიური. თუ ზმნას გააჩნია პირდაპირი ობიექტი, იგი ყოველთვის გარდამავალია, ხოლო თუ მას გააჩნია ირიბი ობიექტი, იგი ყოველთვის ირიბ-გარდამავალია. ორი დანარჩენი ჯგუფი ზმნების პასივიზაციის შედეგს წარმოადგენს. პირველი დიათეზის ზმნის პირდაპირი ობიექტის სუბიექტად გადაქცევის შემთხვევაში, ზმნა მეორე დიათეზაში გადადის. ხოლო თუ პირველი ან მეორე დიათეზის ზმნის გარდაუვალი ობიექტი სუბიექტად გადაიქცევა, მაშინ ზმნა მესამე დიათეზაში გადადის. დიათეზა მსგავსი ფორმისა და ფუნქციის ზმნების ერთმანეთთან დაკავშირების შესაძლებლობას მოიზრებს.

ქართული ენის მოდელირების ამოსავალი ელემენტები, რომლებიც ზეგავლენას ახდენს სისტემის დიზაინზე და რომლებსაც ექცევა განსაკუთრებული ყურადღება შემდეგია:

1. სლოტების მორფოტაქტიკა და რაოდენობა, რომელიც აძლევს მორფემების ურთიერთკავშირისა და თანმიმდევრობის განსაზღვრის შესაძლებლობას;
2. მორფოფონოლოგიური ცვლილებანი, რომლებიც სასაზღვრო მორფემების მონაცვლეობას განაპირობებს და ზოგჯერ ძირშიც ვლინდება;
3. ლინგვისტური თეორია, რომელზეც დაყრდნობით ხორციელდება ქართული ენის მოდელირება ენაზე დამოუკიდებელი კომპიუტერული სამუალებების გამოყენებით;
4. სიტყვათა ინდექსი, სხვაგვარად რომ ვთქვათ, ელექტრონული ლექსიკონები ან სიტყვათა სიები, რომლებიც ზემოხსენებულ პარამეტრებზე დაყრდნობით ამოსავალი ლემის დადგენის შესაძლებლობას იძლევა. მესამე ნაწილი ოთხ ძირითად თავად იყოფა. რომლებსაც წინ უძღვის

შესავალი ქართული ბუნებრივი ენის დამუშავების მიმოხილვითა და სისტემის საერთო დიზაინის აღწერით; 2. მეორე თავი ეხება ქართული ენის მორფოლოგიურ ანალიზატორსა და გენერატორს, რომელიც შეიქმნა სასრული პოზიციის კომპიუტერული საშუალებების გამოყენებით; 3. მესამე თავი ეთმობა ტოკენიზატორის მიმოხილვას და, 4. მეოთხე თავით ხდება ტესტირების კომპონენტების აღწერა შედეგების მიმოხილვით და, 5. მეხუთე თავი კი აჯამებს მესამე ნაწილში განხილულ ინფორმაციას.

თავი 1

შესავალი

ბუნებრივი ენის დამუშავებისას მორფოლოგიური პროცესების აღწერა განსაკუთრებით მნიშვნელოვანია აგლუტინაციური მორფოლოგიურად რთული ენების შემთხვევაში. მათი განხილვა და აღქმა კომპიუტერული სისტემის მიერ განაპირობებს სხვადასხვა დონის ამოცანების გადაწყვეტას როგორცაა ტეგირება, პარსინგი, სამანქანო თარგმანი წესებზე ორიენტირებული სისტემების შემთხვევაში და ა.შ. მორფოლოგიური ანალიზის შედეგად ვლინდება სხვადასხვა დონის გრამატიკული ომონიმია (მეტყველების ნაწილების ტეგირების დროს, სხვადასხვა მეტყველების ნაწილის შესაბამისი მორფემების გამოვლენის დროს და ა.შ.).

ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის შიდა დიზაინი შეესაბამება სასრული პოზიციის ორდონიანი მორფოლოგიის მიდგომებს (ბისლი და კარტუნენი 2003) და ემსახურება სხვადასხვა ტიპის ანალიზის განხორციელებას როგორც ზედა (სიტყვა-ფორმის გენერირების), ასევე ქვედა (ამოსავალი ლემისა და ლინგვისტური ანოტირების) დონეზე. თანამედროვე ქართული ტექსტის დამუშავებისას ვლინდება ისეთი თავისებურებანი, როგორცაა შემოკლებანი (მაგ. *ნატომ, გაერომ* და სხვ.), სხვადასხვა პუნქტუაციის ნიშნის გამოყენებით ციფრებად გაწერილი სიტყვები, რომლებიც იბრუნვის (მაგ. 12-ის, 10-მა და სხვ.), ნასესხობის გზით ენაში დამკვიდრებული სიტყვები, რომლებიც, მიუხედავად უცხოური წარმომავლობისა, მთლიანად ექვემდებარება ქართული ენის მორფოლოგიურ სტრუქტურას (მაგ. *ა-ლაიქ-ებ-ს, ა-შეარ-*

ებ-ს და სხვ.) და ა.შ. შესაბამისად, მსგავსი ერთეულების ამოცნობაც ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის ნაწილს წარმოადგენს.

შესწავლილი მასალის საფუძველზე და ტესტირების შედეგად გამოიკვეთა ორი პრობლემური მიმართულება, რომლის გადანწყვეტა სამომავლო პერსპექტივას წარმოადგენს, კერძოდ:

1. მანქანის მიერ ნაკითხვადი ფორმატის დაცვით თანამედროვე ქართული ტექსტის ელექტრონული სახით დამახინჯებული წარმოდგენა. აღნიშნულ პრობლემას იწვევს სხვადასხვა ტიპის ორთოგრაფიული შეცდომების სიჭარბე, მათ შორის, უნიკოდის კონსორციუმის მიერ მთავრული ასოების სტანდარტიზაციით გამოწვეული შეცდომები⁸. მსგავსი ტიპის ელექტრონულ ტექსტებში ასოები შეიძლება იყოს გამოტოვებული ე.წ. ბეჭდვის შეცდომა, ხოლო წინადადება ან სიტყვა შეიძლება იწყებოდეს დიდი (მთავრული) ასოებით. აღნიშნული პრობლემა განსაკუთრებული სიხშირით ჩნდება კლავიატურიდან ასოების აკრეფის დროს ანდროიდების და ზოგიერთ სხვა ტიპის სისტემაში, რაც ნეგატიურ ზეგავლენას ახდენს ქართულ ორთოგრაფიაზე.

2. ძველი ქართული ტექსტის ტოკენიზაცია, რომელიც ითხოვს ღრმა ფონოტაქტიკური სტრუქტურის განერას სიტყვის სასრული პოზიციის გამოვლენის მიზნით. აღნიშნული პრობლემა გამოწვეულია იმით, რომ ძველ ქართულ ტექსტებში ინტერვალი ხშირ შემთხვევაში ბუნებრივი გამყოფის ფუნქციას არ ასრულებდა და პუნქტუაციის ნიშნებიც იმ სახით, რომლითაც გამოიყენება თანამედროვე ქართულში – არ გამოიყენებოდა. ამას ემატება დაქარაგმებული ერთეულების სიჭარბე. მსგავს პრობლემას აწყდება ევროპულ ხელნაწერებზე მომუშავე თითქმის ყველა ჯგუფი და მსგავს შემთხვევებს *scriptura continua*-ს უწოდებენ. აღნიშნული პრობლემა ნეგატიურ ზეგავლენას ახდენს სიტყვათა შორის ინტერვალის გამოვლენაზე და ტექსტში შემავალი სიტყვა-ფორმების განსაზღვრაზე.

⁸ მხედრული ანბანი უნიკოდის კონსორციუმის მიერ მოთავსებულია იმავე სტანდარტიზებულ სქემაში, რაც ასომთავრული, იხ. დიაპაზონი 10A0–10FF. ზუსტური წარმოდგენს აღნიშნული სქემის დანამატს, იხ. დიაპაზონი 2D00–2D2F. და ბოლო ცვლილება ამ მიმართულებით 2016 წელს განხორციელდა, როცა შემოხსენებულ სქემებს შეემატა დიაპაზონი 1C90–1CBF.

მორფოლოგიური ანალიზატორი და გენერატორი

სიტყვის მორფოლოგიური ანალიზის წარმოდგენა შესაძლებელია ტეგების განსაზღვრული თანმიმდევრობის სახით. ტეგები კი, თავის მხრივ, მორფემებში ასახულ თვისებათა ერთობლიობას წარმოადგენს. თანამედროვე ქართული ენისათვის დამახასიათებელი სახელური პარადიგმის სტანდარტული ფორმა შეიძლება შემდეგნაირად წარმოჩინდეს:

ტოტბე: ტოტ-ი+Noun+Com+Inanim+Sg+Dat+Post (on)

ამოსავალ ფორმას ძირი წარმოადგენს, რომელიც მიიღება ანალიზატორის სალექსიკონო ნაწილში მითითებულ ლემაზე დაყრდნობით: ტოტ-ი. აღნიშნული სიტყვა-ფორმის ამოსავალი ლემა შეესაბამება მეტყველების კონკრეტულ ნაწილს შესაბამისი ტეგების დართვით. ტეგების დონეზე წარმოდგენილი ინფორმაცია შემდეგნაირად ნაწილდება:

- ძირთან ასოცირებული ტეგები, როგორიცაა: ა) მეტყველების ნაწილი (+Noun), ბ) არსებითი სახელის ტიპი (+Com), გ) არსებითი სახელის სულიერება/უსულოების კატეგორია (+Inanim);
- რიცხვის ამსახველი ნულოვანი მორფის სლოტი, რომელიც მხოლოდ ქვედა დონეზე დასტურდება და მხოლოდით რიცხვს (+Sg) ასახავს;
- თანდებულთან ასოცირებული ინფორმაცია შემდეგია: ა) ბრუნვა (+Dat) და, ბ) თანდებულის ტიპი (+Post (on)).

შესაბამისად, მნიშვნელოვანია იმისი განსაზღვრა, თუ რის საფუძველზე და როგორ ხდება აღნიშნული სიტყვაფორმის გენერირება როგორც ზედა, ასევე ქვედა დონეზე და რის საფუძველზე ხდება გენერაციის შემდგომი მორფოსინტაქსური ანალიზი.

ქვეთავი 2.1 ლემატიზაცია

პარადიგმის ამოსავალი ფორმის დადგენის პროცესს ლემატიზაციას უწოდებენ. უზოგადესი სახით, პარადიგმის გენერაციის მიზნით ლემატიზაცია ეხება ელექტრონულ ან ნაბეჭდ ლექსიკონში სიტყვა-სტატიის სატიტულო ფორმის განსაზღვრისა და შემდგომ მისი გამოყენების შესაძლებლო-

ბას. გასათვალისწინებელია, რომ სიტყვათა უმეტესობა რეგულარული პარადიგმული წარმოებით ხასიათდება, თუმცა ზოგიერთ სიტყვა-ფორმას წარმოების ირეგულარული მოდელი ახასიათებს. ირეგულარული მოდელებიც და მორფოტაქტიკური თვალსაზრისით ძირის საწყისი ასოების მეოთხე ან მეხუთე სლოტში წარმოდგენას სხვადასხვა პრობლემას იწვევს ენის გენერირების თვალსაზრისით. ქართულ ენაშიც ამოსავალი სიტყვა-ფორმის დადგენა პრობლემურია.

მორფოსინტაქსური ანოტირების ჩარჩოს (2012) შესაბამისად, ლემა არის ფლექსიური ფორმების მანარმოებელი კლასების საწყისი ერთეული, რომელიც სხვა დანარჩენი ფორმებისაგან მხოლოდ სიტყვაცვლილებითი მორფოლოგიით განსხვავდება. ევროპულ ენებში ლემის ფუნქციას სახელური ფორმის შემთხვევაში, ძირითადად, მხოლოდითი რიცხვის მამრობითი სქესის ფორმა (იმ შემთხვევაში თუ ენაში არსებობს გრამატიკული სქესის კატეგორია) ასრულებს, ხოლო ზმნური ფორმის შემთხვევაში – ინფინიტივი. იმ შემთხვევაში, თუ ენობრივი მოდელი ნაკლებია, მაშინ მხოლოდითი რიცხვის ფორმას მრავლობითის ფორმა ენაცვლება, ხოლო ინფინიტივს – ზმნის მესამე პირის ფორმა როგორც ეს, მაგალითად, ხდება არაბული ენის შემთხვევაში; ან ზმნის პირველი პირის ფორმა, როგორც ეს, მაგალითად, თანამედროვე ბერძნული ენის შემთხვევაში ხდება.

თანამედროვე ქართული ენის შემთხვევაში, სახელური პარადიგმის ლემას მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმა წარმოადგენს, ზმნას კი ინფინიტივი არ გააჩნია და არ არსებობს ზმნური პარადიგმის ლემის განსაზღვრის მკაფიოდ განერილი წესი. თუ მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმა სახელური პარადიგმის ამოსავალი ლემის ფუნქციით უნიფიცირებულია და ლექსიკონთა უმრავლესობაში დასტურდება, ზმნური პარადიგმის ამოსავალი ლემა სხვადასხვანაირად წარმოგვიდგება: ა) საწყისის (მასდარის) ფორმით, ბ) აბსტრაქტული ძირის ფორმით, და გ) მესამე პირის ფორმით ანმცოსა ან მცოფადში.

ყველა აღნიშნულ მიდგომას თავისი ლოგიკური საფუძველი გააჩნია, მათი აღწერა თავისუფლად შეიძლება სლოტების დაკავშირებისა და ენაში მიმდინარე მორფოსინტაქსური ურთიერთობის განსაზღვრის გზით,

როგორც გარდამავალი და გარდაუვალი ზმნებისგან საწყისის წარმოების თვალსაზრისით, ასევე პირუკუ. თუმცა ასეთი მრავალფეროვნება, ერთი მხრივ, ლექსიკონების მომხმარებლებს აბნევს და მათგან ითხოვს ქართული გრამატიკის საფუძვლიან ცოდნას, ხოლო, მეორე მხრივ, ავნებს ქართული ლექსიკოგრაფიის განვითარებას.

პირველი მიდგომა გულისხმობს საწყისის (მასდარის) ფორმაზე დაყრდნობით ზმნური პარადიგმის წარმოებას (ცოცანიძე და სხვ. 2014). მასდარი ზოგჯერ განიხილება როგორც ზმნური პარადიგმის ინფინიტივი იმ მოსაზრებით, რომ აბსტრაქტული ძირის მიღება საწყისიდან ბევრად უფრო მარტივია, ვიდრე ზმნური ფორმიდან (გიპერტი 2016 და სხვ.). აღნიშნულ მოსაზრებას ლოგიკური დასაბუთებაც აქვს. ამ მხრივ, საყურადღებოა საწყისის მორფოლოგიური სტრუქტურა, რომელიც სახელისა და ზმნის მორფოლოგიურ კატეგორიებს ითავსებს და, რაც ბევრად უფრო მნიშვნელოვანია, სინტაქსური ფუნქცია, რომელიც განაპირობებს საწყისის წინადადებაში ყოველთვის სახელისთვის დამახასიათებელი როლების შესრულებას. გარდა ამისა, საწყისის სიტყვანარმოებითი მოდელებიც უნიფიცირებული არ არის, შდრ.

(43) ა-ბ-ამ-ს -> ბ-მ-ა -> და-ბ-მ-ა და სხვ.

prv-ბ-TS-PRES.3SGSBJ

ბმა-TS-NOM

PV-ბმა-TS-NOM

(44) ხატ-ავ-ს -> ხატ-ვ-ა -> მო-ხატ-ვ-ა და სხვ.

ხატ-TS-PRES.3SGSBJ

ხატვა-TS-NOM

PV-ხატვა-TS-NOM

საწყისი სხვადასხვა სუფიქსის დართვით იწარმოება როგორიცაა მაგ. -ნ- (თხოვ-ს -> თხოვ-ნ-ა და სხვ.), -ომ- (დგ-ა-ს -> დგ-ომ-ა და სხვ.) და სხვ. ზმნების ის მოდელები, რომლებიც სერიების მიუხედავად ლოგიკური სუბიექტისგან მიცემით ბრუნვას ითხოვს, ხოლო ლოგიკური ობიექტისგან - სახელობითს, გამართულ წარმოებას აჩვენებს, მაგრამ ზმნათა ნაწილი

საწყისებს საერთოდ არ აწარმოებს, რაც გამორიცხავს საწყისების ამოსავალი ლემის სახით ლექსიკონში წარმოდგენის შესაძლებლობას.

მეორე მიდგომა პირველისგან იმით განსხვავდება, რომ წარმოდგენს აბსტრაქტულ ზმნურ ძირს შესაბამისი პარადიგმის მითითებით (ჩხენკელი 1965). მსგავსი ლემებით შედგენილი ლექსიკონის მთავარი დანიშნულებაა ლინგვისტური სტრუქტურების ჩვენება, რაც მიზანშეწონილია, თუ ლექსიკონის მომხმარებელი არის ლინგვისტი ანდა ენის გრამატიკული სტრუქტურის მცოდნე სპეციალისტი. მსგავსი მიდგომა, ასევე კომპიუტერული მოდელირების შესაძლებლობასაც აადვილებს, სირთულე იქმნება იმ შემთხვევაში, თუ ამოსავალი ლემა შესაბამისი მახასიათებლებით ლექსიკონში წარმოდგენილი არ არის.

მესამე მიდგომას ლექსიკონთა უმრავლესობა (რეიფილდი 2006, ჩიქობავა 1950-1964 და სხვ.) იზიარებს და მოიცავს მინიმალურ ინფორმაციას ზოგიერთი კატეგორიის შესახებ, როგორიცაა ქცევა, კაუზაცია და სხვ. მსგავსი ტიპის ლექსიკონებში თანაბარ დონეზე წარმოდგენილია როგორც ზმნური ფორმები, ასევე საწყისებიც, რაც საბოლოო ჯამში ქმნის როგორც ზმნური, ასევე საწყისისა და მიმღეობის ანალიზატორის სალექსიკონო ნაწილის შემუშავებასა და შევსების შესაძლებლობას.

სხვაობა აღნიშნულ მიდგომებს შორის ართულებს ქართული ენის სტანდარტიზაციას უნიფიცირებული მოდელირების უზრუნველყოფის მიზნით, რადგანაც, კონკრეტული სამუშაო ჯგუფი, თავისი ამოცანიდან გამომდინარე, კონკრეტულ ვარიანტს ირჩევს. შერჩეული ვარიანტი კი, როგორც წესი, მორფოლოგიური ანალიზატორის სალექსიკონო ნაწილში აისახება ხოლმე.

მორფოლოგიური ანალიზატორის ლექსიკონი შეიქმნა *lexc*-ის გამოყენებით და შედგება ორი ნაწილისაგან: ა) ლემებისაგან, რომლებიც წარმოდგენილია კონკრეტული მეტყველების ნაწილის შესაბამისი ლექსიკონების სახით და, ბ) დამატებითი კლასებისაგან. ლემა ამ შემთხვევაში წარმოდგენილია სიტყვის მონიშნავი ფორმის სახით, სხვაგვარად რომ ვთქვათ, ძირის ან სასათაურო სიტყვის იმ სახით, რომლითაც შეტანილია ნაბეჭდ ან ელექტრონულ ლექსიკონში ანდა სიტყვათა სიაში.

სახელური პარადიგმის ლემას მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმა წარმოადგენს, ხოლო ზმნური პარადიგმისა – ზმნის მესამე პირის ფორმა ანმყოსა ან მყოფადში. ზმნის მესამე პირის ფორმა შერჩეულია პირობითად მოდელირებისათვის გამოყენებული ლექსიკონების ფორმატიდან გამომდინარე, თუმცა, უშუალოდ, გენერაციისათვის ხდება მეორე პირის ფორმის გამოყენება, რომელიც ყველაზე უფრო ახლოსაა კონკრეტული ზმნური პარადიგმის აბსტრაქტულ ძირთან.

ლემათა ბლოკს საფუძვლად დაედო არნ. ჩიქობავას ქართული ენის განმარტებითი ლექსიკონი (1950-1964) და დ. მელიქიშვილის ზმნის ინდექსი (2014). შესაბამისად, სანყის ლექსიკონში წარმოდგენილი მასალა სახელის შემთხვევაში 78000 ერთეულს შეადგენდა, ხოლო ზმნის შემთხვევაში – 85000 ერთეულს. შემდგომ მასალა შეივსო კორპუსზე ტესტირების პროცესში მოპოვებული მასალით.

ქვეთავი 2.2 დიაკრიტული ნიშნულები

როგორც სახელური, ასევე ზმნური მორფოლოგია გენერირდება */exc*-ში შემდგომი მონაცვლეობით. ლექსიკონის სტრუქტურის ამოსავალ წერტილს ლემა წარმოადგენს, თუმცა სლოტების თანმიმდევრობა განაპირობებს კონკრეტული სლოტის სხვა სლოტებთან დაკავშირებას და შეზღუდვას აწესებს დანარჩენი სლოტების კონკრეტულ პირობებში გამოყენებაზე. ამ მხრივ, ქართული ენის მორფოლოგიისათვის დამახასიათებელია სლოტებს შორის გრძელმანძილიანი დამოკიდებულებანი, რომლებიც როგორც ცირკუმფიქსული კავშირების დონეზე ვლინდება, ასევე სხვა შემთხვევებშიც.

