

პირითაღი და დამხმარე სახელმძღვანელოში
თანამდებობები მათგანატიპურ ენათმეცნიერებაში
2007 №2

სამეცნიერო-საგანმანათლებლო ჟურნალის
– „ქართული ენა და ლოგიკა“ –
ყოველწლიური დამატება

გამოცემის რედაქტორი
პ. ჭავაძე

რ. ჰაუსერი

გამოთვლითი ენათმეცნიერების საფუძვლები

ცალილი II

გრამატიკული
თეორიები

იგანე ჯავახიშვილის სახელობის თაილისის სახელმძღვანელო უნივერსიტეტი
ილია გევარას სახელობის გამოყენებითი გათმაატიპის ინსტიტუტი

გამოთვლითი ენათმეცნიერების საფუძვლები

მომავალზე ორიენტირებული გამოთვლითი ენათმეცნიერების ძირითადი ამოცანაა ისეთი კოგნიტური მანქანების კონსტრუირება, რომლებთანაც ადამიანებს შეეძლებათ თავისუფლად ისუბრონ თავიანთივე ბუნებრივი ენის ფარგლებში. უკვე დღიდ ხანია ამ ამოცანით ისაზღვრება ენის ფუნქციონალური თეორიის, ვერიფიკაციის თანამედროვე მეთოდებისა და სხვა გამოყენებითი ხასიათის საკითხების ფართო სპექტრი. ბუნებრივი კომუნიკაცია ითხოვს არამხოლოდ სიტყვითი, აგრეთვე არასიტყვითი მოცემულობების აღქმასა და დამუშავებას. შესაბამისად, სახელმძღვანელო კურსში განიხილება ენის ის თეორია, რომელიც აუცილებელია მოსაუბრების კონსტრუირებისთვის. ამასთან, წინამდებარე კურსის ძირითადი თემაა მოლაპარაკესა და შემჩენელს შორის არსებული ბუნებრივენობრივი კომუნიკაცია.

როლანდ ჰაუსერი, 1999 წელი, ივნისი.

ნაწილი II გრამატიკული თეორიები

მთარგმნელ-რედაქტორი –

- კონსტანტინე ჭხაპაშვილი

მთარგმნელები –

- ლაშა აბზიანიძე
- ნიკო ვასარია
- სანდრო მასხარაშვილი
- ნიკო ჭხაპაშვილი
- ბესო ჩიქვიძე

სახელმძღვანელო კურსი ოთხ ნაწილად იყოფა: I. ენობრივი თეორიები, II. გრამატიკული თეორიები, III. მორფოლოგია და სინტაქსი, IV. სემანტიკა და პრაგმატიკა. სახელმძღვანელო კურსის I ნაწილში, რომლის ქართული ვარიანტის მომზადება ჯგუფის უახლოეს გეგმებშია, ასენილია ბუნებრივი კომუნიკაციის მექანიზმი, რომელიც დაფუძნებულია ენის SLIM თეორიის [2+1] დონის სტრუქტურაზე და განსხვავებული ტიპის ენობრივ ნიშნებზე. კურსის II ნაწილში ფურადღება გამახვილებულია სინტაქსის გრამატიკული კომპონენტის ფარგლებში შედგენილი ნიშნების იერარქიული აგების მეთოდებზე. ეს ფორმალური ენების თეორიის ფარგლებში შემუშავებული ის ფართოდ გამოყენებადი მეთოდებია, რომლებიც მნიშვნელოვან როლს თამაშობენ ლოგიკისა და მათემატიკის დაფუძნების პრობლემატიკაშიც. ამასთან, II ნაწილის მიზანია ლინგვისტური ლირებულების მქონე ფორმალური ცნებებისა და მეთოდების რაც შეიძლება მარტივი გადმოცემა და მათი სუსტი და ძლიერი შხარების მიმოხილვა.

ქართული ტექსტის კონვენტორ-რედაქტორები –

- ვილ. მეტრ. ლოგოტორი გახა გაბუნია
- ვილ. მეტრ. ლოგოტორი ბადრი ცხადაშვილი

სამეცნიერო-საგანგაოთლებლო ჟურნალი
– ქართული ენა და ლოგიკა –
glli_pkhakadze@caucasus.net
უნივერსიტეტის ქ. № 2
ISSN 1512 – 2840

2006 — 2007
№3, №4, №5, №6

რედაქტორი
– კონსტანტინ ჭავაძე

რედაქტორის
მოაღმილე
– გახა გაგულია

სარედაქტორი

საბჭო:
ბადრი ჭავაძე
გიორგი ჩიჩუა
ლაშა აბზიანიძე
ნინო ვახანია
სანდრო გასხარაშვილი
ნინო ჭავაძე
ნინო ლაპაძე
ბესო ჩიქვინიძე

SCIENTIFIC-EDUCATIONAL JOURNAL
- THE GEORGIAN LANGUAGE AND LOGIC -
glli_pkhakadze@caucasus.net
2 UNIVERSITY St
ISSN 1512 - 2840

2006 — 2007
№3, №4, №5, №6

Chair

- Konstantine Pkhakadze

Co-Chair

- Kakha Gabunia

Editorial Board:

Badri Tskhadadze

Giorgi Chichua

Lasha Abzianidze

Niko Vachania

Sandro Mascharashvili

Niko Pkhakadze

Nino Labadze

Beso Chikvinidze

გამოიცემა ქართულ-გველიცარული საჭარმოს
„ვისოლის“ ქვეყლობილი მსარლაჭვრით

PUBLISHED WITH THE SPONSORSHIP OF THE
GEORGIAN-SWISS ENTERPRISE “WISOL”

სარჩევი

7	თავი – წარმომქმნელი გრამატიკა	4
7.1	ენა როგორც თავისუფალი მონოიდის ქვესიმრავლე	4
7.2.	წარმომქმნელი გრამატიკის უპირატესობათა განშაზღვრელი მეთოდოლოგიური მიზეზები	9
7.3	წარმომქმნელი გრამატიკის ადექვატურობა	11
7.4	C-გრამატიკის ფორმალიზმი	13
7.5	C-გრამატიკა ბუნებრივი ენისათვის სავარჯიშოები	17 20
8	თავი – ენის სირთულე და ენობრივი იერარქიები	23
8.1.	PS-გრამატიკის ფორმალიზმი	23
8.2.	ენობრივი კლასები და გამოთვლითი სირთულე	26
8.3.	წარმომქმნითი უნარი და ფორმალურ ენათა კლასები	29
8.4.	PS-გრამატიკა ბუნებრივი ენისათვის	35
8.5.	შემადგენელი სტრუქტურის პარადოქსი სავარჯიშოები	40 45
9	თავი – ანალიზის ძირითადი ცნებები	48
9.1	ანალიზის დეკლარაციული და პროცედურული ასპექტები	48
9.2	გრამატიკის მორგვება ენაზე	51
9.3	ტიპობრივი ტრანსფარენტულობის დამოკიდებულება გრამატიკებსა და პარსერებს შორის	56
9.4	შემავალ-გამომავალი ეპივალენცია მსმენელ-მოლაპარაკე სისტემასთან	63
9.5	გრამატიკის უკმარისობა კონვერგენციის მისაღწევად სავარჯიშოები	66 69
10	თავი – მარცხნივ ასოცირებადი გრამატიკა	71
10.1	წესთა ტიპები და დერივაციული რიგი	71
10.2	LA-გრამატიკის ფორმალიზმი	75
10.3	დროში წრფივი ანალიზი	79
10.4	LA-გრამატიკის სრული ტიპობრივი ტრანსფარენტულობა	82
10.5	LA-გრამატიკა ბუნებრივი ენისათვის სავარჯიშოები	85 92
11	თავი – LA გრამატიკის იერარქია	94
11.1	შეუზღუდავი LA -ების წაროქმნითი უნარი	94
11.2	A-, B- და C-LAG-ების LA-იერარქია	94
11.3	ორაზროვნება LA-გრამატიკაში	97
11.4	გრამატიკებისა და ავტომატების სირთულე	100
11.5	C1-, C2-, და C3-LAG-ების ქვეირარქია სავარჯიშოები	107 114
12	თავი – LA- და PS- იერარქიების ურთიერთმიმართება	116
12.1	LA- და PS- გრამატიკების ენობრივი კლასები	116
12.2	ქვესიმრავლების მიმართებები ამ ორ იერარქიაში	118
12.3	LA- და PS- იერარქიების არაეკვივალენტობა	120
12.4	უმცირესი LA- და PS- კლასების შედარება	123
12.5	ბუნებრივი ენების წრფივი სირთულე	125
	სავარჯიშოები	131

7 თავი ტარმომქმედი გრამატიკა

7.1 ქვეთავში ენა ფორმალური მიდგომებით ისაზღვრება ოოგორც სასრული ლექსიკონით მოცემული თავისუფალი მონოიდის ქვესიმრავლე. 7.2 ქვეთავში მიმოხილულია მათემატიკური, პრაქტიკული და გამოთვლითი ხასიათის არგუმენტები იმისა, თუ რატომ უნდა ვისარგებლოთ ბუნებრივი ენის აღსაწერად წარმომქმნელი გრამატიკით. 7.3 ქვეთავი გვარწმუნებს, რომ ბუნებრივი ენების საანალიზოდ წარმომქმნელი გრამატიკის მეთოდების გამოყენება აუცილებელია, მაგრამ არასაკმარისია იმაში დასარწმუნებლად, რომ ამ მიდგომებით მიღებული გრამატიკა სტრუქტურული თვალსაზრისებით იძლენად კარგად მოერგება ენობრივ სისტემას, რომ ფართო კლასის პრაქტიკული ამოცანების წარმატებულად გადაწყვეტისათვის საჭიროება მისი უფრო ღრმა და ვრცელი ანალიზის გაკეთებისა აღარ იქნება. 7.4. ქვეთავში აღწერილია ის ისტორიულად პირველი წარმომქმნელი გრამატიკა, რომელიც ბუნებრივი ენების საანალიზოდ შემუშავდა, სახელდობრ კატეგორიათა გრამატიკა, ანუ **C**-გრამატიკა. 7.5 ქვეთავი წარმოგვიდგენს **C**-გრამატიკის ფორმალურ გამოყენებას ინგლისური ენის მცირე „ფრაგმენტისათვის“.

7.1 მნა როგორც თავისუფალი მონოიდის ქვესიმრავლე

ფორმალური ენების თეორიაში მუშავდება მათემატიკური მეთოდები, რომლებიც იძლევიან გრამატიკული ანალიზის პრაქტიკულ ხერხებსა და შეტყობინებათა დამუშავების რამდენადაც შესაძლებელია დამოუკიდებელ საშუალებებს. ეს ყველაფერი თვალნათლივ ისახება მათი მეშვეობით გააზრებულ ენის აბსტრაქტულ ცნებაში.

7.1.1 მნის განსაზღვრება

ენა არის სიტყვითი მიმდევრობების სიმრავლე.

თავად ეს სიმრავლე (სიმრავლეთა თეორიის გაგებით) დაულაგებელია. მთებედავად ამისა, ამ სიმრავლის თითოეული წევრი არის სიტყვების დალაგებული მიმდევრობა (**ordered sequence of words**).

ფორმალური ენის განსაზღვრისას უმნიშვნელოვანეს საკითხად დგება გრამატიკულად მართებულად აგებული (**grammatically well-formed**) მიმდევრობების გამოყოფის, ანუ ასეთი მიმდევრობების მახასიათებლების დადგენის საკითხი. ამის გამოა, რომ გრამატიკულად მართებულად აგებული მიმდევრობების ამგებ სიტყვებს განიხილავენ ისეთ მარტივ გამოსახულებებად (**simple surfaces**), რომელთაც არ გააჩნიათ არც არანაირი მახასიათებელი კატეგორიები, არც არანაირი განსხვავებული და ძირეული ფორმები და არც არანაირი მნიშვნელობა (**meaning₁**).

მაგალითად, **a** და **b** სიტყვებით წარმოქმნილი **LX={a,b}** სიტყვათა სიმრავლე არის ქვემოთ განსაზღვრული ფორმალური (ე.ი. აბსტრაქტული) ენის ლექსიკონი, ანუ ალფაბეტი¹. **LX** სიტყვათა სიმრავლის წევრების სხვადასხვაგვარი კომბინირება იძლევა ამ სიტყვებით შედგენილი მიმდევრობების უსასრულო ოდენობას. (ამას უზრუნველყოფს ის, რომ ამ მიმდევრობების სიგრძე არ არის შემოსაზღვრული). ასეთი სიტყვების ამ ყველა შესაძლო მიმდევრობათა უსასრულო სიმრავლეს სასრულ ლექსიკონზე განსაზღვრულ თავისუფალ მონოიდს (**free monoid over a finite lexicon**) უწოდებენ.

7.1.2 LX={a,b} სიმრავლეზე განსაზღვრული თავისუფალი მონოიდი

ε

a,b,

aa, ab, ba, bb

aaa, aab, aba, abb, baa, bab, bba, bbb

aaaa, aaab, aaba, aabb, abaa, abba, abbb...

...

LX სიმრავლეზე განსაზღვრული თავისუფალი მონოიდი **LX** სიმრავლის კლინის ჩაკეტვადაც (**Kleene closure**) იწოდება და მას **LX*** გამოსახულებით აღნიშნავენ. **LX*** თავისუფალი მონოიდი **ε** ნეიტრალური წევრის (**neutral element**) გარეშე **LX⁺** გამოსახულებით აღნიშნება და იგი **LX** სიმრავლის დადებით ჩაკეტვად (**positive closure**) იწოდება.² თავისუფალი მონოიდის **ε** ნეიტრალურ წევრს ზოგჯერ ცარიელ (**empty**), ზოგჯერ კი ნულოვან (**zero**) მიმდევრობასაც (**sequence**) უწოდებენ.

თავისუფალი მონოიდის ცნება ერთნაირად მოხერხებულია როგორც ბუნებრივი, ისე ხელოვნური ენაბისათვის. თუმცა, ბუნებრივ ენებს შედარებით დიდი ლექსიკონები აქვთ: ცხადია, რომ თუკი ენის ყველა განსახვებული სიტყვა ფორმა იქნება გათვალისწინებული, მაშინ ეს ლექსიკონები შეიძლება შეიცავდნენ რამოდენიმე მილიონ მონაცემსაც კი. მიუხედავად ამისა, ბუნებრივი ენების ლექსიკონები გვანან ხელოვნური ენების ლექსიკონებს იმით, რომ ორივენი სასრულია. არადა მათთან მიკავშირებული თავისუფალი მონოიდი უსასრულოა. ეს იმიტომ, რომ თავისუფალი მონოიდი შეიცავს სიტყვათა ყველა შესაძლო მიმდევრობას, მაშინ როდესაც ამ მიმდევრობათა უმეტესობა გრამატიკულად არალირებულია ამ მონოიდთან მიკავშირებული ბუნებრივი თუ ხელოვნური ენისათვის. იმისთვის, რომ გაიფილტროს გრამატიკულად არამართებულად აგებული სიტყვითი მიმდევრობები, საჭიროა ფორმალური კრიტერიუმები, რომლითაც ჩვენ ერთმანეთისაგან განვარჩევთ გრამატიკულად მართებულად და არამართებულად აგებულ მიმდევრობებს.

¹ ფორმალურ ენათა თეორიაში ხელოვნური ენის ლექსიკონს ზოგჯერ აღფაბეტს უწოდებენ, სიტყვას – ასოს, წინადადებას – სიტყვას. ლინგვისტური თვალსაზრისით ეს პრაქტიკა შეცდომაში შემყვანია და იგი არც არანაირი აუცილებლობით არ არის განპირობებული. ამგვარად, ჩვენი მიღვომებით, როგორც ბუნებრივი, ისე ხელოვნური ენის ძირითადი გამოსახულება სიტყვად იწოდება (მაშინაც კი, როცა ეს სიტყვა შედგება მხოლოდ ერთ ასოსგან, მაგალითად **a** ასოსგან). ასევე, ორივე სახის ენების მართებულად აგებული გამოსახულებები ჩვენთან წინადადებად იწოდებიან (იმ შემთხვევაშიც, როცა ეს გამოსახულება არის მხოლოდ ერთ ასოიანი სიტყვების მიმდევრობა, მაგალითად **aaabbb**).

² სხვა სიტყვებით: **LX** სიმრავლეზე განსაზღვრული თავისუფალი მონოიდი იგივეა რაც **LX⁺ ∪ { ε }** (იხ. მ. ჰარისონი, 1978, გვ. 3)

ხელოვნურ ენებში გრამატიკულად მართებულად აგებულობის (**grammatical well-formedness**) ცნების მკაცრი ფორმალური განსაზღვრება სირთულეს არ წარმოადგენს. ასეთ ენებში მართებულად აგებულობის ცნებას თავად ამ ენის ამგები ადამიანები საზღვრავენ. ასე მაგალითად: ვთქვათ კვაზიულობის კლასი **a^kb^k (k≥1)** გამოსახულებით აღნიშნულ ხელოვნურ ენას როგორც სიმრავლეს ყველა იმ სიტყვებისა, რომელთაგან ნებისმიერი შესდგება **a** სიმბოლოს გარკვეული ოდენობითა და მის გვერდზე მიწერილი **b** სიმბოლოს იგივე ოდენობით. ეს ენა არის საკუთრივი ქვესიმრავლე 7.1.2. მაგალითით მოცემული **{a,b}** სიმრავლეზე განსაზღვრული სიტყვითი მიმდევრობების სიმრავლისა.

a^kb^k ხელოვნური ენის ამ არაფორმალური აღწერის თანახმად გამოსახულებები **ab, aabb, aaabbb, aaaabbbb** და ა.შ.. მართებულად აგებული გამოსახულებებია, რაც თავის მხრივ იმასაც ნიშნავს, რომ 7.1.2 მაგალითით მოცემული თავისუფალი მონოიდის ყველა სხვა სახის გამოსახულება არ არის შესაბამისობაში **a^kb^k** ენის ზემოაღწერილ სტრუქტურასთან. მაგალითად, ისეთი გამოსახულებები, როგორებიცაა ვთქვათ **a, b, ba, bbaa, abab**, და ა.შ.. არ არიან **a^kb^k** ენის მართებულად აგებული გამოსახულებები. ანალოგიურად **a^kb^k** ხელოვნური ენის ამ აღწერისა შეგვიძლია აღვწეროთ ისეთი ხელოვნური ენებიც როგორებიცაა **a^kb^kc^k, a^kb^mc^kb^m**, და ა.შ..

არამხოლოდ სასრულ ლექსიკონზე განსაზღვრული თავისუფალი მონოიდი შეიცავს უსასრულოდ ბევრ გამოსახულებას: ასევე შესაძლებელია არსებობდეს ენები – როგორც თავისუფალი მონოიდის ქვესიმრავლები – რომლებიც აგრეთვე შეიცავენ უსასრულოდ³ ბევრ მართებულად აგებულ გამოსახულებასა და წინადადებას. ამით არის განპირობებული ის, რომ თავისუფალი მონოიდებისაგან ენის ფილტრაციის კრიტერიუმებს უფრო რთული აღნაგობა აქვთ, ვიდრე უბრალო სიებს.⁴

ის, რაც ამ შემთხვევაში მარტივ სიას ენაცვლება, არის მართებულად აგებული გამოსახულებების ისეთი სტრუქტურული აღწერა, რომ იგი გამოყენებადი იყოს უსასრულო ოდენობის (ანუ ყოველწილიერ განახლებადი) გამოსახულებების სრულიად განსხვავებული ტიპებისათვის. ფორმალურ ენათა თეორიაში ასეთი სტრუქტურული აღწერები წარმომქმნელ გრამატიკებად იწოდება და ისინი გააზრებულია როგორც რეკურსული წესების სისტემები (**recursive rule systems**), რაც უკვე ლოგიკიდანაა ნასესხები.

a^kb^k ხელოვნური ენის წარმომქმნელი გრამატიკის შემდეგი მაგალითი იყენებს ფორმალიზმს, რომელიც დღეს **PS**-გრამატიკის⁵ სახლითაა ცნობილია

7.1.3 PS-ბრამატიკა **a^kb^k** ენისათვის

$$\begin{aligned} S &\rightarrow a \ S \ b \\ S &\rightarrow a \ b \end{aligned}$$

³ ის, რომ უსასრულო სიმრავლეების ქვესიმრავლე თავადვე შეიძლება იყოს უსასრულო მტკიცდება ღუწი რიცხვების მაგალითით. მართლაც 2,4,6.... არის 1,2,3,4..., ნატურალური რიცხვების უსასრულო ქვესიმრავლე. თავად ნატურალური რიცხვების სიმრავლე კი არის 0,1,2,3,4,5,6,7,8,9 ციფრების სასრულ ლექსიკონზე მიწერის (**concatenation**) (მაგ. 1 და 2 ციფრების მიწერით მიიღება 12 და 21) ოპერაციით აგებული უსასრულო სიმრავლე.

⁴ მიზეზი იმისა, რომ შეუძლებელია ნატურალური რიცხვების სრული სიის გაკეთება არის ის, რომ ციფრებისაგან მიწერის ოპერაციისა და სტრუქტურული პრინციპის მეშვეობით ნატურალური რიცხვების უსასრულოდ ბევრი გამოსახულება წარმოიქმნება.

⁵ **PS**-გრამატიკის დეტალური აღწერა მოცემულია მე-8 თავში.

PS-გრამატიკის წარმოქმნითი უნარი ემყარება წარმომქმნელი წესის ისრის მარცხენა მხარეს განთავსებული გამოსახულების ამავე ისრის მარჯვენა მხარეს განთავსებული გამოსახულებით ჩანაცვლების უმარტივეს ფორმალურ პრინციპს. სამაგალითოდ განვიხილოთ ზემოთ $a^k b^k$ ენისათვის განსაზღვრულ **PS**-გრამატიკაში **aaabbb** სტრიქონის წარმოქმნა. პირველად გამოიყენება **S** \rightarrow **aSb** წესი. ისრის მარცხენა მხარეს მდებარე **S** გამოსახულების ისრის მარჯვენა მხარეს არსებული გამოსახულებით ჩანაცვლება იძლევა

aSb სტრიქონს.

შემდეგ ისევ პირველი წესი გამოვიყენოთ: უკვე ნაწარმოებ **aSb** გამოსახულებაში **S** სიმბოლო ჩავანაცვლოთ **aSb** გამოსახულებით. ამით იწარმოება ახალი

aaSbb სტრიქონი.

და ბოლოს, მეორე წესის თანახმად **aaSbb** სტრიქონში **S** სიმბოლო ჩავანაცვლოთ **ab** გამოსახულებით. შედეგად ვლებულობთ

aaabbb სტრიქონს.

ამგვარად, როგორც ვნახეთ, სტრიქონი, რომლის წარმოქმნის საკითხსაც განვიხილავდით, წარმოიქმნება 7.1.3 გრამატიკის წესების მარტივი ფორმალური გამოყენებით. ეს ამტკიცებს იმას, რომ იგი $a^k b^k$ ენის მართებულად აგებული გამოსახულებაა.