გრძელმანძილიან დამოკიდებულებათა ასახვის მიზნით ქსეროქსის კალკულუსი მიმართავს სპეციალური ნიშნულების გამოყენებას, რომლებსაც **დიაკრიტულ ნიშნულებს** (flag diacritics) უწოდებენ. დიაკრიტული ნიშნული მრავალსიმბოლოიან ელემენტს წარმოადგენს, თუმცა გამოიყენება მხოლოდ იმ შემთხვევაში, როცა სისტემას სჭირდება გრძელმანძილიანი კავშირების დამყარება. კომპიუტერული სისტემა აღნიშნულ სიმბოლოებს როგორც თვისებათა უნიფიცირების შესაძლებ-

ლობას აღიქვამს და დიაკრიტულ ნიშნულებზე დაყრდნობით ზღუდავს ქსელში წარმოქმნილ ბილიკთა რაოდენობას არასასურველი ფორმების გენერირებისგან თავდაცვის მიზნით.

ქსელები, რომლებიც საჭიროებენ დიაკრიტული ნიშნულების გამოყენებას, ტიპურად რთულ დამოკიდებულებებს მოიცავს. მსგავს ქსელებში ხდება ირეგულურალი ბილიკების დიდი რაოდენობის გენერირებაც და ამოცნობაც. დიაკრიტული ნიშნულების გამოყენება აძლევს ირეგულურალი ბილიკების ბლოკირების შესაძლებლობას და უზრუნველყოფს გარდამქმნელის სიმცირეს ზომის თვალსაზრისით.

ყველა მეტყველების ნაწილს თავისი დიაკრიტული ნიშნული ახასიათებს, მათი რაოდენობაც განსხვავებულია. ანკერებისაგან განსხვავებით, რომლების ამოქმედება *xfst*-ს მოდულიდან ხდება, დიაკრიტული ნიშნულები *lexc* სინტაქსის განუყრელ ნაწილს წარმოადგენს.

დიაკრიტული ნიშნულების სინტაქსი შემდეგ შაბლონებს მოიცავს:

```
ოპერატორი . თვისება . მნიშვნელობა
@operator.feature.value @
ოპერატორი.თვისება
@operator.feature@
```

არსებობს დიაკრიტული ნიშნულების წინასწარ შერჩეულ ქმედებათა თანმიმდევრობა და კონკრეტული ოპერატორები, რომლებიც დანვრით აღწერილია ბისლისა და კარტუნენტან (2003). ქართული ენის მორფოლოგიურ ანალიზატორში გამოყენებულ ოპერატორებს შორის აღსანიშნავია მხოლოდ შემდეგი:

- დადებითი თვისებების ოპერატორი (P)

```
დადებითი . თვისება . მნიშვნელობა
@P.feature value@
```

როცა ხდება დადებითი თვისებების ოპერატორის ამოქმედება, მითითებული თვისების მნიშვნელობა პირდაპირ ან ხელახლა ენიჭება

ნაჩვენებ მნიშვნელობას. აღნიშნული მინიჭება არასოდეს იწვევს უკუ-გდებას, ანალიზისა ან გენერაციის წარუმატებელ დასრულებას ან ახალი გადაწყვეტილებების ძებნას. R ოპერატორი გულისხმობს თვისების მნიშვნელობის POSITIVE SETTING-ს.

- აუცილებელი შემონმების ოპერატორი (R)

აუცილებელი.თვისება.მნიშვნელობა
@R.feature.value@

აუცილებელი შემონმებისას ტარდება ტესტი და თუ ტესტი წარმატებით სრულდება, თვისება ენიჭება მნიშვნელობას. ხოლო თუ ტესტი წარუმატებელია, მაშინ ხდება ბილიკის ბლოკირება და სისტემა სხვა შესაძლებლობების მოსაძებნად უკან ბრუნდება. R ოპერატორი გულისხმობს შემონმების აუცილებლობას - REQUIRE TEST-ს.

- ამკრძალავი შემონმების ოპერატორი (D)

ამკრძალავი.თვისება.მნიშვნელობა
@D.feature.value@

ამკრძალავი შემონმებისას, ასევე, ტარდება ტესტი, რომელიც წარმატებით მხოლოდ იმ შემთხვევაში სრულდება, თუ თვისება ნეიტრალურია ან მნიშვნელობის შეუსაბამო. D ოპერატორი ამკრძალავის შემონმებას - DISALLOW TEST-ს გულისხმობს.

- გაერთიანების ოპერატორი (U)

გაერთიანება.თვისება.მნიშვნელობა
@U.feature.value@

თუ თვისება ნეიტრალურია, მაშინ ოპერატორის ამოქმედებისას, თვისებას უბრალოდ ენიჭება მნიშვნელობა. თუ თვისება ნეიტრალური არ არის, მაშინ ტესტი წარმატებით მხოლოდ იმ შემთხვევაში სრულდება, როცა მნიშვნელობა თვისების მიმდინარე მნიშვნელობას შეესაბამება. U ოპერატორი გულისხმობს გაერთიანებას - UNIFICATION-ს.

სისტემა ოთხ ტესტს მიმართავს: დადებით შემონმებას, აუცილებელ შემონმებას, ამკრძალავ შემონმებასა და გაერთიანებას. ქართული ენის მორფოლოგიური ანალიზატორის ყველა ლექსიკონი შეიცავს მრავალ-სიმბოლოიანი ელემენტების ბლოკს, დიაკრიტული ნიშნულების სიას, ანკერების სიასა და კლასების ჩამონათვალს. სტანდარტული ლექსიკონის ნაწილი შემდეგნაირად გამოიყურება:

Multichar_Symbols

+Verb +Aux +Intr +Trans +Act +AutAct +IndAct +Pass
+RelStat +AbsStat +Din +RelDin Pfv+ Ipfv+ +Subj1Sg +Sub-
j2Sg +Subj3Sg +Subj1Pl +Subj2Pl +Subj3Pl +SubjBen1Sg
+SubjBen2Sg და ა.შ.

! Subject agreement

@U.SUBJSG.1@ @U.SUBJSG.2@ @U.SUBJSG.3@ @R.SUBJSG.1@
@R.SUBJSG.2@ @R.SUBJSG.3@ @U.SUBJPL.1@ @U.SUBJPL.2@ @U.
SUBJPL.3@ @R.SUBJPL.1@ @R.SUBJPL.2@ @R.SUBJPL.3@ და ა.შ.

! Preverbs

@P.PREV.MI@ @P.PREV.A@ @P.PREV.DA@ @P.PREV.CA@ @P.PREV.
SE@ @P.PREV.GA@ @P.PREV.GADA@ @P.PREV.TSA@ @P.PREV.MO@
@P.PREV.AMO@ @P.PREV.CAMO@ და ა.შ.

! Screeves

@U.PRES.0@ @R.PRES.0@ @U.PRES.A@ @R.PRES.A@ @U.PRES.I@
@R.PRES.I@ @U.PRES.U@ @R.PRES.U@ @U.FUT.0@ @R.FUT.0@ @U.
FUT.A@ @R.FUT.A@ @U.FUT.I@ @R.FUT.I@ და ა.შ.

! Causation

@U.CAUS.0@ @U.CAUS.IN@ @U.CAUS.EVIN@ @R.CAUS.0@ @R.
CAUS.IN@ @R.CAUS.EVIN@ @D.CAUS.0@ @D.CAUS.IN@ @D.CAUS.
EVIN@

! Thematic Suffixes

@U.TS.0@ @U.TS.EB@ @U.TS.AV@ @U.TS.IV@ @U.TS.UV@
@R.TS.0@ @R.TS.EB@ @R.TS.AV@ @R.TS.IV@ @R.TS.UV@ @D.TS.0@
@D.TS.EB@ @D.TS.AV@ @D.TS.IV@ @D.TS.UV@

! Valency

@U.VAL.II@ @R.VAL.II@ @D.VAL.II@ @U.VAL.III@ @R.VAL.
III@ @D.VAL.III@

^Imp ^PresSbj ^Fut ^FutSbj ^FutCond ^Aor ^AorSbj ^AorImp
^Perf ^PluPerf ^PerfSbj და ა.შ.

LEXICON Root

620 ;

და ა.შ.

გამოსახულება 8 მძნური ლექსიკონის შემადგენელი ნაწილების შემოკლებული მაგალითი

ღია კლასების საწარმოებლად კონკრეტული მეტყველების ნაწილის თავისებურების გათვალისწინებით, სალექსიკონო ბლოკები დიაკრიტული ნიშნულების სხვადასხვა რაოდენობას მოიცავს. გამოყენებული ნიშნულების რაოდენობა მერყეობს 61-დან (ნაცვალსახელების შემთხვევაში) 167-მდე (მძნების შემთხვევაში). სისტემაში არსებული დიაკრიტული ნიშნულების საერთო რაოდენობა 314-ს უდრის. დიაკრიტული ნიშნულები მხოლოდ *lexc*-ის სალექსიკონო ნაწილში გამოიყენება; ზეგავლენას ახდენს კონკრეტული მორფემების ურთიერთობაზე და უზრუნველყოფს სლოტების გამართულ მორფოტაქტიკას, მაგრამ მორფოფონოლოგიურ ცვლილებებს არ უზრუნველყოფს. მორფოფონოლოგიური ცვლილებების განხორციელების მიზნით ხდება შენაცვლების წესების ამოქმედება *lexc*-ის სალექსიკონო ნაწილში ჩასმულ ანკერებზე დაყრდნობით.

ქვეთავი 2.3 ანკერები

ანალიზატორში მორფოტაქტიკის გათვალისწინება იმდენად სავალდებულოა, რამდენადაც სასრული პოზიციის კონკრეტული გარდამქმნელი საჭიროებს შენაცვლების წესების ამოქმედებას. ამიტომაც გამგრძელებელ კლასებს *lexc* სინტაქსის დაცვით ახლავს ინფორმაცია ანკერების

შესახებ, მაგ.

+Sg: ^S0 Cs1 ;
+Pl: ^Pებ Cs1 ;

აღნიშნულ ორ ანკერზე დაყრდნობით ხდება, ერთი მხრივ, სახელური პარადიგმის მხოლოდითი და მრავლობითი რიცხვის გენერირება შემდგომი კუმშვისა და კვეცის უზრუნველყოფის მიზნით, ხოლო, მეორე მხრივ, კონკრეტული მორფემის საზღვრების დადგენა.

ქართული ენისთვის დამახასიათებელი სახელური და ზმნური ფლექსიის თავისებურების გათვალისწინებით, სახელური და ზმნური პარადიგმის სეგმენტირება გულისხმობს სხვადასხვა გრამატიკული თვისების კონკატენაციური მორფემების სახით წარმოდგენის შესაძლებლობას და მათ შემდგომ დამუშავებას სლოტების სახით. როგორც სახელური, ასევე ზმნური ფლექსიაც, ხშირად ფუძის ცვლილებას ექვემდებარება, მათ შორის ძირისა (45) და/ან ძირისეული ხმოვნის (46) მონაცვლეობას. ამავდროულად, როცა ხდება აფიქსების ძირზე დართვა, ხდება კონკრეტული მორფოფონოლოგიური პროცესების ამოქმედება, როგორიცაა მეტათეზისი (47), ხმოვნის მონაცვლეობა (48), ხმოვნის დაკარგვა (49) და სხვ.

(45) ე-ტყვ-ი-ს და სხვ.

prv-უბნ-TS-FUT.3SGSBJ

Ipfv+უბნებ-ა+V+Trans+Act+Fut+Subj 3Sg+Obj Rec 3+Obj 3

(46) მო-დრიკ-ა და სხვ.

PV-დრეკ-AOR.3SGSBJ

Pfv+დრეკ-ს+V+Din+Trans+Act+Aor+Subj 3Sg+Obj 3

(47) ხნ-ავ-ს -> ხვნა

ხნ-TS-PRES.3SGSBJ

ხნ-TS-FUT.3SGSBJ

{ Ipfv+ხნავ-ს+V+IDt+Din+Trans+Act+Pres+Obj3+Subj3Sg
| Ipfv+ხნავ-ს+V+IDt+Din+Trans+Act+Fut+Obj3+Subj3Sg }

ხვნა<TS>-NOM

Ipfv+ხნავ-ს+VN+Sg+Nom

(48) ნიორ-ი -> ნივრ-ის და სხვ.

ნიორ-GEN

ნიორ-ი+Noun+Com+Sg+Gen

(49) და-ქერ-ი და სხვ.

PV-ქრ-AOR.2SGSBJ

PV-ქრ-AOR.IMP.2SGSBJ

{ Pfv+ქრი-ს+V+IDt+Din+Trans+Act+Aor+Subj2Sg+Obj3
| Pfv+ქრი-ს+V+IDt+Din+Trans+Act+AorImp+Subj2Sg+Obj3 }

ხმოვნების მონაცვლეობა, ბოლოკიდური ცვლილებები და სხვა პროცესები, როგორც ზემოთ აღინიშნა, ხორციელდება საპეციალური ანკერების საშუალებით (^S, ^P და სხვ.) ე.წ. მონაცვლეობის წესების ამუშავებაზე ორიენტირებული მომხმარებლებისთვის ხელმიუწვდომელი ტევებით, რომლებიც განაპირობს სისტემურ გენერირებას. სამაგალითოდ ავიღოთ -ებ-იანი მრავლობითის ხმოვანფუძიანი სახელების წარმოება (მაგ. მამა -> მამ-ებ-ი და სხვ.), კვეცის უზრუნველყოფის მიზნით საჭიროა ბოლოკიდური -ა- ხმოვნის ჩამოცილება. შესაბამისად, მონაცვლეობის წესიც (50) შემდეგნაირია:

ა-> [] || _ % ^P

(50) მამ-ებ-ი და სხვ.

მამა-PL-NOM

მამა+Noun+Com+Pl+Nom:მამა^Pები

ანკერების დანიშნულება და რაოდენობა ისევე, როგორც დიაკრიტიკული ნიშნულებისა, სხვადასხვა მეტყველების ნაწილთან სხვადასხვანაირია და დამოკიდებულია ბრუნების ან უღლების ტიპზე. მაგალითად, მხოლოდ არსებითი სახელის ბრუნება სხვადასხვა ტიპს იძლევა, რომელიც სხვადასხვა შედეგზე გასვლის წინაპირობას წარმოადგენს. არსებითი სახელის ლექსიკონი 22 კლასად იყოფა და ორ ძირითად ნაწილს მოიცავს: სამოგადო და საკუთარი სახელებისა. სამოგადო არსებითი სახელი მოიცავს 9 განსხვავებული გრამატიკული თავისებურების მქონე კლასს შესაბამისი კომპიუტერული რეალიზაციით, ხოლო საკუთარი – 13.

შესაბამისად, არსებითი სახელის გრამატიკული თავისებურების ანალიზატორში რეალიზაციის უზრუნველყოფის მიზნით, სისტემაში გათვალისწინებულია 31 ანკერი. მათი ამოქმედება შუალედურ დონეზე შემდეგნაირად ხდება:

ქვედა დონე	შუალედური დონე	ზედა დონე
პურ-ი+Noun+Com+Inanim+Sg+Erg	პური^S^Eმა	პურმა
პურ-ი+Noun+Com+Inanim+Sg+Erg+IndS peech1	პური^S^Eმა- მეთქი	პურმა- მეთქი
პურ-ი+Noun+Com+Inanim+Sg+Erg	პური^S^Eმან	პურმან
პურ-ი+Noun+Com+Inanim+Sg+Nom	პური^S^Nი	პური
პურ-ი+Noun+Com+Inanim+Sg+Nom+Aux	პური^S^Nია	პურია
პურ-ი+Noun+Com+Inanim+Sg+Nom+Aux+ IndSpeech1	პური^S^Nია- მეთქი	პურია- მეთქი

პურ-ი+Noun+Com+Inanim+Sg+Nom+Post (like)	პური^S^Nივით	პურივით
პურ-ი+Noun+Com+Inanim+Sg+Nom+Post (like) +IndSpeech1	პური^S^Nივით'- მეთქი	პურივით- მეთქი
პურ-ი+Noun+Com+Inanim+Sg+Nom+IndS peech1	პური^S^ი-მეთქი	პური-მეთქი
პურ-ი+Noun+Com+Inanim+Pl+Erg	პური^Pებ^Eმა	პურებმა
პურ-ი+Noun+Com+Inanim+Pl+Erg+IndS peech1	პური^Pებ^Eმა- მეთქი	პურებმა- მეთქი
პურ-ი+Noun+Com+Inanim+Pl+Erg	პური^Pებ^Eმან	პურებმან
პურ-ი+Noun+Com+Inanim+Pl+Erg	პური^Pებ^Nი	პურები
პურ-ი+Noun+Com+Inanim+Pl+Nom	პური^Pებ^Nია	პურები
პურ-ი+Noun+Com+Inanim+Pl+Nom+Aux	პური^Pებ^Nია-	პურები.
პურ-ი+Noun+Com+Inanim+Pl+Nom+Aux+	მეთქი	პურები-
IndSpeech1	პური^Pებ^Nივით	მეთქი
პურ-ი+Noun+Com+Inanim+Pl+Nom+Post (like)	პური^Pებ^Nივით-	პურებივით-
პურ-ი+Noun+Com+Inanim+Pl+Nom+Post (like) +IndSpeech1	პური^Pებ^Nი-	მეთქი
პურ-ი+Noun+Com+Inanim+Pl+Nom+IndS peech1	მეთქი	პურები-
და ა.შ.		მეთქი

*ცხრილი 15 სახელის ქვედა და ზედა დონის შემოკლებული მაგალითი
შუალედური დონის ჩვენებით*

ანკერების ნახვა არც ქვედა და არც ზედა დონეზე არ ხდება. როგორც მორფოლოგიური ანალიზის, ასევე გენერაციის შედეგად მიღებული სიტყვა-ფორმიდან ხდება ანკერების ამოღება სკრიპტის ფაილში შესაბამისი წესის განწერის გზით:

```
define R12 [ %^N -> [] .o. %^NT -> [] და სხვ. ] ;
```

აღნიშნული სკრიპტის საშუალებით ხდება სისტემაში გამოყენებული ანკერების გაუქმება ზედა დონის სრულყოფილი ფორმების წარმოდგენის მიზნით.

ქვეთავი 2.4 სისტემის საერთო დიზაინი

როგორც ზემოთ აღინიშნა, ქართული ენის მორფოლოგიური ანალიზატორი და გენერატორი შეიქმნა ორი რესურსის გამოყენებით როგორიცაა *xfst* და *lexc*. ენის მორფოტაქტიკა წარმოდგენილია ლექსიკონის სახით. სასრული პოზიციის ავტომატებისა და გარდამქმნელების განსაზღვრის მიზნით ლექსიკონის კომპილატორში გამოიყენება *lexc* - მაღალი დონის დეკლარაციული ენა, ხოლო *xfst* არის რეგულარული გამოსახულებების კომპილატორი, რომელიც გამოიყენება *lexc*-ში აღწერილი ქსელების (ავტომატებისა და გარდამქმნელების) მართვის მიზნით. ლინგვისტური თვალსაზრისით, *lexc* - ენის მორფოტაქტიკასთან ასოცირდება, ხოლო *xfst* - ფონოლოგიური და ორთოგრაფიული შენაცვლების წესებთან.

მთლიანად მორფოლოგიური ანალიზატორი შედგება სალექსიკონო გარდამქმნელებისა და შენაცვლების წესების გარდამქმნელებისაგან (იხ. გრაფიკი 3). მეტყველების ნაწილებს თავისი ლექსიკონები ახასიათებთ. აღნიშნულ ლექსიკონებზე დაყრდნობით ხდება მორფოლოგიური პროცესების ამოქმედება. ღია კლასების შემთხვევაში, როგორიცაა სახელური და ზმნური პარადიგმები სისტემური შენაცვლების წესები კოდირებულია *xfst*-ს სკრიპტების სახით.

3.2.4.1 ლექსიკონები და შენაცვლების წესები

ენობრივი მოდელის სრულყოფილი ანოტირებისათვის, სისტემაში, ასევე, წარმოდგენილია მეტყველების ნაწილების სხვა კლასები, როგორიცაა კავშირის, ნაწილაკის, ზმნიზედის, შორისდებულის, თანდებულის, პუნქტუაციისა და შემოკლებანის. დახურული კლასის ზოგიერთი ერთეული სიტყვანარმოებითი მოდელის ფარგლებში იწარმოება ღია კლასის ერთეულებზე დარყდნობით, ასე მაგალითად ზმნიზედის ზოგიერთი ტიპი იწარმოება სახელურ პარადიგმაზე -ად სუფიქსის დართვით:

LEXICON AdjPrefixes

0:უ@U.DEGREE.SUP@ AdjRoots ;

0:მ@U.DEGREE.COM@ AdjRoots ;

LEXICON AdjRoots

გემრიელ-ი:გემრიელი A1 ;

LEXICON A1

+Adj+Posit: ^S0@D.DEGREE.SUP@@D.DEGREE.COM@ Nmbr_1 ;

+Adj+Sup: ^სეს@R.DEGREE.SUP@ Nmbr_1;

+Adj+Comp: ^სო@R.DEGREE.COM@ Nmbr_2 ;

+Adv: ^სად@D.DEGREE.SUP@@D.DEGREE.COM@ # ;

+Adv: ^სად@D.DEGREE.SUP@@D.DEGREE.COM@ Post ;

და ა.შ.

გამოსახულება 9 მზნიშედის წარმოების დერივაციული მოდელის მაგალითი

მორფოლოგიური გარდამქმნელის დიზაინის მთავარ ამოცანას სისტემური სიმარტივე წარმოადგენდა, რომლის საფუძველზე, ლინგვისტი ან ენის მცოდნე სპეციალისტი ზედმეტი პრობლემების გარეშე შეძლებდა ლექსიკონში ამოსავალი ფორმის ჩასმას შემდგომი გენერაციის უზრუნველყოფის მიზნით. აღნიშნული ასპექტი ძალიან მნიშვნელოვანია ფორმების გენერაციის შემდგომი მორფოსინტაქსური ამოცნობის თვალსაზრისით.

ლექსიკონში მოთავსებული ლექსიკური ერთეულები ნაწილდება კლასების მიხედვით. კლასებში განაწილება ხდება აფიქსური წარმოებისა და შემდგომი შიდა ცვლილებების საფუძველზე. ყველა ერთეული კლასების მიხედვით ნაწილდება ისე, რომ ქვედა (ლინგვისტურ) დონეზე მიჰყვება ამოსავალი ლემისა და მორფოლოგიური ტეგების სქემას. ზედა დონის ცვლილებები კი ხორციელდება ლექსიკონზე დაყრდნობით შენაცვლების წესების განხორციელებით.

შენაცვლების წესები წარმოდგენილია სპეციალური რეგულარული

გამოსახულებების სახით და ყველა წესი ზეგავლენას ახდენს ზედა დონის წარმოებაზე. თუ ლექსიკონში წარმოდგენილი ანკერების შეჯერების გზით ხდება კონკრეტული პირობების უზრუნველყოფა, სკრიპტში წარმოდგენილი თანმიმდევრობა აძლევს მორფოფონოლოგიური წესის განხორციელებას. ახალი სიტყვისთვის ყველა შესაძლო ფლექსიური ფორმის გენერაციის უზრუნველყოფის მიზნით, ანალიზატორის ლექსიკონს ემატება კონკრეტული სიტყვის ამოსავალი ლემა შესაბამისი კლასის მითითებით.

ქართული სახელური პარადიგმის მოდელი მორფოფონოლოგიური თვალსაზრისით საზიაროა მეტყველების სხვა ნაწილებს შორის: არსებითი სახელი, ზედსართავი სახელი, რიცხვითი სახელი და ნაცვალსახელი. მორფოტაქტიკური მოდელიც მცირე ცვლილებებით საკმაოდ რეგულარულია, თუმცა განასხვავებს სლოტების სხვადასხვა რაოდენობასა და გრძელმანძილიან დამოკიდებულებებს სხვადასხვა მორფოლოგიური კატეგორიის აღწერის თვალსაზრისით (მაგ. ხარისხის ფორმები ზედსართავი სახელების შემთხვევაში და სხვ.). როგორც სახელური, ასევე ზმნური მოდელის მორფოლოგია წარმოდგენილია ორი ტიპის რესურსის სახით:

1. *lexc*-ში წარმოდგენილია ლექსიკონი დიაკრიტული ნიშნულებით გრძელმანძილიანი დამოკიდებულებების ასახვის მიზნით და ანკერებით მორფოფონოლოგიური შენაცვლების წესების *xfst*-დან ამოქმედების მიზნით;
2. *xfst*-ში წარმოდგენილია სკრიპტი შენაცვლების წესებით, რომლებზეც დაყრდნობით ხდება მორფოფონოლოგიური ცვლილებების უზრუნველყოფა სწორი ფორმების გენერირების მიზნით.

სალექსიკონო ფაილებში წარმოდგენილი კლასების რაოდენობა ყველა მეტყველების ნაწილს თავისი აქვს:

- არსებითი სახელის სალექსიკონო ფაილი 22 კლასისგან შედგება, რომელიც საზოგადო და საკუთარ სახელების ორ დიდ ჯგუფად იყოფა: 9 კლასი - საზოგადო სახელებს შეესაბამება, ხოლო დანარჩენი - საკუთარს;

- ზედსართავი სახელის ლექსიკონი ორ ნაწილად იყოფა, რომლებსაც ეწარმოებათ ხარისხის ფორმები და რომლებსაც – არ ეწარმოებათ. ვითარებით ზედსართავებს ხარისხი ეწარმოებათ, მიმართებით ზედსართავებს – არა. კლასების რაოდენობა კი 10-ს შეადგენს: 5 კლასი – ვითარებით ზედსართავებს აწარმოებს და 5 კლასი – მიმართებითებს;
- ნაცვალსახელის ლექსიკონი ყველაზე დიდია მასში წარმოდგენილი ჯგუფების თვალსაზრისით, რაც საბოლოო ჯამში ზეგავლენას ახდენს კლასების რაოდენობაზეც. ლექსიკონში წარმოდგენილი ჯგუფების ყველაზე დიდი პრობლემა უკავშირდება იმას, რომ თითოეული ჯგუფი თავისთავში მოიაზრებს მორფოლოგიური თვალსაზრისით განსხვავებული ბრუნების ტიპების გაერთიანებას. მაგალითად, განსაზღვრებითი ნაცვალსახელის წარმოებითი მოდელი მოიცავს არა მხოლოდ ხმოვანფუძიან ან თანხმოვანფუძიან ბრუნებას, არამედ თანხმოვანზე ფუძეგათავებულ ერთეულებსაც; ბრუნების ტიპიც ითვალისწინებს კვეცას, კუმშვას ან არც ერთს (მაგ. *თვით, თვითეული, ყველა, მავანი* და სხვ.) და შედეგად წარმოებითი კლასების რაოდენობაც, ერთი მხრივ, დამოკიდებულია ჯგუფის ტიპზე, ხოლო, მეორე მხრივ, უშუალოდ ფლექსიური წარმოების თავისებურებაზე. შედეგად ნაცვალსახელის ლექსიკონი 49 კლასს მოიცავს;
- რიცხვითისახელის ლექსიკონი წარმოებითი მოდელის თვალსაზრისით მიჰყვება სახელური პარადიგმის წარმოების წესებს. თუმცა რიცხვითი სახელის დერივაციული მოდელი ეყრდნობა რა ოცობით წარმოებით სისტემას, განაპირობებს მისი გენერაციის შესაძლებლობას. შესაბამისად, ლექსიკონიც გულისხმობს არა მხოლოდ ფლექსიურ, არამედ დერივაციულ წარმოებასაც და მოიაზრებს რაოდენობითი (*ათ-ერთ-მეტ-ი, თ-ერთ-მეტ-ი* და სხვ.), რიგობითი (*მე-ორ-ე, მე-სამ-ე* და სხვ.), წილობითი (*მე-ორ-ედ-ი, მე-სამ-ედ-ი* და სხვ.), ოდნაობითი (*ორ-ი-ოდ-ე, სამ-ი-ოდ-ე* და სხვ.) და ჯერობითი (*ორ-ჯერ, სამ-ჯერ* და სხვ.) რიცხვითი სახელების ჯგუფებს. ლექსიკონში წარმოდგენილი კლასების რაოდენობა კი, საერთო ჯამში ოთხს არ აღემატება;

- წარმოებითი კლასების თვალსაზრისით ზმნის მანარმოებელი მოდელი ყველაზე რთულია. იგი მოიცავს 66-ს რეგულარული და 1-ს ირეგულარული წარმოების ფლექსიურ კლასებს. ირეგულარული წარმოების კლასებს განეკუთვნება სუპლექტიური ზმნები, როგორიცაა მწკრივების მიხედვით ძირმონაცვლე ზმნები (მაგ. არის - იქნება - იყო - ყოფილა და სხვ.), რიცხვში მონაცვლე ზმნები (მაგ. მის - სხედან და სხვ.), სულიერება/უსულოების მიხედვით ძირმონაცვლე ზმნები (მაგ. აქვს - ჰყავს და სხვ.). ყველა აღნიშნული კლასი შერჩეულია დ. მელიქიშვილის (2001) სისტემის მიხედვით, ყველა კლასი სხვადასხვა ზმნურ ძირს იყენებს, თუმცა შიგადაშიგ სხვადასხვა პარადიგმაში იდენტური ზმნები ჩნდება. ზემოხსენებული ზმნური პარადიგმების კომპიუტერული რეალიზაცია კი გულისხმობს 80 კლასის არსებობას, რომლებიც ფარავენ როგორც რეგულარულ, ასევე ირეგულარულ შემთხვევებს.
- შედეგად, ზმნისწინების გარეშე ზმნურ ძირზე დაყრდნობით წარმოებული ფორმების მინიმალური რაოდენობა 54-ს უდრის, ხოლო მაქსიმალური - 1076-ს. კლასების უმეტესობა ორ ნაწილად იყოფა მ-ტიპის და ვ-ტიპის პარადიგმებად და ავლენს გრძელმანძილიან დამოკიდებულებათა სიმრავლეს, რომელიც გენერირდება დიაკრიტულ ნიშნულებზე დაყრდნობით. და თუ, ევროპულ ენათა უმეტესობაში ზმნა ზმნური ძირით იწყება, ქართულ ენაში ზმნური პარადიგმა პრეფიქსებით იწყება, რომლებიც თანამედროვე ქართულში ძირამდე ოთხ სლოტს იკავებს, ხოლო ძველ ქართულში - ექვსს;
- ღია კლასებს, ასევე, განეკუთვნება საწყისი და მიმღეობა. აღნიშნული ერთეულების შექმნაზე ორიენტირებული გარდამქმნელები მიმღეობის შემთხვევაში - ზმნისა და ზედსართავი სახელის, ხოლო საწყისის შემთხვევაში - ზმნისა და არსებითი სახელის წარმოებით მოდელებს იზიარებს. საწყისისა და მიმღეობის შესაბამისი ტეგსეტი აღწერილია ზმნის მორფოსინტაქსური ტეგების ჩამონათვალში. შესაბამისად, საწყისის გარდამქმნელი მოიცავს ინფორმაციას ხუთი კლასის შესახებ, ხოლო მიმღეობისა - სამი კლასის შესახებ;

- ფუნქციონალური სიტყვების უმეტესობა განეკუთვნება ერთეულთა დახურულ კლასს. გრამატიკული ურთიერთობების ჩვენების მიზნით აღნიშნული სიტყვები წინადადებაში სიტყვებს შორის გამოიყენება. ფუნქციონალურ სიტყვებს განეკუთვნება: ზმნიზედა, კავშირი, ნანი-ლაკი, შორისდებული და თანდებულები. ფუნქციონალური სიტყვების წარმოებითი მოდელი მინიმალურია, თუმცა მათი გამოყენების სიხ-შირე ძალიან მაღალია. შესაბამისად, ფუნქციონალური სიტყვების გარდამქმნელი კონკრეტული მეტყველების ნაწილის შესაბამის ხუთ კლასს მოიცავს, თუმცა ზმნიზედების ნაწილი სახელურ პარადიგმაზე -ად სუფიქსის დართვით გენერირდება;
- შემოკლებების, სხვაგვარად რომ ვთქვათ აბრევიატურების წარმოები-თი მოდელი მიჰყვება სახელური პარადიგმის წარმოების წესებს და ითვალისწინებს აღნიშნული წესების ამოქმედებას, როცა შემოკლება დეფისით იწერება ან მის გარეშე. არსებობს შემოკლებების ისეთი ტიპი, როცა შემოკლება ქართულ ტექსტში ჩნდება უცხო ანბანის სიმბოლოების გამოყენებით, მსგავს შემთხვევებში იგი ხდება უცხო სიტყვების გარდამქმნელის ნაწილი. უცხო სიტყვების გარდამქმნელის მთავარი დანიშნულებაა უცხოენოვანი ანბანის სიმბოლოებით გა-ფორმებული სიტყვა-ფორმების ამოცნობა და ქართული ანბანის სიმ-ბოლოებით გაფორმებული სიტყვა-ფორმებისაგან გამიჯვნა შესაბა-მისი ტეგის მინიჭებით. მსგავს შემთხვევებს სალექსიკონო ფაილი არ სჭირდება, ხოლო გამიჯვნა ხდება მხოლოდ *xfst*-ს სკრიპტის დონეზე შემდეგნაირად:

```
define EnglishLetter A|B|C|D|E|F|G|H|I|
J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z|a|b|c|d|e|f|g|h|i|
j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z ;
define GeorgianLetter ს| და ა.შ. ;
define Punct %-|%«|%»|%, ;
define Words [EnglishLetter|Punct EnglishLetter|English
Letter Punct|EnglishLetter Punct Punct GeorgianLetter]+ ;
read regex [ %+Foreign:Words ] ;
save stack foreignEN.fst
```

გამოსახულება 10 უცხო სიტყვების ამოცნობის გარდამქმნელის მაგალითი

მსგავსი სახით ანალიზატორში სამი ენის (ინგლისური, ბერძნული და რუსული) გამოყვანა ხდება, რადგანაც სამივე ენა სხვადასხვა დროს ჩნდება ქართულენოვან ტექსტში უცხოენოვანი ანბანის სიმბოლოების გამოყენებით.

3.2.4.2 მორფოლოგიური ანალიზის მაგალითები **მეტყველების ნაწილების მიხედვით**

სასრული პოზიციის მორფოლოგიური ანალიზატორისა და გენერატორის შემავალი სიგნალი ორ ნაწილად იყოფა: 1. უცვლელად დარჩენილი შემავალი სიგნალი, სხვაგვარად რომ ვთქვათ, ტექსტში წარმოდგენილი სიტყვა-ფორმა; 2. ტექსტში წარმოდგენილი სიტყვა-ფორმის ამოსავალი ლემა და მისი გრამატიკული სტრუქტურა შესაბამისი ტეგებით. სიტყვა-ფორმების ნაირგვარობისა და თავისებურების გათვალისწინებით, როგორც ზემოთ აღინიშნა, ხშირია გრამატიკული ომონიმების შემთხვევები. თვალსაჩინოებისთვის შემავალ-გამავალი სიგნალების მეტყველების ნაწილების მიხედვით ზოგიერთი სიტყვა-ფორმის გარჩევის ვარიანტი შემდეგნაირია:

არსებითი სახელი

ბედაურზე: ბედაურ-ი+N+Com+Anim+Sg+Dat+Post (on)

გამოსახულება 11 თანხმოვანფუძიანი არსებითი სახელის თანდებულებიანი ფორმის გარჩევის მაგალითი

ღობესთან: ღობე+N+Com+Inanim+Sg+Dat+Post (with/at)

გამოსახულება 12 ხმოვანფუძიანი არსებითი სახელის თანდებულებიანი ფორმის გარჩევის მაგალითი

ზედსართავი სახელი

ლამაზად: ლამაზ-ი+A+Posit+Sg+Advb

ლამაზად: ლამაზ-ი+Adv

გამოსახულება 13 ვითარებითი ბრუნვის ნიშნით გამოწვეული გრამატიკული ომონიმების მაგალითი

რიცხვითი სახელი

12-ზე: 12+Num+Card+Dig+Ins+Post (on)

XX: 20+Num+Card+Rom

XX: +Foreign

თორმეტზე: თორმეტ-ი+Num+Card+Sg+Dat+Post (on)

თორმეტიოდე: თორმეტ-ი+Num+Approx+Sg+Nom

გამოსახულება 14 ზოგიერთი რიცხვითი სახელის გარჩევის მაგალითი

ნაცვალსახელი

შე: შენ+Pron+Pers+2+Sg+Voc

გამოსახულება 15 ნაცვალსახელის გარჩევის მაგალითი

ზმნა

შემოვიდა: შემოდის+V+Intr+Aor+<NomSubj>+Subj2Sg

დაინახეს: ხედავს+V+Trans+Act+Aor+<ErgSubj>+<NomObj>+Sub-
j3Pl+Obj3

დაინახეს: Pfv+ნახავ-ს+V+IDt+#26+Din+Trans+Act+Aor+<ErgSub-
jBen>+<NomObj>+SubjBen3Pl+Obj3

დაინახეს: Pfv+ნახულობ-ს+V+IDt+#32+Din+Trans+Act+Aor+<Erg-
SubjBen>+<NomObj>+SubjBen3Pl+Obj3

დაინახეს: Pfv+ანახებ-ს+V+IDt+#28+Din+Trans+Act+Aor+<ErgSub-
jBen>+<NomObj>+SubjBen3Pl+Obj3

გადმოაყარა: Pfv+ყრი-ს+V+IDt+#24+Din+Trans+Act+Aor+<ErgSub-
j>+<NomObj>+<DatObjLoc>+Subj3Sg+ObjLoc3+Obj3

გამოსახულება 16 ერთპირიანი, ორპირიანი და სამპირიანი ზოგიერთი ზმნის გარჩევის მაგალითი ომონიმის შემთხვევებით

ზმნიზედა

ფეხდაფეხ : ფეხდაფეხ+Adv+Mod

გამოსახულება 17 ზმნიზედის გარჩევის მაგალითი

კავშირი

და: და+N+Com+Anim+Sg+Nom

და: და+Conj+Coord

გამოსახულება 18 კავშირის გარჩევის მაგალითი გრამატიკული ომონიმიის ჩვენებით

ნანილაკი

განა: განა+Part+Q

არც: არც+Part+IntP

გამოსახულება 19 ზოგიერთი ნანილაკის გარჩევის მაგალითი

შორისდებული

ყოჩაღ: ყოჩაღ+Itj

გამოსახულება 20 შორისდებულის გარჩევის მაგალითი

აბრევიატურა

აშშ-ის: ამერიკის შეერთებული შტატები+Abbr+N+Com+Sg+Gen

გამოსახულება 21 აბრევიატურის გარჩევის მაგალითი

უცხო სიტყვა

χρυσά: +Foreign

გამოსახულება 22 უცხო სიტყვის გარჩევის მაგალითი

პუნქტუაცია

, : +F+Comma

გამოსახულება 23 პუნქტუაციის გარჩევის მაგალითი

3.2.4.3 ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის სტრუქტურა

ქართული მორფოლოგიური ანალიზატორისა და გენერატორის ორმომართულებიანი სტრუქტურა, რომელიც შეიქმნა ქსეროქსის სასრული პოზიციის ტექნოლოგიის (ბისლი და კარტუნენი 2003) გამოყენებით გულისხმობს დამუშავების კონკრეტულ თანმიმდევრობასა და სტრუქტურის არსებობას. აღნიშნული ხელსაწყოების გამოყენებით ქართული ენის მორფოსინტაქსურ წესებზე დაყრდნობით სისტემას ორი ტიპის ფაილი მიეწოდება: ამოსავალი ლექსიკონის ფაილი - ლემისა და ქართული ენის მორფოტაქტიკის ჩვენებით და სკრიპტის ფაილი - ქართული ენის მორფოლოგიური პროცესების აღწერით. ფაილების გაერთიანების შემდეგ შეიქმნა სასრული პოზიციის გარდამქმნელი, რომელიც უზრუნველყოფს ენის ზედა დონის გენერირებასა და ქვედა დონეზე გენერირებული ფორმის მორფოსინტაქსურ ანალიზს. გამავალი სიგნალის გარჩევა სხვადასხვა თვალსაზრისით შეიძლება:

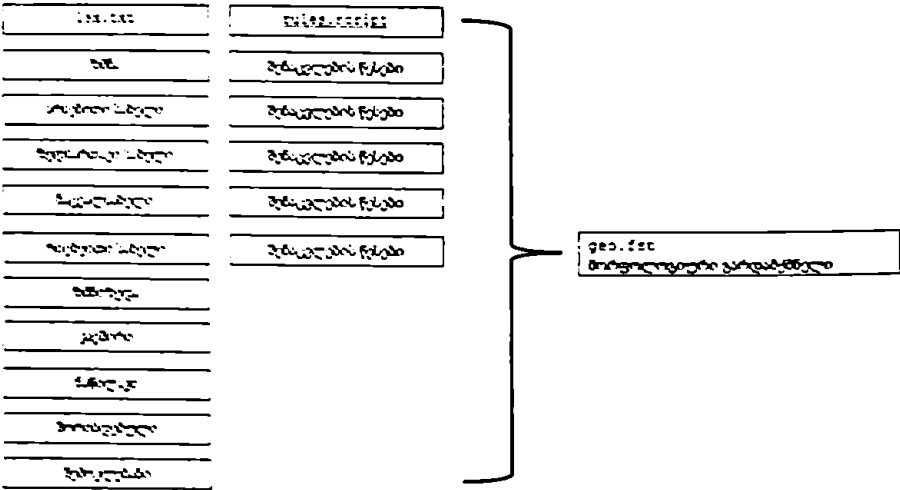
ა) მორფოლოგიური თვისებების სანახავად, რადგანაც ანალიზატორში წარმოდგენილია მორფოლოგიური თვისებების ამსახველი ტეგები;

ბ) მორფემების სანახავად, რადგანაც ანალიზატორის მცირე ცვლილებებით შეუძლია აღნიშნული ინფორმაციის ზედა დონეზე წარმოდგენა აფიქსების ძირისგან გამიჯვნის მიზნით;

გ) მორფოსინტაქსური თვისებების სანახავად, რადგანაც მზნის ანალიზი გულისხმობს არა მხოლოდ მისი მორფოლოგიური მახასიათებლების, არამედ მასში წარმოდგენილი სუბიექტისა და ობიექტების ბრუნვებისა და სემანტიკური როლების ჩვენებას და;

დ) კონკრეტული სიტყვა-ფორმების ლექსიკური ტიპების განსასაზღვრელად, რადგანაც მორფოლოგიური ანალიზატორის ნაწილს არა მხოლოდ კონკრეტული სიტყვების მორფოსინტაქსური მახასიათებლები წარმოადგენს, არამედ სიტყვების ტიპების მიხედვით ჯგუფებად გაერთიანების შესაძლებლობაც შესაბამისი ტეგების ჩვენებით.