მიუხედავად იმისა, რომ 7.1.3 გრამატიკა ეყრდნობა სასრულ **{a,b}** ლექსიკონსა და იყენებს სასრულ რაოდენობა წესებს ის წარმოქმნის $a^k b^k$ ენის უსასრულოდ ბევრ გამოსახულებას. ეს ფორმალურად განპირობებულია პირველი წესის **რეკურსულობით (recursion)**: **S** ცვლადს, რომელიც პირველი წესის ისრის მარცხენა მხარეს მდებარე გამოსახულებას წარმოადგენს, შემოსვლა აქვს ამავე წესის ისრის მარჯვენა მხარეს მდებარე გამოსახულებაშიც, რაც განაპირობებს იმას, რომ ამ წესით შეიძლება ვიმოქმედოთ ამავე წესის გამოსაგალზე (**output**) (ანუ, ამავე წესით გამოყვანილ (ე.ი. უკვე წარმოქმნილ) გამოსახულებაზე). მეორე მხრივ, 7.1.3 გრამატიკის მეორე წესი ნიმუშია **PS**-გრამატიკის არარეცურსული (**non-recursion**) წესისა.

ხელოვნური და ბუნებრივი ენების ანალიზის ზემოთ ნაწილობრივ აღწერილი წარმომქმნელი მეთოდოლოგია არ არის შეზღუდული ერთი რომელიმე კონკრეტული გრამატიკული ფორმალიზმით. უფრო მეტიც, ეს მეთოდოლოგია საშუალებას იძლევა ავირჩიოთ ერთი რამოდენიმესაგან.⁶ ფორმალურ ენათა თეორიაში ასეთი ძირეული (ე.ი. ელემენტარული) ფორმალიზმები უკვე დიდი ხანია განსაზღვრულია.

7.1.4 ჭარმომქმნელი გრამატიკის ელემენტარული ფორმალიზმები

- კატეგორიათა, ანუ **C**-გრამატიკა
- ფრაზათა სტრუქტურული, ანუ **PS**-გრამატიკა
- მარცხნივ-ასოცირებადი, ანუ **LA**-გრამატიკა

ეს ელემენტარული ფორმალიზმები ერთმანეთისაგან განსხვავდებიან თავთავიანთი საბაზისო ცნებებითა და გამოყვანის წესებით (იხ. 10.1.3, 10.1.4 და 10.1.5) და აგრეთვე გამოყვანის წესების რიგითობით (იხ. 10.1.6). ელემენტარული ფორმალიზმების ფორმალურ საფუძვლებს იძლევა მათი ალგებრული განსაზღვრებები (**algebraic definition**).

⁶ სამაგალითოდ, შეადარეთ ერთმანეთს $a^k b^k$ ენის **PS**- და **C**- გრამატიკული ანალიზები, რომელიც 7.1.3 და 7.4.4 განაწერებშია გაკეთებული.

7.1.5 ალგებრული განსაზღვრება

წარმოქმნელი გრამატიკის ალგებრული განსაზღვრება ცხადად წარმოგვიდგენს სისტემის ძირითად შემადგენლებს იმით, რომ იძლევა ამ შემადგენლებისა და მათი სტრუქტურული ურთიერთობიმართებების განსაზღვრებებს მხოლოდ სიმრავლეთა თეორიის ცნებების მეშვეობით.

C-გრამატიკის ალგებრული განსაზღვრება მოცემულია 7.4.2, **PS**-გრამატიკისა 8.1.1, ხოლო **LA**-გრამატიკის 10.2.1 განაწერში. – ელემენტარული ფორმალიზმი გარდა ზემოაღნიშნულისა ითხოვს მისი ყველაზე უფრო მეტად ღირებული მათემატიკური თვისებების განსაზღვრას. კერძოდ, ეს გულისხმობს ამ ელემენტარული ფორმალიზმების ქვეტიპებით წარმოქმნად ენათა კლასის იერარქიულ დალაგებას მათი ურთიერთ ქვეტიპობისა და სირთულის მიხედვით. აგრეთვე უნდა დადგინდეს ამ ელემენტარულ ფორმალიზმებს შორის არსებული ფორმალური დამოკიდებულებები.

ამ ძირეულ ელემენტარულ ფორმალიზმებზე დაყრდნობით უკვე განსაზღვრულია მრავალი მათგან წარმოებული ფორმალიზმი (**derived formalisms**). ეს პროცესები წლებია გრძელდება. მიზანი ამ ახალ-ახალი წარმოებული ფორმალიზმების განსაზღვრისა არის ელემენტარული ფორმალიზმების სიღრმისეულად დამახასიათებელი სისუსტეების დაძლევა. ზოგიერთი მაგალითი ასეთი წარმოებული ფორმალიზმებისა, რომლებიც ამა თუ იმ დროს პოპულარობით სარგებლობდნენ, ჩამოთვლილია 7.1.6 და 7.1.7 განაწერებში.

7.1.6 PS-გრამატიკის წარმოებული ფორმალიზმები

სინტაქსური სტრუქტურები, წარმოქმნელი სემანტიკები, სტანდარტული თეორია (**ST**), გაფართოებული სტანდარტული თეორია (**EST**), შემოწმებული გაფართოებული სტანდარტული თეორია (**REST**), მართვა და კავშირი (**GB**), ბარიერები, განზოგადებულ ფრაზათა სტრუქტურული გრამატიკა (**GPSG**), ლექსიკურად ფუნქციონალური გრამატიკა (**LFG**), თავით მართული ფრაზათა სტრუქტურული გრამატიკა (**HPSG**).

7.1.7 C-გრამატიკის წარმოებული ფორმალიზმი

მონტეგიუს გრამატიკა (**MG**), ფუნქციონალურად უნიფიცირებადი გრამატიკა (**FUG**), კატეგორიათა უნიფიცირებადი გრამატიკა (**CUG**), კომბინატორული კატეგორიათა გრამატიკა (**CCG**), უნიფიცირებაზე დაფუძნებული კატეგორიათა გრამატიკა (**UCG**)

ხაზგასასმელია ისც, რომ უკვე იყო გამოთქმული მოსაზრებები **PS**- და **C**- გრამატიკების ერთ წარმოებულ ფორმალიზმი კომბინირებაზე, როგორც ეს არის კუპერის გრამატიკაში, რათა მოხდეს თითოეული მათგანის სუსტი მხარეების კომპენსაცია და დამატებითი სარგებლის მიღება მათივე ძლიერი მხარეებიდან.

წარმოებული ფორმალიზმების მათემატიკური თვისებები უნდა დახასიათდეს ისევე ცხადად და ზუსტად როგორც ეს არის მოთხოვნილი ძირეული ელემენტარული ფორმალიზმებისათვის. ერთი შესველით, ივარაუდება, რომ წარმოებული ფორმალიზმები იმყოფებან უპირატეს მდგომარეობაში და რომ მათი შესაბამისი ელემენტარული ფორმალიზმების უკვე შესწავლილმა თვისებებმა უნდა გაგვიაღვილოს მათი ანალიზი. მიუხედავად ამისა, აღმოჩნდა, რომ წარმოებული ფორმალიზმების მათემატიკური ანალიზი მათი შესაბამისი ელემენტარული ფორმალიზმების მათემატიკურ ანალიზზე

არანაკლებ როტულია. ამ სახის კვლევითი პროცესები ხშირად ათწლეულები გრძელდებოდა და, შესაბამისად, არც თუ იშვიათად იყო შეცდომებიც.⁷

ძირული და წარმოებული ფორმალიზმების გარდა არსებობენ აგრეთვე წარმომქმნელი გრამატიკის ნახევრად ფორმალური სისტემები (**semi-formal systems**). რომელთა აღწერაც არ არის დაუუმნებული აღგებრულ განსაზღვრებებზე. მაგალითად, ასეთებია: **დამოკიდებულებათა გრამატიკა (dependency grammar)** (ტენიერი, 1959) და **სისტემური გრამატიკა (systemic grammar)** (ჰალიდე, 1985). ასეთი სისტემების მკვლევარები ეყრდნობინ იმ ჰიპოტეზურ მოსაზრებას, რომ შესასწავლი სისტემის თვისებები გარკვეული ხარისხით სიმილარულია რომელიმე უკვე ცნობილი ფორმალური სისტემის თვისებების. მაგალითად, დამოკიდებულებათა გრამატიკასა და C-გრამატიკას შორის უდავოდ არის გარკვეული სტრუქტურული მსგავსება.

გასაგებ მიზეზთა გამო მათემატიკური თვისებების ანალიზისა და პრაქტიკული ვარგისიანობის საკითხი განსაკუთრებული ყურადღებით განხილული იქნება მხოლოდ ძირული, ანუ ელემენტარული ფორმალიზმებისთვის. საკითხი დაისმის იმის თაობაზე, თუ ამ ძირული ფორმალიზმებიდან რომელია უფრო ვარგისი ბუნებრივი ენის (i) ლინგვისტური ანალიზისა და (ii) ავტომატური პარსერის კონსტრუირების მიზანთა ჭრილში. ამავე მიზნებით იხსნება კონცეპტუალური და ფორმალური მიზეზები იმისა, თუ რატომ იგება ასე მრავლად ძირული ფორმალიზმების ახალ-ახალი წარმოებული ფორმალიზმები.

7.2 წარმომხელი გრამატიკის უპირატესობათა განსაზღვრელი მეთოდოლოგიური მიზანები

ისეთი ხელოვნური ენებისგან განსხვავებით, როგორებიცაა მაგალითად **a^kb^k** და **a^kb^kc^k**, რომლებიც წინასწარ განსაზღვრული მიზნებითაა შექმნილი (მათ შორის, გარკვეული სირთულის მქონე თეორიული საკითხების განხილვის მიზნებით), ბუნებრივი ენები მოიცემა მათი შესაბამისი ენობრივი საზოგადოებით. ეს იმას ნიშნავს, რომ ბუნებრივი ენის გამოსახულების გავება ან გრამატიკულად მართებულად, ან გრამატიკულად არამართებულად აგებულ გამოსახულებად დამოკიდებულია მთქმელის ენობრივ ინტუიციაზე. მაგალითად: ის, რომ 7.2.1 განაწერით მოცემული გამოსახულება ინგლისური ენის გრამატიკულად მართებულად აგებული გამოსახულებაა, არა სადაო ინგლისური ენის მცოდნეთათვის.

7.2.1 გრამატიკულად განთვალისწინებული გამოსახულება

the little dogs have slept earlier

მეორე მხრივ კი, შემდეგ გამოსახულებას, რომელიც მიიღება პირველი გამოსახულების სიტყვების რიგის შეცვლით, ინგლისური ენის ნებისმიერი მცოდნე უარყოფს და მას გრამატიკულად არამართებულად აგებულ გამოსახულებად მიიჩნევს.

⁷ მაგალითად: 6. ჩომსკი თავიდან ფიქრობდა რომ წაშლების აღდგენის პირობა (**recoverability condition of deletions**) შეინარჩუნებდა ტრანსფორმაციული გრამატიკის ამოხსნადობის (**decidable**) თვისებას (იხ. ქვეთავი 8.5), რაც უარყოფილ იქნა 1972 წელს ს. პეტერსისა და რ. რიჩის მიერ. გ. გაზდარი თავიდან თვლიდა, რომ მეტაწესების შემთხვევას **GPSG** სისტემაში არ უნდა გაეზარდა ამ სისტემის თავდაპირველი კონტექსტისგან თავისუფალი სირთულე, რაც უარყოფილი იქნა 1986 წელს ჰ. უშკორეულისა და ს. პეტერსის კვლევებით.

7.2.2 გრამატიკულად არამართებულად აგნაზული გამოსახულება

*earlier slept have dogs little the⁸

ენობრივად მართებულად და არამართებულად აგებულ გამოსახულებებს შორის არსებული ინტუიციური განსხვავებების საფუძველზე წარმომქმნელი გრამატიკები ბუნებრივი ენებისთვისაც შეიძლება დაიწეროს. ისევე როგორც ფიზიკური კანონები იძლევიან საშუალებას წინასწარ გამოვთვალოთ დროის გარკვეულ მომენტში ციური სხეულების სივრცული მდებარეობა, ასევე ბუნებრივი ენის შესატყვისი, ანუ სრულად აღმწერი წარმომქმნელი გრამატიკის წესები უნდა იძლეოდეს ფორმალურ გადაწყვეტებს იმისა, არის თუ არა ესა თუ ის ენობრივი გამოსახულება ამ ენის გრამატიკულად მართებულად აგებული გამოსახულება.

ერთი შეხედვით, თეორიული ლინგვისტიკის ეს მიზანი შეიძლება ზედმეტად აკადემიურადაც კი შეფასდეს. თუმცა, სინამდვილეში, გასაგები ხდება, რომ წარმომქმნელი გრამატიკების გამოყენება თანამედროვე ლინგვისტურ მეთოდოლოგიებში სრულიად აუცილებელია.

7.2.3 ფარმომქონელი გრამატიკის მეთოდოლოგიური მიღზომები

ემპირიკული: იძლევა ცხადი ჰამოყალიბების საშუალებას:

ნებისმიერი წარმომქმნელი გრამატიკული ანალიზი ჯამდება ისეთი ფორმალური წესების სისტემით, რომელიც შეიცავს ცხად ჰამოციებებს, იმის შესახებ თუ რომელი შესავალი (**input**) გამოსახულებაა მართებულად აგებული და რომელი არა. ეს ცხადი სახის ჰამოციებები ნათელს ხდის თუ რომელი ფორმალური გრამატიკა არის ემპირიკულად ადეკვატური და რომელი არა. — ეს კი, თავის მხრივ, მნიშვნელოვანი წინაპირობაა ემპირიკული აღწერის შემდეგი საფეხურებრივი გაუმჯობესებისა.

მათემატიკური: ფორმალური თვისებების განსაზღვრა:

მხოლოდ მკაცრად ფორმალიზებული აღწერა გვაძლევს ამ აღწერების ისეთი მათემატიკური თვისებების⁹ ანალიზის საშუალებას, როგორიცაა ამოხსნადობა, სირთულე და წარმოქმნითი უნარი (ე. ი. შესაძლებლობა). გრამატიკული ფორმალიზმის მათემატიკური თვისებები კი, თავის მხრივ, საშუალებას იძლევიან განვსაზღვროთ რამდენად გამოყენებადია იგი კონკრეტული პრაქტიკული აღწერებისა და გამოთვლითი რეალიზაციებისათვის.

გამოთვლითი: აპარატის აღწერილობითი განსაზღვრა:

მხოლოდ ფორმალურ წესთა სისტემა შეიძლება იქნეს გამოყენებული პარსერის (**parser**), ანუ სინტაქსური ანალიზატორის დეკლარაციული დახსასიათებისათვისა (**declarative specification**)¹⁰ და მისი იმ აუცილებლად საჭირო თვისებების განსასაზღვრავად, რომლებიც განსხვავდება პროგრამისტის მიერ სისტემაში შემთხვევით ჩადებული

⁸ ლინგვისტიკაში არაგრამიტიკული სტრუქტურის მქონე მაგალითები მონიშნულია ვარსკვლავით *. ეს შეთანხმება დამკვიდრა და ბლუმფილდმა 1933 წელს.

⁹ ცხადია, რომ არაფორმალური აღწერის მათემატიკური თვისებები შეუძლებელია გამოკვლეული იქნას მათი სტრუქტურების არასაქმარისი სიცხადის გამო

¹⁰ ჯერ კიდევ არსებობენ პროგრამები, რომლებიც არ არიან დაუუძნებული აღწერილობით განსაზღვრებაზე (დეკლარაციულ სპეციფიკაციაზე). თუმცა, მიუხედავად იმისა, რომ ეს პროგრამები უკვე დიდი ხანია მუშაობენ ჯერ კიდევ არ არის ცხადი მათი რომელი თვისებაა თეორიულად აუცილებელი და რომელი არის შემთხვევითი შედეგი პროგრამისტის თვითშემოქმედებისა. ასეთ პროგრამებს უროებს ებახიან, და ისინი თეორიული თვალსაზრისებით მხოლოდ მცირედ არიან საინტერესონი, პრაქტიკული თვალსაზრისებით კი ძნელად ექვემდებარებიან თანდათანობითი გაუჭვიობებების პროცესებს, რაც მათ სწრაფ მივიწყებას განაპირობებს. კავშირი გრამატიკულ სისტემებსა და მათ რეალიზაციებს შორის მოგვიანებით განიხილება 15.1 ქვეთავში.

თვისებებისაგან. ამასთან, პარსერის კონსტრუირება განპირობებულია და, შესაბამისად, იძლევა იმ ავტომატური ენობრივი ანალიზის გაკეთების შესაძლებლობას, რომელიც აუცილებელია წარმომქმნელი ფორმალიზმის ფარგლებში დაწერილი ამ თუ იმ კონკრეტულ გრამატიკის შემოწმებისათვის.

რამდენადაც კარგი მეთოდოლოგია აუცილებელი წინაპირობაა იმისა, რომ მივიღოთ სოლიდური აღწერილობითი შედეგი, ბუნებრივი ენების ანალიზისას წარმომქმნელი გრამატიკების სისტემური გამოყენება როგორც წესი იძლევა დადგებით შედეგებს, რაც უფრო ეფექტურს ხდის ლინგვისტური ანალიზის პრაქტიკულ გამოყენებებსაც.

არ არის გამორიცხული არსებობდეს წარმომქმნელი გრამატიკის რამდენიმე ერთმანეთისაგან განსხვავებული ფორმალიზმი, რომელთაგან თითოეული ერთნაირად სწორადაა განსაზღვრული. თუმცა, ამავდროულად, ისინი შეიძლება არათანაბრად ვარგისიანები აღმოჩნდნენ ერთი რომელიმე ბუნებრივი ენის აღსაწერად. მართლაც, გრამატიკის ფორმალიზმის ერთადერთი წინაპირობა ისაა, რომ ის უნდა იყოს საკვლევი ენის ადეკვატურად კონსტრუირებული. ამგვარად, მიუხედავად იმისა, რომ წარმომქმნელი გრამატიკის გამოყენება მეთოდოლოგიურად აუცილებელია ლინგვისტური ანალიზის მათემატიკური, ემპირიკული, და გამოთვლითი თვისებების გამოსაკვლევად, ეს იმაში, რომ ეს გრამატიკული ფორმალიზმი სტრუქტურულად გამოსადევი იქნება განხილვის ქვეშ მყოფი მიზნებისათვის, არ გვარწმუნებს.

ესა თუ ის ფორმალიზმი ამ თუ იმ კონკრეტულ ენასთან მიმართებაში ფასლება ემპირიკულად ვარგისად, თუ მას აქვს საქმარისი გამომხატველობითი ძალა იმისათვის, რომ აღწეროს ამ ენის ყველა შესაძლო სტრუქტურა. მათემატიკური თვალსაზრისით ფორმალიზმი რაც შეიძლება ნაკლებ როგორც უნდა იყოს. გამოთვლითი თვალსაზრისით გამომყვანი ალგორითმის ბიჯების თანმიმდევრობა შეთანხმებადი უნდა იყოს ეფექტურ და გამჭვირვალე პროგრამულ სტრუქტურასთან. ამის გარდა, ფორმალიზმი უნდა იყოს ფუნქციონალურად ვარგისი ისეთი ხელოვნური შემცნებითი მოქმედის (**agent**) სარეალიზაციოდ, რომელთანაც შესაძლებელი იქნება რეალურ დროში თავისუფალი და ეფექტური ურთიერთობის დამყარება მხოლოდ ბუნებრივი ენობრივი საშუალებებით.

ის, რომ რომელიმე გრამატიკული თეორია დადასტურდა როგორც არაოპტიმალური ამა თუ იმ ენის მათემატიკური ანალიზისათვის, არ ნიშნავს იმას, რომ ამის მიზეზი წარმომქმნელი მეთოდოლოგია. პირიქით, რომ არა ეს მეთოდოლოგია, ჩვენ ვერც ვერანაირ დასკვნას ვერ გავაკეთებდით ამა თუ იმ ფორმალიზაციის არავარგისიანობის თაობაზე. ამგვარად, ასეთი შემთხვევები ფორმალური გრამატიკული თეორიების ღირებულებათა უარყოფითად შემფასებელ ფაქტორად ვერ გამოდგება. უფრო მეტიც, ცხადია, რომ კონკრეტული ფორმალური გრამატიკის არსებობა არის აუცილებელი წინაპირობა იმისა, რომ კონკრეტულ ენასთან მიმართებაში დადგინდეს ამ ფორმალური გრამატიკის ადეკვატურობის, ან არაადეკვატურობის საკითხი.

7.3 წარმომხმარელი გრამატიკის აღეპვატურობა

წარმომქმნელ გრამატიკას უწოდებენ მოცემული ენის სრულად აღმწერს, ანუ დესკრიფიულად აღეკვატურს (**descriptively adequate**), თუ ის წარმოქმნის ამ ენის გრამატიკულად მართებულად აგებულ ყველა გამოსაზულებას და მხოლოდ მათ. არააღეკვატური გრამატიკა არის ან არასწორი (**incorrect**), ან არასრული (**incomplete**), ან ორივე ერთად. წარმომქმნელი გრამატიკა არის

არასწორი, ანუ არაკორექტული, თუ ის წარმოქმნის ისეთ გამოსახულებებს, რომლებიც არ არიან გრამატიკულად მართებულად აგებულნი (ზემოარმოებლობა (**overgeneration**)). წარმოქმნელი გრამატიკა არასრულია, თუ არსებობს ერთი მაინც ისეთი გრამატიკულად მართებულად აგებული გამოსახულება, რომელსაც იგი ვერ წარმოქმნის (ნაკლულმწარმოებლობა (**undergeneration**)).

დავუშვათ უკვე გვაქვს ისეთი ბუნებრივი ენის ფორმალური გრამატიკა, რომელზეც ჩვენ არ ვლაპარაკობთ (მაგ. კეჩუას ენის¹¹) და გვინდა გავარკვიოთ ამ ენის რომელიმე მოცემულ გამოსახულებაზე არის თუ არა იგი მართებულად აგებული. ამისთვის გასაანალიზებელი გამოსახულების სიტყვა ფორმებს ვუსაბამებთ მათ შესაბამის კატეგორიებს კეჩუას ენის ამ ფორმალური გრამატიკის ლექსიკონის მიხედვით. შემდეგ ვცდილობთ გამოვიყვანოთ (ე.ი. წარმოვებნათ) გასაანალიზებელი გამოსახულება ამავე ფორმალური გრამატიკის წესების გამოყენებით. გამოსახულება, რომელსაც ვაანალიზებთ, მხოლოდ იმ შემთხვევაში იქნება მართებულად აგებული, თუ ასეთი მიღვომებით მისი გამოყვანის არსებობა დადასტურდება.

ამგვარად, ბუნებრივი ენის გამოსახულების ფორმალური გამოყვანის არსებობა გამოსახულების გრამატიკულად მართებულად აგებულობის დამადასტურებელია, თუ ცნობილია, რომ ეს წარმოქმნელი გრამატიკა არის ამ ენის სრულად აღმწერი, ანუ დესკრიფიულად აღეკვატური. ამდენად, როცა არსებობს რაიმენაირი ეჭვი იმ გრამატიკის დესკრიფიულ აღეკვატურობაზე, რომელზედაცაა საუბარი, მაშინ გრამატიკით წარმოქმნილი გამოსახულებები უნდა წარვუდგინოთ ენის ბუნებრივად მატარებელს და მისი მეშვეობით უნდა გავარკვიოთ არიან თუ არ ისინი მართებულად აგებულნი (გრამატიკის კორექტულობა). აგრეთვე, ენის ბუნებრივად მცოდნე პიროვნების მეშვეობით უნდა გადამოწმდეს ენის ახალ-ახალი მართებულად აგებული გამოსახულებები და უნდა გაირკვევს წარმოიქმნება თუ არა ისინი განსახილველი გრამატიკით (გრამატიკის სისრულე).