ქართული ენის მორფოსინტაქსური ანალიზატორის საერთო სტრუქტურა მოიცავს კონკრეტული მეტყველების ნაწილის ლექსიკონის ფაილს (მაგ. nounCom.txt და სხვ.) და მორფოფონოლოგიური წესების შესაბამის სკრიპტს (მაგ. nounCom.script). აღნიშნულ ორ ფაილზე დაყრდნობით იქმნება კონკრეტული მეტყველების ნაწილის შესაბამისი სასრული პოზიციის გარდამქმნელის ფაილი (მაგ. nounCom.fst და სხვ.). მეტყველების ნაწილების შესაბამისი სასრული პოზიციის გარდამქმნელების შეკვრის გზით იქმნება საერთო გარდამქმნელი, რომელიც სრულყოფილად ქართული ენის მორფოსინტაქსურ სტრუქტურას მოიცავს:



გრაფიკი 3 მორფოლოგიური ანალიზატორის გამართივებული სტრუქტურა

როგორც ზემოთ აღინიშნა ზმნური მოდული ყველაზე რთული სტრუქტურისაა და მოიცავს 66 რეგულარულ და ირეგულარულ კლასებს, რომლების საფუძველზე ხდება საერთო სასრული პოზიციის გარდამქმნელის გენერირება შემდეგი სქემის მიხედვით:

```

read regex @"verb1.fst" ;
read regex @"verb2.fst" ;
read regex @"verb3.fst" ;
Etc.
read regex @"irregulars.fst" ;
read regex @"participle.fst" ;
read regex @"masdar.fst" ;
union net
save stack verb.fst

```

გამოსახულება 24 ზმნური მოდულის სასრული პოზიციის გარდამქმნელის შექმნის თანმიმდევრობა

თანამედროვე ქართული ენის ლექსიკონის საწყის ბაზას არნ. ჩიქობავას ქართული ენის განმარტებითი ლექსიკონი (1950-1964) და დ. მელიქიშვილის ზმნების ინდექსი (2001) წარმოადგენს, ხოლო ძველი და საშუალო ქართული ენის - ი. აბულაძის (1973). ტესტირების პროცესში აღნიშნული ლექსიკონები ქართული ენის კორპუსიდან მოპოვებული მასალით შეიცვალა.

თავი 3

ტოკენიზატორი

ტექსტის დამუშავებისა და ტექსტში წარმოდგენილი სიტყვა-ფორმების მორფოლოგიური ანალიზის თანმიმდევრობა მკაცრად დაცულია. აღნიშნულ თანმიმდევრობაში იგულისხმება, უპირველეს ყოვლისა, ტექსტის მოპოვება, სტრუქტურული დონის ანოტირება (მეტა-ინფორმაციისა და ტექსტის სტრუქტურის აღწერა), და, მხოლოდ შემდეგ ტექსტის ტოკენებად დაშლა და მორფოლოგიური ანალიზი. ტექსტის სიტყვა-ფორმებად დაშლას სპეციალური კომპიუტერული საშუალება სჭირდება. ასეთ საშუალებას **ტოკენიზატორი** წარმოადგენს. ტოკენიზატორიც სასრული პოზიციის გარდამქმნელია, რომლის ამოცანას შემავალი ტექსტის ჯერ წინადადებად და შემდგომ სიტყვა-ფორმებად დაშლა წარმოადგენს.

ქართული ენის ტოკენიზატორი ორი ნაწილისაგან შედგება. საწყის ეტაპზე ტექსტის წინადადებად დაშლა ხდება ტექსტში დადასტურებული პუნქტუაციის ნიშნებზე დაყრდნობით (*sentence.fst*) და შემდგომ ტოკენებად ტექსტში დადასტურებული ინტერვალების, პუნქტუაციის ნიშნების, ზოგიერთი შესიტყვების, ინიციალებისა და აბრევიატურების გათვალისწინებით (*tok.fst*). მიღებული განმარტებების მიხედვით ტექსტში დამუშავებაზე ორიენტირებულ სიმბოლოთა თანმიმდევრობას ტოკენი ეწოდება (მანინგი 2008 და სხვ.).

ტოკენიზატორის სკრიპტი ძირითადად ანა შილერის მიერ შემუშავებულ სასრული პოზიციის ტოკენიზატორს მიჰყვება (ბისლი და კარტუნენი 2003, კარტუნენი და სხვ. 1997), თუმცა კონკრეტული ამოცანების გადასაწყვეტად საჭირო გახდა აღნიშნულ სკრიპტის დახვეწა ქართული ენის სტრუქტურის გათვალისწინებით და თუ თანამედროვე ქართულში ტოკენებად ტექსტის დაშლა საკმაოდ რეგულარულ ხასიათს ატარებს სიტყვებს შორის ინტერვალისა და პუნქტუაციის ნიშნების გამართული სისტემის გამო, ძველი ქართული ტექსტის ტოკენიზაცია ირეგულარული ხასიათისაა ინტერვალისა და პუნქტუაციის ნიშნების არასტანდარტული გამოყენების გამო. ქართულ ტექსტში ტოკენების ამოცნობის მიზნით სისტემაში ჩადებულია ქართული ენისთვის დამახასიათებელი სამი სკრიპტი *ასომთავრული*, *ნუსხური* და *მხედრული*:

```
define LETTER
[აბგდევჟზთიკლმნოპჟრსტუფქღყშჩცძწჭხჟჴქჷღჶჷ
|ჲ|ႠႡႢႣႤႥႦႧႨႩႪႫႬႭႮႯႰႱႲႳႴႵႶႷႸႹ
|ႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻ
|ႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻ
|ႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻႼႽႾႿႺႻ]
```

გამოსახულება 25 ქართული ანბანის ტოკენიზატორში წარმოდგენის ტიპი

თანამედროვე ქართული ენის თვალსაზრისით შედარებით რთულია წილადების, ინიციალებისა და აბრევიატურების ამოცნობა, რომლებშიც ხდება მძიმეებისა და წერტილების გამოყენება. ამიტომაც მსგავსი

9 ასომთავრული და მხედრული წარმოდგენილია უნიკოდის სტანდარტის ვერსია 12.1-ში, სქემა 10A0-10FF, ხოლო ნუსხური სკრიპტი წარმოდგენს უნიკოდის დანართს, სქემა 2D00-2D2F (ევერსონი 1991-2019).

შემთხვევები წინასწარ განსაზღვრულია ტოკენიზატორის კოდში:

```
define INIT [LETTER %.]+;
define ABBR [{ა.შ.}|{ე.ო.}|{სხვ.}|{იხ.}|{გვ.}]; ! და ა.შ.
```

გამოსახულება 26 ტოკენიზაცია ინიციალებისა და შემოკლებების ამოცნობის მიზნით

```
define DIGIT [%0|%1|%2|%3|%4|%5|%6|%7|%8|%9] ;
define NUMOP [%-|%+|%*|%/|%=|%=] ;
define NUMSEP [%.|%,] ;
define NUM [[DIGIT|NUMOP|NUMSEP]+ & ${DIGIT}] ;
```

გამოსახულება 27 ტოკენიზაცია წილადების ამოცნობის მიზნით

ტექსტის წინადადებად დაშლის მოდული დამოკიდებულია პუნქტუაციის ნიშნებსა და სიტყვებს შორის ინტერვალზე და ტაბულაციამზე პარაგრაფების დასაწყისში. ანალიზატორის კოდში წინადადებების დასრულების აღსანიშნავად შემდეგი სიმბოლოები გამოიყენება: წერტილი (.), ძახილის ნიშანი (!), კითხვის ნიშანი (?), წერტილ-მძიმე (;), კითხვის ნიშანი ძახილის ნიშანთან ერთად (?!), ძახილის ნიშანი ორ წერტილთან ერთად (!..), სამი წერტილი (...) და პარაგრაფების გამყოფი (:) ძველი და საშუალო ქართულისათვის.

ტოკენიზატორის ბლოკის შემონშება ხდებოდა საჭიროებიდან გამომდინარე კოდის გაუმჯობესებისა და დახვეწის გზით. ძველი ქართული ხელნაწერი-ტექსტის თავისებურების გათვალისწინებით რაღაც მომენტში დამატებით განიხილებოდა ძველი ქართული ტექსტის სტრუქტურირება სახელური და ზმნური ფორმების გამოვლენის მიზნით. აღნიშნული სტრუქტურირება გულისხმობდა ტოკენების გამოვლენას თანხმოვანთა კლასტერებისა და მარცვლების დადგენის გზით, თუმცა ამ ეტაპზე, კვლევის აღნიშნული მონაკვეთი საჭიროებს დამატებით შესწავლას, ტესტირებასა და შემონშებას.

```

define ST [ბ|ფ|პ|დ|თ|ტ|გ|ქ|კ|ყ|წ] ;           ! მსკდომები
define AF [ძ|ც|ხ|ჩ|რ|ჭ] ;                       ! აფრიკატები
define FR [ზ|ს|ქ|შ|ღ|ხ|პ] ;                     ! ფრიკატივები
define SN [მ|ნ|რ|ლ|ვ|ჟ|ფ] ;                     ! სონანტები
define V [ა|ე|ო|უ|ი|მ|ღ|შ] ;                     ! ხმოვნები

# ბუცხრიკიძესთან (2002) წარმოდგენილი სქემის მიხედვით ძირის
დასაწყისში შეიძლება იყოს თანხმოვანთა შემდეგი კომპლექსები
define ISEQ2 [ [ ST ST | ST AF | ST FR | ST SN | AF ST |
AF FR | AF SN | FR ST | FR AF | FR FR | FR SN | SN ST |
SN AF | SN FR | SN SN ] ] ;
# სამი თანხმოვნის შემთხვევაში შემდეგი
define ISEQ3 [ [ ST ST ST | ST AF ST | ST FR SN | ST SN
ST | ST SN AF | ST SN SN | AF ST SN | AF FR SN | AF SN
ST | AF SN AF | AF SN FR | AF SN SN | FR ST SN | FR AF
SN | FR FR SN | FR SN ST | FR SN AF | FR SN FR | FR SN
SN | SN ST SN | SN AF SN | SN AF FR | SN FR SN | SN SN
AF ] ] ;
და ა.შ.
# განსაზღვრავს ერთმარცვლიან სახელურ ძირს ეთერლიშვილის (1970)
მიხედვით
Define N4 [ [ ISEQ2 | ISEQ3 | ISEQ4 | ISEQ5 | ISEQ6 ] [
a | e ] %Guess%N:0 ] ;
# განსაზღვრავს ერთმარცვლიან ხმოვანფუძიან სახელებს
Define N5 [ [ ISEQ2 | ISEQ3 | ISEQ4 | ISEQ5 | ISEQ6 ] [
o | u ] %Guess%N:0 ] ;
read regex [ N4 | N5 ] ;

```

გამოსახულება 28 ერთმარცვლიანი სახელური ფუძის ამომცნობის სტრუქტურა

ამ ეტაპზე შემონშების შედეგად ტოკენიზატორში შემდეგი პრობლემური საკითხები გამოვლინდა:

- ძველი და საშუალო ქართული დიგიტალიზირებული ხელნაწერი ტექსტების შემთხვევაში ინტერვალის არათანმიმდევრული გამოყენება;

- სათაურებისა და სახელწოდებების ამოცნობის პრობლემა, რომელიც გამომწვეულია თანამედროვე ქართულში დიდი ასოების არარსებობით, ხოლო ძველ ქართულში ასომთავრული ანბანის არათანმიმდევრული გამოყენებით;
- სათაურების შემდეგ პუნქტუაციის ნიშნების არარსებობა, რაც მათ გამოვლენას ართულებს;
- ინიციალებსა და აბრევიატურებში პუნქტუაციის ნიშნების გამოყენება;
- სხვადასხვა სახის ტიპოგრაფიული და ორთოგრაფიული შეცდომები.

თავი 4

ბასტირება

სასრული პოზიციის ტექნოლოგიის გამოყენებით ქართული ენის ტოკენიზატორისა და მორფოლოგიური ანალიზატორის შექმნა ისევე, როგორც ნებისმიერი სხვა ენის შემთხვევაში გულისხმობს შედეგების შეჯერებასა და შემოწმებას რესურსის გაუმჯობესებისა და სუსტი მხარეების გამოვლენის მიზნით. ბუნებრივი ენის დამუშავების ნებისმიერი რესურსი ყოველთვის ასოცირდება ტესტირებასთან, შეცდომათა ანალიზთან და თეორიული მონაცემების გაუმჯობესებასთან. ყველა აღნიშნული პროცედურა შემოწმების შემდეგ ტიპებს გულისხმობს: წესების მთლიანობის, ტეგების თანმიმდევრობისა და ამოცნობის პროცენტის შემოწმებას.

I. **წესების მთლიანობის შემოწმება** გულისხმობს სისტემის სხვადასხვა დროს შექმნილი ვერსიების რეგრესულ ტესტირებას გარდამქმნელის ზედა და ქვედა დონეების შედარების გზით. ქართული ენის მორფოლოგიური ანალიზატორის რეგრესული ტესტირება მუდმივად მიმდინარეობს სისტემაში დამატებული და სისტემიდან ამოღებული სიტყვა-ფორმების კონტროლის გზით. რეგრესული ტესტის პროცედურული თანმიმდევრობა კარგად აღწერილია ბისლი და კარტუნენტან (2003), პირველ შემთხვევაში ხდება ორი ვერსიის შედარება სისტემაში დამატებული (words-added.txt) და სისტემიდან ამოღებული (words-lost.txt) სიტყვა-ფორმების გამოვლენის მიზნით;

2. **ტეგების თანმიმდევრობის შემონმება** პირდაპირ კავშირშია სიტყვა-ფორმების ზედა დონესთან. ტეგების თანმიმდევრობა უნდა იყოს ფიქსირებული ანალიზატორის სხვა სისტემებში ინტეგრირების მიზნით. უნდა არსებობდეს წინასწარ შემუშავებული მორფოსინტაქსური მახასიათებლების სია შესაბამისი ტეგებით. ბისლი და კარტუნენთან (2003) ტესტირების პროცედურული თანმიმდევრობა გულისხმობს შემონმებისას ტეგების სპეციალური გრამატიკის შექმნასა და აღნიშნულ გრამატიკაზე დაყრდნობით ქსელის შემონმებას. ქართული ენის მორფოლოგიური ანალიზატორის შექმნის პროცესში სპეციალურად შემუშავდა ტეგების ლექსიკური გრამატიკის ვერსია:

```
[ა|ე|ი|ო|უ|ჟ|მ|ღ|მ|ბ|გ|დ|ვ|ზ|თ|კ|ლ|მ|ნ|პ|ჟ|რ|ს|ტ|ფ|ქ|ღ|ყ|შ|ჩ|
  ც|ძ|ც|ჭ|ხ|ჯ|პ|ვ|ჯ|ფ|ღ|მ|ნ|რ|]+
(%+OGE| %+MGE| %+GE)
[%+Noun [%+Prop (%+Name| %+Geog)
[%+Anim [%+Sg| %+Pl]
[%+Nom| %+Erg| %+Dat| %+Gen| %+Ins| %+Advb| %+Voc]

(%+Emph) ([%+Post% (like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1| %+IndSpeech2| %+IndSpeech3]) ]
|
[%+Inanim [%+Sg| %+Pl]
და სხვ.)
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1| %+IndSpeech2| %+IndSpeech3]) ] ]
|
[%+Com [%+Anim [%+Sg| %+Pl]
[%+Nom| %+Erg| %+Dat| %+Gen| %+Ins| %+Advb| %+Voc]
(%+Emph)
```

```

([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1|+%IndSpeech2|+%IndSpeech3]])
|
[%+Inanim [%+Sg|+%Pl]
[%+Nom|+%Erg|+%Dat|+%Gen|+%Ins|+%Advb|+%Voc]
(%+Emph)

([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1|+%IndSpeech2|+%IndSpeech3]])
]]]
|
%+Adj (%+Pos|+%Com|+%Sup)
[%+Sg|+%Pl]
[%+Nom|+%Erg|+%Dat|+%Gen|+%Ins|+%Advb|+%Voc]
(%+Emph)

([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1|+%IndSpeech2|+%IndSpeech3]])
|
%+Num (%+Alpha|+%Roman|+%Arabic)
[[%+Card [%+Sg|+%Pl]
[%+Nom|+%Erg|+%Dat|+%Gen|+%Ins|+%Advb|+%Voc]

(%+Emph)
([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1|+%IndSpeech2|+%IndSpeech3]])
|
[%+Ord [%+Sg|+%Pl]
[%+Nom|+%Erg|+%Dat|+%Gen|+%Ins|+%Advb|+%Voc]
(%+Emph)
([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1|+%IndSpeech2|+%IndSpeech3]])
|

```

```

[%+Fract [%+Sg| %+Pl]
[%+Nom| %+Erg| %+Dat| %+Gen| %+Ins| %+Advb| %+Voc]
(%+Emph)
([%+Post%(like%) | და სხვ. ])

(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1| %+IndSpeech2| %+IndSpeech3]) ]
|

[%+Approx [%+Sg| %+Pl]
[%+Nom| %+Erg| %+Dat| %+Gen| %+Ins| %+Advb| %+Voc]
(%+Emph)
([%+Post%(like%) | და სხვ. ])
(%+Emph) (%+Ptcl) (%+Emph) (%+Aux)
([%+IndSpeech1| %+IndSpeech2| %+IndSpeech3]) ]
|

[%+Rep (%+Emph) (%+Aux)
([%+IndSpeech1| %+IndSpeech2| %+IndSpeech3]) ]
]
| და სხვ.
];

```

გამოსახულება 29 შემოკლებული ამონარიდი ტეგების ლექსიკური გრამატიკიდან

ტეგეტის შემადგენლობა და ტეგების თანმიმდევრობა მონმდებოდა პროეცირებისა და გამოკლების გზით. პროეცირების შედეგად ხდება ქსელის ზედა ან ქვედა დონეების ამოღება და შემდგომ ზემოხსენებულ გრამატიკასთან შედარება. გამოკლებისას კი შესაბამისობის შემთხვევაში გარდამქმნელიდან ვილებდით ლექსიკურ გრამატიკასა და ცარიელ ენას.

3. **ამოცნობის პროცენტის შემოწმება** უკავშირდება ციპფის კანონს (ციპფი 1932), რომლის თანახმად, მსოფლიოში არსებულ ნებისმიერ ენაში მხოლოდ სიტყვათა მცირე რაოდენობას გამოყენების მაღალი სიხშირე ახასიათებს, საშუალო რაოდენობას – საშუალო სიხშირე, ხოლო უმეტესობას – დაბალი სიხშირე. და, შესაბამისად, მორფოლოგიური ანალიზატორის ამოცნობის პროცენტის განსაზღვრა

სიტყვათა სიებისა და კორპუსების მონაცემების ტესტირების გზით ხდება. სიტყვათა სიების ტესტირება მუდმივი პროცესია, რომელიც თანმიმდევრულად მიმდინარეობს მორფოლოგიური ანალიზატორის სხვადასხვა ბლოკზე მუშაობისას, ხოლო კორპუსულ მონაცემებზე დაყრდნობით ამოცნობის პროცენტის განსაზღვრა და ანალიზატორის გაუმჯობესება გულისხმობს კორპუსის შექმნას, კორპუსული მონაცემების ტოკენიზატორით სიტყვა-ფორმებად დაშლას და შემდგომ მათ ანალიზს *lookup*-ის გამოყენებით. მორფოლოგიური ანალიზატორის შექმნა და გაუმჯობესება ქართული ენის კორპუსის (დობორჯინიძე და სხვ. 2012-2020) შექმნის პარალელურად მიმდინარეობდა და ამით იქმნებოდა ანალიზატორის მუდმივად ტესტირების შესაძლებლობა.

ქართული ენის კორპუსში (ქეგ) მოთავსებული მასალა მოიცავს ქართულ სალიტერატურო ენას მეექვსე საუკუნიდან მოყოლებული 2014 წლამდე. თუმცა თვისობრივად კორპუსი არ განეკუთვნება არც ისტორიულ და არც პარალელურ კორპუსთა ტიპებს. ამ ეტაპზე ქართული ენის კორპუსი პუნქტუაციის ნიშნების გარეშე დაახლოებით 15 მილიონ სიტყვას მოიცავს. კორპუსში წარმოდგენილი მასალა ჟანრობრივად განსხვავებულია, იგი მოიცავს როგორც პროზაულს, ასევე პოეტურ ნაწარმოებებს, როგორც სამეცნიერო, ასევე საგაზეთო სტატიებს და ა.შ. კორპუსის წინასწარი დამუშავება მოიცავდა არა მხოლოდ მონაცემების მოგროვებას, არამედ მათ აღწერას და ტექსტის გამზადებას შემდეგი თანმიმდევრობის დაცვით:

- *ტექსტური ფორმატის უზრუნველყოფა*: ტექსტების უმეტესობა საჭიროებდა კონვერტირებას უნიკოდის UTF-8 ფორმატის უზრუნველყოფის მიზნით. თითოეული ფაილის დამუშავება ხდებოდა ცალ-ცალკე, რადგანაც, ფაილების ფორმატი სხვადასხვანაირი იყო;
- *ტექსტის ბედა და ქვედა ნაწილების მოხსნა*: ძალიან ხშირად ნაბეჭდი გამოცემიდან ელექტრონული ტექსტის შექმნის დროს, ტექსტში ხშირად ისეთი ინფორმაცია ხდებოდა როგორიცაა გვერდის ნომერი, რაიმე წარწერა და ა.შ. განსაკუთრებით აღნიშნული წესი PDF-დან გადმოტანილ ტექსტურ ფაილებს ეხებოდა. მსგავსი ტიპის

ინფორმაცია საჭიროებდა მოხსნას განსაკუთრებით იმ შემთხვევებში, როცა არღვევდა ტექსტის სტრუქტურას:

ქველს სელაჰი სე აისიოგის ძიოლოდ 10%-ის დახაზვას შესაძლებელი. ხოლო აისიოგის 90 % წყალშია დაბალქული. აისიოგის მსგავსად, კელტურის ქველასე შინაარსობრივი ნაწილია შექმნეველი და მას კელტურის

4

ინყება ახალი გვერდი

სეზიექტურ ნაწილს უწოდებენ, ხოლო ხილულ ნაწილს – ობიექტურს.

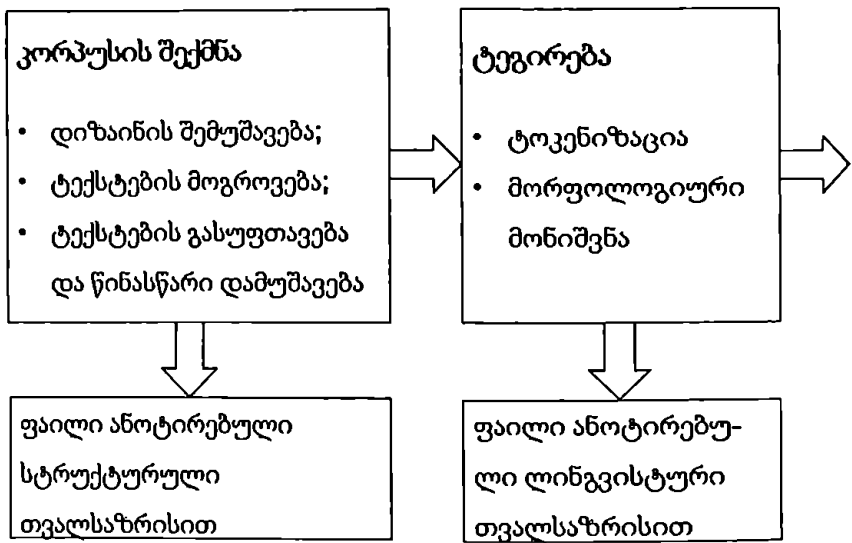
გრაფიკი 4 ამონარიდი PDF ფაილიდან

- *ტექსტების სანყისი და ბოლო ნაწილების მოხსნა:* ნებისმიერი სახის მასალა, რომელიც წინ უსწრებდა სათაურს, შესავალს ან წინასიტყვაობას საჭიროებდა მოხსნას. ინფორმაცია ავტორის, ილუსტრატორის, გამომცემლის, საავტორო უფლებების და ა.შ. შესახებ ტექსტის მეტა-ანოტირების ნაწილს წარმოადგენს. მეტა-ანოტირება, ძირითადად, მიჰყვებოდა ტექსტის კოდირების საინიციატივო ჩარჩოს, სამაგალითოდ, ძველი ქართული ტექსტის მეტა-ანოტირების ტეგების სიის ნახვა შესაძლებელია დანართ 2-ში. ასევე, საჭიროებდა მოხსნას სხვადასხვა ინდექსი, სიტყვათა სია, ბიბლიოგრაფია, შენიშვნები და ა.შ.;
- *შიდატექსტური ცვლილებები:* ძირითად ტექსტს ცვლილებები არ ეხებოდა გარდა იმ შემთხვევებისა, როცა ტექსტში იყო ცხრილები, სურათები, ფოტოები, ფორმულები და ა.შ. უცხოენოვანი ტექსტის მოხსნა არ ხდებოდა. ტექსტის წინასწარი დამუშავების ძირითად მიზანს სტრუქტურულად გამართული აბსოლუტურად სუფთა ტექსტის შექმნა წარმოადგენდა. მსგავსი ტექსტები შედგებოდა სრული წინადადებებისა და არა მათი ფრაგმენტებისაგან. ტექსტის სტრუქტურა (თავი, ქვეთავი, პარაგრაფი და ა.შ.) არ ირღვეოდა;

- გაზეთებისა და პერიოდული გამოცემების გასუფთავება: გაზეთებში, პერიოდულ გამოცემებში, ჟურნალებში სხვადასხვა ტიპის ინფორმაციაა, მათ შორის კროსვორდები, სახელები, მისამართები, თარიღები, სურათები, განცხადებანი, ფეხბურთის მატჩების შედეგები, სატელევიზიო გადაცემების პროგრამები და ა.შ. აღნიშნული ინფორმაციის შეტანა კორპუსში გათვალისწინებული არ იყო;
- გარდა ამისა, გაზეთში და ნებისმიერ სხვა პერიოდულ გამოცემაში, ტექსტებს სვეტური განლაგება ახასიათებს, რაც განაპირობებდა გადატანის ნიშნის მქონე სიტყვების სიჭარბეს. აღნიშნული სიტყვების ლინგვისტური ანალიზი რთულდება, მაგ. თუ გვაქვს სიტყვა „გებუ-ნებრივი“, რომელიც ტექსტში გვხვდება შემდეგი სახით „გებუ-“ „ნებ-რივი“, გარდამქმნელის ლექსიკონში მისი მოძებნა შეუძლებელია. შესაბამისად, ხდებოდა შიდატექსტური გადატანის ნიშნების მოშორება. აღნიშნული ამოცანის დაძლევა ნაწილობრივ ხელით და ნაწილობრივ პროგრამულად ხდებოდა.

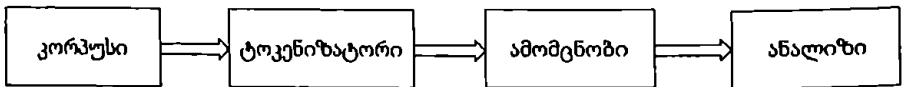
კორპუსული მონაცემების დამუშავების შემდგომ საფეხურს ტექსტების მეტა- და შიდა-ანოტირება წარმოადგენდა, რაც განსაკუთრებულად აქტუალური იყო ძველი ქართული ენის შემთხვევაში. აღნიშნული ანოტირება სრულდებოდა ტექსტის კოდირების საინიციატივო ჩარჩოს რეკომენდაციების გათვალისწინებით (TEI:P5 2019).

ენის დაფარვის ტესტი განხორციელდა გარდამქმნელის ლექსიკონის შეფასების მიზნით სიტყვათა სიებთან და კორპუსულ მონაცემებთან შედარების გზით. ორივე შემთხვევაში განსაკუთრებული ყურადღება გამახვილდა გამოყენების მაღალი სიხშირის მქონე სიტყვებზე. ამოუცნობი სიტყვა-ფორმები ინახებოდა ცალკე, შემდგომი ანალიზის უზრუნველყოფის მიზნით. პროცედურული თანმიმდევრობა იყო შემდეგი:



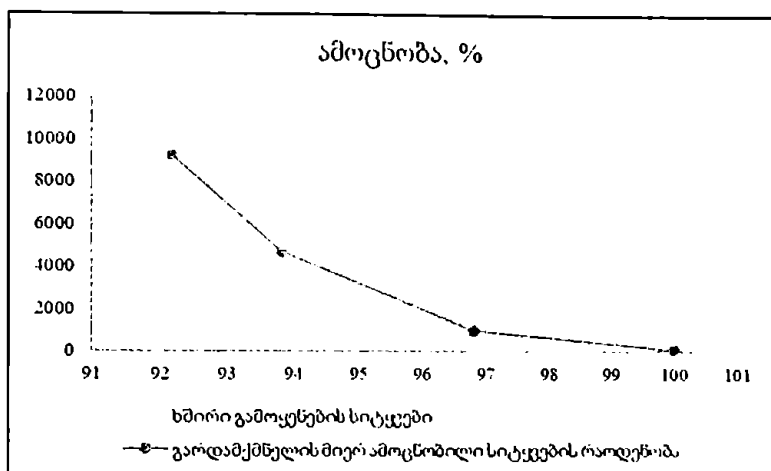
გრაფიკი 5 მონიშვნის პროცედურული თანმიმდევრობა

აღნიშნული თანმიმდევრობა გულისხმობდა ტექსტური ფაილის შემავალი სიგნალის სახით ტოკენიზატორში და შემდგომ ამომცნობის მოდულში გატარებას გამავალი სიგნალის სახით მორფოსინტაქსურად გარჩეული სიტყვების მისაღებად.



გრაფიკი 6 მორფოლოგიური ანალიზის პროცედურული თანმიმდევრობა

კორპუსულ მონაცემებზე დაყრდნობით შეიქმნა ქართული ენის კორპუსში ყველაზე ხშირად გამოყენებული სიტყვა-ფორმების სია ცალკე თანამედროვე ქართული ენისთვის და ცალკე ძველი და საშუალო ქართული ენისთვის. აღნიშნულ სიებში მოხვედრილი სიტყვა-ფორმების რაოდენობა როგორც ძველი, ასევე თანამედროვე ქართული ენისთვის 10000 შეადგენდა. მოხდა აღნიშნული სიების შესწავლა ამომცნობის პროცენტის დადგენის მიზნით. განსაკუთრებული ყურადღება მიექცა სახელურ და ზმნურ ფლექსიურ პარადიგმებს.



გრაფიკი 7 თანამედროვე ქართული ენის ამოცნობის პროცენტული მაჩვენებელი

როგორც გრაფიკიდან ვლინდება, თანამედროვე ქართული ენის ყველაზე ხშირად გამოყენებადი 10000 სიტყვის ამოცნობის პროცენტი 92.17% შეადგენს. ტესტირების პროცესში ამოუცნობი სიტყვების ამოსავალი ლემები გარდამქმნელის ლექსიკონს ემატებოდა და იზრდებოდა ამოცნობილი სიტყვების რაოდენობაც. ამავდროულად, ზმნური ფორმების რაოდენობა 1000 ყველაზე ხშირად გამოყენებად სიტყვა-ფორმაზე 530-ს შეადგენდა, მათ შორის, 521 – ამოცნობილი და 9 – ამოუცნობი, რაც მეტყველებს იმაზე, რომ თანამედროვე ქართული ენის შემთხვევაში შესაძლებელია ნებისმიერი ტექსტის დახლოებით 85-90 % ფორმის ავტომატური ამოცნობა.

ზმნური სიტყვა-ფორმების შემთხვევაში აღსანიშნავია პარადიგმებს შორის გამოვლენილი თანხვედრაც, კერძოდ, პირველი დიათეზის მე-19 და 26-ე კლასებსა (51) და სხვადასხვა დიათეზაში მოხვედრილ 28-ე, 29-ე, 30-ე და 47-ე კლასებს (52) შორის.

(51) კეტ-ავ-ს და სხვ.

კეტ-TS-3SGSbj:PRS.IND

კეტავს: Ipfv+კეტავ-ს+Verb+-

Main+IDt+#19+Din+Trans+Act+Pres+<NomSubj>

+<DatObj>+Subj3Sg+Obj3

კეტავს: Ipfv+კეტავ-ს+Verb+-

Main+IDt+#26+Din+Trans+Act+Pres+<NomSubj>

+<DatObj>+Subj3Sg+Obj3

გამოსახულება 30 გრამატიკული გადაფარვის მაგალითი მე-19 და 26-ე კლასებს შორის

(52) და-ი-ნთ-ო და სხვ.

prev.PFV-ნთ-3SGSbj:AOR.IND ან

prev.PFV-ნთს-3SGSbj:AOR.SUBJ

დაინთო: Pfv+ინთებ-ს+Verb+Main+IDt+#29+Din+Trans

+Act+AorSbj+<ErgSubjBen>+<NomObj>+SubjBen2Sg+Obj3

დაინთო: Pfv+ინთებ-ს+Verb+Main+IDt+#30+Din+Trans

+Act+Aor+<ErgSubjBen>+<NomObj>+SubjBen2Sg+Obj3+Ind-Speech3

დაინთო: Pfv+ინთებ-ს+Verb+Main+IDt+#30+Din+Trans

+Act+AorSbj+<ErgSubjBen>+<NomObj>+SubjBen2Sg+Obj3

დაინთო: Pfv+ინთებ-ა+Verb+Main+IIDt+#47+Din+Intr

+Pass/AutAct+AorSbj+<NomSubjBen>+SubjBen2Sg

დაინთო: Pfv+ანთებ-ს+Verb+Main+IDt+#28+Din+Trans

+Act+AorSbj+<ErgSubjBen>+<NomObj>+SubjBen2Sg+Obj3

დაინთო: Pfv+ინთებ-ს+Verb+Main+IDt+#29+Din+Trans

+Act+Aor+<ErgSubjBen>+<NomObj>+SubjBen3Sg+Obj3

დაინთო: Pfv+ინთებ-ს+Verb+Main+IDt+#30+Din+Trans

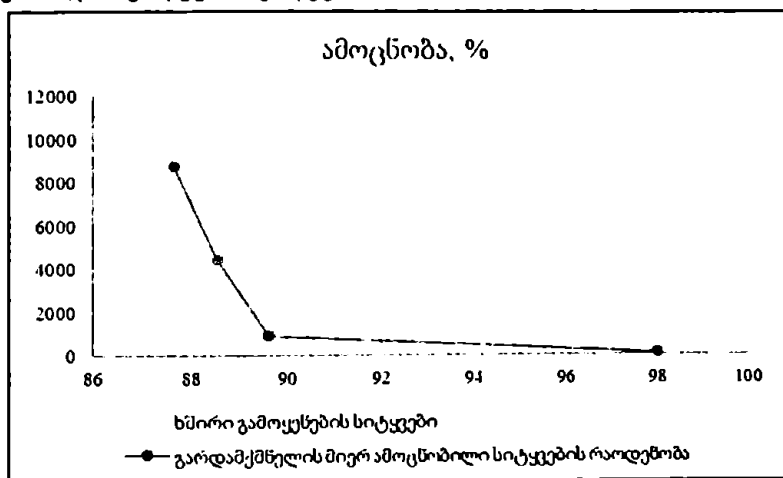
+Act+Aor+<ErgSubjBen>+<NomObj>+SubjBen3Sg+Obj3

დაინთო: Pfv+ინთებ-ა+Verb+Main+IIDt+#47+Din+Intr

+Pass/AutAct+Aor+<NomSubjBen>+SubjBen3Sg

გამოსახულება 31 გრამატიკული გადაფარვის მაგალითი 28-ე, 29-ე, 30-ე და 47-ე კლასებს შორის

ანალიზატორის თანამედროვე ქართული ენის ბლოკთან შედარებით ძველი და საშუალო ქართული ენის შემთხვევაში ამოცნობის პროცენტი ნაკლებია და შეადგენს შემდეგს:



გრაფიკი 8 ძველი და საშუალო ქართული ენის ამოცნობის პროცენტული მაჩვენებელი

როგორც გრაფიკიდან ვლინდება, ძველი და საშუალო ქართული ენის ყველაზე ხშირად გამოყენებადი 10000 სიტყვის ამოცნობის პროცენტი 87.64% შეადგენს. ამოცნობი სიტყვები ტესტირების პროცესში დროდადრო ემატება გარდამქმნელის ლექსიკონს. აღნიშნული ტესტირება ჩატარდა კორპუსის ხელნაწერების გამოცემულ ტექსტებზე დაყრდნობით. გამოცემული ხელნაწერების ტექსტების მორფოსინტაქსური ამოცნობის პროცენტი ბევრად ნაკლებია; ხელნაწერი ტექსტების დიგიტალურ ვერსიებში ყველაზე დიდი პრობლემა პუნქტუაციის ნიშნების ნაკლებობას, სიტყვებს შორის ინტერვალის არარსებობასა და დაქარაგმებული სიტყვების სიმრავლეს უკავშირდება. აღნიშნული პრობლემის აღმოფხვრის მიზნით გახდა საჭირო მარცვლებში შემავალი თანხმოვნებისა და ხმოვნების ფონოტაქტიკის გათვალისწინებით სახელური და ზმნური ფუძეების ამოცნობაზე ორიენტირებული ამომცნობების საწყისი ტესტირება.

ამოცნობის აღნიშნული პროცენტი მეტყველებს იმაზე, რომ ძველი და საშუალო ქართული ენის გამოცემული ტექსტების შემთხვევაში,

შესაძლებელი იქნება 80-85% ამოცნობა. ძველი ქართული ენის მორფოლოგიური ანალიზატორის ლექსიკონის შევსება ახლაც მიმდინარეობს და, ბუნებრივია, რომ აღნიშნული პროცენტიც გაიზრდება.

თავი 5

დასკვნები

ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორის ამოცნობის პროცენტის გაზრდა დამოკიდებულია ლექსიკონების გამდიდრებაზე. მუშაობის პროცესში ძველი, საშუალო და თანამედროვე ქართული ენის ლექსიკონები, რომლებიც გარდამქმნელებს დაედო საფუძვლად შეივსო კორპუსიდან მოპოვებული მასალით. ამ დროისათვის, ლექსიკონების გამდიდრება შესაძლებელია ლექსიკონის ფლექსიური კლასის განსაზღვრის გზით ლინგვისტის ან ქართული ენის სტრუქტურის მცოდნე სპეციალისტის მიერ.

დანართი 1: მორფოლოგიური ტეგსეტი

აღნიშნულ ცხრილებში წარმოდგენილია ტეგსეტი, რომელიც ქართული ენის მორფოლოგიური ანალიზატორისა და გენერატორში გამოყენებულ მორფოსინტაქსურ ტეგებს შეესაბამება. აღნიშნული ტეგსეტი მიჰყვება სასრული პოზიციის მორფოლოგიის ფარგლებში სხვა ენებისათვის შემუშავებულ ჩარჩოს, რომელსაც ქართული ენის სპეციფიკური მახასიათებლები ემატება.

ცხრილი 1 ძირითადი ტეგები

განმარტება	სასრული პოზიციის (სპ) ტეგები ¹⁰
ძველი ქართული	+OGE
საშუალო ქართული	+MGE
თანამედროვე ქართული	+GE

ცხრილი 2 მეტყველების ნაწილები

მეტყველების ნაწილი	სპ ტეგები	თვისებათა რაოდენობა
არსებითი სახელი	+Noun	5
ზმნა	+Verb	11
ზედსართავი სახელი	+Adj	5
რიცხვითი სახელი	+Num	4
ნაცვალსახელი	+Pron	5
კავშირი	+Conj	1
ნაწილაკი	+Part	1

¹⁰ აღნიშნული ტეგსეტი სპეციალურად შემუშავდა შოთა რუსთაველის საქართველოს ეროვნული სამეცნიერო ფონდის მიერ დაფინანსებული პროექტების (#DP2016_23 და #AR/320/4-105/11) ფარგლებში.

გმნიზება	+Adv	2
თანდებული	+Post	1
შორისდებული	+Itj	-
აბრევიატურა	+Abr	4
პუნქტუაციის ნიშანი	+F	-

ცხრილი 3 არსებითი სახელის მორფოსინტაქსური ტეგები

არსებითი სახელი		+Noun
თვისება	დამ. თვის. მნიშვნელობა	სპ ტეგები
ტიპი	სამოგადო	+Com
	საკუთარი	+Prop
	ქვეტიპი პირადი სახელი	+Name
	გეოგრაფიული სახელი	+Geog
სულიერება/	სულიერი	+Anim
უსულოება	უსულო	+Inanim
რიცხვი	მხოლოდითი	+Sg
	მრავლობითი	+Pl
ბრუნვა	სახელობითი	+Nom
	მოთხრობითი	+Erg
	მიცემითი	+Dat
	ნათესაობითი	+Gen
	მოქმედებითი	+Ins

	ვითარებითი	+Advb
	წოდებითი	+Voc
კლიტიკები	თანდებული	+Post
ქვეტიპი	თანდებულები, -ვით, -ებრ, ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)
	თანდებული, -კენ	+Post (to)
	თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
	არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
	არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
	არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3
	ნაწილაკი	+Ptcl
	დამხმარე გმნა	+Aux
	ემფატიკური ხმოვანი	+Emph

ცხრილი 4 ზედსართავი სახელის მორფოსინტაქსური ტევები

ზედსართავი სახელი			+Adj
თვისება	დამ. თვის.	მნიშვნელობა	სპ ტევები
ხარისხი		დადებითი	+Pos
		ოდნობითი	+Com
		უფროობითი	+Sup
რიცხვი		მხოლოდითი	+Sg
		მრავლობითი	+Pl
ბრუნვა		სახელობითი	+Nom
		მოთხრობითი	+Erg
		მიცემითი	+Dat
		ნათესაობითი	+Gen
		მოქმედებითი	+Ins
		ვითარებითი	+Advb
		წოდებითი	+Voc
კლიტიკები		თანდებული	+Post
	ქვეტიპი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
		თანდებული, -თან	+Post (with/at)

თანდებული, -ზე	+Post (on)
თანდებული, -ში	+Post (in)
თანდებული, -თვის	+Post (for)
თანდებულები, -გან, -დან	+Post (from)
თანდებული, -ურთ	+Post (with)
თანდებული, -კენ	+Post (to)
თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3
ნაწილაკი	+Ptcl
დამხმარე ზმნა	+Aux
ემფატიკური ხმოვანი	+Emph

ცხრილი 5 რიცხვითი სახელის მორფოსინტაქსური ტეგები

		რიცხვითი სახელი	+Num
ფორმები		ანბანური	+Alpha
		რომაული	+Roman
		არაბული	+Arabic
თვისება	დამ. თვის.	მნიშვნელობა	სპ ტეგები
ტიპი		რაოდენობითი	+Card
		რიგობითი	+Ord
		წილობითი	+Fract
		ოდნაობითი	+Approx
		ჯერობითი	+Rep
რიცხვი		მხოლოდობითი	+Sg
		მრავლობითი	+Pl
ბრუნვა		სახელობითი	+Nom
		მოთხრობითი	+Erg
		მიცემითი	+Dat
		ნათესაობითი	+Gen
		მოქმედებითი	+Ins
		ვითარებითი	+Advb