ზოგჯერ იმის გარკვევა, მოცემული გამოსახულება წარმოიქმნება თუ არა მოცემული ფორმალური გრამატიკით, დიდ დროს ითხოვს, რიგ შემთხვევებში უსასრულოდ დიდ დროსაც კი, თუნდაც გამოყვანას სუპერკომპიუტერი აწარმოებდეს. ყველა ასეთ შემთხვევაში მითხვდა, რომ გამოყენებადი ფორმალური გრამატიკა მაღალი მათემატიკური სირთულისაა. წარმოქმნელი გრამატიკის მათემატიკური სირთულე იზომება წარმოქმნის (ანუ, გამოყვანის) ყველაზე ცუდ შემთხვევაში გამოყენებული წესების მაქსიმალური რიცხვისა და შესავალი სიტყვის სიგრძის თანაფარდობით (იხ. 8.2 ქვეთავი).

ბუნებრივი ენების წარმოქმნელი გრამატიკების თანამედროვე სისტემების უმეტესობა მაღალი მათემატიკური სირთულისაა და, ცხადია, ეს ასე იქნება სანამ გრამატიკულად მართებულად აგებულობის საკითხი კომუნიკაციის ფუნქციონალური თეორიის გარეშე განიხილება. ამდენად, დღეს განხილვადი უმეტესი სისტემებისა არაეფექტურად აანალიზებს ბუნებრივ ენებს, რადგან ისინი პრაქტიკულ მიზნებზე ჯერ კიდევ არასრულყოფილად არიან მორგებულნი. ამგვარად, თავი რომ დავიზღვიოთ ამგვარი თეორიული ნაკლებისაგან, ბუნებრივი ენის სინტაქსური ანალიზი უნდა იყოს:

- განსაზღვრული მათემატიკურად, როგორც დაბალი სირთულის ფორმალური თეორია,
- აგებული ფუნქციონალურად, როგორც ბუნებრივი კომუნიკაციის კომპონენტი, და

¹¹ სამხრეთ ამერიკელი ინდიელების ენა

- რეალიზებული მეთოდოლოგიურად, როგორც ეფექტურად განხორციელებული კომპიუტერული პროგრამა, რომელშიც ფორმალური ენის თეორიისა და ბუნებრივი ენის ანალიზის თვისებები წარმოდგენილია მოდულარული და გამჭვირვალე ხერხებით.

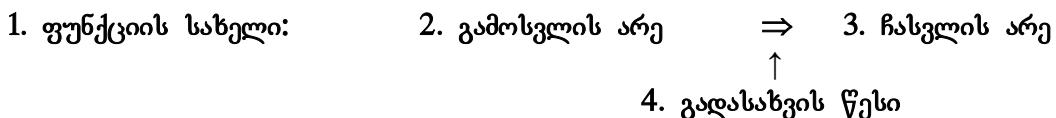
ცხადია, რომ ეს მოთხოვნები უნდა შესრულდეს ერთდროულად: რა სარგებლობა აქვს კარგად ჩამოყალიბებულ სინტაქსის, თუ მისი სირთულე ან ამოუხსნადი, ან ექსპონენციალურია? რა მიზანი აქვს მათემატიკურად სწორად ფორმირებულსა და ეფექტურ ფორმალიზმს, თუ ის აღმოჩნდება სტრუქტურულად არათაგსებადი ბუნებრივი კომუნიკაციის მექანიზმებისათვის? რამდენად საიმედოა მათემატიკურად და ფუნქციონალურად სწორად მორგებული (ანუ, ვარგისიანი) გრამატიკა, თუ ის გამოთვლადობის თვალსაზრისით არარეალურია, ანუ თუ ის მოითხოვს მონაცემების ძალიან დიდ მოცულობას?

7.4 C-გრამატიკის ფორმალიზმი

ისტორიულად პირველი წარმომქმნელი გრამატიკის ფორმალიზმი არის **C-გრამატიკა**, ანუ კატეგორიათა გრამატიკა (**categorial grammar**). ის შეიმუშავეს პოლონელმა ლოგიკოსებმა ლეშნისკიმ (1929) და ეკლესიერიმა (1935) იმ მიზნით, რომ ფორმალური ენების ანალიზისას თავი დაეღწიათ რასელის პარადოქსისაგან. პირველად **C-გრამატიკა** ბუნებრივი ენის დამუშავების მიზნით ი. ბარ-ჰილელმა¹² გამოიყენა 1953 წელს.

რასელის პარადოქსის კონტექსტში **C-გრამატიკის** წარმოშობა ემყარებოდა ლოგიკური სემანტიკური მიმართულების უკვე მინამდელი არსებობას (იხ. 19.3). შესაბამისად, **C-გრამატიკის** კომბინატორიკა დაფუძნებულია ფუნქტორ-არგუმენტის ლოგიკურ სტრუქტურაზე. ფუნქტორი არის ფუნქცია, რომელიც დასაშვებ არგუმენტებს გადასახავს მათ მნიშვნელობებში. ლოგიკური თვალსაზრისებით ფუნქცია მოიცემა სახელის (**name**), გამოსვლის არის (**domain**), ჩასვლის არისა (**range**) და იმ გადასახვის (**assignment**) წესის მინიშნებით, რომელიც ყოველ არგუმენტს უსაბამებს არა უმეტეს ერთი მნიშვნელობისა.

7.4.1 ლოგიკური ფუნქციის სტრუქტურა



ფუნქციის მარტივი მაგალითია ნატურალური რიცხვის კვადრატში აყვანა. ამ ფუნქციის სახელია რიცხვის_კვადრატი, მისი გამოსვლის არე, ისევე როგორც ჩასვლის არე ნატურალურ რიცხვთა სიმრავლეა. ამასთან, გადასახვის წესი უზრუნველვყოფს იმას, რომ ნებისმიერი შესაძლო არგუმენტი, მაგალითად: 1, 2, 3, 4, 5, ... შესაბამისად აისახება ზუსტად ერთ და კერძოდ 1, 4, 9, 16, 25, ... მნიშვნელობებზე¹³.

¹² ამ საკითხების კარგი არაფორმალური შეჯვამება არის პ. გიჩის (**P. Geach, 1972**), ასევე ჯ. ლამბეკის (**J. Lambek, 1958**) წიგნებში და ი. ბარ-ჰილელთან (**Y. Bar-hillel, 1964**) თავი 14, პარაგრაფები 185-189.

¹³ კვადრატული ფესვი არაა ფუნქცია. რადგან იგი განსაზღვრის არის ზოგიერთ არგუმენტს მნიშვნელობათა არეში ორ განსხვავებულ მნიშვნელობას უთმადებს. მაგალითისთვის 4-დან კვადრატულ ფესვს აქვს ორი მნიშვნელობა, ესენია: 2 და -2. ამის გმირა, რომ კვადრატულ ფესვს უწოდებენ არა ფუნქციას, არამედ მიმართებას (**relation**) უწოდებენ.

C-გრამატიკის კატეგორიებისა და წესების ფორმალური განსაზღვრებები ფუნქტორისა და არგუმენტის ზემომოყვანილ ცნებებს ეფუძნება. სამაგალითოდ განვიხილოთ შემდეგი ალგებრული განსაზღვრება.¹⁴

7.4.2 C-გრამატიკის ალგებრული განსაზღვრება

C-გრამატიკა არის ხუთეული **<W, C, LX, R, CE>**. სადაც

1. **W** არის სიტყვითი გამოსახულებების სასრული სიმრავლე.
2. **C** არის ისეთი კატეგორიების სიმრავლე, როგორებიცაა
 - ა) ბაზისი: **u** და **v** $\in \mathbf{C}$ ¹⁵;
 - ბ) ინდუქცია: თუ **X** და **Y** $\in \mathbf{C}$, მაშინ **(X/Y)** და **(X\Y)** $\in \mathbf{C}$;
 - გ) ჩაკეტვა: არაფერი სხვა გარდა (ა) და (ბ) წესებით განსაზღვრულებისა არ არის **C** სიმრავლის წევრი.
3. **LX** სასრული სიმრავლეა. ამასთან, **LX $\subset (W \times C)$** .
4. **R** არის სიმრავლე, რომელიც შეიცავს შემდეგ ორ სქემატურ წესს:

$$\alpha_{(X/Y)} \circ \beta_{(Y)} \Rightarrow \alpha \beta_{(X)}$$

$$\beta_{(Y)} \circ \alpha_{(Y/X)} \Rightarrow \beta \alpha_{(X)}$$
5. **CE** არის სიმრავლე, რომელიც შეიცავს მხოლოდ დასრულებული გამოსახულებების (**complete expressions**) კატეგორიებს; **CE $\subseteq C$** .

ალგებრული განსაზღვრების ზოგადი ცნების შესაბამისად **C**-გრამატიკის ძირითადი კომპონენტები ცხადად არის წარმოდგენილი **<W, C, LX, R, CE>** ხუთეულით და, ამასთან, 1-5 წინადადებებით მოცემულია ამ კომპონენტების სიმრავლური განსაზღვრებები.

უფრო კონკრეტულად: **W** არის აღსაწერი ენის ძირითადი სიტყვითი გამოსახულებების სასრული სიმრავლე. მაგალითად **a^kb^k** ხელოვნური ენის შემთხვევაში **W** სიმრავლე შესდგება მხოლოდ **a** და **b** სიტყვებისაგან.

C სიმრავლე განსაზღვრულია რეკურსულად. შესაბამისად, საბაზისო **u** და **v** წევრებთან ერთად ინდუქციური პირობის ძალით **C** სიმრავლეში არიან აგრეთვე **(u|v)**, **(v|u)**, **(u/v)** და **(v/u)** კატეგორიებიც. თავის მხრივ, ისევ და ისევ ინდუქციური პირობის ძალით **((u/v)/v)**, **((u/v)|v)**, **((v/u)/u)**, **((v/u)|u)**, **(u/(u/v))**, **(v/(u/v))** და ა.შ. კატეგორიები აგრეთვე **C** სიმრავლის წევრები არიან. ამგვარად, **C** კატეგორიათა უსასრულო სიმრავლეა, რადგან მისი ახალი წევრები მიიღება მისი ძველი წევრებისაგან არაგანმეორებადი რეკურსის გზით.

LX სიმრავლე არის დალაგებული წყვილების სასრული სიმრავლე. ამასთან, თითოეული ეს დალაგებული წყვილი აგებულია (i) **W** სიმრავლის ერთი წევრისა და (ii) **C** სიმრავლის ერთი წევრისაგან. ჩანაწერობრივად ამ დალაგებული წყვილის მეორე კომპონენტი, რომელიც თავის მხრივ რაიმე კატეგორიაა, ინდუქსად მიეწერება წყვილის პირველ კომპონენტს, როგორც ეს არის მაგალითად **a((u/v)|v)** გამოსახულებაში. ამ გამოსახულების ძირითადი ნაწილი (ე.ი. **W** სიმრავლის

¹⁴ მსგავსი განსაზღვრებების ნახვა შეიძლება ი. ბარ-ჰილელის წიგნში (1964), პარაგრაფი 188.

¹⁵ ელემენტარული, ანუ ძირული კატეგორიების სახელები და რაოდენობა (აქ **u** და **v**) აუცილებლობით და ცალსახობით არ ისაზღვრება. მაგალითისათვის ეჯლუპვერი იყენებდა მხოლოდ ერთ ელემენტარულ კატეგორიას, გრი და მონტეგიუ ორს, სხვები სამს.

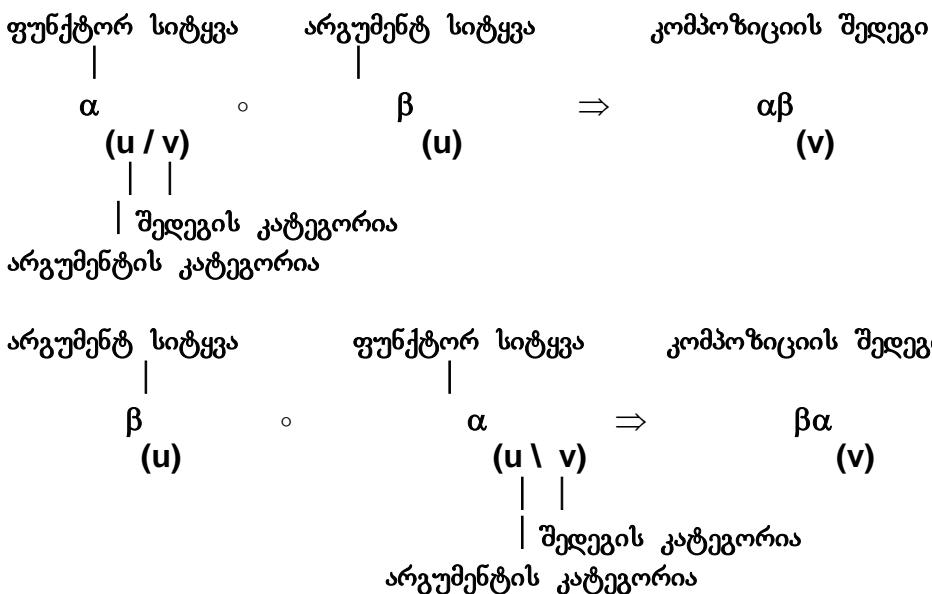
წევრი) ინდექსად იღებს **C** სიმრავლის რომელიდაც წევრს, ანუ კატეგორიას, რაც იმას ნიშნავს, რომ გამოსახულების ძირითადი წევრი ამ კატეგორიის გამოსახულებაა, რაც მთლიანობაში, ცხადად მოიცემა **LX** სიმრავლის შემადგენელი წყვილებისა და აქ გაკეთებული შეთანხმებების საფუძველზე.

R სიმრავლე შეიცავს ორ სქემატურ წესს. ამ წესებში **α** და **β** ცვლადები შესაბამისად გამოიყენება ფუნქტორისა და არგუმენტის ჩანაწერობრივი სახით წარმოსადგენად. შესაბამისად **X** და **Y** ცვლადები წარმოგვიდგენებ მათ კატეგორიალურ მახასიათებლებს. პირველი სქემა აკომბინირებს ფუნქტორსა და არგუმენტს **αβ** რიგით. ამასთან, **Y**, როგორც **β** არგუმენტის კატეგორიის მანიშნებელი, აუქმებს **α** ფუნქტორის კატეგორიალურ მაჩვენებელში **Y** კატეგორიას. მეორე სქემა აკომბინირებს ფუნქტორსა და არგუმენტს პირიქითი, ანუ **βα** რიგით. ამ დალაგების ფორმალური გასაღებია **α** ფუნქტორის კატეგორიალურ მაჩვენებელში უკუხაზად (**backslash**) წოდებული **** სიმბოლო (ნაცვლად წინა წესში გამოყენებადი **ხაზად** (**slash**) წოდებული **/** სიმბოლოსი). ამ ტიპის **C**-გრამატიკებს ბარ-ჰილელმა ორმიმართულებიანი (**bidirectional**) **C**-გრამატიკები უწოდა, რადგან ეს სქემატური წესები საშუალებას გვაძლევენ ფუნქტორი დავსვათ როგორც არგუმენტის წინ, ისე მის შემდეგ.

და ბოლოს, **CE** წარმოგვიდგენს მხოლოდ იმ გამოსახულებების კატეგორიების სიმრავლეს, რომლებიც დასრულებულ გამოსახულებებად მიიჩნევა. **C**-გრამატიკისა და ენის სპეციფიკების მიხედვით ეს სიმრავლე შეიძლება იყოს სასრული და განისაზღვროს სიის პირდაპირი ჩამოთვლით, ან შეიძლება იყოს უსასრულო და დახასიათდეს ცვლადების შემცველი ზოგადი მოდელებით.

ორი კატეგორიზებული ენობრივი გამოსახულების კანონიერი კომბინირება დამოკიდებულია იმაზე, შესაძლებელია თუ არა მათი კატეგორიების დაწყვილება სისტემაში არსებული რომელიმე სქემატური წესით. კატეგორიათა დაწყვილების ეს წესი C-გრამატიკაში ცხადად აღწერილად ითვლება, რადგან ამ გრამატიკის ასეთი სქემატური წესების რაოდენობა ორით იზღუდება. — მკაცრადაც თუ განვსჯით, უნდა ჩაითვალოს, რომ ეს სქემატური წესები ცხადად იძლევანან კატეგორიათა დაწყვილების ნიმუშებს (იხ. თავი 17 და 18).

7.4.3 ლატვიულების ცხადი ნიმუშები C-გრამატიკის პომაზინაციები



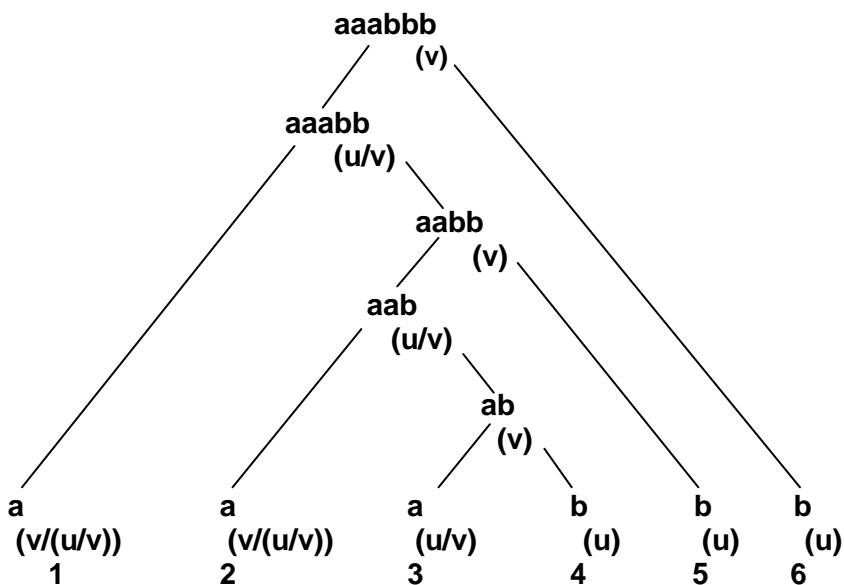
ცხადია, რომ ეს სქემატური წესები და ამ წესებით განსაზღვრული ეს კომბინაციები ნაგულისხმევი უნდა იყოს **C**-გრამატიკის კატეგორიალურ სტრუქტურებში. აქედან გამომდინარე, ენა ითვლება საკმარისად განსაზღვრულად, თუ მოცემულია ამ ენის (i) **LX** ლექსიკონი და (ii) დასრულებული გამოსახულებების **CE** სიმრავლე. სამაგალითოდ განვიხილოთ ჩვენთვის უკვე ნაცნობი $a^k b^k$ ხელოვნური ენის (იხ. 7.1.3) ქვემოთ მოცემული განსაზღვრება.

7.4.4 C-გრამატიკა $a^k b^k$ ენისათვის

$$\begin{aligned} LX &=_{\text{def}} \{a_{(u/v)}, b_{(u)}, a_{(v/(u/v))}\} \\ CE &=_{\text{def}} \{(v)\} \end{aligned}$$

a სიტყვა ლექსიკურად ორი **(u/v)** და **(v/(u/v))** კატეგორიით ისაზღვრება. შესაბამისად იხსნება ქვემოთ წარმოდგენილი გამოყვანის ხის განმსაზღვრელი მიზეზები.

7.4.5 $a^k b^k$ გამოსახულების წარმოქმნა $k=3$ შემთხვევისთვის



წარმოქმნა იწყება 3 და 4 სიტყვების კომბინირებით. მხოლოდ ეს საწყისი კომპოზიცია იყენებს **(u/v)** კატეგორიის **a** სიტყვას. შედეგი არის **(v)** კატეგორიის **ab** გამოსახულება. ეს გამოსახულება **CE** სიმრავლის განსაზღვრების თანახმად დასრულებული გამოსახულებაა. შემდეგ, სიტყვა 2 კომბინირდება წინა კომბინაციის შედეგთან და წარმოქმნის **(u/v)** კატეგორიის **aab** გამოსახულებას. ამ შედეგთან რიგის მიხედვით კომბინირდება სიტყვა 5, რის შედეგადაც წარმოიქმნება **(v)** კატეგორიის **aabb** გამოსახულება, და ა.შ.. ასეთი გამოყვანებით წარმოიქმნებიან **(v)** კატეგორიის დასრულებული გამოსახულებები, რომელთაგან ნებისმიერი შესდგება **a** სიტყვის გარკვეული ოდენობისაგან და მარცხნიდან მათზე უშუალოდ მიწერილი **b** სიტყვის ზუსტად იმავე ოდენობისაგან. ამგვარად, გამოდის, რომ სასრული 7.4.4 განსაზღვრება წარმოქმნის $a^k b^k$ ენის უსასრულოდ ბევრ გამოსახულებას.

სტრუქტურული თვალსაზრისებით **C**-გრამატიკას მრავალი ნაკლი აქვს. პირველი: კორექტული შეაღებური გამოსახულებების გამორჩევა შეიძლება მხოლოდ ცდისა და შეცდომის მეთოდით. მაგალითად, ყოველთვის როდია ცხადი რა რიგით უნდა მოხდეს შემავალი სტრიქონებით განსაზღვრული მონაცემების კომბინირება. მეორე: **C**-გრამატიკა, იმისთვის რომ მოხდეს განსხვავებული სიტყვათა წყობების სრული კატეგორიზაცია, იმისთვის ლექსიკური ორაზროვნების მაღალ ხარისხის. ამის შედეგია ის, რომ ხშირად ექსპერტებსაც კი უჭირთ **C**-გრამატიკის მოცუმული სტრიქონისთვის 7.4.5 გამოყვანის მსგავსი გამოყვანის აგება, ანდა იმის დამტკიცება, რომ ასეთი გამოყვანა არ არსებობს. — განსაკუთრებით ეს ჭირს ისეთ არამარტივ ენეტში, როგორიც **a^kb^k** ენაა. შესაბამისად, **C**-გრამატიკაზე დაფუძნებული ავტომატური ანალიზი გამოთვლითი თვალსაზრისით არაეფექტურია, რადგან გადასამოწმებელია ყველა შესაძლო კომბინაციების საკმაოდ დიდი ოდენობა. ამასთან, **C**-გრამატიკა არ იძლევა დროში წრფივი გამოყვანების (**time-linear derivation**) კეთების საშუალებას. მაგალითად, 7.4.5 პუნქტში, იძულებულნი ვართ გამოყვანა დავიწყოთ 3 და 4 სიტყვებით. კატეგორიების მახასიათებელი სტრუქტურებისა და **C**-გრამატიკის სქემატური წესების თვისებების გამო შეუძლებელია **a^kb^k** ენისათვის ისეთი **C**-გრამატიკის შექმნა, რომელსაც შეეძლება ნებისმიერი წინადაღების წარმოქმნა დროში წრფივი მეთოდით.

7.5 C-გრამატიკა ბუნებრივი ენისათვის

C-გრამატიკა არის პროტოტიპი ლექსიკური მიღეომისა: ენის ყველა კომბინაციური თვისება კოდირებულია მისი ძირითადი გამოსახულებების კატეგორიებში. ეს დასტურდება როგორც 7.4.4 პუნქტში გაკეთებული **a^kb^k** ხელოვნური ენის **C**-გრამატიკული განსაზღვრებით, ასევე ინგლისური ენის მცირე ფრაგმენტის¹⁶ ქვემოთ მოყვანილი განსაზღვრებითაც.