	წოდებითი	+Voc
კლიტიკები	თანდებული	+Post
ქვეტიპი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)
	თანდებული, -კენ	+Post (to)
	თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
	არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
	არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
	არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3
	ნაწილაკი	+Ptcl

დამხმარე ზმნა	+Aux
ემფატიკური ხმოვანი	+Emph

ცხრილი 6 ნაცვალსახელის მორფოსინტაქსური ტეგები

ნაცვალსახელი		+Pron	
თვისება	დამ. თვის.	მნიშვნელობა	სპ ტევები
ტიპი		პირისა	+Pers
		ჩვენებითი	+Dem
		კუთვნილებითი	+Poss
		განუსაზღვრელობითი	+Indf
		კითხვითი	+Int
		მიმართებითი	+Rel
		ურთიერთობითი	+Recp
		უარყოფითი	+Neg
		განსაზღვრებითი	+Det
პირი		პირველი	+1
		მეორე	+2
		მესამე	+3

რიცხვი	მხოლოდითი	+Sg
	მრავლობითი	+Pl
ბრუნვა	სახელობითი	+Nom
	მოთხრობითი	+Erg
	მიცემითი	+Dat
	ნათესაობითი	+Gen
	მოქმედებითი	+Ins
	ვითარებითი	+Advb
	წოდებითი	+Voc
კლიტიკები	თანდებული	+Post
ქვეტიკი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/ at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)

თანდებული, -კენ	+Post (to)
თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3
ნაწილაკი	+Ptcl
დამხმარე ზმნა	+Aux
ემფატიკური ხმოვანი	+Emph

ცხრილი 7 ზმნის მორფოსინტაქსური ტევები

	ზმნა	+Verb
თვისება	დამ. თვის. მნიშვნელობა	სპ ტევები
ტიპი	ძირითადი	+Main
	დამხმარე	+Aux
ქვეტიპი	აბსოლუტურ-სტატიკური	+AbsStat
	რელატიურ-სტატიკური	+RelStat

	დინამიკური	+Dyn
	რელატიურ დინამიკური	+RelDyn
ფორმა	თხრობითი	+Ind
	კავშირებითი	+Subj
	ბრძანებითი	+Imp
	კაუზატივი	+Caus
	მიმღეობა	+VerbalAdj
	მასდარი (საწყისი)	+VerbalNoun
დრო და	ანწყო	+Pres
კილო	უნწყვეთელი (ნამყო უსრული)	+Imperf
	ანწყოს კავშირებითი	+PresSbj
	მყოფადი	+Fut
	ხოლმეობით-პირობითი	+FutCond
	მყოფადის კავშირებითი	+FutSbj
	წყვეტილი	+Aor
	მეორე კავშირებითი	+AorSbj
	ბრძანებითი	+AorImp
	პირველი რეზულტატივი/ თურმეობითი	+Res1

	მეორე რეზულტატივი/ თურმეობითი	+Res2
	მესამე კავშირებითი	+PerfSbj
ასპექტი	უსრული	Ipfv+
	სრული	Pfv+
რიცხვი	მხოლოდითი	+Sg
	მრავლობითი	+Pl
გარდამავლობა	გარდაუვალი	+Intrans
	ირიბ-გარდამავალი	+IndTrans
	გარდამავალი	+Trans
სუბიექტისა და	პირველი პირის სუბიექტი, მხ. რიცხვი	+Subj1Sg
ობიექტების კორელაცია,	მეორე პირის სუბიექტი, მხ. რიცხვი	+Subj2Sg
პირი, რიცხვი	მესამე პირის სუბიექტი, მხ. რიცხვი	+Subj3Sg
	პირველი პირის სუბიექტი, მრ. რიცხვი	+Subj1Pl
	მეორე პირის სუბიექტი, მრ. რიცხვი	+Subj2Pl
	მესამე პირის სუბიექტი, მრ. რიცხვი	+Subj3Pl

პირველი პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen1Sg
მეორე პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen2Sg
მესამე პირის სუბიექტი, მხ. რიცხვი, ბენეფიციარი	+SubjBen3Sg
პირველი პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen1Pl
მეორე პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen2Pl
მესამე პირის სუბიექტი, მრ. რიცხვი, ბენეფიციარი	+SubjBen3Pl
პირველი პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus1Sg
მეორე პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus2Sg
მესამე პირის სუბიექტი, მხ. რიცხვი, კაუზატორი	+SubjCaus3Sg
პირველი პირის სუბიექტი, მრ. რიცხვი, კაუზატორი	+SubjCaus1Pl
მეორე პირის სუბიექტი, მრ. რიცხვი, კაუზატორი	+SubjCaus2Pl
მესამე პირის სუბიექტი, მრ. რიცხვი, კაუზატორი	+SubjCaus3Pl
პირველი პირის ობიექტი, მხ. რიცხვი	+Obj1Sg

მეორე პირის ობიექტი, მხ. რიცხვი	+Obj 2Sg
მესამე პირის ობიექტი, მხ. რიცხვი	+Obj 3Sg
პირველი პირის ობიექტი, მრ. რიცხვი	+Obj 1Pl
მეორე პირის ობიექტი, მრ. რიცხვი	+Obj 2Pl
მესამე პირის ობიექტი, მრ. რიცხვი	+Obj 3Pl
პირველი პირის ობიექტი, ბენეფიციარი	+Obj Ben1
პირველი პირის ობიექტი, მხ. რიცხვი, ბენეფიციარი	+Obj Ben1Sg
პირველი პირის ობიექტი, მრ. რიცხვი, ბენეფიციარი	+Obj Ben1Pl
მეორე პირის ობიექტი, ბენეფიციარი	+Obj Ben2
მეორე პირის ობიექტი, მხ. რიცხვი, ბენეფიციარი	+Obj Ben2Sg
მეორე პირის ობიექტი, მრ. რიცხვი, ბენეფიციარი	+Obj Ben2Pl
მესამე პირის ობიექტი, ბენეფიციარი	+Obj Ben3
პირველი პირის ობიექტი, ლოკატივი	+Obj Loc1

პირველი პირის ობიექტი, მხ.
რიცხვი, ლოკატივი +ObjLoc1Sg

პირველი პირის ობიექტი, მრ.
რიცხვი, ლოკატივი +ObjLoc1Pl

მეორე პირის ობიექტი,
ლოკატივი +ObjLoc2

მეორე პირის ობიექტი, მხ.
რიცხვი, ლოკატივი +ObjLoc2Sg

მეორე პირის ობიექტი, მრ.
რიცხვი, ლოკატივი +ObjLoc2Pl

მესამე პირის ობიექტი,
ლოკატივი +ObjLoc3

პირველი პირის ობიექტი,
რეციპიენსი +ObjRec1

პირველი პირის ობიექტი, მხ.
რიცხვი, რეციპიენსი +ObjRec1Sg

პირველი პირის ობიექტი, მრ.
რიცხვი, რეციპიენსი +ObjRec1Pl

მეორე პირის ობიექტი,
რეციპიენსი +ObjRec2

მეორე პირის ობიექტი, მხ.
რიცხვი, რეციპიენსი +Obj2RecSg

მეორე პირის ობიექტი, მრ.
რიცხვი, რეციპიენსი +Obj2RecPl

მესამე პირის ობიექტი,
რეციპიენსი +Obj3Rec

პირველი პირის ობიექტი, კაუზაციის ადრესატი	+ObjRecCaus1
პირველი პირის ობიექტი, მხ. რიცხვი, კაუზაციის ადრესატი	+ObjRecCaus1Sg
პირველი პირის ობიექტი, მრ. რიცხვი, კაუზაციის ადრესატი	+ObjRecCaus1Pl
მეორე პირის ობიექტი, კაუზაციის ადრესატი	+ObjRecCaus2
მეორე პირის ობიექტი, მხ. რიცხვი, კაუზაციის ადრესატი	+Obj2RecCausSg
მეორე პირის ობიექტი, მრ. რიცხვი, კაუზაციის ადრესატი	+Obj2RecCausPl
მესამე პირის ობიექტი, კაუზაციის ადრესატი	+Obj3RecCaus
პირველი პირის ობიექტი, პაციენსი	+Obj1Pat
პირველი პირის ობიექტი, მხ. რიცხვი, პაციენსი	+Obj1PatSg
პირველი პირის ობიექტი, მრ. რიცხვი, პაციენსი	+Obj1PatPl
მეორე პირის ობიექტი, პაციენსი	+Obj2Pat
მეორე პირის ობიექტი, მხ. რიცხვი, პაციენსი	+Obj2PatSg
მეორე პირის ობიექტი, მრ. რიცხვი, პაციენსი	+Obj2PatPl

	მესამე პირის ობიექტი, პაციენსი	+Obj 3Pat
გვარი	I დიათეზა	+IDt
	II დიათეზა	+IIDt
	III დიათეზა	+IIIDt
ქვეტიპი	აქტივი	+Act
	ავტოაქტივი	+AutAct
	ინაქტივი	+Inact
	პასივი	+Pass
	მედიოპასივი	+MPass
ბრუნვა	სახელობითი	+Nom
	მოთხრობითი	+Erg
	მიცემითი	+Dat
	ნათესაობითი	+Gen
	მოქმედებითი	+Ins
	ვითარებითი	+Advb
	წოდებითი	+Voc
სუბიექტისა და	სუბიექტი სახელობითში	+<NomSubj>
ობიექტის	სუბიექტი მოთხრობითში	+<ErgSubj>

ბრუნვები	სუბიექტი მიცემითში	+<DatSubj>
	სუბიექტი სახელობითში, ბენეფიციარი	+<NomSubjBen>
	სუბიექტი მოთხრობითში, ბენეფიციარი	+<ErgSubjBen>
	სუბიექტი მიცემითში, ბენეფიციარი	+<DatSubjBen>
	ობიექტი სახელობითში	+<NomObj>
	ობიექტი მიცემითში	+<DatObj>
	ობიექტი მიცემითში, ბენეფიციარი	+<DatObjBen>
	ობიექტი მიცემითში, რეციპიენსი	+<DatObjRec>
	ობიექტი მიცემითში, ლოკატივი	+<DatObjLoc>
	ობიექტი მიცემითში, პაციენსი	+<DatObjPat>
	ობიექტი ნათესაობითში	+<GenObj (for)>
	სუბიექტი სახელობითში, კაუზატორი	+<NomSubjCaus>
	სუბიექტი მოთხრობითში, კაუზატორი	+<ErgSubjCaus>

	სუბიექტი მიცემითში, კაუზატორი	+<DatSubjCaus>
	ობიექტი მიცემითში, კაუზაციის ადრესატი	+<DatObjRecCaus>
კლიტიკები	თანდებული	+Post
ქვეტიპი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)
	თანდებული, -კენ	+Post (to)
	თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
	არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
	არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2

არაპირდაპირი თემის ნანილაკი, 3 პირი	+IndSpeech3
ნანილაკი	+Ptc1
დამხმარე ზმნა	+Aux
ემფატიკური ხმოვანი	+Emph

ცხრილი 8 ზმნიზედის მორფოსინტაქსური ტეგები

	ზმნიზედა	+Adv
თვისება	დამ. თვის. მნიშვნელობა	სპ ტეგები
ტიპი	ადგილის	+Loc
	დროის	+Temp
	ვითარების	+Mod
	ზომა-ოდენობის	+Quan
	მიზეზის	+Caus
	მიზნის	+Spec
	კითხვითი	+Q
	მიმართებითი	+Rel

კლიტიკები	თანდებული	+Post
ქვეტიპი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)
	თანდებული, -კენ	+Post (to)
	თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
	არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
	არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
	არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3

ცხრილი 9 კავშირის მორფოსინტაქსური ტეგები

	კავშირი	+Conj
თვისება დამ. თვის.	მნიშვნელობა	სპ ტეგები
ტიპი	მაერთებელი	+Coord
	მაქვემდებარებელი	+Subord

ცხრილი 10 ნანილაკის მორფოსინტაქსური ტეგები

	ნანილაკი	+Part
თვისება დამ. თვის.	მნიშვნელობა	სპ ტეგები
ტიპი	კითხვითი	+Q
	მიმართებითი	+Rel
	უკუთქმითი	+Proh
	დადებითი	+Aff
	გაძლიერებითი	+Int
	განუსაზღვრელობითი	+Inf
	უარყოფითი	+Neg

ცხრილი 11 შორისდებულის მორფოსინტაქსური ტეგი

	შორისდებული	+Itj
თვისება დამ. თვის.	მნიშვნელობა	სპ ტეგები

ცხრილი 12 თანდებულის მორფოსინტაქსური ტეგები

თანდებული		-
თვისება	დამ. თვის.	მნიშვნელობა
ტიპი	თანდებული	სპ ტეგები
		+Post

ცხრილი 13 შემოკლებების მორფოსინტაქსური ტეგები

შემოკლება		+Abr
თვისება	დამ. თვის.	მნიშვნელობა
ტიპი	არსებითი	სპ ტეგები
	გმნა	+Noun
	გედსართავი	+Verb
	გმნიზედა	+Adj
	გმნიზედა	+Adv
რიცხვი	მხოლოდითი	
	მრავლობითი	+Sg
		+Pl
ბრუნვა	სახელობითი	
	მოთხრობითი	+Nom
	მიცემითი	+Erg
	ნათესაობითი	+Dat
	მოქმედებითი	+Gen
		+Ins

	ვითარებითი	+Advb
	წოდებითი	+Voc
კლიტიკები	თანდებული	+Post
ქვეტიპი	თანდებულები, -ვით, -ებრ, -ებრივ, -მებრ	+Post (like)
	თანდებული, -თან	+Post (with/at)
	თანდებული, -ზე	+Post (on)
	თანდებული, -ში	+Post (in)
	თანდებული, -თვის	+Post (for)
	თანდებულები, -გან, -დან	+Post (from)
	თანდებული, -ურთ	+Post (with)
	თანდებული, -კენ	+Post (to)
	თანდებულები, -დმი, -მდე, -მდი, -მდინ, -მდის, -და, -დამ, -დამი	+Post (till)
	არაპირდაპირი თქმის ნაწილაკი, 1 პირი	+IndSpeech1
	არაპირდაპირი თქმის ნაწილაკი, 2 პირი	+IndSpeech2
	არაპირდაპირი თქმის ნაწილაკი, 3 პირი	+IndSpeech3

ნაწილაკი	+Ptcl
დამხმარე ზმნა	+Aux
ემფატიკური ხმოვანი	+Emph

ცხრილი 14 პუნქტუაციის ნიშნების მორფოსინტაქსური ტეგები

პუნქტუაციის ნიშნები		+F	
თვისება	დამ. თვის.	მნიშვნელობა	სპ ტეგები
ტიპი		წერტილი	+Period
		მძიმე	+Comma
		ფრჩხილები	+Paren
		ტირე	+Hyphen
		ძახილის ნიშანი	+ExclPoint
		კითხვის ნიშანი	+IntPoint
		ორი წერტილი	+Colon
		წერტილ-მძიმე	+Semicolon
		სამი წერტილი	+Ellipsis
		ბრჭყალები	+Quote
		ვარსკვლავი	+Star
		ნებისმიერი სიმბოლო	+Symbol

დანართი 2: მეტა-ანოტირების ტეგსები

საშუალო და ძველი ქართული ენის კოლექციები, ძირითადად, მიჰყვება ტექსტის კოდირების საინიციატივო ჩარჩოს (TEI:P5 2019). აღნიშნული ჩარჩოს ფარგლებში არსებული ყველა ტეგის აღწერა კვლევის ამოცანას არ წარმოადგენს, ამიტომაც აღნიშნულ დანართში წარმოდგენილია მხოლოდ ზოგიერთი ტეგი, რომელიც გამოიყენებოდა კორპუსის სტრუქტურირებული ანოტირების დროს.

ცხრილი 1 მეტა-ანოტირების ზოგიერთი ტეგის მნიშვნელობის აღწერა

თვისება	XML ტეგები
პროექტის დახასიათების ბლოკი	
დამფინანსებელი ორგანიზაცია	<funder>
ნამყვანი ორგანიზაცია	<principal>
პასუხისმგებელი პირის სახელი, გვარი	<name>
პასუხისმგებელი პირის ფუნქცია	<resp>
პასუხისმგებელი პირი, ორგანიზაცია	<respStmt>
პროექტის სახელწოდება	<projectDesc>
ფაილის დახასიათების ბლოკი	
ფაილის ავტორის გვარი, სახელი	<name>
ფაილის წყარო	<sourceDesc>
ფაილის ენა	<langUsage>
ფაილის ზომა, კბ	<measure quantity="" unit="kB">
ფაილის შექმნის თარიღი	<date value="2010-2020">
ფაილის შექმნის ადგილი	<rs type="city">

ინფორმაცია ფაილში შეტანილი ცვლილებების შესახებ	<revisionDesc>
I ცვლილების თარიღი	<date>
I ცვლილების ტიპი	<change>
I კორექტორის გვარი, სახელი	<name>
სტროფების რაოდენობა	<measure type="stanzas" quantity="">
სიტყვების რაოდენობა	<measure type="words" quantity="">
ნაბეჭდი ტექსტის დახასიათების ბლოკი	
ტექსტის სახელწოდება	<title>
ავტორის სახელი, გვარი	<author>
წყარო ენა	<textLang>
ტექსტის შექმნის თარიღი	<date when="1937"></date>
ტექსტის შექმნის ადგილი	<origPlace>
გამომცემელი	<publisher>
გამოცემის ადგილი	<pubPlace>
გამოცემის თარიღი	<date when="1937"></date>
რედაქტორი	<editor>
მთარგმნელი	<editor role="translator">
ილუსტრატორი	<editor role="illustrator">
ტომების/ნომრების რაოდენობა	<idno type="vol" n="1.2" />
გვერდების რაოდენობა	<extent>
ტექსტის გვერდები -დან -მდე	<extent from="1" to="2" />

ISBN/ISSN	<idno type="ISBN" />
წვდომა	<availability>
გამავრცელებელი	<distributor>
უფლებამოსილი ორგანიზაცია	<authority>
შენიშვნები	<note>
ხელნაწერის დახასიათების ბლოკი	
შენახვის ადგილი	<settlement>
საცავის სახელწოდება	<repository>
საცავის ნომერი	<idno>
კოლექციის სახელწოდება	<collection>
დამატებითი საიდ. კოდი, კატალოგის ნომერი	<altIdentifier>
ხელნაწერის ავტორის, გადამწერის ან შემდგენელის გვარი, სახელი	<name>
ხელნაწერის ავტორის ფუნქცია	<editor role="compiler">
ხელნაწერის სახელწოდება	<title>
ხელნაწერის ენა და საწერი სისტემა	
ხელნაწერის ენა და საწერი სისტემა (ასომთავრული, ნუსხური, მხედრული)	<textLang mainLang="ka" otherLangs="HEL">ენა: ძველი ქართული; შრიფტი: მხედრული</textLang>
ხელნაწერის ფიზიკური მდგომარეობის აღწერა	
ობიექტის ფორმა	<objectDesc form="codex">
მასალა, ქაღალდი	<supportDesc material="paper">
ფურცლების რაოდენობა	<extent>

ფურცლის ზომის ტიპი	<dimensions scope="all" type="leaf" unit="inch">
სიმაღლე	<height>
სიგანე	<width>
ხელნაწერის მდგომარეობა (ცვლილებების, დაზიანებების აღწერა)	<condition>
ხელნაწერის ფურცლების დანომრვის აღწერა (რექტო, ვერსო და ა.შ.)	<foliation>
ფურცლების თანმიმდევრობის დახასიათება (არეული თანმიმდევრობა და ა.შ.)	<collation>
ხელნაწერის ფორმის აღწერა	
ხელის აღწერა	<handDesc>
სკრიპტის აღწერა	<scriptDesc>
მინიატურებისა და დეკორაციების აღწერა	<decoDesc>
მინაწერები	<additions>
ისტორია	
ხელნაწერის შექმნის ადგილი	<origin>
ხელნაწერის შექმნის თარიღი-დან	<origDate notAfter="1600" notBefore="1700">
ხელნაწერის შექმნის თარიღა-მდე	<origDate notAfter="1600" notBefore="1700">
ინფორმაცია ხელნაწერის შესახებ საცავში მოხვედრამდე (არსებობის შემთხვევაში)	<provenance>
ინფორმაცია ხელნაწერის შეძენისა ან ჩუქების შესახებ	<acquisition>

Summary

This book covers the principles of morphosyntactic annotation and the main structure of the morphological analyzer and generator of Georgian language. These natural language processing tools have been developed for the corpus annotation purposes within the projects implemented under the financial support of the Shota Rustaveli National Science Foundation (grants No DP2016_23 and № AR/320/4-105/11). The results of the research are important for the computational processing of Old, Middle and Modern Georgian.

The main purpose of the monograph is to describe automatic annotation principles of Georgian morphology from the viewpoint of the finite-state technology. The book consists of three parts. The first part describes derivational and inflectional morphologies of Georgian paying special attention to some aspects important from the natural language processing perspectives. The second part includes information on computational morphology from the viewpoint of finite-state calculus and on principles of morphosyntactic annotation of Georgian language. The third part covers the development of the tokenizer and the morphological analyzer of Georgian including description of their structures and testing issues, annotation procedures of Georgian Language Corpus (GLC) and output results.

The appendices comprise two tagsets: a) finite state morphosyntactic tagset of Georgian language and, b) some tags from the meta-annotation scheme of the GLC.

Keywords: finite state morphological analysis, finite state transducer, Georgian language

Aarts, J. (1991). Intuition-based and observation-based grammars. *English Corpus Linguistics, English Corpus Linguistics.*, 44-62.

Amiridze, N. (2018). Accommodating loan verbs in Georgian: Observations and questions. *Journal of Pragmatics*, 133, 150-165.