7.5.1 C-გრამატიკა ინგლისური ენის მცირე ფრაგმენტისათვის

$$\begin{aligned} LX &=_{\text{def}} \{W_{(e)} \cup W_{(elt)}\}, \\ \text{სადაც} \\ W_{(e)} &= \{\text{Julia, Peter, Mary, Fritz, Suzy ...}\} \\ W_{(elt)} &= \{\text{sleeps, laughs, sings, ...}\} \\ CE &=_{\text{def}} \{(t)\} \end{aligned}$$

7.4.4 გრამატიკასთან ამ გრამატიკის შედარება გვიჩვენებს ორ ნოტაციურ განსხვავებას: პირველი, ლექსიკური მონაცემები თავმოყრილია სიტყვათა კატეგორიზებულ (ანუ, **W_{cat}**) კლასებში. ეს მიღეომა, რომელიც მონტეგიუდან მოდის, ააღვილებს ლექსიკონის აგებას, რადგან კატეგორიები მიეწერება მთლიან კლასებს ნაცვლად ცალკეული სიტყვა ფორმებისა. მეორე, 7.4.2 პუნქტში **u**, **v** სიმბოლოებით მონიშნული ძირებული კატეგორიები **e**, **t** სიმბოლოებად გადასახელდნენ. მონტეგიუს თანახმად **e** ანზოგადებს საგობრივ, **t** კი – ჭეშმარიტულ მნიშვნელობებს.

(**elt**) კატეგორიის სიტყვები: **sleeps** (სძინავს), **laughs** (იცინის), **sings** (მღერის) ... 7.4.1 პუნქტის მიხედვით მოტივირებულნი არიან არა მარტო სინტაქსურად, არამედ სემანტიკურადაც: (**elt**) ინტერპრეტირებულია როგორც მახასიათებელი ფუნქცია საგანთა სიმრავლიდან ჭეშმარიტულ

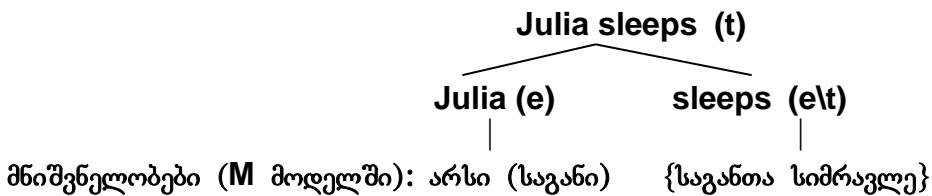
¹⁶ რ. მონტეგიუ იყენებდა ფრაგმენტის ცნებას იმ მიზნით, რომ გამოეყო ბუნებრივი ენის ის ქვესიმრავლე, რომელიც უკვე შემუშავებული ფორმალური გრამატიკით უნდა გაანალიზებულიყო.

მნიშვნელობათა სიმრავლეში. ამასთან, **sleeps** (სძინავს) სიტყვის მახასიათებელი ფუნქცია ამოწმებს ყველა **e** სახის ობიექტს და იმის მიხედვით სძინავს მას, თუ არა, ამ **e** ობიექტს უთანადებს ან 1 (ჭეშმარიტ), ან 0 (მცდარ) ჭეშმარიტულ მნიშვნელობას. ამგვარად, სიმრავლე, რომელიც აღნიშნულია სიტყვით სძინავს, შეიცავს მხოლოდ იმ ობიექტებს, რომლებსაც ამ სიმრავლის შესაბამისი მახასიათებელი ფუნქცია ჭეშმარიტში (ანუ, 1 სიმბოლოში) გადასახავს.

რადგანაც ისეთ სიტყვებს როგორებიცაა სძინავს და სეირნობს აქვთ ერთნაირი კატეგორია მათ აქვთ ერთნაირი განსაზღვრისა და მნიშვნელობების არც. ამასთან, მათ შინაარსებს განასხვავებს მათი სხვადასხვაგვარი გადასახვის წესი (იხ. 7.4.1 განაწერი). ის, თუ როგორ განვსაზღვროთ ეს სხვადასხვა გადასახვის წესი არატრივიალური გზით (ე.ი. არადალაგებული წყვილების სის ჩამოწერით) არის ერთ-ერთი ძირითადი პრობლემა მოდელურ-თეორიული სემანტიკებისა.¹⁷

Julia sleeps (ჯულიას სძინავს) წინადადების **C**-გრამატიკული ანალიზის შინაარსული ინტერპრეტაცია ნებისმიერად აღებულ **M** მოდელში შემდეგნაირად ისაზღვრება: წინადადება ჭეშმარიტია თუ **M** მოდელში სიტყვა ჯულიას მნიშვნელობას იმავე **M** მოდელში სძინავს სიტყვის მნიშვნელობა გადასახავს 1-ში, ანუ ჭეშმარიტში, მცდარია წინააღმდეგ შემთხვევაში.

7.5.2 მრთლოული სინტაქსური და სემანტიკური ანალიზი



ზემოთ ხაზგასმულის ეკვივალენტური სიმრავლურ-თეორიული თვალსაზრისების შესაბამისად, ეს წინადადება ჭეშმარიტია, როცა **M** მოდელში ჯულია სიტყვის აღსანიშნი (**denotation**) არის იმ სიმრავლის წევრი, რომელიც იმავე **M** მოდელში სძინავს სიტყვით აღინიშნება.

C-გრამატიკაში, ბუნებრივი ენის გამოსახულებათა კატეგორიები ორგვარ მოტივირებულნი არიან: კერძოდ, (i) დენოტაციური (სემანტიკა) და (ii) კომბინატორული (სინტაქსი) თვალსაზრისებით. მაგალითად, არსებითი სახელები, რომლებიც სიმრავლეებს აღნიშნავენ, მონტეგიუმ¹⁸ გადაიაზრა საგანთა სიმრავლის ჭეშმარიტულ მნიშვნელობათა სიმრავლეში ამსახველ მახასიათებელ ფუნქციად, ანუ **(e/t)** კატეგორიად. აქ მან გადახრილი ხაზი გამოიყენა არსებითი სახელებისა და ინტრანზიტული ზმნების სინტაქსური განსხვავებულების ხაზგასამელად.

დეტერმინერი (**determiner**) კომბინირდება ზოგად სახელთან (**noun**) და ამგვარი კომბინირებით მიღებული სახელური ფრაზა (**noun phrase**), მსგავსად საკუთარი სახელებისა (**proper name**), შეიძლება განვიხილოთ საგნის აღმნიშვნელად. ეს საკმარის საფუძველს იძლევა იმისა, რომ იგი **((e/t)/e)** კატეგორიად, ანუ ისეთ ფუნქტორად გავიაზროთ, რომელიც მოქმედებს **(e/t)** კატეგორიის ზოგად სახელზე და მას **(e)** კატეგორიის საკუთრივ სახელად გარდაქმნის. ზედსართავი სახელი (**adjective**) კი, რომელიც მოქმედებს ზოგად სახელზე და შედეგად ისევ ზოგად სახელს იძლევა,

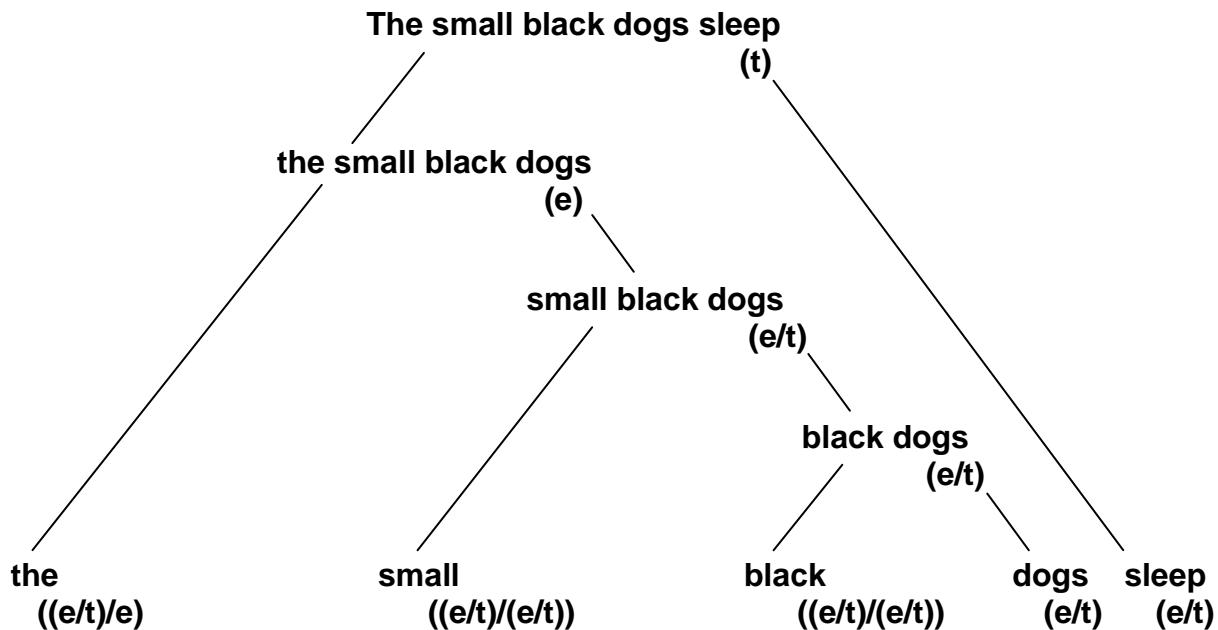
¹⁷ იხილეთ 19.4 ქვეთავი და **Col.**, გვ 292-295.

¹⁸ სიმპარტივისა და შეთანხმებულობისათვის: მონტეგიუს აღნიშვნებისგან განსხვავდით, ჩენ სემანტიკურ ტიპებსა და სინტაქსურ კატეგორიებს განვასხვავებთ არგუმენტისათვის ხაზამდე გამოყოფილი ადგილის მეშვეობით.

((e/t)/(e/t)) კატეგორიად უნდა გადააზრდეს. ზედსართავი სახელისა და ზოგადი სახელის კომბინირების ასეთი გაგება განაპირობებს მათი ურთიერთმიმართების რეპურსიულ გაგრძელებადობას, ანუ იმას, რომ ზოგად სახელს შეუძლია ზედსართავი სახელების მიერთება განუსაზღვრელი ოდენობით.

ზემოთ განხილულ კატეგორიზაციაზე დაფუძნებული მიდგომებით **The small black dogs sleep** (ამ პატარა შავ ძაღლებს სძინავთ) წინადადების ანალიზი შემდეგ გამოყვანის ხეს იძლევა:

7.5.3 ბუნებრივი ენის ფინალადებების C-ანალიზი



როგორც წინა მაგალითებში (იხ.7.4.4 და 7.5.1), ასევე აქაც, **C**-გრამატიკული ანალიზი, ანუ წარმოქმნა ემყარება მხოლოდ **LX** ლექსიკონსა და **CE** სიმრავლეს.

7.5.4 C-გრამატიკა 7.5.3 მაგალითისათვის

$$LX =_{\text{def}} \{W_{(e)} \cup W_{(e)t} \cup W_{(e/t)} \cup W_{((e/t)/(e/t))} \cup W_{((e/t)/t)}\},$$

სადაც

$$W_{(e)} = \{\text{Julia, Peter, Mary, Fritz, Suzy ...}\}$$

$$W_{(e)t} = \{\text{sleeps, laughs, sings, ...}\}$$

$$W_{(e/t)} = \{\text{dog, dogs, cat, cats, table, tables....}\}$$

$$W_{((e/t)/(e/t))} = \{\text{small, black...}\}$$

$$W_{((e/t)/t)} = \{\text{a, the, evry}\}$$

$$CE =_{\text{def}} \{(t)\}$$

ამ **C**-გრამატიკის ლინგვისტური მოტივაციაა ფუნქტორ-არგუმენტის სტრუქტურების სემანტიკური მიმართებებისა და სინტაქსური მაკომბინირებელი უნარების მახასიათებლების წარმოჩენა. თეორიის მოტივაციის მაღალი დონე და მისგან მომდინარე შეზღუდვები ლინგვისტური კვლევებისას დადებით

ფაქტორად მიიჩნევა. თუმცა, მიუხედავად იმისა, რომ **C**-გრამატიკა ერთდროულად იზღუდება სინტაქსური და სემანტიკური ასპექტებით, მისი მოტივაცია მხოლოდ ნაწილობრივია. ეს იმიტომ, რომ არაა ნათელი **C**-გრამატიკა და მისი მოდელურ-თეორიული სემანტიკა როგორ იფუნქციონირებს ბუნებრივ ენობრივ კომუნიკაციაში.¹⁹

პრაქტიკული გადაწყვეტებისას **C**-გრამატიკის სტრუქტურები მიიღება საკმაოდ რთული გამოყვანებით, რომლებსაც იგივე მახასიათებლები აქვთ, რაც თეორემათა მატტკიცებელ სისტემებს. ამ ნაკლს აძლიერებს ის ფაქტი, რომ ბუნებრივი ენის დიდი ფრაგმენტისთვის საჭირო **C**-გრამატიკა ითხოვს ძალიან მაღალი ხარისხის ლექსიკურ არაცალსახობას, რათა შესაძლებელი გახდეს კატეგორიების აუცილებელი კომბინაციური შეზღუდვების კოდირება.

ეს დასტურდება 7.5.4 პუნქტში აღწერილი **C**-გრამატიკით, რომელიც არ არის ადეკვატური ინგლისური ენის იქ განხილული მცირე ფრაგმენტის, და რომელიც, ამავდროულად, ხასიათდება ჭარბწარმომქნელობით: მაგ. სიტყვების კომბინაცია **dog(e/t) ° Peter(e)** „წინადადებაში” ***dog Peter** (ძალი პიტერი) არ არის დაბლოკილი. უფრო მეტიც, აქ არაა განხილული ის კერძო ხერხი, რომელიც დეტერმინერებსა და ზოგად სახელებს შეათანხმებდა, რის გამოც 7.5.3 დაშვებულია ისეთი არაგრამატიკული კომბინაციების წარმოქმნაც, როგორებიცა ***every dogs** (თითოეული ძალები) და ***all dog** (ყველა ძალი).

თეორიაში ყოველი ასეთი სირთულე დაძლეულ უნდა იქნას ფორმალიზმის შესაბამისი გაფართოებით. თუმცა, პრაქტიკაში, ნებისმიერი გაფართოება ითხოვს საკმარისი მოცულობის აღწერითი სამუშაოების²⁰ ჩატარებას. გაფართოებული სისტემის გადაქცევა ტრანსფარენტულ და მათემატიკურად ადვილად აღწერად **C**-გრამატიკულ ფორმალიზმად მაღალი სირთულის ამოცანაა და დიდ ძალისხმევებს მოითხოვს, რადგან ამგვარი გაფართოებების კეთების განუწყვეტელი საჭიროება ითხოვს ისეთ ახალ გადაწყვეტებს, რომელთა თანმხლები მოვლენები არსებითად აქვეითებს სისტემის ტრანსპარენტულობის ხარისხს. ამით არის განპირობებული ის, რომ თეორიული ლინგვისტიკის სიღრმისეული მიზნის მიღწევის, ანუ ბუნებრივი ენის სრული წარმოქმნითი აღწერის შემუშავების თვალსაზრისით **C**-გრამატიკა არასასურველ კანდიდატად განიხილება.

სავარჯიშოები

7.1 ქვეთავი

1. როგორ განისაზღვრება ენა ფორმალურ გრამატიკაში?
2. ახსენით თავისუფალი მონოიდის ცნება და მისი კავშირები წარმომქნელ გრამატიკასთან.
3. რა განსხვავება დადებით და კლინის ჩაკეტვებს შორის?
4. რა გაგებით შეიძლება იქნეს განხილული წარმომქნელი გრამატიკა, როგორც ფილტრი?

¹⁹ იხ. დისკუსია მეტაენაზე დაფუძნებული სემანტიკის თაობაზე 19-ე და 20-ე თავებში.

²⁰ საცდელად იხილეთ **SCG**.

5. ახსენით რეკურსის როლი 7.13 მიდგომებით **aaaabb** გამოსახულების გამოყვანაში.
6. რა არის ალგებრული განსაზღვრება და რა არის მისი მიზანი?
7. რა განსხვავებაა ძირეულ (ე.ო. ელემენტარულ), წარმოებულ, და ნახევრად-ფორმალურ ფორმალიზმებს შორის?
8. რა არის წარმოებული ფორმალიზმების განვითარების მიზანი?

7.2 ქვეთავი

1. მართებულად აგებულობის თვალსაზრისებით რით განსხვავდებიან ხელოვნური და ბუნებრივი ენები?
2. ახსენით რატომ გაიგება გრამატიკულად მართებულად აგებულობის მახასიათებლების აღწერა თეორიული აღწერითი ლინგვისტიკის საბოლოო მიზნად?
3. ჩამოთვალეთ თანამედროვე ლინგვისტიკაში წარმომქმნელი გრამატიკის გამოყენების სამი მიზეზი.
4. რატომაა წარმომქმნელი გრამატიკის გამოყენება აუცილებელი, მაგრამ არასაკამარისი პირობა ენის სრული და წარმატებული ანალიზისათვის?

7.3 ქვეთავი

1. რა შემთხვევაშია წარმომქმნელი გრამატიკა დესკრიფციულად აღექვატური?
2. რა იგულისხმება ფორმალური გრამატიკის მათემატიკურ სირთულეში და რა არის მნიშვნელოვანი პრაქტიკული მიზნებისათვის?
3. რა განსხვავებაა ფუნქციონალურ და არაფუნქციონალურ გრამატიკულ თეორიებს შორის?
4. რა სამი ასპექტი უნდა იქნეს გათვალისწინებული წარმომქმნელი გრამატიკის შესაქმნელად და რატომ?

7.4 ქვეთავი

1. ვის მიერ, როდის და რა მიზნით იქნა **C**-გრამატიკა შემუშავებული?
2. როდის და პირველად ვის მიერ იქნა **C**-გრამატიკა ბუნებრივ ენაზე მიყენებული?
3. როგორია ლოგიკური ფუნქციის სტრუქტურა?
4. ჩამოაყალიბეთ **C**-გრამატიკის ალგებრული განსაზღვრება.
5. ახსენით **C**-გრამატიკის რთული კატეგორიების ფუნქტორებად ინტერპრეტირების საკითხი?
6. რატომა **C**-გრამატიკის კატეგორიათა სიმრავლე უსასრულო, ლექსიკონი კი - სასრული?
7. რომელი ფორმალური პრინციპით განპირობდა ის, რომ 7.4.4 პუნქტში აღწერილი **C**-გრამატიკა მიუხედავად მისი წესებისა და ლექსიკონის სასრულობისა წარმოქმნის უსასრულოდ ბევრ გამოსაზულებას?
8. რატომ იწოდება 7.4.4 პუნქტში განსაზღვრული გრამატიკული ფორმალიზმი ორმიმართულებიან **C**-გრამატიკად?
9. შესაძლებელია თუ არა **C**-გრამატიკის განხილვა ენის **SLIM** თეორიის სინტაქსურ კომპონენტად?

7.5 ქვეთავი

1. რატომ მიიჩნევა **C**-გრამატიკა ლექსიკური მიდგომების პროტოტიპულად?
2. როგორ გაიგება წარმომქმნელ გრამატიკაში ბუნებრივი ენის ფრაგმენტი?

3. ახსენით რა დამოკიდებულებაა **C**-გრამატიკის როლი კატეგორიების ფუნქციონალურ ინტერპრეტაციებსა და ბუნებრივი ენის მოდელურ-თეორიულ ინტერპრეტაციებს შორის?
4. ახსენით რეკურსული სტრუქტურა 7.5.4 პუნქტში აღწერილი **C**-გრამატიკისა
5. ახსენით **C**-გრამატიკის სემანტიკური ინტერპრეტაციის პრინციპები.
6. გააფართოვთ 7.5.4 პუნქტში აღწერილი **C**-გრამატიკა ისე, რომ მან წარმოქმნას წინადადებები **The man send the girl a letter** (კაცმა გაუგზავნა გოგოს წერილი), **The girl received a letter from The man** (გოგომ მიიღო წერილი კაცისგან), **The girl was send a letter by the man** (გოგოს გაეგზავნა წერილი კაცისგან). ახსენით დამატებული კატეგორიების სემანტიკური მოტივაცია?
7. რატომ არ არის ბუნებრივი ენების დიდ მასშტაბიანი **C**-გრამატიკული აღწერები?
8. რატომ არ არის ეფექტური **C**-გრამატიკები პრაქტიკული თვალსაზრისებით?
9. რატომ არ არსებობს **C**-გრამატიკის მნიშვნელოვანი მეთოდოლოგიური პრობლემების ეფექტური გადაწყვეტები?
- 10 შეუძლია თუ არა **C**-გრამატიკას უზრუნველყოს ბუნებრივი კომუნიკაციის მექანიზმი, ან, ხომ არ გამოდგებოდა იგი ასეთი მექანიზმის ერთ-ერთ კომპონენტად.

8 თავი ენის სირთულე და ენობრივი იმრარჩივი

წარმომქნელი გრამატიკის მეორე ელემენტარული ფორმალიზმი 1936 წელს ამერიკელმა ლოგიკოსმა ე. პოსტმა გამოაქვეყნა. ეს ფორმალიზმი ცნობილი იყო სახელწოდებით პოსტის მარტომობელი სისტემა, მოკლედ გადაწერა. იგი ჩაისახა რეკურსის თეორიის მათემატიკურ ნაწილში. ამასთან, მჭიდროდ იყო დაკავშირებული ავტომატებისა და გამოთვლების სირთულის თეორიებთან.

ბუნებრივი ენისათვის პირველი გადამწერი სისტემა, სახელწოდებით ფრაზის სტრუქტურული გრამატიკა¹, მოკლედ **PS**-გრამატიკა, 1957 წელს ნ. ჩომსკის მიერ იქნა აღწერილი. **PS**-გრამატიკაზე დაყრდნობით ჩომსკიმ და სხვებმა ააგეს მთელი რიგი წარმოქმნული ფორმალიზმებისა, რომლებსაც თავდაპირველად ტრანსფორმაციულ გრამატიკებს უწოდებდნენ.

8.1 თავში მოყვანილია **PS**-გრამატიკის ალგებრული განსაზღვრება და აღწერილია ის შეზღუდვები, რომელთა **PS**-წესებზე დამატებით მიიღება წესიერი, კონტექსტუალი, კონტექსტზე დამოკიდებული და შეუზღუდავი **PS**-გრამატიკები. 8.2 თავში ახსნილია სირთულის ხარისხის ოთხი ძირითადი დონე და მათი კავშირები სხვადასხვა ტიპის **PS**-გრამატიკებთან. 8.3 თავში განხილულია წარმოქმნის უნარის ცნება და გაშუქებულია **PS**-გრამატიკისა და ხელოვნური ენების ურთიერთმიმართებათა გამოყენებითი მხარეები. 8.4 თავში განიხილება კონტექსტისგან თავისუფალი **PS**-გრამატიკის გამოყენებითი მხარე ბუნებრივ ენებთან მიმართებაში და განსაზღვრულია შემადგენლობითი სტრუქტურის ლინგვისტური ცნება. 8.5 თავში ახსნილია შემადგენლობითი სტრუქტურის პარადოქსი და ის, თუ რატომ აყალიბებს გარკვეული ტრანსფორმაციები **PS**-გრამატიკულ ფორმალიზმებს არაამოხსნადი სისტემის სახით.

8.1 PS-გრამატიკის ფორმალიზმი

PS-გრამატიკის ალგებრული განსაზღვრება გვაგონებს **C**-გრამატიკის (იხ. 7.4.2) ალგებრულ განსაზღვრებას და კერძოდ იმით, რომ სისტემის ძირითადი კომპონენტები გადანომრილია და თითოეული მათგანი სიმრავლეთა თეორიის ტერმინებითაა დახასიათებული.

¹ ჩომსკი არ იხსენიებს პოსტის წვლილს. ამას 1960 წელს ხაზს უსვამს ბარ-ჰილელი:

ეს მიდგომა (ე.ი. გადამწერი სისტემები) სტანდარტულია კომბინატორული სისტემებისათვის გაცილებით ადრე იქნა შემუშავებული პოსტის მიერ (1936). ეს იყო ფორმალური ალრიცხვების ზოგად სტრუქტურებში მისი კარგამზნელი გამოკვლევების შედეგი. თუმცა, ჩომსკიმ, თავისი მოღვაწეობის ბოლო წლებში, გააცნობიერა პოსტის იდეებთან მისი იდეების განსაკუთრებული სიახლოვე.

ბარ-ჰილელი, 1964, გვ.103

აქ გასათვალისწინებელია ისიც, რომ ჩომსკის სადოქტორო დისერტაციის ხელმძღვანელს ზ. პარესსა და ი. ბარ-ჰილელს 1947 წლიდან ახლო მეცნიერული კონტაქტები ჰქონდათ. გარდა ამისა, გასათვალისწინებელია 1951 წლის დასწესები ბარ-ჰილელისა და ჩომსკის დისკუსიაც „დაუსრულებელი საუბრები ლინგვისტიკაზე, ლოგიკასა და მეთოდოლოგიაზე“. (ბარ-ჰილელი, 1964, გვ.16).

8.1.1 PS-გრამატიკის ალგორიტმი განსაზღვრება

PS-გრამატიკა (PS-grammar) არის $\langle V, V_T, S, P \rangle$ ოთხეული, სადაც:

1. **V** არის ნიშნების (**signs**) სასრული სიმრავლე.
2. **V_T** არის **V** სიმრავლის საკუთრივი ქვესიმრავლე. ამასთან, მისი წევრები ტერმინალურ სიმბოლოებად (**terminal symbols**) იწოდება,
3. **S** არის ნიშანი, რომელიც **V** და **V_T** სიმრავლეების სხვაობის წევრია და მას დამწყებ სიმბოლოს (**start symbol**) უწოდებენ, და
4. **P** არის $\alpha \rightarrow \beta$ სახის გადამწერის წესების (**rewrite rules**) სიმრავლე, სადაც α არის **V** სიმრავლის, β კი – **V^{*}** სიმრავლის² წევრი.

ამგვარად, **PS**-გრამატიკის ძირითადი კომპონენტებია **V**, **V_T** და **P** სიმრავლეები და **S** დამწყები სიმბოლო. **V_T** სიმრავლის შემადგენელი ტერმინალური სიმბოლოები არიან ენის სიტყვების გარეზედაპირები (**word surfaces**), ანუ სიტყვები, როგორც მხოლოდ გამოსახულებები, ხოლო **V** და **V_T** სიმრავლეების სხვაობის შემადგენელ არატერმინალურ სიმბოლოებს (**nonterminal symbols**) ცვლადებს უწოდებენ. ჩვენ გამოვიყენებთ ბერძნულ ასოებს **V^{*}** სიმრავლის მიმდევრობების აღსანიშნავად, დიდი ლათინური ასოებით ინდივიდუალურ ცვლადებს (**individual variables**) აღვნიშნავთ, პატარა ლათინური ასოებით კი – ინდივიდუალურ ტერმინალურ სიმბოლოებს (**individual terminal symbols**).

PS-გრამატიკა გადამწერი წესების მეშვეობით ენის გამოსახულებებს შემდეგნაირად წარმოქმნის: წესის მარცხენა მხარეს არსებული სიმბოლოთა მიმდევრობა იცვლება წესის მარჯვენა მხარეს არსებული სიმბოლოთა მიმდევრობით. მაგალითად, თუ $\alpha \rightarrow \beta$ არის **P** სიმრავლეში არსებული გადამწერი წესი, ხოლო γ და δ არიან **V^{*}** სიმრავლის მიმდევრობები, მაშინ

$$\gamma\alpha\delta \Rightarrow \gamma\beta\delta$$

არის $\gamma\alpha\delta$ მიმდევრობის პირდაპირი ჩანაცვლება $\gamma\beta\delta$ მიმდევრობით. სხვა სიტყვებით: $\gamma\beta\delta$ მიმდევრობა მითლება $\gamma\alpha\delta$ მიმდევრობაზე $\alpha \rightarrow \beta$ წესის პირდაპირი მიმართვით, ანუ გამოყენებით.

PS-გრამატიკის გადამწერი წესების ზოგადი სახე გვიბიძგებს შემდეგი სახის სისტემური შეზღუდვების შემოტანისკენ: (რიგითი თანამიმდევრობა განპირობებულია ტრადიციით):

8.1.2 შეზღუდვები PS-სემატურ ჭესებზე

0. **შეუზღუდავი PS-წესები (unrestricted PS-rules):** 0 ტიპის წესის მარჯვენა და მარცხენა მხარე შეიძლება იყოს ტერმინალური და არატერმინალური სიმბოლოების ნებისმიერი მიმდევრობა.
1. **კონტექსტური დამოკიდებული PS-წესები (context sensitive PS-rules):** 1 ტიპის წესის მარჯვენა და მარცხენა მხარე შეიძლება იყოს ტერმინალური და არატერმინალური სიმბოლოების ნებისმიერი მიმდევრობა. თუმცა, აუცილებელია, რომ მარცხენა მხარეს განთავსებული გამოსახულების სიგრძეს. მაგალითი: **ABC→ADEC**

² **V^{*}** არის **V** სიმრავლის დადებითი ჩაკეტვა, **V^{*}** კი – კლინის ჩაკეტვა.

2. კონტექსტისგან თავისუფალი **PS-წესები (context-free PS-rules)**: 2 ტიპის წესის მარცხენა მხარე მხოლოდ ერთი ცვლადისგან უნდა შედგებოდეს, ხოლო მარჯვენა მხარე შეიძლება იყოს V^+ სიმრავლის ნებისმიერი მიმდევრობა. მაგალითი: $A \rightarrow BC$, $A \rightarrow bBCc$ და $.A\vec{}$.³

3. წესიერი **PS-წესები (regular PS-rules)**: 3 ტიპის წესის მარცხენა მხარე მხოლოდ ერთი ცვლადისგან უნდა შესდგებოდეს, მარჯვენა მხარე კი უნდა შეიცავდეს ერთ, და მხოლოდ ერთ ტერმინალურ სიმბოლოს და არ უნდა შეიცავდეს ერთზე მეტ ცვლადს. მაგალითი⁴: $A \rightarrow b$, $A \rightarrow bC$

რადგან წესის ტიპის რიგითი ნომრის ზრდასთან ერთად (დაწყებული 0 ტიპიდან და დამთავრებული 3 ტიპით) შემზღვეულ პირობათა მოცულობა სულ უფრო ფართოვდება, გასაგები ხდება, რომ ნებისმიერი ტიპის წესი აკმაყოფილებს ყველა მასზე დაბალი ტიპის წესის მოთხოვნებსაც, რაც, ბუნებრივია, იმასაც ნიშნავს, რომ ნებისმიერი ტიპის წესი ამავდროულად არის ნებისმიერი მასზე დაბალი ტიპის წესიც. მაგალითად, 3 ტიპის წესი

$A \rightarrow bC$

აკმაყოფილებს 0, 1, და 2 ტიპის წესების პირობებს. ამასთან, კონტექსტისგან თავისუფალი ტიპის (ე.ი. 2 ტიპის) წესი

$A \rightarrow BC$

არ აკმაყოფილებს 3 ტიპის წესის შემზღვდავ პირობებს (თუმცა იგი აკმაყოფილებს 0 და 1 ტიპის წესების პირობებს). გასაგებია, რომ სხვა ტიპის წესებთან დაკავშირებითაც ვითარება ანალოგიურია.

8.1.2 განაწერში **PS-გრამატიკის წესების განსხვავებული შეზღუდვებით PS-გრამატიკის ოთხი განსხვავებული ტიპი ისაზღვრება: PS-გრამატიკა, რომელიც მხოლოდ 3 ტიპის, შესაბამისად 2 ტიპის, შესაბამისად 1 ტიპის წესებს შეიცავს, წესიერ (**regular**), შესაბამისად კონტექსტისგან თავისუფალ (**context-free**), შესაბამისად კონტექსტზე დამოკიდებულ (**context sensitive**) PS-გრამატიკად (**PS-grammar**) იწოდება; PS-გრამატიკა, რომლის წესებიც არანაირად არ იზღუდება, შეუზღუდავ **PS-გრამატიკად (unrestricted PS-grammar)** იწოდება.**

ამ ოთხი განსხვავებული ტიპის **PS-გრამატიკით** წარმოიქმნება ოთხი განსხვავებული კლასის ენები, კერძოდ: წესიერი **PS-გრამატიკით** წარმოიქმნება წესიერი ენები (**regular languages**); კონტექსტისგან თავისუფალი **PS-გრამატიკით** წარმოიქმნება კონტექსტისგან თავისუფალი ენები (**context-free languages**); კონტექსტზე დამოკიდებული **PS-გრამატიკით** წარმოიქმნება კონტექსტზე დამოკიდებული ენები (**context sensitive languages**); შეუზღუდავი **PS-გრამატიკით** წარმოიქმნება რეკურსულად გადათვლადი ენები (**recursively enumerable languages**).

³ კონტექსტისგან თავისუფალი გრამატიკა ხანდახან იყენებს $A \rightarrow \epsilon$ ფორმის ე.წ. „ეფსილონ წესებსაც“. თუმცა, ირკვევა, რომ ნებისმიერ შემთხვევაში შესაძლებელია ეფსილონ წესების ელიმინირება (იხ. .J.E.Hopcroft & J.D.Ullman, 1979, პ.9.0, თეორემა 4.3). ჩვენი მიღომებით 2 ტიპის წესებისთვის ვითხოვთ მარჯვენა მხარეს განთავსებული მიმდევრობის არაცარიელობას, რათა ფორმლურად შესაძლებელი იყოს კონტექსტისგან თავისუფალი წესების განხილვა კონტექსტზე დამოკიდებული წესების კერძო შემთხვევად.

⁴ ეს არის მარჯვნივ წრფივი **PS-გრამატიკის (right linear PS-grammar)** მაგალითი. $A \rightarrow b$, $A \rightarrow Cb$ წესებით გნესაზღვრული **PS-გრამატიკა** არის ზემოაღწერილი მარჯვნივ წრფივი გრამატიკის ინვერსული მარცხნივ წრფივი **PS-გრამატიკა (left linear PS-grammar)**. მარცხნივ და მარჯვნივ წრფივი გრამატიკები ურთიერთებებისათვის არა განკუთხებულია (იხ. J. E.Hopcroft & J. D. Ullman, 1979, გვ. 219, თეორემა 9.2).

ამასთან, განსხვავებული ტიპის ენების ეს კლასები ერთმანეთშია ჩალაგებული, კერძოდ: წესიერი ენების კლასი საკუთრივი ქვესიმრავლეა კონტექსტისგან თავისუფალი ენების კლასის, რომელიც, თავის მხრივ, საკუთრივი ქვესიმრავლეა კონტექსტზე დამოკიდებული ენების კლასის, ეს უკანასკნელი კი, ასევე საკუთრივი ქვესიმრავლეა რეკურსულად გადათვლადი ენების კლასის.

ენათა ამ კლასებს შორის არსებული განსხვავებები განპირობებულია **შესაბამისი PS-გრამატიკების** განსხვავებული წარმოქმნითი უნარით (შესაძლებლობით). გრამატიკული ტიპის წარმოქმნის უნარი მაღალია, თუ ამ ტიპის გრამატიკას შეუძლია არა მარტო რეკურსული წარმოქმნა ენის მრავალი განსხვავებული ფორმალური სტრუქტურისა, არამედ, ამავდროულად, თუ მას შეუძლია გამორიცხოს (**exclude**) ის სტრუქტურებიც, რომლებიც არ არიან ენის ნაწილი. მეორეს მხრივ, გრამატიკული ტიპის წარმოქმნის უნარი დაბალია, თუ ამ ტიპის შესაბამისი გრამატიკები იძლევაან მხოლოდ შეზღუდული კონტროლის დამყარების საშუალებას წარმოსაქმნელ სტრუქტურებზე⁵.

8.2 მონაცემი კლასები და გამოთვლითი სირთულე

გრამატიკული ტიპით წარმოქმნილი ენების კლასზე და ამ გრამატიკული ტიპის წარმოქმნის უნარზეა დამოკიდებული გამოთვლითი სირთულის ხარისხი, ანუ ენობრივი კლასის გამოსახულებების საანალიზოდ საჭირო გამოთვლის დრო და/ან მეხსიერების მოცულობა. გამოთვლითი სირთულე განუწყვეტლივ იზრდება წარმოქმნის უნარის ზრდასთან ერთად. მოკლედ: სქემატური წარმოქმნელი წესების განსხვავებული შეზღუდვები შედეგად იძლევა

- განსხვავებულ გრამატიკულ ტიპებს (**types of grammar**), რომლებსაც გააჩნიათ
- განსხვავებული ხარისხის წარმოქმნითი უნარები (**degrees of generative capacity**) და რომლებიც წარმოქმნიან
- განსხვავებულ ენობრივ კლასებს (**language classes**), რომლებიც, თავის მხრივ, ნიმუშებია
- განსხვავებული ხარისხის გამოთვლითი სირთულეებისა (**degrees of computational complexity**).

ზემოთ გაშუქებული სტრუქტურული ურთიერთდამოკიდებულება არ იზღუდება **PS-გრამატიკით**, არამედ იგი სამართლიანია წარმომქმნელი გრამატიკის ნებისმიერი მკაცრად განსაზღვრული (**well-defined**) ფორმალიზმისათვის (იხ. 8.2.3 და 11.5.10).

⁵ მაგალითად, 7.1.3 პუნქტში $a^k b^k$ ხელოვნური ენისათვის აღწერილი **PS-გრამატიკის** წარმოქმნის უნარი უფრო მაღალია, ვიდრე 8.3.2 პუნქტში **{a, b}** სიმრავლეზე განსაზღვრული თავისუფალი მონოიდისათვის აღწერილი წესიერი **PS-გრამატიკის** წარმოქმნის უნარი (იხ. 7.1.2). თავისუფალი მონოიდი შეიცავს $a^k b^k$ სახის ყველა გამოსახულებას, მაგრამ მის შესაბამის წესიერ **PS-გრამატიკას** არ შეუძლია გამორიცხოს ის გამოსახულებები, რომლებიც არ ეკუთვნიან $a^k b^k$ ენას.

წარმოქმნითი გრამატიკული ფორმალიზმის სირთულე იზომება იმ ალგორითმის სირთულით, რომლის საფუძველზეც ეს გრამატიკა, როგორც ოპერაციული პროცედურა, რეალიზდება ასტრაქტულ ავტომატზე (მაგალითად ისეთზე, როგორებიცაა ტიურინგის მანქანა, წრფივად ბმული ავტომატი, ძირს მდიდები ავტომატი, სასრულ მდგომარეობისანი ავტომატი დ.ა.შ.). ამ ალგორითმის სირთულე კი იმ პრიმიტიული ოპერაციების⁶ (**primitive operations**) რიცხვით ისაზღვრება, რომელსაც ითხოვს დასამუშავებლად შემოსული გამოსახულების ანალიზი ყველაზე უარეს შემთხვევაში (ზედა ზღვარი⁷ (**upper bound**)). ამასთან, ამ პრიმიტიული ოპერაციების ეს რიცხვი ისაზღვრება დასამუშავებელი გამოსახულების სიგრძესთან ურთიერთკავშირში.

წარმომქმნელი გრამატიკების ელემენტარულ ფორმალიზმებში, ვთქვათ ისეთებში, როგორებიცაა **PS**-გრამატიკა ან **LA**-გრამატიკა, სირთულე ისაზღვრება (ე.ი. ითვლება) შხოლოდ მათი ისეთი მკაცრად განსაზღვრული ქვეტიპებისათვის როგორებიცაა, ვთქვათ, **PS**-გრამატიკის წესიერი, კონტექსტისგან თავისუფალი, კონტექსტზე დამოკიდებული და შეუზღუდული ქვეტიპები, ანდა **LA**-გრამატიკის **C1-, C2-, C3-, B-** და **A-LAG** ქვეტიპები. ამის გარდა, ენობრივი კლასის სირთულე გაიგივებულია მასთან მიკავშირებული გრამატიკის ტიპის სირთულესთან. მაგალითად, ამბობენ, რომ კონტექსტისგან თავისუფალი ენების სირთულე არის n^3 (**the context-free languages have a complexity of n^3**), ეს იმიტომ, რომ ცნობილია ალგორითმი, რომელსაც შეუძლია კონტექსტისგან თავისუფალი ნებისმიერი **PS**-გრამატიკის ნებისმიერი გამოსახულება გაანალიზოს n^3 პრიმიტიული ოპერაციის გამოყენებით, სადაც **n** არის დასამუშავებელი (შესავალი) გამოსახულების სიგრძე.

გამოთვლითი სირთულის განსხვავებული ხარისხები შედეგად შემდეგ ოთხ ძირითად კლასს იძლევა.

8.2.1 სირთულის ძირითადი ხარისხები

1. წრფივი სირთულე (**Linear complexity**):
 n , $2n$, $3n$, დ.ა.შ.
2. პოლინომიალური სირთულე (**Polynomial complexity**):
 n^2 , n^3 , n^4 , დ.ა.შ.
3. ექსპონენციალური სირთულე (**Exponential complexity**):
 2^n , 3^n , 4^n , დ.ა.შ.
4. არაამოხსნადი (ამოუხსნადი) (**Undecidable**):
 $n \times \infty$

⁶ ჯ ერლეიმ (J. Earley) 1970 წელს დაახასიათა პრიმიტიული ოპერაცია, როგორც „რადაც აზრით ურთულესი ოპერაცია, რომელსაც ალგორითმი აწარმოებს და რომლის სირთულეც დამოუკიდებელია გრამატიკისა და შეყვანილი მიმდევრობის ზომისაგან“. პრიმიტიული ოპერაციის გაგება იცვლება ერთი გრამატიკული ფორმალიზმიდან მეორეში გადასვლისას, მაგალითად, ერლეიმ თავისი ცნობილი კონტექსტისგან თავისუფალ გრამატიკებში ალგორითმის პრიმიტიულ ოპერაციად აირჩია ოპერაცია: ერთი სტრუქტურის დამატება სტრუქტურათა სიმრავლეზე (იხ. 9.3). სხვა მნიშვნელოვანი, **LA**-გრამატიკის **C-LAG** კლასის ქვეკლასი პრიმიტიულ ოპერაციად იყენებს წესის გამოყენებას (იხ. 11.4).

⁷ აյ საუბარია დროით სირთულეზე.

იმის გასაზრებლად თუ რა გავლენას ახდენს სირთულის ხარისხი პრაქტიკულ გადაწყვეტებზე, განვიხილოთ გერმანული ენის ლიმას კორპუსი (**Limas corpus**).⁸ ლიმას კორპუსში წინადადებების საშუალო სიგრძე არის 17.54 სიტყვა ფორმა (პუნქტუაციის ნიშნების ჩათვლით). ეს იმას ნიშნავს, რომ **3n** სირთულის წრფივი ალგორითმი საშუალო სიგრძის წინადადების ($n=17$) ანალიზისთვის მოგვთხოვს არაუმეტეს **51** ოპერაციისა, n^3 სირთულის პოლინომიალური ალგორითმი მოგვთხოვს არაუმეტეს **4913** ოპერაციისა, **3^n** სირთულის ექსპონენციალური ალგორითმი მოგვთხოვს არაუმეტეს **127 362 132** ოპერაციისა, ამოუხსნადი სირთულის ალგორითმი კი – არაუმეტეს **17 × ω** (= ∞) ოპერაციისა.

მ. რ. გარეიმ და დ. ს. ჯონსონი (**M.R. Garey & D.S. Johnson**) 1979 წელს შეადარეს, თუ რა დროს მოანდომებდნენ n^3 (პოლინომიალური) და 2^n (ექსპონენციალური) სირთულის მქონე ალგორითმები სხვადასხვა ზომის ერთი და იგივე პრობლემების გადაჭრას. შედეგები ასახულია ქვემოთ მოყვანილ ცხრილში.

8.2.2 პოლინომიალური და ესპარენციალური ალგორითმების შედარება

	პრობლემის ზომა n		
დროითი სირთულე	10	50	100
n^3	.001 წამი	.125 წამი	1.0 წამი
2^n	.001 წამი	35.7 წელი	10^{15} საუკუნე

ამ მაგალითში გარეი და ჯონსონი შემდეგი სიტყვის დამატებას თავიანთი ალგორითმის პრიმიტიულ ოპერაციად განიხილავდნენ.

მთლიანად ლიმას კორპუსში 71148 წინადადებაა, აქედან ზუსტად **50** წინადადება შეიცავს არანაკლებ **100** სიტყვა ფორმას. ამასთან, კორპუსის ყველაზე გრძელი წინადადება 165 სიტყვისაგან შესდგება.⁹ ამგვარად, გარეი და ჯონსონის გაზომვებით გამოდის, რომ 2^n ხარისხის ექსპონენციალური სირთულის გრამატიკული ალგორითმით კორპუსის ავტომატური ანალიზი ყველაზე ცუდ შემთხვევაში 10^{15} საუკუნეზე მეტ ხანს უნდა გაგრძელდეს. ეს დრო ბევრად მეტია

⁸ როგორც 15.3 ქვეთავშია ახსნილი, ლიმას კორპუსი აიგო **Brown** და **LOB-ის** კორპუსის ანალოგიურად და შეიცავს 500 ტექსტს, რომელთაგან თითოეული 2000 სიტყვა ფორმისგან შესდგება. ამასთან, ეს ტექსტები აირჩა სრულიად ნებისმიერად ტექსტების იმ 15 განსხვავებული ტიპიდან, რომლებიც ასევე სასაფუძვლო იყო **Brown** და **LOB-ის** კორპუსისათვის. გასაგებადა, რომ 1973 წელს განხორციელებული ამ მცდელობის მიზანი იყო რაც შეიძლება მეტი მიახლოვება იმ სასურველ და ბალანსირებულ კორპუსთან, რომელიც, თავის მხრივ, მთელი გერმანული ენის წარმიმადგრენელი იქნებოდა.

⁹ ეს მონაცემები გამოაქვეყნა მარკუს შულცმა **CLUE-ში**.

სამყაროს ასაკზე და სწრაფი მანქანებითაც კი მისი პრაქტიკულად ღირებულ დრომდე შემცირება არარეალურია.

PS-გრამატიკისათვის შემზღვდავი წესების, გრამატიკული ტიპების, ენობრივი კლასებისა და სირთულის ხარისხების ურთიერთკავშირები ასახულია ქვემოთ მოყვანილ ცხრილში.