Amiridze, N., Asatiani, R., Baratashvili, Z. (2019). Compounds or Phrases? Pattern Borrowing from English into Georgian. *Lecture Notes in Computer Science (LNCS) 11456*, 1-20.

Anderson, S. (1976). On the notion of subject in ergative. L. C. N.-ში, *Subject and topic* (გვ. 1-23). New York: Academic Press.

Anderson, S. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Antidze, J., Gulua, N. (2010). Software tools for computer realization of morphological and syntactic models of Georgian texts. *Computer Science and Telecommunications No 1*, 54-63.

Apridonidze, Ah., Chkhaidze, L. (2004 წლის 4 June). *From Georgian and into Georgian. Transliteration of Georgian Alphabet*. მოპოვებული 2020 წლის 15 January, <http://laag.iatp.org.ge/article/008.htm>-დან

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Aronson, H. (1969). Towards a formal analysis of the Georgian declension. *General linguistics* 9, 173-184.

Aronson, H. (1984). On homonymy in the Georgian verbal system. *Folia Slavica* 7, 21-37.

Aronson, H. (1989). Inflection vs. derivation in Georgian conjugation. H. Aronson-ში, *Non-Slavic languages of the USSR: linguistic studies* (გვ. 1-19).

Chicago: Chicago Linguistic Society.

Aronson, H. (1990). *Georgian: A reading grammar*. Bloomington: Slavica Publishers, Inc.

Atwell, E. (2008). Development of tag sets for part-of-speech tagging. *HSK Corpus Linguistics*, 1-26.

Bakker, D., König, E., Dahl, Ö., Haspelmath, M., Koptjevskaja-Tamm, M., Lehmann, C., Siewierska, A. (1993). *Eurotyp Guidelines*. European Science Foundation in Language Typology.

Baratashvili, Z. (2019). The Types of the Causative Construction in Georgian. *Bulletin of the Georgian National Academy of Sciences*, 13, 126-136.

Bauer, L. (2003). *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Bauer, L. (2006). Morphology: Overview. K. A. Brown-შენ, *Encyclopedia of Language & Linguistics (Second Edition)* (გვ. 719-726). Boston: Elsevier.

Beesley, K., R. & Karttunen, L. (2003). *Finite-State Morphology: Xerox Tools and Techniques*. Stanford: CSLI Publications.

Boeder, W. (1968). Über die Versionen des georgischen Verbs. *Folia Linguistica* 2, 82–152.

Boeder, W. (1969). Nachträge zum Aufsatz: Über die Versionen des georgischen Verbs. <http://www.staff.uni-oldenburg.de/winfried.boeder/download/>, 1-40.

Boeder, W. (1979). Ergative syntax and morphology in language change: the South Caucasian languages. F. Plank-შენ, *Ergativity: towards a theory of grammatical relations* (გვ. 435-480). Orlando: Academic Press.

Boeder, W. (1989). Verbal person marking, noun phrase and word order in Georgian. P. M. ed. Marácz-შენ, *Configurationality: the typology of asymmetries*

(83. 159-184). Dordrecht: Foris.

Boeder, W. (2000). Evidentiality in Georgian. L. U. Johanson-ძნ, *Evidentials: Turkic, Iranian and Neighbouring Languages* (83. 275-328). Berlin, New York: Mouton de Gruyter.

Boeder, W. (2002). Speech and thought representation in the Kartvelian (South Caucasian) languages. *Reported Discourse. A Meeting-Ground of Different Linguistic Domains. Typological Studies in Language*, 3-48.

Boeder, W. (2005). The South Caucasian languages. *Lingua* 115, 5-89.

Booij, G. (2006). Inflection and Derivation. K. A. Brown-ძნ, *Encyclopedia of Language & Linguistics (Second Edition)* (83. 654-661). Boston: Elsevier.

Booij, G. (2010). *Construction morphology*. Oxford: Oxford University Press.

Booij, G., Van Kemenade, Ans. (2003). Preverbs: An introduction. *Yearbook of Morphology 2003*, 1-11.

Bournaud, L. (2005). Metadata for Corpus Work. D. L. Practice-ძნ, Wynne, M. (83. 30-46). Oxford: Oxford Books.

Bresnan, J. (2016). *Lexical-functional syntax*. Chichester: Wiley Blackwell.

Butskhrikidze, M. (2002). *The Consonant Phonotactics of Georgian*. Utrecht: LOT.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory* (2), 113-124.

Chomsky, N., Schützenberger, M. (1963). The Algebraic Theory of Context-Free Languages. P. H. Braffort-ძნ, *Computer Programming and Formal Systems* (83. 118-161). Amsterdam: North-Holland.

Coates, R. (1999). *Word structure*. London and New York: Routledge.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and Morphology*. Oxford: Blackwell Publisher.

Datukischvili, K. L. (2005). Morphologic Processor of Georgian Language. *Tbilisi Symposium Language, Logic, Computation*. Batumi: Batumi State University.

Doborjginidze, N., Lobzhanidze, I., Gunia, I. (2012 წლის 19 12). *Georgian Language Corpus*. მოპოვებული 2019 წლის 30 10, <http://corpora.iliauni.edu.ge/-დან>

Doborjginidze, N., Lobzhanidze, I., Mirianashvili, G. (2014 წლის 24 02). *Corpus of Georgian Chronicles*. მოპოვებული <http://corpora.iliauni.edu.ge/-დან>

Erjavec, T. (2004 წლის 10 May). *MULTEXT-East Morphosyntactic Specifications*. მოპოვებული Version 3.0: <http://nl.ijs.si/ME/Vault/V3/msd/html/-დან>

Everson, M. (1991-2019). *Georgian Supplement, Range: 2D00-2D2F*. მოპოვებული 2019 წლის 16 July, <https://unicode.org/charts/PDF/U2D00.pdf-დან>

Everson, M. (1991-2019). *Georgian, Range: 10A0-10FF*. მოპოვებული 2019 წლის 16 July, <https://unicode.org/charts/PDF/U10A0.pdf-დან>

Everson, M., Gujejiani, N., Razmadze, A. (2016 წლის 24 January). *Proposal for the addition of Georgian characters to the UCS*. მოპოვებული 2019 წლის 16 July, <https://unicode.org/L2/L2016/16034-n4707-georgian.pdf-დან>

Gamkrelidze, T., Gudava, T. (1998). Kartvelian (South Caucasian) Languages. *Encyclopedia Britannica*, retrieved from <https://www.britannica.com/topic/Caucasian-languages#ref75090>.

Gerlach, B., Grijzenhout, J. (2000). *Clitics in phonology, morphology and syntax*. Amsterdam: John Benjamins.

Gippert, J. (2016). *Complex Morphology and Its Impact on Lexicology*:

The Kartvelian Case. *The XVII EURALEX International Congress* (33. 16-37). Tbilisi: Ivane Javakhishvili Tbilisi University Press.

Gippert, J., Meurer, P., Tandashvili, M. (2011 – 2019). *The Georgian national corpus*. Bergen: <http://gnc.gov.ge/gnc/page>.

Gurevich, O. (2004). On mismatches between syntax and morphology in Georgian. *International Symposium on the Typology of Argument Structure and Grammatical Relations in Languages Spoken in Europe and North and Central Asia (LENCA-2)* (33. 61-64). Kazan: Kazan State University.

Gurevich, O. (2006). A Finite-State Model of Georgian Verbal Morphology. *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL* (33. 45-48). New York: Association for Computational Linguistics.

Gurevich, O. (2006). *Constructional Morphology: the Georgian Version*. Berkeley: PhD Dissertation, University of California.

Harris, A. (1981). *Georgian Syntax: a study in Relational Grammar*. Cambridge: Cambridge University Press.

Harris, A. (1985). *Diachronic syntax: the Kartvelian Case*. New York: Academic Press.

Harris, A. (2003). Preverbs and their origin in Georgian and Udi. G. J. Boo-ji-ᄁᄁ, *Yearbook of Morphology 2003* (33. 61-87). Dordrecht: Kluwer Academic Publishers.

Harris, A., Xu, Z. (2006). Diachronic Morphological Typology. K. A. Brown-ᄁᄁ, *Encyclopedia of Language & Linguistics (Second Edition)* (33. 509-515). Boston: Elsevier.

Hewitt, B. (1995). *Georgian: A Structural Reference Grammar*. Amsterdam: John Benjamins.

Hewitt, B. (2005). *Georgian: A Learner's Grammar*. New York: Routledge.

Holisky, D. (1981). *Aspect and Georgian Medial Verbs*. New York: Caravan Books.

Holisky, D. (1981). Aspect theory and Georgian aspect. P. Z. Tedeschi-ში, *Tense and Aspect (Syntax and Semantics)* (გვ. 127-144). New York: Academic Press.

Hulden, M. (2009). მოპოვებული Foma - a finite-state compiler and C library: <https://fomafst.github.io/>-დან

Hulden, M. (2009). Foma: a finite-state compiler and library. *Association for Computational Linguistics*, 29-32.

Iacobini, C. (2006). Morphological Typology. K. A. Brown-ში, *Encyclopedia of Language & Linguistics (Second Edition)* (გვ. 278-282). Boston: Elsevier.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Delhi: Pearson Education.

Kapanadze, O. (2009). Describing Georgian Morphology with a Finite-State System. *Proceedings of the 8th international conference on Finite-State Methods and Natural Language Processing* (გვ. 114-122). Pretoria: Springer.

Kaplan, R. & Bresnan, J. (1982). Lexical-Functional Grammar: A formal system for grammatical representation. *The Mental Representation of Grammatical Relations*, 173-281.

Kaplan, R. & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20 (3), 331-378.

Karlsson, F. (1994). Computational Morphology. R. S. Asher-ში, *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon.

Karlsson, F., & Karttunen, L. (1997). Sub-Sentential Processing. R. M. Cole-ში, *Survey of the state of the art in human language technology*. (გვ. 96-100).

New York: Cambridge University Press.

Karttunen, L. (2009). Finite-State Technology. R. Mitkov-ში, *The Oxford Handbook of Computational Linguistics* (გვ. 339-358). Oxford: Oxford University Press.

Karttunen, L., Chanod, J., Grefenstette, G., Schiller, A. (1997). Regular Expressions for Language Engineering. *Natural Language Engineering*, 1-24.

Kay, P. (2002). An informal sketch of a formal architecture for Construction Grammar. *Grammars* 5(1), 1-19.

Kay, P., Fillmore, Ch. (1999). Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. *Language* 75(1), 1-33.

Kleene, S. (1956). Representation of Events in Nerve Nets and Finite Automate. *Automata Studies*, 3-42.

Koskenniemi, K. (1983). *Two-level morphology: A general computational model for word-form recognition and production*. Helsinki: University of Helsinki.

Leech, G., Wilson, A. (1996 წლის March). *EAGLES - Recommendations for the Morphosyntactic Annotation of Corpora*. მოპოვებული Browse the EAGLES Guidelines: <http://www.ilc.cnr.it/EAGLES/browse.html>-დან

Linden, K. P. (2009). Weighted Finite-State Morphological Analysis of Finnish Compounding with HFST-LEXC. *Proceedings of the NODALIDA* (გვ. 89-95). Odense: Northern European Association for Language Technology.

Lobzhanidze, I. (2019). Computational Model of the Modern Georgian Language and Search Patterns for an Online Dictionary of Idioms. In: Silva A., Stanton S., Sutton P., Umbach C. (eds) *Language, Logic, and Computation. TbiLLC 2018. Lecture Notes in Computer Science*, vol 11456, 187-208.

Lobzhanidze, I. (2019 წლის 15 November). *Wardrops' Collection Online*. მოპოვებული 2019 წლის 10 December, Wardrops' Collection Online, v1.:

Makharoblidze, T. (2009). *A Short Grammar of Georgian*. Munich: Lincom Europe.

Makharoblidze, T. (2018). On Georgian Preverbs. *Open Linguistics*, 163-183.

Manning, C., Raghavan, P. & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Manning, C., Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press.

Margvelani, L. (1999-2001). A Subsystem Analyzing Georgian Word-Forms and Its Application to Spellchecking. *Proceedings of the 3rd and 4th International Symposium on Language, Logic and Computation* (33. 1-7). Borjomi: ILLC Scientific Publications.

Mealy, G. (1955). A Method for Synthesizing Sequential Circuits. *Bell System Technical Journal I*, 1045–1079.

Melikishvili, D., Humphries, J.D., Kupunia, M. (2010). *The Georgian Verb: A Morphosyntactic Analysis*. Hyattsville, MD: Dunwoody Press.

Meurer, P. (2007). A Computational Grammar for Georgian. *Lecture Notes in Computer Science* (33. 1-15). Berlin: Springer.

Mohri, M. (1997). Finite-state transducers in language and speech processing. *Computational Linguistics* 23 (2), 269-311.

Mohri, M., Sproat, R. (1996). An efficient compiler for weighted rewrite rules. *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics* (33. 231-8). Santa-Cruz: ACL' 96.

Moravcsik, E. (1978). Reduplicative constructions. *Universals of human language, Volume 3, Word Structure*, 297–334.

Oflazer, K. (1996). Error-tolerant finite state recognition with applications to morphological analysis and spelling correction. *Computational Linguistics*, 22 (1), 73-90.

Oflazer, K., Saraclar, M. (2018). *Turkish Natural Language Processing*. Cham: Springer International Publishing.

Peterson, D. (2007). *Applicative constructions*. Oxford: Oxford University Press.

Rayfield, D. (2006). *A Comprehensive Georgian-English Dictionary*. London: Garnett.

Schmid, H. (2005). მოპოვებული Stuttgart Finite State Transducer Tools: <https://www.cis.uni-muenchen.de/~schmid/tools/SFST/>-დან

Schmid, H. (2005). A Programming Language For Finite State Transducers. *Proceedings of the 5th International Workshop on Finite State Methods in Natural Language Processing (FSMNLP 2005)* (გვ. 308-309). Helsinki: Springer.

Schützenberger, M. (1961). A remark on finite transducers. *Information and Control* 4, 185-196.

Spencer, A., Luis, A. (2012). *Clitics: an introduction*. New York: Cambridge University Press.

Sproat, R. (1992). *Morphology and Computation*. Cambridge: MIT.

Standards, I. (2012). *Language Resource Management – Morpho-syntactic Annotation Framework (MAF)*, No 24611. <https://www.iso.org/standard/51934.html>.

Stump, G. (2001). *Inflectional morphology: a theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, G. (2002). Morphological and syntactic paradigms: arguments for a theory of paradigm linkage. G. v. Booij-ში, *Yearbook of morphology 2001* (გვ.

147–180). Dordrecht: Kluwer.

TEI Consortium, e. (2019 წლის 16 07). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. მომოვეებული 2019 წლის 29 11, <https://tei-c.org/guidelines/p5/-დან>

Trost, H. (2009). Morphology. R. Mitkov-ში, *The Oxford Handbook of Computational Linguistics* (გვ. 25-48). Oxford: Oxford University Press.

Tschenkeli, K. (1965). *Georgisch-Deutsches Wörterbuch*. Zürich: Amira-ni-Verlag.

Tuite, K. (1984). Case attraction and case agreement. *Eastern States Conference on Linguistics, 1* (გვ. 110-121). Ohio: Ohio State University.

Tuite, K. (1998). *Kartvelian morphosyntax: Number Agreement and Morphosyntactic Orientation in the South Caucasian Languages*. Munich: LINCOM.

Tuite, K. (2019). On the origin of Kartvelian “version”. https://www.academia.edu/40077656/On_the_origin_of_Kartvelian_version_, 1-50.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proc. London Math. Soc.* 2 (42), 230-265.

Turing, A. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, (გვ. 230-265).

Tuskia, M. (2010). *saxeluri da saxelzmnuri carmoeba k'art'ulši (Noun and verbal noun derivation in Georgian)*. Tbilisi: Georgian Academy of Sciences.

Uí Dhonnchadha, E. (2002). *An Analyser and Generator for Irish Inflectional Morphology using Finite-State Transducers*. Dublin: DCU Online Research Access Service.

Uí Dhonnchadha, E. (2009). *Part-of-Speech Tagging and partial parsing*

for Irish using finite-state transducers and constraint grammar. Dublin: DCU Online Research Access Service.

Vogel, P., Comrie, B. (2000). *Approaches to the typology of word classes*. Berlin: Mouton de Gruyter.

Vogt, H. (1968). *brunvat'a sistema žvel k'art'ulši* (Case system in Old Georgian). *Reviewer*, 251-284.

Weir, T. (2011). *Georgian morphosyntax and feature hierarchies in natural language*. Chicago: ProQuest LLC.

Wynne, M. (2005). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books for the Arts and Humanities Data Service.

Zipf, G. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge: Harvard University Press.

Zwicky, A. (1977). *On clitics*. Bloomington: Indiana University Linguistics Club.

Zwicky, A. (1985). Clitics and Particles. *Language* 61, 283-305.

Zwicky, A. (1985). How to describe inflection. *Proceedings of the eleventh annual meeting of the Berkeley Linguistics Society*. (გვ. 372–386). Berkeley: Berkeley.

Zwicky, A. (1990). Inflection as a (sub)component of morphology. W. e. Dressler-ში, *Contemporary morphology*. (გვ. 217-236). Berlin - New York: Mouton de Gruyter.

აბესაძე, ნ. (1956). პირის (ადამიანის) სახელთა ბრუნება ძველსა და ახალ ქართულში. სახელის ბრუნების ისტორიისათვის ქართველურ ენებში, 129-137.

აბულაძე. (1973). *ძველი ქართული ენის ლექსიკონი*. თბილისი: მეცნიერება.

ანთიძე, ჯ. (2009). *ფორმალური ენები და გრამატიკები*. თბილისი: ნეკერი.

არაბული, ა. (1984). *მესამე სერიის ნაკვეთელთა წარმოება და მნიშვნელობა ძველ ქართულში*. Tbilisi: მეცნიერება.

არაბული, ა. (2011). ინვერსია. გ. გოგოლაშვილი-ში, *თანამედროვე ქართული ენის მორფოლოგია* (გვ. 296-301). თბილისი: მერიდიანი.

ასათიანი, რ. (1989). კაუზაცია და კონტაქტი ქართველურ ენებში. *მაცნე (ბიულეტენი)*, 1, 119-129.

აფრიდონიძე, შ. (1986). *სიტყვათგანლაგება ახალ ქართულში*. თბილისი: მეცნიერება.

ბერიძე, მ. (1998). სტატიკურ ზმნათა ერთი ჯგუფისა და გარდამავალ ზმნათა ოთხეულის წარმოების საკითხისათვის. *საენათმეცნიერო ძიებანი*, 68-74.

გაბუნია, კ. (2016). *ქართული ენის გრამატიკის ზოგადი კურსი*. თბილისი: საქართველოს განათლებისა და მეცნიერების სამინისტრო.

გამყრელიძე, თ., კიკნაძე, ზ., შადური, ი., შენგელაია, ნ. (2008). *თეორიული ენათმეცნიერების კურსი*. თბილისი: თბილისის უნივერსიტეტის გამომცემლობა.

გიგაშვილი, ქ. (2004). თემის ნიშანთა და ზმნისწინთა ისტორიული ურთიერთმიმართებისათვის ქართულში. *საენათმეცნიერო ძიებანი*, 17, 50-60.

გოგოლაშვილი, გ., არაბული, ა., სუხიშვილი, მ., მანჯგალაძე, მ., ჭუმბურიძე, მ., ჯორბენაძე, ნ. (2011). *თანამედროვე ქართული ენის მორფოლოგია*. თბილისი: მერიდიანი.

დათუკიშვილი, ქ. (1992). ზმნური კატეგორიების კლასიფიკაცია უღლების სისტემასთან მიმართებით. *საენათმეცნიერო ძიებანი*, 52-61.

დათუკიშვილი, ქ. (1996). ვნებიოთის ყალიბის მქონე სტატიკური ფორმები ქართულში. *საენათმეცნიერო ძიებანი*, 73-77.

დათუკიშვილი, ქ. (1997). პირის ნიშანთა სისტემა ქართულში. *გურაბ ქუმბურიძეს (დაბადების 70 წლისთავისადმი მიძღვნილი კრებული)*, 66-76.

დანელია, კ. (1975). უცხო ენათა გავლენის კვალი ძველი ქართული წერილობითი ძეგლების ენაში. *მაცნე*, 79-90.

ერთელიშვილი, ფ. (1965). ქცევის საკითხისათვის ქართულში. *თბილისის სახელმწიფო უნივერსიტეტის შრომები*, 14 (გვ. 177-198).-ში თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

ერთელიშვილი, ფ. (1980). *სახელურ ფუძეთა ფონემატური სტრუქტურისა და ისტორიის საკითხები ქართულში*. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

ზურაბიშვილი, თ. (1956). ემფატიკური -ა ახალ ქართულში. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 224-234.

ზურაბიშვილი, თ. (1972). ემფატიკური ა ხმოვანი ახალ ქართულში. *ქართული სიტყვის კულტურის საკითხები*, 44-57.

თოფურია, ვ. (1956). -მე, -ვე, -ლა, -ლაც(ა) ნაწილაკიან სახელთა ბრუნება. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 131-139.

თოფურია, ვ. (1956). წოდებითი ბრუნვისათვის. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 36-48.

იმნაიშვილი, დ. (1952). უარყოფითი ნაცვალსახელები და უარყოფითი ზმნისართები იბერიულ-კავკასიურ ენებში. *იბერიულ-კავკასიური ენათმეცნიერება*, 4, 53-71.

იმნაიშვილი, ი. (1956). წრფელობითი ბრუნვის საკითხი საკუთარ სახელებში. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 59-76.