8.2.3 ფორმალური ენების **PS**-გრამატიკული იმრარჩია

შეზღუდვები	PS -გრამატიკის ტიპები	ენათა კლასები	სირთულის ხარისხები
ტიპი 3	წესიერი PSG	წესიერი ენები	წრფივი
ტიპი 2	კონტ. თავისუფალი PSG	კონტ. თავისუფალი ენები	პოლინომური
ტიპი 1	კონტ. დამოკიდებული PSG	კონტ. დამოკიდებული ენები	ექსპონენციალური
ტიპი 0	შეზღუდვავი PSG	რეკურ. გადათვლადი ენები	არამოწინადი

ამ **PS**-გრამატიკულ იერარქიას ჩომსკის იერარქიასაც უწოდებენ. ამ იერარქიის ალტერნატიულ იერარქიად განიხილება ფორმალური ენების კლასების **LA**-გრამატიკული იერარქია (იხ. 11-ე თავის 11.5.10 და 11.5.11 ქვეთავები). ეს ალტერნატიული **LA**- და **PS**- გრამატიკული იერარქიები შედარებულია მე-12 თავში.

8.3 წარმოქმნითი უნარი და ფორმალურ ენათა კლასები

ლინგვისტური თვალსაზრისებით ისმის კითხვა: არსებობს თუ არა ისეთი ტიპის **PS**-გრამატიკა, რომელიც წარმოქმნის იმ და მხოლოდ იმ სტრუქტურებს, რომლებიც ბუნებრივი ენებისთვისაა დამახასიათებელი. ამ კითხვის ინტერესებიდან გამოდინარე უფრო ახლოს მივიდეთ იმ სტრუქტურებთან, რომლებსაც განსხვავებული ტიპის **PS**-გრამატიკები წარმოქმნიან.

PS-გრამატიკის ტიპებს შორის ყველაზე მეტად შეზღუდული წესები, წარმოქმნის უმდაბლესი უნარი და გამოთვლითი სირთულის უმდაბლესი ხარისხი აქვთ წესიერი ტიპის გრამატიკებს.¹⁰ წესიერი გრამატიკის წარმოქმნითი უნარი ერთი წარმოქმნითი ოპერაციის ფარგლებში ემყარება ყოველგვარი თანამდევი რეკურსული მოვლენების გარეშე მხოლოდ ერთი, ცალკე აღებული სიტყვის რეკურსული გამორების შესაძლებლობას.

¹⁰ წესიერი ენების კლასი არ შედის **LA**-გრამატიკულ იერარქიაში, თუმცა ირკვევა, რომ ამ ენობრივ იერარქიაშიც შეიძლება მისი აწყობა (**CoL**, თეორემა 3, გვ. 138). **LA**-გრამატიკულ იერარქიაში არსებობს წესიერი ენების ალტერნატიული **C1**-ენების წრფივი კლასი. როგორც ნაჩვენება 11.5 ქვეთავში, **C1**-ენების კლასი შეიცავს როგორც წესიერ ენებს, ისე ისეთ დეტერმინისტულ კონტექსტისაგან თავისუფალ ენებს, რომლებიც გამოცნობადია ეფსილონ-თავისუფალი **DPDA**-ს მიერ, და, ასევე, ბევრ კონტექსტზე დამოკიდებულ ენებს.

მაგალითად, \mathbf{ab}^k წესიერი ენის ნებისმიერი გამოსახულება შედგება მხოლოდ ერთი \mathbf{a} სიმბოლოსა და მასზე მიღევნებული ან ერთი, ან ორი, ანდა უფრო მეტი \mathbf{b} სიმბოლოსაგან. \mathbf{ab}^k ენის წესიერი **PS**-გრამატიკა განისაზღვრება შემდეგნაირად:

8.3.1 \mathbf{ab}^k ენის ჭესიერი **PS**-გრამატიკა ($k \geq 1$)

$$\begin{aligned} V &=_{\text{def}} \{\mathbf{S}, \mathbf{B}, \mathbf{a}, \mathbf{b}\} \\ V_T &=_{\text{def}} \{\mathbf{a}, \mathbf{b}\} \\ P &=_{\text{def}} \{\mathbf{S} \rightarrow \mathbf{aB}, \mathbf{B} \rightarrow \mathbf{bB}, \mathbf{B} \rightarrow \mathbf{b}\} \end{aligned}$$

წესიერი ენის კიდევ ერთი მაგალითია $\{\mathbf{a}, \mathbf{b}\}$ სიმრავლეზე განსაზღვრული თავისუფალი მონოიდისაგან ნულოვანი წევრის გამოკლებით მიღებული სიმრავლე, რომელიც წარმოიქმნება შემდეგი **PS**-გრამატიკით:

8.3.2 $\{\mathbf{a}, \mathbf{b}\}^+$ ენის ჭესიერი **PS**-გრამატიკა

$$\begin{aligned} V &=_{\text{def}} \{\mathbf{S}, \mathbf{a}, \mathbf{b}\} \\ V_T &=_{\text{def}} \{\mathbf{a}, \mathbf{b}\} \\ P &=_{\text{def}} \{\mathbf{S} \rightarrow \mathbf{aS}, \mathbf{S} \rightarrow \mathbf{bS}, \mathbf{S} \rightarrow \mathbf{a}, \mathbf{S} \rightarrow \mathbf{b}\} \end{aligned}$$

იმაში, რომ წესიერ გრამატიკას არ შეუძლია წარმოქმნას სისტემური შესაბამისობა ნებისმიერად აღებულ რიცხვებს შორის, გვარწმუნებს ჩვენთვის უკვე ნაცნობი კონტექსტისაგან თავისუფალი $\mathbf{a}^k \mathbf{b}^k$ ენისა და წესიერი $\mathbf{a}^m \mathbf{b}^k$ ენის დაპირისპირება.

8.3.3 $\mathbf{a}^m \mathbf{b}^k$ ენის ჭესიერი **PS**-გრამატიკა ($k, m \geq 1$)

$$\begin{aligned} V &=_{\text{def}} \{\mathbf{S}, \mathbf{S}_1, \mathbf{S}_2, \mathbf{a}, \mathbf{b}\} \\ V_T &=_{\text{def}} \{\mathbf{a}, \mathbf{b}\} \\ P &=_{\text{def}} \{\mathbf{S} \rightarrow \mathbf{aS}_1, \mathbf{S}_1 \rightarrow \mathbf{aS}_1, \mathbf{S}_1 \rightarrow \mathbf{bS}_2, \mathbf{S}_2 \rightarrow \mathbf{b}\} \end{aligned}$$

$\mathbf{a}^m \mathbf{b}^k$ ენა წესიერია, რადგანაც \mathbf{a} და \mathbf{b} სიმბოლოების რიცხვი $\mathbf{a}^m \mathbf{b}^k$ გამოსახულებაში ერთმანეთისაგან დამოუკიდებელია. — ამას ერთმანეთისაგან განსხვავებული ზედა m და k ინდექსები უზრუნველყოფენ. მეორე მხრივ, $\mathbf{a}^k \mathbf{b}^k$ ენა წესიერი **PS**-გრამატიკის წარმოქმნითი უნარის, ანუ შესაძლებლობის მიღმა რჩება, რადგანაც ის მოითხოვს \mathbf{a} და \mathbf{b} სიმბოლოების რიცხვის არა ერთმანეთისაგან დამოუკიდებლობას, არამედ, პირიქით, მათ ერთმანეთთან შესატყვისობას. — ეს მათი ერთი და იგივე ზედა k ინდექსით არის განპირობებული.

ინტუიციური მოსაზრებებით, წესიერი **PS**-გრამატიკის წარმოქმნის უნარის შეზღუდულობის მიზეზი უნდა მომდინარეობდეს იმ შეზღუდვებიდან რითაც მისი წესები ხასიათდება: იმის გამო, რომ 3 ტიპის წესის მარჯვენა მხარე შესდგება ზუსტად ერთი ტერმინალური სიმბოლოსა და მაქსიმუმ ერთი ცვლადისაგან, მას, მისი ამ შეზღუდული რეკურსული საშუალებებით, უბრალო დაწყვილების წარმოქმნაც კი არ შეუძლია. — რასაც ბუნებრივია ითხოვს $\mathbf{a}^k \mathbf{b}^k$ ენა.

ფორმალური დამტკიცება იმისა, რომ წესიერი **PS**-გრამატიკის წარმოქმნის უნარი უფრო დაბალია, ვიდრე კონტექსტისაგან თავისუფალი **PS**-გრამატიკისა, არ არის ტრიგიალური და ემყარება

წესიერი ენებისთვის არსებულ მტუმბავ ლემას (**pumping lemma**)¹¹, რომელიც ფორმალურად ამტკიცებს, რომ არსებოს ენები, რომელთა წარმოქმნა წესიერი **PS**-გრამატიკებით შეუძლებელია.

მტუმბავი ლემა გარკვეული ენობრივი კლასისათვის გვიჩვენებს, თუ როგორი სტრუქტურები შეიძლება პქნოდეს მას. ეს კეთდება ამ ენობრივი კლასის იმ ძირითადი სტრუქტურული ნიმუშების (**patterns**) ისეთი ცხადი ამოწერით, რომ შესაძლებელი ხდება მათგან ყველა შესაძლო დამატებითი გამოსახულებების უსასრულო სიმრავლის ამოტუმბავა, რითაც საბუთდება ის, რომ ისინი შეიძლება შედგებოდნენ მხოლოდ ამ ძირითადი სტრუქტურული ნიმუშების გარკვეული სახის გამორჩებით.

PS-გრამატიკულ იერარქიაში გრამატიკის შემდეგი იერარქიული ტიპი არის კონტექსტისგან თავისუფალი **PS**-გრამატიკა. კონტექსტისგან თავისუფალი ენის მაგალითებია: $a^k b^k$ ენა, რომლის წარმომქმნელი **PS**-გრამატიკა 7.1.3 ქვეთავში, ხოლო **C**-გრამატიკა 7.4.3 ქვეთავშია განსაზღვრული, და $a^{k_1} b^{3k}$ ენა, რომლის განსაზღვრებაც აქვე კეთდება.

8.3.4 $a^k b^{3k}$ ენის პონტექსტისგან თავისუფალი **PS**-გრამატიკა

$$\begin{aligned} V &= \text{def } \{S, a, b\} \\ V_T &= \text{def } \{a, b\} \\ P &= \text{def } \{S \rightarrow aSbbb, S \rightarrow abbb\} \end{aligned}$$

2 ტიპის გრამატიკებს კონტექსტისგან თავისუფალს უწოდებენ, რადგანაც, განსაზღვრების თანახმად (იხ. 8.1.2), ამ ტიპის წესების მარცხნა მხარე შესდგება მხოლოდ ერთი გამხოლოებული ცვლადისაგან. – ე.ი. სხვა „კონტექსტურად“ მოსაზღვრე ნიშნებისა და სიმბოლოების გარეშე¹².

კონტექსტისგან თავისუფალი წესის ფორმას (იხ. 7.1.3 და 8.3.4) გავყავართ წარმოქმნის უნარის სხვა შეზღუდვაზე: კონტექსტისგან თავისუფალ გრამატიკებს შეუძლიათ რეკურსულად წარმოქმნან ურთიერთშესაბამისობები, მაგრამ მხოლოდ ისეთები, როგორებიცაა ვთქვათ **abc...cba** მიმდევრობის მსგავს ინვერსულ წყვილებში (**inverse pair**).¹³

ინვერსული წყვილის სტრუქტურა, რითაც კონტექსტისგან თავისუფალი ენები ხასიათდებიან, კარგად ჩანს კონტექსტისგან თავისუფალი გამოსახულებების გამოყვანებში. სამაგალითოდ განვიხილოთ კონტექსტისგან თავისუფალი ენა **WW^R**, სადაც **W** აღნიშნავს სიტყვათა ნებისმიერ მიმდევრობას (მაგ. **abcd**), ხოლო **W^R** შესაბამის ინვერსულ მიმდევრობას (ე.ი. **dcb**).¹⁴

8.3.5 **WW^R** ენის პონტექსტისგან თავისუფალი **PS**-გრამატიკა

$$\begin{aligned} V &= \text{def } \{S, a, b, c, d\} \\ V_T &= \text{def } \{a, b, c, d\} \\ P &= \text{def } \{S \rightarrow aSa, S \rightarrow bSb, S \rightarrow cSc, S \rightarrow dSd, S \rightarrow aa, S \rightarrow bb, S \rightarrow cc, S \rightarrow dd\} \end{aligned}$$

¹¹ იხ. Hopcroft and Ullman, 1979, გვ.55.

¹² „კონტექსტის“ გაგება სპეციფიკურია **PS**-გრამატიკის ტერმინოლოგიისათვის და არაფერი კავშირი არ აქვს მოსაზღუდებელი ინტერნალურ კონტექსტურ გაგებასთან (იხ. თავი 3–6 თავები).

¹³ ყოველი კონტექსტისგან თავისუფალი ენა არის წესიერი სიმრავლის და ნახევრად დიკ სიმრავლის გადაკვეთის პომობრიჯული (ჩომსკი-შუტცენბერგის თეორება). იხ. M. Harrison 1978, გვ. 317.

¹⁴ ზედა **WW^R** ინდექსი **WW^R** გამოსახულებაში სიმბოლურად განიშნებს ინვერსიაზე.

წესიერი ენების კლასთან შედარებით კონტექსტისგან თავისუფალი ენების კლასის გაზრდილი წარმოქმნითი უნარი იწვევს გამოთვლითი სირთულის ზრდასაც. მართლაც, მაშინ როცა წესიერი ენები ანალიზდება წრფივ დროში, კონტექსტისგან თავისუფალი ენების გაანალიზებას პოლინომური დრო სჭირდება (იხ. 8.2.3).

კონტექსტისგან თავისუფალი **PS**-გრამატიკების წარმოქმნითი უნარი საკმაოდ შეზღუდულია¹⁵. კლასიკურ მაგალითად იმისა, რომ ენა სცდება კონტექსტისგან თავისუფალი **PS**-გრამატიკის წარმოქმნითი უნარით განსაზღვრულ შესაძლებლობებს, განვიხილოთ $a^k b^k c^k$ ხელოვნური ენა, რომლის გამოსახულებები **a**, **b** და **c** სიმბოლოებით აგებული ტოლი სიგრძის მიმდევრობებისგან შესდგება. მაგალითად: **a b c**, **a a b b c c**, **a a a b b b c c c**, და ა.შ.

კონტექსტისგან თავისუფალ **PS**-გრამატიკას არ შეუძლია $a^k b^k c^k$ ენის წარმოქმნა, რადგანაც ის მოითხოვს სამი განსხვავებული ნაწილის ურთიერთკავშირს – ეს არ ჯდება კონტექსტისგან თავისუფალი ენების დაწყვილებული ინვერსულობის სტრუქტურულ ბუნებაში, რაც ჩვენთვის ცნობილია **a^kb^k** და **WW^R** ენების უკვე განხილული მაგალითებით.

კიდევ ერთი ენა, რომელიც სცდება კონტექსტისგან თავისუფალი **PS**-გრამატიკების წარმოქმნითი უნარით განსაზღვრულ შესაძლებლობებს არის **WW** ენა. აյ **W** სიტყვების ნებისმიერი მიმდევრობაა. თუ 8.3.5 პუნქტით განსაზღვრული კონტექსტისგან თავისუფალი **WW^R** ენა შეიცავდა შემდეგი სახის გამოსახულებებს

a a
a b b a
a b c c b a
a b c d d c b a
...

რომლებიც დაწყვილებული ინვერსულობის აშკარად გამოკვეთილი სტრუქტურით ხასიათდებოდნენ, **WW** ენა შეიცავს შემდეგი სახის გამოსახულებებს

a a
a b a b
a b c a b c
a b c d a b c d
...

რომლებსაც არ გააჩნიათ ინვერსულობის სტრუქტურა. ამგვარად, მოუხედავად **WW^R** და **WW** ენების აშკარა მსგავსებისა, **WW** ენისთვის 8.3.5 სახის **PS**-გრამატიკა არ არსებობს.

a^kb^kc^k და **WW** ენების მეშვეობით ინტუიციურად დასაბუთდა, რომ არსებობს ენები, რომელთა წარმოქმნაც შეუძლებელია კონტექსტისაგან თავისუფალი **PS**-გრამატიკით. – ამის ფორმალური

¹⁵ კონტექსტისგან თავისუფალი ენების კლასი არ შედის **LA**-გრამატიკულ იერარქიაში, თუმცა შეიძლება იქ მისი აწყობა (**CoL**, თეორემა 4, გვ. 138). იხ. აგრეთვე 11.2, შენიშვნა 12. სამაგიეროდ **LA**-გრამატიკული იერარქია წარმოშობს ალტერნატიულ **C2**-ენების პოლიონომურ კლასს. როგორც ნაჩვენებია 12.4 ქვეთავში **C2**-ენების კლასი შეიცავს უმეტეს, მაგრამ არა ყველა კონტექსტისგან თავისუფალ ენას, ასევე ბევრ კონტექსტურ დამოკიდებულ ენას.

დამტკიცება არანაირად არ არის ტრივიალური. როგორც წესიერი ენების შემთხვევაში გვქონდა, დამტკიცება ამ შემთხვევაშიც მტუმბავ ლემას ემყარება. თუმცა, ამჯერად, ეს ლემა ყალიბდება კონტექსტისგან თავისუფალი ენებისათვის.¹⁶

შემდეგი უფრო ფართო ენობრივი კლასი **PS**-გრამატიკულ იერარქიაში არის კონტექსტზე დამოკიდებული ენები, რომელსაც **PS**-გრამატიკა წარმოქმნის 1 ტიპის წესების გამოყენებით.

თითქმის ნებისმიერი ენა, რომელზეც შეგვიძლია ვიფიქროთ, კონტექსტზე დამოკიდებულია; ერთადერთი ცნობილი დამტკიცება, რომ გარკვეული სახის ენები არ არიან **CSL** ტიპის, საბოლოო ჯამში დაფუძნებულია დიაგონალიზაციაზე.

J.E. Hopcroft and J.D. Ullman, 1979, გვ. 224

რადგანაც პრაქტიკულად შეუძლებელია კონტექსტზე დამოკიდებული ენების საბაზისო სტრუქტურების ნიმუშების ამომწურავი ჩამოწერა „ვინმებ შეიძლება იგარაუდოს“, რომ კონტექსტზე დამოკიდებული ენებისათვის მტუმბავი ლემა არ მუშაობს. 1 ტიპის, ანუ კონტექსტზე დამოკიდებული წესების სტრუქტურა აღწერილია 8.3.6 განაწერში:

8.3.6 პონტექსტზე დამოკიდებული ღვამების სტაციარატული სემა

$\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2$, სადაც β არ არის ცარიელი მიმდევრობა.

PS-გრამატიკაში კონტექსტზე დამოკიდებული და კონტექსტისგან თავისუფალი ტერმინების ცნებითი შინაარსები გაიგება მათ ურთიერთდაპირისპირებულობაში. მაშინ როდესაც კონტექსტისგან თავისუფალი 2 ტიპის წესის მარცხენა მხარეს დაიშვება მხოლოდ ერთი გამხოლობული ცვლადის განთავსება, კონტექსტზე დამოკიდებული 1 ტიპის წესის მარცხენა მხარეში ნებადართულია ცვლადის განთავსება გარემზობელ (ე.ი. მის ორივე მხარეს მდებარე) ტერმინალურ სიმბოლოებთან ერთად. ეს, როგორც ამას 8.3.6 პუნქტში აღწერილი სქემაც გვიჩვენებს, 1 ტიპის წესის კონტექსტზე დამოკიდებულს ხდის, რადგან ამ სქემის მიხედვით, **A** ცვლადი შეიძლება ჩანაცვლდეს (ე.ი. გადაიწეროს) β სიმბოლოთი მხოლოდ $\alpha_1 — \alpha_2$ ‘კონტექსტში’.

1 ტიპის წესის მარცხენა მხარეს განთავსებული ცვლადის გარემომცველი კონტექსტის სხვადასხვაგვარი განსაზღვრის შესაძლებლობები დიდად ზრდის გადაწერების გაკონტროლების საშუალებებს, რაც, თავის მხრივ, ასევე დიდად ზრდის კონტექსტზე დამოკიდებული **PS**-გრამატიკების წარმოქმნით უნარს. ამაში გვარწმუნებს $a^k b^k c^k$ ენის ქვემოთ აღწერილი კონტექსტზე დამოკიდებული **PS**-გრამატიკა:

8.3.7 პონტექსტზე დამოკიდებული $a^k b^k c^k$ ენის **PS**-პრამატიკა

$$V =_{\text{def}} \{S, B, C, D_1, D_2, a, b, c\}$$

$$V_T =_{\text{def}} \{a, b, c\}$$

$$\begin{aligned} P =_{\text{def}} & \{ S \rightarrow a S B C, & \\ & S \rightarrow a b C, & \text{წეს 1} \\ & C B \rightarrow D_1 B, & \text{წეს 2} \\ & D_1 B \rightarrow D_1 D_2, & \text{წეს 3a} \\ & & \text{წეს 3b} \end{aligned}$$

¹⁶ იხ. Hopcroft and Ullman , გვ.125.

$D_1 D_2 \rightarrow B D_2,$	წესი 3c
$B D_2 \rightarrow B C,$	წესი 3d
$b B \rightarrow b b,$	წესი 4
$b C \rightarrow b c,$	წესი 5
$c C \rightarrow c c \}$	წესი 6

3a – 3d წესების თანმიმდევრული გამოყენება იძლევა იგივე შედეგს რასაც

წესი 3

CB→BC

8.3.7 პუნქტში აღწერილი **PS**-გრამატიკა იყენებს **3a – 3d** წესებს, რადგანაც მათი ეკვივალენტური წესი 3 არ აკმაყოფილებს კონტექსტზე დამოკიდებული წესის იმ მოთხოვნას, რომლის თანახმადაც წესის მარცხენა მხარეს განთავსებულ გამოსახულებაში მხოლოდ ერთი ცვლადის ჩანაცვლებაა დაშვებული.

a^kb^kc^k ენის გამოსახულებებში არსებული სამჯერადი ურთიერთშესაბამისობის გაკონტროლებასა და რეალიზაციაში კონტექსტის გადამწყვეტ როლს თვალნათლივ წარმოაჩენს **aaabbbcc** გამოსახულების ქვემოთ წარმოდგენილი გამოყვანა. პროცესის გამარტივების მიზნით ამ გამოყვანაში 8.3.7 გრამატიკის **3a – 3d** წესების კომბინაცია მათი ეკვივალენტური 3 წესით ნაცვლდება.

8.3.8 a a a b b b c c c გამოსახულების გამოყვანა

შუალედური ჯაჭვი წესები

- | | |
|-----------------------|-----|
| 1. S | |
| 2. a S B C | (1) |
| 3. a a S B C B C | (1) |
| 4. a a a b C B C B C | (2) |
| 5. a a a b B C C B C | (3) |
| 6. a a a b B C B C C | (3) |
| 7. a a a b B B C C C | (3) |
| 8. a a a b b B C C C | (4) |
| 9. a a a b b b C C C | (4) |
| 10. a a a b b b c C C | (5) |
| 11. a a a b b b c c C | (6) |
| 12. a a a b b b c c c | (6) |

კონტექსტზე დამოკიდებული **PS**-გრამატიკის მაღალი წარმოქმნითი უნარი განპირობებულია მის მიერვე უკვე წარმოებული მიმდევრობების გადალაგების (**changing the order of sequences**) უნარით. წარმოებული მიმდევრობების ასეთ გადალაგებას ადგილი ჰქონდა 4-დან 7-მდე განთავსებულ შუალედურ მიმდევრობებში.

კონტექსტზე დამოკიდებული წესებით მიმდევრობის გადალაგების შესაძლებლობა უზრუნველყოფს რეგულირების ხარისხს, რომელიც გაცილებით მაღალია ვიდრე ეს კონტექსტისგან თავისუფალ

PS-გრამატიკებს გააჩნიათ. თუმცა, იგივე მიზეზები ამავდროულად განაპირობებს გამოთვლითი სირთულის მაღალ ხარისხს. იმისათვის, რომ ავტომატური საშუალებებით ზუსტად განვსაზღვროთ კონტექსტზე დამოკიდებული წესების ის მიმღევრობა, რომელიც მოცემულ გამოსახულებას იძლევა, ცხადია უნდა გადამოწმდეს შესაძლო გადალაგებების პოტენციურად ექსპონენციალური რაოდენობის შემთხვევები.

ამ სახის ძიებათა სფერო ისეთი ფართოა, რომ არ შეიძლება არსებობდეს არანაირი პრაქტიკული ლირგულების მქონე პარსერი კონტექსტზე დამოკიდებული **PS**-გრამატიკებისათვის. სხვა სიტყვებით რომ ვთქვათ, კონტექსტზე დამოკიდებული ენების კლასი პრიმიტიულ ავტომატურ გამოთვლას ძნელად ექვემდებარება.

კონტექსტზე დამოკიდებული ენების კლასი არის რეკურსული ენების კლასის საკუთრივი ქვესიმრავლე.¹⁷ რეკურსული ენები არ არის ასახული **PS**-გრამატიკულ იერარქიაში, რაც განპირობებულია იმით, რომ არ არსებობს **PS**-წესებზე (იხ. 8.1.2) მორგებადი შეზღუდვათა ისეთი სისტემა, რომ შესაბამისმა **PS**-გრამატიკულმა კლასმა წარმოქმნას მხოლოდ რეკურსული ენები¹⁸.

ენა მაშინ და მხოლოდ მაშინაა რეკურსული, თუ ის ამოხსნადია, ე. ი. თუ არსებობს ალგორითმი, რომლითაც ნებისმიერი შესავალი მოცემულობისათვის სასრული ბიჯების მეშვეობით ისაზღვრება, ეკუთვნის თუ არა იგი ამ ენას. ისეთი რეკურსული ენის მაგალითს, რომელიც, ამავდროულად, არ არის კონტექსტზე დამოკიდებული, წარმოადგენს აკერძანის ფუნქცია.¹⁹

PS-გრამატიკულ იერარქიაში ყველაზე ფართო ენათა კლასს რეკურსულად გადათვლადი ენათა კლასი წარმოადგენს. ენათა ამ კლასს შეუზღუდავი ანუ 0 ტიპის **PS**-გრამატიკები წარმოქმნის. შეუზღუდავ **PS**-გრამატიკებში წესის მარჯვენა მხარე შეიძლება მოკლე იყოს მარცხენა მხარესთან შედარებით. 0 ტიპის წესების ეს მახასიათებელი თვისება იძლევა უკვე წარმოქმნილი მიმდევრობების ნაწილების წაშლის შესაძლებლობას. – ამის გამოა, რომ რეკურსულად გადათვლადი ენების კლასი არაამოხსნადია. იმის გარკვევამ, არის თუ არა რეკურსულად გადათვლადი ენის გამოსახულება სწორად აგებული, შეიძლება წაიღოს არა ძალიან დიდი დრო, არამედ სამუდამო.²⁰

8.4 **PS**-გრამატიკა პუნქტუაციის მნიშვნელობის

ბუნებრივ ენაში **PS**-გრამატიკის მარტივი გამოყენების დემონსტრირებას ქვემოთ მოყვანილი განსაზღვრების მეშვეობით გავაკეთებთ. ამასთან, შედარებითი ინტერესების გათვალისწინებით აქ განხილული გრამატიკა წარმოქმნის იმავე წინადადებას, რომელსაც წარმოქმნიდა 7.5.3 განაწერში აღწერილი **C**-გრამატიკა.

¹⁷ Hopcroft and Ullman ,1979, გვ.228, ოთორემა 9.8.

¹⁸ **LA**-გრამატიკულ იერარქიაში რეკურსული ენების კლასი ფორმალურად განსაზღვრულია როგორც **A**-ენები, წარმოქმნილი შეუზღუდავი **LA**-გრამატიკებისაგან (იხ. 11.2.2). უფრო მეტიც, კონტექსტზე დამოკიდებული ენების კლასი ფორმალურად განსაზღვრულია როგორც **B**-ენების კლასი, წარმოქმნილი ბმული **LA**-გრამატიკებისაგან (**bounded LA-grammar**).

¹⁹ Hopcroft and Ullman , გვ.175, 7.4.

²⁰ მართალია რეკურსულად გადათვლადი ენების კლასი არ შედის **LA**-გრამატიკის იერარქიაში, თუმცა შეიძლება მისი აწყობა (იხ. 15-ე შენიშვნა 11.2 განაწერის ბოლოს).

8.4.1 PS-გრამატიკა 7.5.3 მაგალითისათვის

$V = \text{def} \{ S, NP, VP, V, N, DET, ADJ, \text{black}, \text{dogs}, \text{little}, \text{sleep}, \text{the} \}$

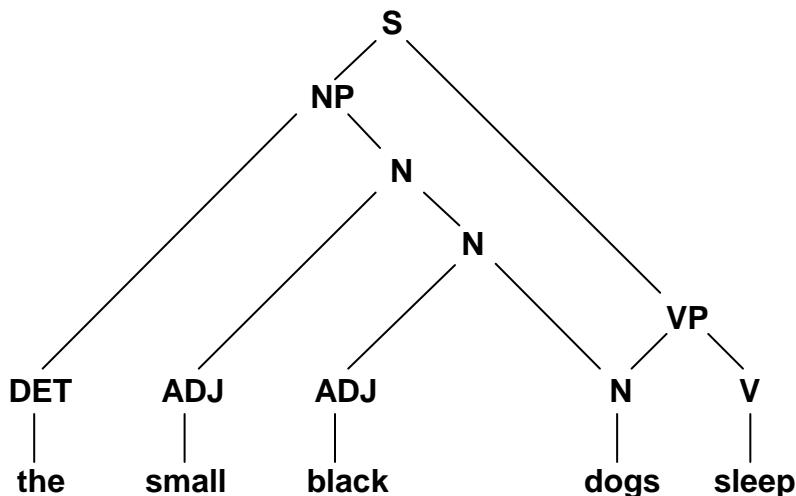
$V_T = \text{def} \{ \text{black}, \text{dogs}, \text{little}, \text{sleep}, \text{the} \}$

$P = \text{def} \{ S \rightarrow NP\ VP, \quad VP \rightarrow V, \quad NP \rightarrow DET\ N, \quad N \rightarrow ADJ\ N,$
 $N \rightarrow \text{dogs}, \quad ADJ \rightarrow \text{little}, \quad ADJ \rightarrow \text{black}, \quad DET \rightarrow \text{the}, \quad V \rightarrow \text{sleep} \}$

ამ **PS**-გრამატიკის ფორმა კონტექსტისგან თავისუფალია: ის ჯერ კიდევ არ არის კონტექსტზე დამოკიდებული. ეს იმიტომ, რომ წესების მარცხნა მხარეები შეიცავენ მხოლოდ ერთ ცვლადს, და უკვე აღარაა წესიერი, იმიტომ, რომ ზოგიერთი წესის მარჯვენა მხარე შეიცავს ერთ ცვლადზე მეტს.

მსგავსად 7.5.4 პუნქტში აღწერილი **C**-გრამატიკული გამოყვანისა, 8.4.1 განსაზღვრებაზე დაფუძნებული **PS**-გრამატიკული გამოყვანა შეიძლება წარმოდგენილი იქნეს შემდეგი ხის სახით.

8.4.2 7.5.3 ფინანდადების PS-გრამატიკული ანალიზის გაგალითი



ასეთ ხეებს **PS**-გრამატიკაში ფრაზათა სტრუქტურებს უწოდებენ. ფრაზის სტრუქტურის ხეში კატეგორიათა აღმნიშნელი სიმბოლოები კვანძებად იწოდება. კვანძებს შორის ორი სახის ფორმალური კავშირი არსებობს: დომინანტობის (**dominance**) და პრესიდენსობის (**precedence**). მაგალითად, 8.4.1 გრამატიკის $S \rightarrow NP\ VP$ წესის თანახმად **S** კვანძი დომინირებს **NP** და **VP** კვანძებზე. ამავდროულად ეს წესი განსაზღვრავს პრესიდენსობას: ამ ხეზე **NP** კვანძი უფრო მარცხნივ ზის (ე.ი. უფრო წინ ზის) ვიდრე **VP** კვანძი.

C-გრამატიკისგან განსხვავებით, რომელიც ენობრივ კომბინაციებს იაზრებს მისი სიტყვა ფორმების რთული კატეგორიების მეშვეობით და რომელიც კომპოზიციისთვის იყენებს მხოლოდ ორ სქემატურ წესს, **PC**-გრამატიკა იყენებს მხოლოდ ელემენტარულ კატეგორიებს. ამასთან, ამ ელემენტარული კატეგორიების მაკომბინირებელი შესაძლებლობები სხვადასხვა გადამწერი წესებით მოიცემა. უფრო მეტიც, ლექსიკონიც კი, რომელიც **C**-გრამატიკაში კატეგორიზებული სიტყვა ფორმების სიმრავლედ განიხილება (იხ. 7.5.5), **PC**-გრამატიკაში გარკვეული სახის გადამწერი წესების შემადგენელი

ნაწილების სახითაა წარმოდგენილი. ამ წესებს ტერმინალურ წესებს უწოდებენ, რადგანაც მათი მარჯვენა მხარე ტერმინალურ სიმბოლოს, ანუ სიტყვას წარმოადგენს. დანარჩენი წესები, რომელთაც არატერმინალურ წესებს უწოდებენ, წარმოქმნიან **PS**-გრამატიკის წინადადებების ყალიბებს (**frame**), რომლებშიც ტერმინალური წესებით სხვადასხვა სიტყვები და სიტყვათა ფორმები ისმება.

გარდა ამისა, როდესაც საუბარია **C**- და **PS**- გრამატიკებს შორის ფორმალურ განსხვავებებზე, არსებითი ხდება ბუნებრივი ენის მათეული ანალიზის განსხვავებული მიზნები: **C**-გრამატიკული ანალიზის მიზანია ბუნებრივი ენის ფუნქტორ-არგუმენტული სტრუქტურის (**functor-argument structure**) შესწავლა, მაშინ როდესაც **PS**-გრამატიკული ანალიზი იმიზნებს ბუნებრივი ენის შემადგენელი სტრუქტურების (**constituent structure**) ამომწურავ გამოკვლევას.

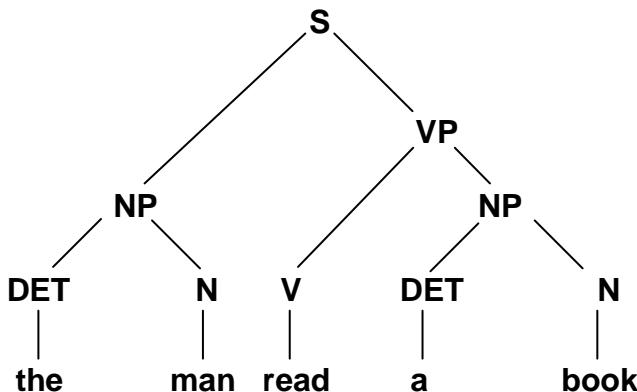
შემადგენელი სტრუქტურები ასახავენ ლინგვისტურ ინტუიციებს, იმის შესახებ, თუ წინადადების რომელ ნაწილებს შორის არის ყველაზე უფრო მჭიდრო შინაარსობრივი კავშირები. შემადგენელი სტრუქტურების ეს ინტუიციური პრინციპი ფორმალურად გადააზრებულია ფრაზის სტრუქტურული ხეების თვისებებად.

8.4.3 შემადგენელი სტრუქტურის განსაზღვრება

1. სიტყვები და/ან შემადგენლები, რომლებიც შინაარსობრივად ერთმანეთს მიეკუთვნებიან, უნდა იყვნენ უშუალოდ და სრულად დომინირებული რომელიმე კვანძის მიერ.
2. შემადგენელი სტრუქტურის ხაზები არ უნდა იკვეთებოდნენ (**არაჩლართულობის პირობა (nontangling condition)**).

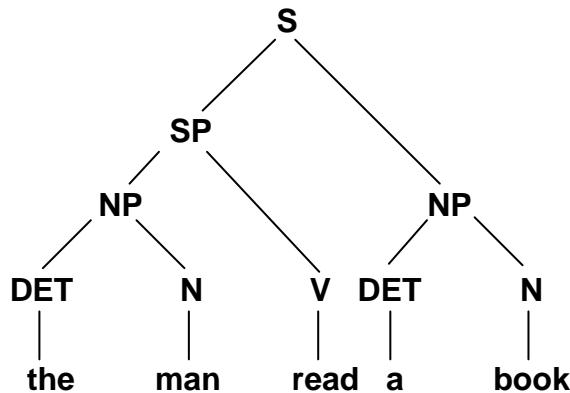
ამ განსაზღვრების თანახმად, **the man read a book** წინადადების შემდეგი ანალიზი ლინგვისტურად კორექტულია.

8.4.4 შემადგენელი სტრუქტურის კორექტული ანალიზი



დასაპირისპირებლად განვიხილოთ იმავე წინადადების ალტერნატიული ანალიზი (იხ. 8.4.5), რომელიც, თუმცა კი ფორმალურად დასაშვებია, არღვევს შემადგენელი სტრუქტურის განმსაზღვრელ პრინციპს.

8.4.5 შემადგენელი სტრუქტურის არაპორექტული ანალიზი



ეს არაკორექტული ანალიზი არღვევს წინადადების შემადგენელ სტრუქტურას: **PS**-გრამატიკოსების აზრით, ინტუიციურად, **read** და **a book** შინაარსობრივად ერთმანეთს მიეკუთვნებიან (ე.ი. მათი აზრით **read a book** უფრო მჭიდრო შინაარსობრივ კავშირშია, ვიდრე **the man read**) და, ამიტომ, ისინი უშუალოდ და სრულად ერთი კვანძით უნდა იყვნენ დომინირებულნი (როგორც ეს არის 8.4.4 კორექტულ ანალიზში).

ისტორიულად, შემადგენელი სტრუქტურის ცნება არც ისე ძველია. ის წარმოიშვა ამერიკელი სტრუქტურალისტის ლ. ბლუმფილდის (**L. BLOOMFIELD (1887-1949)**) უშუალო შემადგენელთა ანალიზისა და მისი სტუდენტის – ზ ჰარისის (**Z. Harris**) დისტრიბუციული ტესტირების საფუძველზე შემუშავებული მიღეომების შეჯერებით. ბლუმფილდის მთავარ ნაშრომში ენა, რომელიც 1933 წელს გამოქვეყნდა, უშუალო შემადგენლების საკითხს არ უკავია ცენტრალური ადგილი: 549 გვერდიდან ეს საკითხი მხოლოდ 4 გვერდზეა მოხსენიებული. იგი მოკლედ მიმოიხილება 161-ე და 167-ე გვერდებზე, და მოგვიანებით ის გამოყენებულია მორფოლოგიაში (გვ. 209/10, 221/2).

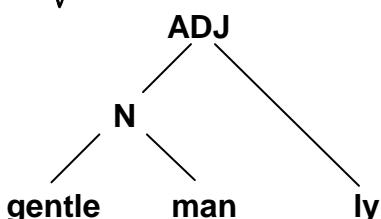
უშუალო შემადგენლების პრინციპს მივყავართ, მაგალითად, **gentlemanly** სიტყვის მსგავს ფორმათა კლასთან, სადაც **gentlemanly** გაიგება არა როგორც შედგენილი სიტყვა, არამედ როგორც მეორადი დერივაციული სიტყვა, სადაც უშუალო შემადგენლები არიან დაბმული ფორმა **-ly** და შინაგანი ფორმა **gentleman**.

L. Bloomfield, ენა, გვ. 210.

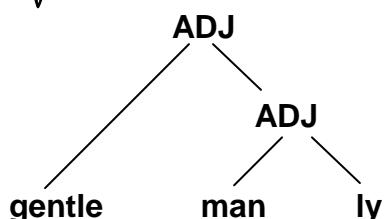
ეს მოსაზრება შემდეგი წისებრი სტრუქტურით გამოიხატება:

8.4.6 უშუალო შემადგენლები **PS**-გრამატიკაში:

სწორი:



არასწორი:



gentlemanly სიტყვის ეს მაგალითი არის აგრეთვე პარისის წიგნში 1951 (გვ. 278-280), სადაც დისტრიბუციული ტესტირების (**distribution tests**) ინოვაციური მეთოდოლოგიაა განხილული. დისტრიბუციული ტესტირება იშლება და რეალიზდება ან როგორც ჩანაცვლების ტესტირება (**substitution tests**), ანდა როგორც გადანაცვლების ტესტირება (**movement tests**). მათი მიზანია ერთმანეთისაგან განასხვაოს გრამატიკულად კორექტული და გრამატიკულად არაკორექტული ჩანაცვლებები და გადანაცვლებები.

8.4.7 ჩანაცვლების ტესტირება

კორექტული ჩანაცვლება:

Suzan has [eaten] an apple



არაკორექტული ჩანაცვლება:

Suzan has [eaten] an apple



Suzan has [cooked] an apple

* Suzan has [desk] an apple

მარცხენა მხარეს არსებული ჩანაცვლება კორექტულად არის მიჩნეული, რადგანაც ის შედეგად გვაძლევს წინადადებას, რომელიც, ისევე როგორც თავდაპირველი წინადადება, გრამატიკულად სწორადა ფორმირებული. მარჯვენა ჩანაცვლება არაკორექტულია, რადგანაც იგი დასამუშავებლად შემოსულ გრამატიკულად სწორად ფორმირებულ წინადადებას არაგრამატიკულად გარდაქმნის.

ანალოგიურ ვითარებასთან გვაქვს საქმე გადანაცვლების ტესტირებისას.

8.4.8 გადანაცვლების ტესტირება

კორექტული გადანაცვლება:

Suzan [has] eaten an apple ⇒ [has] Suzan eaten an apple (?)

არაკორექტული გადანაცვლება:

Suzan has eaten [an] apple ⇒ * [an] Suzan has eaten apple

ამერიკული სტრუქტურალისტებისთვის დისტრიბუციული ტესტირება მეთოდოლოგიურად მნიშვნელოვანი იყო. ამით მათ მიეცათ საშუალება ობიექტურად დაეფუძნებინათ თავიანთი ინტუიცია წინადადების კორექტული სეგმენტაციის (**correct segmentation**) თაობაზე. წინადადების სეგმენტაცია და მისი თანამდევი პიპოთეზები წინადადებების ერთმანეთთან უფრო მეტად, თუ უფრო ნაკლებად დაკავშირებულ ნაწილებზე საჭირო გახდა რათა ერთმანეთისაგან განესხვავებინათ ლინგვისტურად კორექტული და არაკორექტული ფრაზის სტრუქტურის ხეები.

ასეთი განსხვავება აუცილებელია, რადგანაც ნებისმიერი სასრული სტრიქონით განსაზღვრული შესაძლო ფრაზის სტრუქტურების რაოდენობა უსასრულოა²¹. ეს ქმნის ხელისშემშლელი სიჭარბის

²¹ თუ **A-B-C...A** სახის ფრაზის სტრუქტურებს გამოვრიცხავთ, განსხვავებული ფრაზის სტრუქტურული ხეების რაოდენობა შეყვანილი მიმდევრობის სიგრძის ზრდასთან ერთად მანც ექსპონენციალურად გაზრდება. ფორმალური თვალსაზრისით, კონტექსტისგან თავისუფალ **PS**-გრამატიკებში ასეთი სახის სტრუქტურები ლეგიტიმურად მიიჩნევა.

ეფექტს: ყველა ამ შესაძლო ფრაზის სტრუქტურებიდან შეუძლებელია ყველა თანაბრად ლინგვისტურად კორექტული იყოს.

შესაძლო ხეების საკმაოდ დიდი რაოდენობა ეხება ფორმალური გრამატიკის გარეთ დარჩენილ იზოლირებულ წინადადებებს. რადგან ენის სტრუქტურული აგებულების პრინციპები ცნობილია და იგი **PS**-გრამატიკის სტრუქტურალურ პრინციპებს ემთხვევა, წინადადებების შესატყვისი ფრაზული სტრუქტურები ამ გრამატიკის მეშვეობითვე იგება. ამასთან, როგორც წესი, იზოლირებული წინადადებების შესატყვისი ყველა შესაძლო სტრუქტურული ხეების რაოდენობა ბევრად უფრო მეტია ვიღრე იმ ხეების რაოდენობა, რომელიც მათ ამ ენის გრამატიკით შეესაბამება. არადა, მოცემული არაწინააღმდეგობრივი **PS**-გრამატიკისთვის შეიძლება არსებობდეს მხოლოდ ერთი სწორად განსაზღვრული სტრუქტურა ნებისმიერი სწორად ფორმირებული წინადადებისთვის.

კონტექსტისგან თავისუფალი ხელოვნური ენების სტრუქტურული პრინციპების საკმარისი სიმარტივის გამო, როგორც წესი, საკმარისად მარტივად ისაზღვრება მათი ადეკვატური ფორმალური **PS**-გრამატიკები (მაგ. 8.3.1, 8.3.2, 8.3.3, 8.3.4, 8.3.5). როდესაც ერთი და იგივე კონტექსტისგან თავისუფალი ენისათვის რამოდენიმე არსებითად განსხვავებული **PS**-გრამატიკა იგება, მაშინ არ არსებობს არანაირი რაციონალური საფუძველი იმის გარკვევისა, თუ რომელი მათგანი იძლევა ფრაზის სტრუქტურის ‘სწორ’, ანუ ‘მართებულ’ ხეებს.

მეორეს მხრივ, ბუნებრივი ენების ადეკვატური **PS**-გრამატიკის ამგები სტრუქტურალური პრინციპები ჯერ კიდევ უცნობია. ამგვარად, ღიაა შეკითხვა თუ რაიმე წინადადებისთვის, ან წინადადებათა რაიმე მცირე სიმრავლესთვის, რომელი **PS**-გრამატიკა აღმოჩნდება უფრო შესაფერი მისი შემდგომი ისეთი გაფართოებისთვის, რომ თანდათანობით დაიფაროს მთლიანი ენა. – იმისთვის, რომ ბუნებრივი ენის სრულფასოვნად შესასწავლად სასაფუძვლოდ ავირჩიოთ **PS**-გრამატიკა მისი შემდგომი ხანგრძლივი განვითარებისა და გაფართოების წინასწარგანსაზღვრული მიზნებით, ემპირიული კრიტერიუმებია საჭირო.

შემადგენელი სტრუქტურის ინტუიციური პრინციპი, ჩანაცვლებისა და გადანაცვლების ტესტირებები ამ მიზნით იქნა შემუშავებული. მაგრამ ეს არ არის საკმარისი. ამიტომაც არის მუდმივი დისკუსია ბუნებრივი ენის მატარებელთა შორის არის თუ არა ფრაზის სტრუქტურა ლინგვისტურად სწორი, და თუ არის რატომ.