იმნაიშვილი, ი. (1957). *სახელთა ბრუნება და ბრუნთა ფუნქციები ძველ*

ქართულში. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

იმნაიშვილი, ი. (1968). ვნებითი გვარის ზმნათა თავისებურებანი ძველ ქართულში. *ძველი ქართული ენის კათედრის შრომები II*, 27-54.

იმნაიშვილი, ივ., იმნაიშვილი, ვ. (1996). *ზმნა ძველ ქართულში*, 2 ტ. მაინის ფრანკფურტი.

მარტიროსოვი, ა. (1958). აბსტრაქტულ სახელთა წარმოება და საწარმოებელ აფიქსთა შედგენილობა ძველ ქართულში. *იბერიულ-კავკასიური ენათმეცნიერება*, 121-127.

მარტიროსოვი, ა. (1964). *ნაცვალსახელი ქართველურ ენებში*. თბილისი: მეცნიერება.

მელიქიშვილი, დ. (1979). მოქმედებითი გვარის ზმნის უღლების სისტემა თანამედროვე ქართულში. *მაცნე (ბიულეტენი)*, 1, 84-87.

მელიქიშვილი, დ. (1980). პირისნაცვალსახელთა ფუძის აგებულებასთან დაკავშირებული ზოგიერთი საკითხი. *ნარკვევები იბერიულ-კავკასიურ ენათა მორფოლოგიიდან*, 48-58.

მელიქიშვილი, დ. (2001). ზმნის სტრუქტურა და კონსტრუქცია ქართულ ენაში დიათეზისა და გვარების თეორიის კონტექსტში. *ვარლამ თოფურია 100*, 135-142.

მელიქიშვილი, დ. (2001). *ქართული ზმნის უღლების სისტემა*. თბილისი: ლოგოსი.

მელიქიშვილი, დ. (2009). ინვერსია ქართულ ზმნაში დიაქრონიული და სინქრონიული ასპექტით. *ფილოლოგიური ძიებანი*, 589-592.

მელიქიშვილი, დ. (2009). ინკლუზივ-ექსკლუზივის კატეგორიის გამოხატვის ისტორიისათვის ქართულ ზმნაში. *ფილოლოგიური ძიებანი*, 116-123.

მელიქიშვილი, დ. (2014). *ქართული ზმნის სისტემური მორფო-სინტაქ-*

სური ანალიზი. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

რუხაძე, გ., კობლაძე, გვ. (2008). *ძველი ქართული ენის შეერთებული ლექსიკონი*. თბილისი: საქართველოს საპატრიარქო.

სარჯველაძე, ზ. (1997). *ძველი ქართული ენა*. თბილისი: თბილისის სახელმწიფო პედაგოგიური უნივერსიტეტი.

სუხიშვილი, მ. (1986). *გარდამავალი გზები ქართულში, სისტემისა და ისტორიის ზოგი საკითხი*. თბილისი: საქართველოს მეცნიერებათა აკადემია.

უთურგაიძე, თ. (1976). *ქართული ენის ფონემატური სტრუქტურა*. თბილისი: მეცნიერება.

უთურგაიძე, თ. (1986). *ქართული ენის სახელის მორფონოლოგიური ანალიზი*. თბილისი: მეცნიერება.

უთურგაიძე, თ. (2001). *გრამატიკული კატეგორიებისა და მათი ურთიერთმიმართებისათვის ქართულ გზაში*. თბილისი: ქართული ენა.

ფოჩხუა, ბ. (1974). *ქართული ენის ლექსიკოლოგია*. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

ღლონტი, ა. (1964). *ქართული ლექსიკოლოგია*. თბილისი: ცოდნა.

შანიძე, ა. (1942). *გზაზე გარდამავლობის საკითხისათვის ქართველურ ენებში*. *მეცნიერებათა აკადემიის ბიულეტენი*, 3, 182-189.

შანიძე, ა. (1956). *წოდებითი ფორმის ადგილისათვის გრამატიკაში*. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 48-56.

შანიძე, ა. (1961). *გრამატიკული სუბიექტი ზოგიერთ გარდაუვალ გზასთან ქართულში*. *ძველი ქართული ენის კათედრის შრომები*, 207-238.

შანიძე, ა. (1967). *ორობითი რიცხვის საკითხისათვის ხევსურულში*. *თბილისის სახელმწიფო უნივერსიტეტის შრომები*, 23-27.

შანიძე, ა. (1973). *ქართული გრამატიკის საფუძვლები, მორფოლოგია, I. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.*

შანიძე, ა. (1976). *ძველი ქართული ენა. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.*

შანიძე, ა. (1984). *ქართული კილოები მთაში, I. თბილისი: მეცნიერება.*

შანიძე, მ. (1956). *რომელ ნაცვალსახელის ფუნქციისა და ადგილისათვის ძველ ქართულში. სახელის ბრუნების ისტორიისათვის ქართველურ ენებში, 140-143.*

ჩართოლანი, ნ. (1985). *ჩვენებით ნაცვალსახელთა სისტემები ქართულში სხვა ქართველურ ენებთან შედარებით. თბილისი: მეცნიერება.*

ჩიქობავა, ა. (1934). - *გან თანდებულის ხმარებისათვის ნათესაობითთან და მოქმედებითთან. ქართველურ ენათა სტრუქტურის საკითხები, 13-17.*

ჩიქობავა, ა. (1936). *ქანურის გრამატიკული ანალიზი ტექსტებითურთ. თბილისი: სსრკ მეცნიერებათა აკადემია-საქართველოს ფილიალი .*

ჩიქობავა, ა. (1937). *ერთი უცნობი თანდებული ახალ ქართულში. ენიმკის მოამბე, 1, 55-65.*

ჩიქობავა, ა. (1940). *მესამე პირის სუბიექტის უძველესი ნიშანი ქართველურ ენებში. ენიმკი-ს მოამბე V-VI, 13-46.*

ჩიქობავა, ა. (1942). *სახელის ფუძის უძველესი აგებულება ქართველურ ენებში. თბილისი: მეცნიერება.*

ჩიქობავა, ა. (1946). *ინვერსიული ზმნები და სახელთა კლასიფიკაცია. თბილისი: სახალხო განათლება.*

ჩიქობავა, ა. (1950-1964). *ქართული ენის განმარტებითი ლექსიკონი. თბილისი: მეცნიერებათა აკადემია.*

ჩიქობავა, ა. (1953). *მასდარისა და მიმღეობის ისტორიული ურთიერთობისათვის ქართულში. იბერიულ-კავკასიური ენათმეცნიერება V, 33-49.*

ჩიქობავა, ა. (1954). მრავლობითის სუფიქსთა გენეზისათვის ქართულში . *იბერიულ-კავკასიური ენათმეცნიერება*, 67-76.

ჩიქობავა, ა. (1956). მრავლობითის სუფიქსთა გენეზისისათვის ქართულში. ვ. თოფურია-ში, *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში* (გვ. 313-323). თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

ჩიქობავა, ა. (1956). მრავლობითობის სუფიქსთა გენეზისათვის ქართულში. *სახელის ბრუნების ისტორიისათვის ქართველურ ენებში*, 313-325.

ჩიქობავა, ა. (1961). თანდებულიან ბრუნვათა საკითხისათვის ქართულში. *ქართული ენის სტრუქტურის საკითხები II* , 197-208.

ჩიქობავა, ა. (1968). *მარტივი წინადადების პრობლემა ქართულში*. თბილისი: მეცნიერება.

ჩიქობავა, ა. (2008). *ენათმეცნიერების შესავალი*. თბილისი: თბილისის სახელმწიფო უნივერსიტეტი.

ჩიქობავა, ა. (2013). გრამატიკული კლას-კატეგორია და ზმნის უღვლილებების ზოგი საკითხი ძველ-ქართულში. *იბერიულ-კავკასიური ენათმეცნიერება*, 397-410.

ცოცანიძე, გ., ლოლაძე, ნ., დათუკიშვილი, ქ. . (2014). *ქართული ლექსიკონი*. თბილისი: ბაკურ სულაკაური.

აბლაუტი	16, 17, 21	აპოკოპე.....	17
აბრევიატურა.79, 82, 86, 89, 102		არაბული ფორმა.....	35, 62, 106, 136
აბსოლუტური დასასრული	30	არაპირდაპირი თქმის ნაწილაკი..	37, 38, 46
აბსოლუტური დასაწყისი.....	30, 53	არსებითი სახელი	7, 12, 20, 34, 36, 51, 55, 61, 72, 76, 78, 80, 101, 102
აბსოლუტურ-სტატიკური ზმნა ..	39, 110	ასომთავრული	35, 60, 86, 89, 128
აბსტრაქტული ძირი.....	62	ასპექტი.....	1, 6, 9, 19, 21, 23, 39, 40, 42, 44, 46, 50, 75, 112, 137
აგენსი.....	20, 40, 44, 48	ატომური გამოსახულება.....	30
აგლუტინაციური	9, 45, 49, 59	ატრიბუტი.34, 35, 39, 40, 41, 42, 43	
ადგილის ზმნიზედა	120, 138	აუცილებელი შემონმების ოპერატორი	67
ავტოაქტივი.....	19, 40, 41, 117	აფიქსაცია.....	10, 47
ავტოაქტიური	45, 58	აფიქსებად სიტყვის დაშლის მეთოდოლოგია	23
ავტოაქტიური სუბიექტი.....	40	აფრიკატი	88
ავტოპასივი	19, 40, 41	აქტივი	19, 40, 41, 117
ავტოპასიური.....	45	აქტიური.....	31, 35, 40, 45, 58
ამკრძალავი შემონმების ოპერატორი	67	ანმყო.19, 42, 43, 46, 47, 48, 62, 65, 111	
ამომცნობი	24, 96, 99	ანმყოს კავშირებითი	111
ამოცნობის პროცენტი.89, 92, 93, 96, 97, 99, 100			
ანბანური ფორმა.....	106		
ანკერი33, 52, 54, 66, 68, 69, 70, 71, 72, 73, 74, 76			
ანოტირება.1, 33, 39, 59, 62, 74, 85, 94, 95, 126			

ბენეფიციარი..20, 41, 42, 44, 48, 113, 114, 118	გარდამქმნელი 28
ბილიკი 67	გარდაუვალი.....40, 58, 63, 112
ბმული მორფემა.....10	გაძლიერებითი ნაწილაკი.....122
ბოლო პოზიცია 27, 28	გეოგრაფიული სახელი 34
ბრუნება 36	გვარი 19
ბრუნვის ნიშანი 37	გრამატიკული ომონიმია.....59
ბრძანებითი 19, 43, 111	დადებითი თვისებების ოპერატორი 66
ბრძანებითი კილო 19	დადებითი ნაწილაკი 122
ბრჭყალები 125	დადებითი ხარისხი 104
ბუნებრივი ენის დამუშავება ..1, 6, 7, 22, 23, 24, 25, 33, 49, 55, 58, 59, 89	დამხმარე გმნა 13, 37, 38, 39, 46
გადაკვეთა 31, 32	დატიური კონსტრუქცია.....43
გაერთიანება.....32	დერივაციული 11, 13, 24, 43, 57, 77
გაერთიანების ოპერატორი.....67	დიათეზა .19, 40, 45, 58, 97, 117
განსაზღვრებითი ნაცვალსახელი . 36, 108	დიაკრიტული ნიშნული38, 51, 52, 65, 66, 68, 69, 72, 76
განუსაზღვრელობითი	დიეზის ნიშანი 30
ნაცვალსახელი 36, 108	დინამიკური 111
განუსაზღვრელობითი ნაწილაკი.. 122	დინამიკური გმნა 39, 111
გარდამავალი....40, 58, 63, 112, 137	დრო 42, 44
გარდამავალი მატრიცა 27	დროის გმნიზედა.....120
გარდამავლობა 40, 41, 112	ემფატიკური ხმოვანი..13, 36, 37, 38
	ენის დაფარვის ტესტი 95

კითხვითი ნაწილაკი	122	ლოკალიზაცია	31
კითხვის ნიშანი	125	ლოკატივი.....	44, 114, 115, 118
კილო ...	19, 21, 42, 44, 111, 138	მაერთებელი კავშირი.....	122
კლინის ვარსკვლავი	30, 31	მარტივი ავტომატი	28
კლინის პლუსი	31	მასდარი	62, 63, 111, 138
კლიტიკი ..	13, 36, 103, 104, 107, 109, 119, 121, 124	მატება	31, 36, 43, 46, 60, 76
კომბინირებული მორფი	16	მაქვემდებარებელი კავშირი..	122
კომპიუტერული მორფოლოგია....	21, 22, 26	მეორე დიათეზა.....	40, 41
კომპოზიტურობა.....	10, 14	მეორე თურმეობითი.....	19
კონკატენცია	31	მეორე კავშირებითი	43, 111
კონტექსტზე დამოკიდებული ენა..	26	მეორე რეზულტატივი	43, 112
კონტექსტზე დამოუკიდებელი ენა	26	მესამე დიათეზა.....	40, 41
კუთვნილებითი ნაცვალსახელი ..	36	მესამე კავშირებითი ..	19, 43, 112
კუთვნილებითი ნაცვალსახელი ...	108	მიელის მანქანა.....	29
კუმშვა.....	17	მიზეზის ზმნიზედა	120
ლემა .	23, 33, 46, 47, 48, 50, 55, 57, 60, 61, 62, 64, 65, 76, 77, 80, 89, 99, 139	მიზნის ზმნიზედა	120
ლემატიზაცია	61	მიმართებითი ზმნიზედა	120
ლემატიზაციის პრობლემა	46	მიმართებითი ნაცვალსახელი.	36, 108
		მიმართებითი ნაწილაკი	122
		მიმღეობა.....	12, 43, 78, 111
		მიცემითი ბრუნვა.....	19, 36, 102, 104, 106, 109, 117, 123
		მოთხრობითი ბრუნვა	19, 102, 104, 106, 109, 111, 117

მორფოლოგიური პროცესი ..1, 9, 59	ნათესაობითი ბრუნვა 13, 17, 19, 36, 44, 53, 102, 104, 106, 109, 117, 118, 123, 138
მორფოსემანტიკური..... 7	ნაცვალსახელი 12
მორფოსინტაქსური ..1, 8, 15, 16, 19, 20, 22, 23, 24, 25, 33, 34, 38, 39, 45, 55, 58, 61, 62, 75, 78, 83, 84, 90, 96, 99, 101, 102, 104, 106, 108, 110, 120, 122, 123, 125	ნაწილაკი.....12
მორფოტაქტიკა....23, 38, 57, 58, 69, 74	ნომინატიური კონსტრუქცია 43
მორფოფონემური 7	ნულოვანი ენა 28
მოქმედებითი ბრუნვა .17, 18, 19, 36, 53, 102, 104, 106, 109, 117, 123, 137, 138	ნულოვანი მორფი 15, 61
მრავლობითი რიცხვი .17, 18, 19, 36, 52, 53, 62, 70, 71, 102, 104, 106, 109, 112, 123, 138, 139	ნუსხური 35, 60, 86, 128
მსკდომი 88	ობიექტი...19, 40, 41, 44, 45, 48, 58, 63, 113, 114, 115, 116, 117, 118, 119, 128
მყოფადი .19, 42, 43, 46, 47, 48, 111	ოდნაობითი რიცხვითი სახელი19, 77, 106
მყოფადის კავშირებითი 111	ოდნაობითი ხარისხი 104
მძიმე 125	ორი წერტილი 125
მწკრივი 42	ორპირიანი 45
მხედრული 86	პარსინგი 59
მხოლოდობითი რიცხვი...15, 17, 18, 19, 36, 47, 48, 51, 52, 53, 62, 65, 70, 102, 104, 106, 109, 112, 123	პასივი.....19, 40, 41, 58, 117
	პასივის ნიშანი 46
	პასიური 45
	პაციენსი 20, 40, 41, 44, 48, 116, 118
	პირადი სახელი.....34
	პირდაპირი ობიექტი.....19, 58
	პირველი დიათეზა 40, 41

პირველი რეზულტატივი ..43, 111	რეციპიენსი 41, 44, 115, 118
პირთა კორელაციის ნიშანი 41	რიგობითი რიცხვითი სახელი .35, 77, 106
პირიანობა 45	რიცხვი...7, 9, 15, 17, 18, 19, 20, 21, 34, 36, 37, 41, 42, 45, 48, 51, 52, 53, 61, 62, 65, 70, 76, 77, 81, 101, 102, 104, 106, 109, 112, 123, 138
პირის ნაცვალასხელი 35	რიცხვითი სახელი 12, 34, 35, 36, 77
პირის ნაცვალსახელი 108	რკალი.....28
პირის ნიშანი.....45, 46	რომაული ფორმა.....35, 106
პირობითობა.....31	სავრცობი.....46
პოლიპერსონალიზმი.....48	საზოგადო არსებითი სახელი..34, 51, 72, 76, 102
პრეფიქსი.....38	საკუთარი არსებითი სახელი...34, 72, 102
პროექცია.....32	სამი წერტილი.....125
რაოდენობითი რიცხვითი სახელი 35, 77, 106	სამპირიანი 45
რეგრესული ტესტირება.....89	სასრული პოზიციის ენა.....28
რეგულარული გამოსახულება.27, 28, 29, 30, 31, 32, 33, 49, 52, 53, 55, 74, 76	სასრული პოზიცია 27
რეგულარული გრამატიკა.....26	სასრული პოზიციის გარდამქმნელი 1, 23, 25, 29, 30, 39, 47, 48, 49, 50, 51, 55, 83, 84, 85
რეგულარული ენა 27, 29	სასრული პოზიციის მორფოლოგიური ანალიზი.1, 23, 55, 80
რეგულარული რელაცია.....29	
რეკურსულად თვლადი გრამატიკა 26	
რელატიურ დინამიკური ზმნა ..39, 111	
რელატიურ-სტატიკური ზმნა....39, 110	
რელაცია .30, 31, 32, 41, 44, 45, 112	

სასრული პოზიციის ოპერატორი	23	სუბიექტი	19, 36, 41, 42, 43, 44, 48, 63, 83, 112, 113, 117, 118, 119, 138
სასრული პოზიციის რელაცია ..	28	სუბიექტ-ობიექტის კორელაციის ნიშნები	45
სასრული პოზიციის ტეგი	35	სულიერება/უსულოების კატეგორია.....	61
სასრული პოზიციის ქსელი.29, 52		სულიერი.....	102
საწყისი	63	სუპლექტივიზმი	18, 21
საწყისი პოზიცია.....	27	სუფიქსი	63, 74, 79
სახელობითი ბრუნვა ..19, 36, 51, 62, 65, 102, 104, 106, 109, 117, 123		ტეგების ლექსიკური გრამატიკა	90
სახელური პარადიგმა .34, 74, 79		ტეგირება	7, 59
სემანტიკა	7	ტეგსეტი8, 24, 38, 39, 78, 90, 92, 101, 126	
სერია.....	19	ტირე.....	125
სიგმა-ვარსკვლავი	31	ტოკენი.....	49, 60
სიგმა-პლუსი	31	ტოკენიზატორი.1, 59, 85, 86, 87, 88, 89, 93, 96	
სინკოპე	17	ტოკენიზაცია	49, 60, 86
სიტყვაცვლილება.....	9, 12	ტურინგის მანქანა.....	26
სიტყვანარმოება.....	9, 12, 15	უარყოფითი ნაცვალსახელი ...	36, 108
სიტყვანარმოებითი	11, 12	უარყოფითი ნაწილაკი.....	122
სიტყვანარმოებითი მორფოლოგია	14	უკუთქმითი ნაწილაკი	122
სლოტი	15, 61, 65	უნიკოდი	60, 86, 93
სონანტი.....	88	ურთიერთობითი ნაცვალსახელი..	
სრული ასპექტი	112		
სტატისტიკური მეთოდოლოგია	23		

36, 108	ცარიელ რიგი 28
უსრული ასპექტი 39, 112	ცარიელი მორფი.....16
უსულო.....102	ცარიელი რიგის ენა 28, 31
უფრობითი ხარისხი 104	ციფრის კანონი.....92
უცხო სიტყვების შესება 13	ცირკუმფიქსი.....11, 34, 37, 65
უნწყვეთელი.....42, 46, 111	ძახილის ნიშანი.....125
ფლექსიური 57	ძირი 45
ფლექსიური მორფოლოგია 9, 10, 11, 13, 24, 33, 45, 50, 62, 76, 77, 100	ძირითადი ზმნა 39
ფორმალური გრამატიკა 26	ძირის მონაცვლეობა 18, 47
ფრიკატივი.....88	წერტილი 125
ფრჩხილები 125	წერტილ-მძიმე 125
ქართული ენის კორპუსი 1, 85, 93	წილობითი რიცხვითი სახელი.35, 77, 106
ქსელი.....28, 51, 90, 92	წოდებითი ბრუნვა 19, 36, 103, 104, 107, 109, 117, 124, 137, 138
ქსეროქსის კალკულუსი38, 53, 65	წრე.....28
ქცევის კატეგორია.....45	წყვეტილი 19, 43, 46, 111
შეკავშირება.....32	ხარისხი 36
შემაჯალი სიგნალი.....26	ხმოვანი 88
შორისდებული .8, 12, 13, 21, 56, 74, 79, 82, 102, 122	ხმოვანფუძიანი 20, 21, 71
ჩვენებითი ნაცვალსახელი.....36	ხმოვნის მონაცვლეობა.....47, 70
ჩვენებითი ნაცვალსახელი.....108	ხოლმეობით-პირობითი....19, 42, 111
ჩომსკი-შუტცენბერგის	ჯერობითი რიცხვითი სახელი..77, 106
იერარქია.....26	