ფორმალურ ენათა თეორიის თვალსაზრისით, ამ საბოლოო ჯამში უშედეგო დებატების მიზეზი ბუნებრივი ენების სრული **PS**-გრამატიკების არ არსებობაა – როგორც ელემენტარულის, ისე წარმოებული ფორმალიზმების. ორმოცდათი წლის მნიშვნელოვნად ფინანსირებადმა კვლევებმა ვერ მოგვცა ვერცერთი ბუნებრივი ენის სრული ანალიზი, რაც შეიძლება მივიღოთ მინიშნებად, რომ ეს მიდგომა არასაკმარისად ოპტიმალურია.

8.5. შემადგენელი სტრუქტურის აარადოესი

ენათა **SLIM** თეორიის თვალსაზრისით, არსებობს შემადგენელი სტრუქტურის ინტუიციური პრინციპის საპირისპირო არგუმენტები (იხ. 8.2.1). პირველი: შემადგენელი სტრუქტურა და დისტრიბუციული ტესტირება ეწინააღმდეგება ბუნებრივი ენის დროში **წრფივ** (*time linear*) სტრუქტურას. მეორე: ფრაზების სტრუქტურულ ხეებს არ გააჩნიათ საკომუნიკაციო მიზნები. მესამე: შემადგენელი სტრუქტურის პრინციპები არაყოველთვის მართლდება.

ეს იმიტომ, რომ შემადგენელი სტრუქტურის 8.2.1 პუნქტში აღწერილი პირობები მოითხოვს, რომ ის ნაწილები, რომლებიც შინაარსობრივად ერთმანეთს მიეკუთვნებიან, ბუნებრივი ენის გარეზედაპირულ გამოსახულებაში ერთმანეთის გვერდი-გვერდ იყვნენ განლაგებულნი. თუმცა, ბუნებრივ ენაში არის გამოსახულებები ე. წ. დისკონტინიალური ელემენტები (**discontinuous elements**), რომელთათვისაც ეს პირობა არ სრულდება.

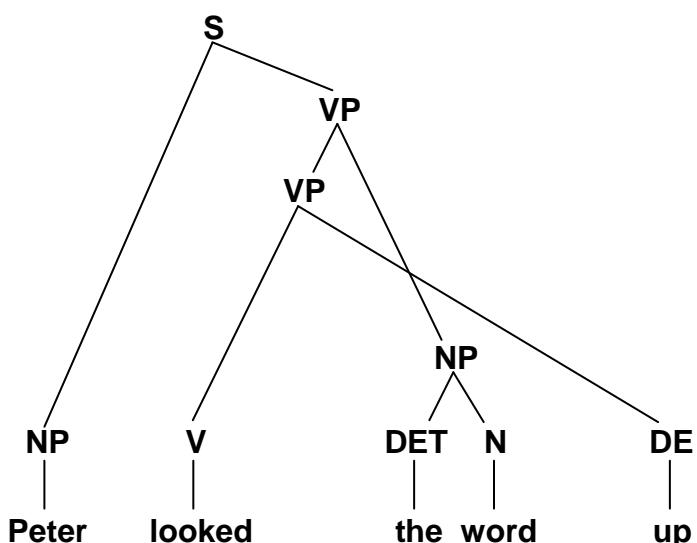
მაგალითად, საერთო შეთანხმებით წინადადებაში

Peter looked the word up.

დისკონტინიალური ელემენტები **looked** და **up** შინაარსობრივად უფრო ახლოს არიან, ვიდრე უშუალო მეზობლებით შედგენილი გამოსახულებები **looked – the word** და **the word – up**.

ამგვარი დისკონტინიალური ელემენტების არსებობა არის სტრუქტურული მიზეზი იმისა, თუ რატომ არ შეიძლება კონტექსტისგან თავისუფალ **PS**-გრამატიკაში შემადგენელი სტრუქტურის პრინციპის სრული და წარმატებული განხორციელება (ეს ნაჩვენებია 8.5.1 და 8.5.2 ალტერნატიული სტრუქტურული ხეებით). ამ სტრუქტურულ პრობლემას შემდგენელი სტრუქტურის პარადოქსი (**constituent structure paradox**) ეწოდა²².

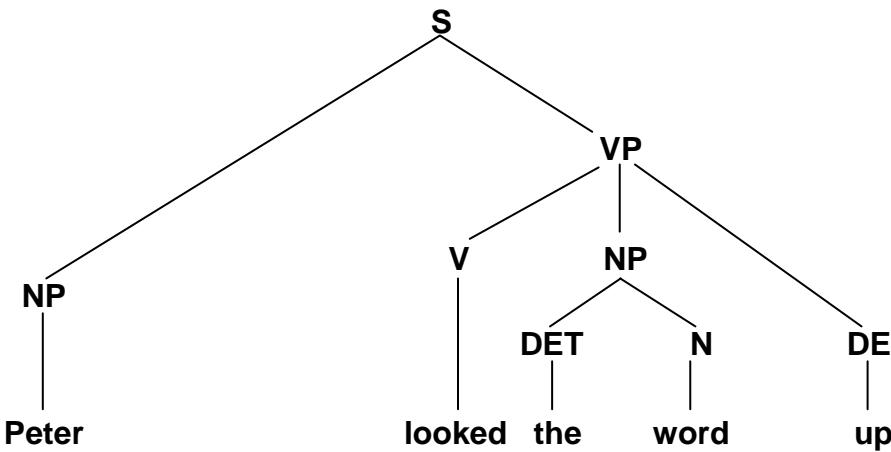
8.5.1 8.4.3 განაცვრის მეორე პირობის უარმყოფელი მაგალითი



აქ შინაარსობრივად ერთმანეთთან მიკავშირებული ქვეგამოსახულებები **looked** და **up** უშუალოდ და სრულად ერთი კვანძის დომინირების ქვეშ არიან. ასე რომ, ამ ანალიზისას 8.2.1 პუნქტის პირველი პირობა კმაყოფილდება. თუმცა, ანალიზი არღვევს 8.2.1 განაწერის მეორე პირობას, რადგანაც აქ ხის ტოტები ერთმანეთს კვეთენ.

²² CoL, გვ. 24.

8.5.2 8.4.3 განაწერის პირველი პირობის უარყოფელი მაგალითი



ამ ხის ტოტები არ იკვეთებიან და ამით იგი აქმაყოფილებს 8.2.1 განაწერის მეორე პირობას. თუმცა, ამ ანალიზისას ირღვევა პუნქტის პირველი პირობა, რაღაც სემანტიკურად მიკავშირებული **looked** და **up** გამოსახულებები, ანუ **V** და **DET** კვანძები სრულად და უშუალოდ არ დომინირდებიან ხის სხვა კვანძით. მართლაც, კვანძი, რომელიც უშუალოდ დომინირებს **V** და **DET** კვანძებზე ასევე დომინირებს **NP** კვანძზე, ანუ **the word** ფრაზაზე.

ფორმალური ენების თეორიის თვალსაზრისით, შემადგენელი სტრუქტურების პარადოქსი განპირობებულია იმ ფაქტით, რომ კონტექსტისგან თავისუფალი **PS**-გრამატიკების წარმოქმნის უნარი არ არის საკმარისი დისკონტინიალური ელემენტების დასამუშავებლად. ეს ემპირიული პრობლემა გასული საუკუნის 50-იანი წლებიდანაა ცნობილი²³.

ყველა ბუნებრივ ენას გააჩნია ამა თუ იმ სახის დისკონტინიალური ელემენტი. ამგვარად, 1957 წელს, ნ. ჩომსკიმ, იმისათვის, რომ რამდენადაც ეს შესაძლებელი იყო დაეცვა შემადგენელი სტრუქტურის პრინციპი, ზ. პარისის მეთოდოლოგით მოტივირებული ჩანაცვლებისა და გადანაცვლების ტესტები გადაამუშავა წარმოქმნელ წესებად. ამ წესებს ტრანსფორმაციები ეწოდა და ისინი აუცილებელი, ანუ სისტემაში ჩადგმულ წესებად გამოცხადდა (ნ. ჩომსკი 1965, გვ. 47).

ტრანსფორმაციული წესი მოქმედებს შესასვლელზე დასამუშავებლად შემოსული ფრაზის სტრუქტურულ ხეზე და გამოსავალზე იძლევა მოდიფიცირებული ფრაზის სტრუქტურულ ხეს. ტრანსფორმაციულ გრამატიკებში რიგი ტრანსფორმაციები გარკვეული რიგით ერთიანდებიან ერთ ტრანსფორმაციულ კომპონენტში და ერთი მეორეს მიყოლებით გამოიყენებიან დასამუშავებელი ფრაზის სტრუქტურაზე. ტრანსფორმაციულ წესებში შეტანისა (**input-**) და გამოტანის (**output-**) პირობები ფორმალურად მოიცემა ცვლადების შემცველი გამოსახულებების სახით, რომლებსაც ინდექსირებულ ფრჩხილებიან ფორმებს (**indexed bracketings**) უწოდებენ.

²³ ი. ბარ-პილელი 1960 [1964, გვ.102] წელს წერდა, რომ მან მიატოვა **C**-გრამატიკის 1953 წლის ნაშუშვრები დისკონტინიალურ შემადგენლების ანალოგიური სინქლის გამო **He gave it up** წინადადებაში.

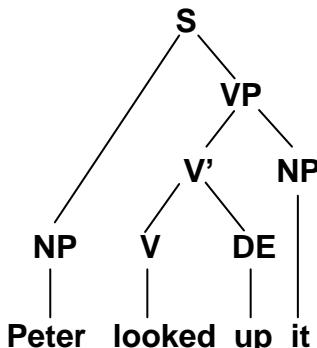
8.5.3 ფრანსფორმაციის მაგალითი

$$[[V \ DE]_V \ [DET \ N]_{NP}]_{VP} \Rightarrow [V \ [DET \ N]_{NP} \ DE]_{VP}$$

ამ ტრანსფორმაციის გამოყენება ნაჩვენებია 8.5.4 პუნქტში.

8.5.4 8.5.3 ტრანსფორმაციის გამოყენება

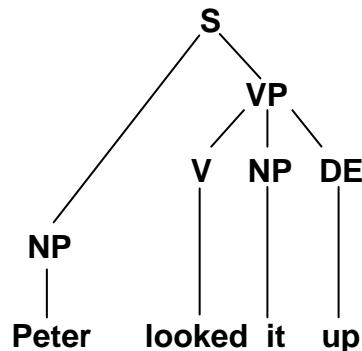
სიღრმისეული სტრუქტურა:
(deep structure)



ტრანსფორმაცია

⇒

ზედაპირული სტრუქტურა:
(surface structure):



სტანდარტულ თეორიის (**ST**, ჩოშიკი, 1965) ტრანსფორმაციულ კომპონენტში შესატანი ინფორმაცია წარმოიქმნება კონტექსტისგან თავისუფალი **PS**-გრამატიკით. ეს სიღრმისეული სტრუქტურები (**deep structures**) უნდა აკმაყოფილებდნენ შემადგენელი სტრუქტურის 8.4.3 პუნქტში აღწერილ პირობებს, თუმცა აუცილებელი არაა ისინი შესაბამებოდნენ გრამატიკულ მიმღევრობებს (როგორც ეს არის 8.5.4 პუნქტში აღწერილი მარცხნა ფრაზის სტრუქტურულ ხეში).

თუ ტრანსფორმაციულ წესში შემავალი გამოსახულება (ინდექსირებული ფრაზისილებანი ფორმა) წარმატებულად დაწყვილდა რომელიმე ფრაზულ სტრუქტურასთან, ის ამ ტრანსფორმაციის წესის გამოსავალი გამოსახულების შესაბამისად ახალ ფრაზულ სტრუქტურად ტრანსფორმირდება. ერთი ან რამდენიმე ტრანსფორმაციის წესის გამოყენების შედეგად მიიღება ზედაპირული სტრუქტურა. ფრაზის სტრუქტურის ეს ზედაპირული ფორმა უნდა იყოს შესაბამისობაში გრამატიკულ მიმღევრობასთან, თუმცა აუცილებელი არაა იგი აკმაყოფილებდეს შემადგენელი სტრუქტურის 8.4.3 პუნქტში მოცემულ პირობებს (როგორც ეს არის 8.5.4 პუნქტში აღწერილი მარჯვენა ფრაზის სტრუქტურულ ხეში).

8.5.4 პუნქტში აღწერილი ტრანსფორმაცია განიხილება როგორც მნიშვნელობის შემნახველი (**meaning preserving**) ზუსტად ისევე როგორც 4.5.2 პუნქტში განხილული მაგალითები. ივარაუდება, რომ აღწერება მოსაუბრე-შემენელის თანდაყოლილი ცოდნა კომუნიკაციური ფუნქციონირების გარეშე.

მათემატიკური თვალსაზრისით, მექანიზმი, რომლის მიზანია რეკურსულად გარდაქმნას მასში შემავალი სტრუქტურების ღია სიმრავლე ყოველთვის ხასიათდება მაღალი ხარისხის სირთულით. ეს ნაჩვენები იყო 8.3.7 პუნქტში კონტექსტზე დამოკიდებული **a^kb^kc^k** ხელოვნური ენისათვის აღწერილი **PS**-გრამატიკით. მანამ სანამ, კონტექსტზე დამოკიდებული ენები არის მხოლოდ

ექსპონენციალური, ტრანსფორმაციული გრამატიკა ტიურინგის მანქანის ეკვივალუნტურია, წარმოქმნის რეკურსულად გადათვლად ენებს და, შესაბამისად, არაამოხსნადია.

6. ჩომსკი იმედოვნებდა, რომ გვერდს აუკლიდა ამ არასასურველ ვითარებას აღდგენის უნარიან წაშლებად (**recoverability of deletion**) წოდებული ფორმალური შეზღუდვების დამატებით. ამ შეზღუდვების თანახმად, ტრანსფორმაციას უფლება აქვს წაშალოს კვანძი მხოლოდ იმ შემთხვევაში, თუ შესაძლებელია მისი აღდგენა ისეთი მკაცრად განსაზღვრული გზით (**well-defined access**), რომელიც ამ კვანძთან ერთად მის მიერ დომინირებული ქვე-ზის აღდგენის საშუალებასაც იძლევა.

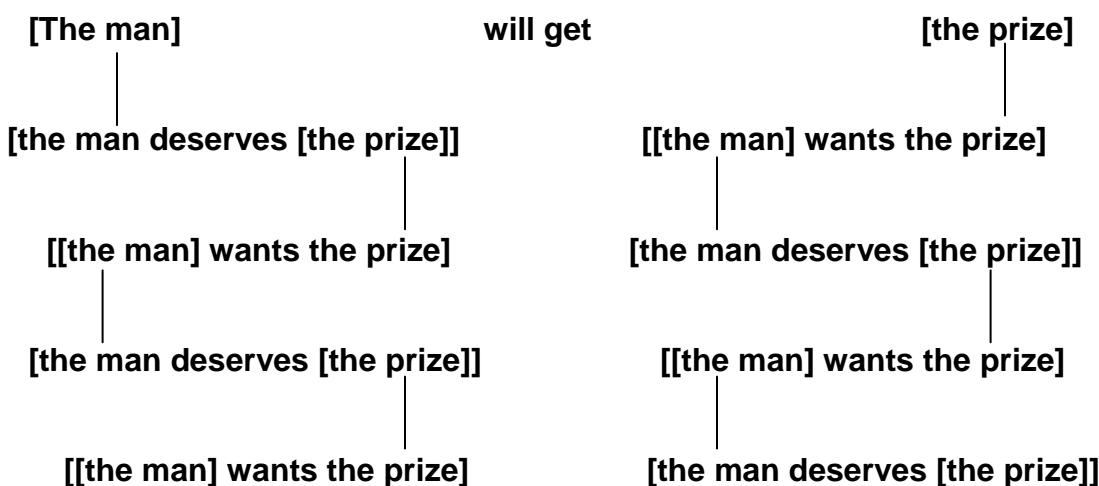
ბახ-პეტერსის წინადადებები გვარწმუნებენ, რომ აღდგენის უნარიანი წაშლები ყოველთვის არ იძლევა სასურველ ეფექტს.

8.5.5 ბახ-პეტერსის ფინანსურის მრთი მაგალითი

The man who deserves it will get the prize he wants.

წინადადება შეიცავს ორ სახელურ ფრაზას მათთან მიკავშირებული კლაუზებითურთ (**noun phrase with relative clauses**). თითოეული ეს კლაუზა შეიცავს თითო ნაცვალსახელს, რომლებიც მათ შესაბამის სახელურ ფრაზასთან ბუნებრივ ანტეცედენტურ – პოსტცედენტურ (**postcedent - antecedent**) მიმართებაში არიან (იხ. ქვეთავი 6.3). დაშვება იმისა, რომ ნაცვალსახელი ტრანსფორმაციულად გამოიყვანება მასთან ახლოს მდებარე იმ სახელური ფრაზებისაგან, რომელიც, თავის მხრივ, კორეფერენტულია მისსავე ან ანტეცედენტურ, ან პოსტცედენტურ ნაცვალსახელთან, 8.5.5 მაგალითთან მიმართებაში გვაძლევს შემდეგი სახის სიღრმისეულ სტრუქტურას.

8.5.6 ბახ-პეტერსის ფინანსურის სიღრმისეული სტრუქტურა



ტრანსფორმაციული ალგორითმი გამიზნულია მოიძიოს 8.5.5 პუნქტით მოცემული ზედაპირული წინადადების ის სიღრმისეული სტრუქტურა, რომლისგანაც ეს ზედაპირული ფორმა გამოიყვანება კორექტული ტრანსფორმაციული საშუალებებით. მაგალითის სტრუქტურიდან გამომდინარე ალგორითმი შემოგვთავაზებს სულ უფრო და უფრო სიღრმისეულ სტრუქტურებს აღდგენის უნარიანი წაშლების ბიჯის მოლოდინში.

ასე მაგალითად, დავუშვათ ალგორითმმა სრული სახელური ფრაზა **[the man who deserves it]** სიღრმისეულ სტრუქტურად და **he** ნაცვალსახელის ანტეცედენტად გააპოსტულატა. ეს სიღრმისეული სტრუქტურა თავად შეიცავს **it** ნაცვალსახელს, რომლისთვისაც ალგორითმი სრულ სახელურ ფრაზას **[price which he deserves]** აპოსტულატებს პოსტცედენტად. ახლა უკვე ეს სიღრმისეული სტრუქტურა შეიცავს **he** ნაცვალსახელს, დ.ა.შ.

ამგვარად, ეს პროცედურა ორივე მიკავშირებული კლაუზასთვის უსასრულოდ გრძელდება ყოველგვარი გაჩერებების გარეშე (გაჩერების პრობლემა (**halting problem**)). ფორმალურად ის, რომ ტრანსფორმაციული გრამატიკა არაამოხსნადია და წარმოქმნის რეკურსულად გადათვლადი ენების კლასს, დამტკიცდა 1969 წელს პეტერსის და რიტჩის²⁴ მიერ (გამოქვეყნდა 1973 წელს).

ნეიტივიზმის (**nativism**) გვიანი ვარიანტები, როგორებიცაა **LFG**, **GPSG** და **HPSG**, ტრანსფორმაციული მიღვომების გარეშე ცდილობენ გადაწყვიტონ შემადგენელი სტრუქტურის პარადოქსი. **PS**-გრამატიკის ამ წარმოებულ ფორმალიზმებში პირობები შემადგენელი სტრუქტურებისგან ითხოვენ მხოლოდ იმას, რაც დაშვებულია ენობრივად შესატან მოცემულობაზე. სხვაგვარად რომ ვთქვათ (როგორც დისკონტინიალური შემადგენლების შემთხვევაში), აქაც ინტუიცია იმისა, თუ რომელი ნაწილებია შინაარსობრივად უფრო მჭიდროდ მიკავშირებულნი, გამოისახება არა ფრაზის სტრუქტურის ხეებით, არამედ მათი ალტერნატიული საშუალებებით, კერძოდ კი – მახასიათებელი სტრუქტურებით (**feature structure**).

ამგვარად, შემადგენელ სტრუქტურათა მეთოდმა ვერ შეძლო გაემართლებინა მისი ნავარაუდევი სტატუსი: იგი ვერ დასაბუთდა ვერც როგორც ადამიანთათვის თანდაყოლილი ენობრიობის უნივერსალურად განმსაზღვრელი და მაწარმოებელი სისტემა და, შესაბამისად, ვერც როგორც მრავალმნიშვნელოვანი და უნაკლო მეთოდოლოგი. ვინმემ შეიძლება სრულიად სამართლიანად იკითხოს კიდეც, თუ რატომ დგანან ისეთი სისტემები, როგორებიცაა **GPSG**, **LFG** და **HPSG**, შემადგენელი სტრუქტურების გრამატიკებთან ერთად. ამის გარდა, ეს გვიანი სისტემები სირთულის თვალსაზრისით ტრანსფორმაციულ გრამატიკებზე ნაკლებ სრულყოფილებია: ისინი წარმოქმნიან რეკურსულად გადათვლად ენათა კლასს და არიან არაამოხსნადები²⁵.

²⁴ ამ პერიოდში ბობ რიტჩის ჰქონდა პირადი კავშირები და აქტიური კონსულტაციები ნ. ჩოშსკისთან.

²⁵ **იხ. B. Barton, R.C. Berwick da E.S. Ristad** 1987, თუ კონტექსტისგან თავისუფალი წესის ამდაგვარი $A \rightarrow B \rightarrow \dots \rightarrow A$ გაკვნება იკრძალება, მაშინ **LFGs** სირთულე ექსპონენციალურია, რადგანაც ასეთი კვნები ფორმალურად ლეგალურია კონტექსტისგან თავისუფალ **PS**-გრამატიკებში. სირთულის თეორიის თვალსაზრისით ეს შეზღუდვა არ არის დაკანონებული. უფრო მეტიც, მაჩვენებლიანი სირთულე გაცილებით მეტია გამოთვლითი გამოყენებებისათვის.

სავარჯიშოები

8.1 ქვეთავი

- ჩამოაყალიბეთ **PS**-გრამატიკის ალგებრული განსაზღვრება.
- რა განსხვავებაა **PS**-გრამატიკის ტერმინალურ და არატერმინალურ სიმბოლოებს შორის?
- პირველად ვინ და როდის აღწერა **PS**-გრამატიკა, რა სახელით, და რა მიზნებით?
- პირველად როდის და ვის მიერ იქნა გამოყენებული **PS**-გრამატიკა ბუნებრივი ენის აღსაწერად?
- აღწერეთ **PS**-გრამატიკის წესების სტანდარტული შეზღუდვები.
- ახსენით ტერმინი წარმოქმნის უნარი.

8.2 ქვეთავი

- ახსენით **PS**-გრამატიკის ტიპებს, ფორმალურ ენათა კლასებს და განსხვავებული დონის სირთულეებს შორის არსებული ურთიერთკავშირები.
- დაასახელეთ სირთულის მთავარი კლასები. რატომ არ არიან ისინი დამოკიდებულნი წარმოქმნელი გრამატიკის განმსაზღვრელ ფორმალიზმზე?
- რა არის ენათა კლასების სირთულე **PS**-იერარქიაში?
- რას უდრის საშუალო წინადადების სიგრძე ლიმას კორპუსში?
- რისი ტოლია მაქსიმალური წინადადების სიგრძე ლიმას კორპუსში?
- ყველაზე ცუდ შემთხვევაში რა დრო სჭირდება ექსპონენციალურ ალგორითმს ლიმას კორპუსის ანალიზისათვის?
- ახსენით ფორმალური ენების **PS**-გრამატიკული იერარქია.
- გამოთვლითი ლინგვისტიკის თვალსაზრისით **PS**-გრამატიკულ იერარქიაში რომელი ენათა კლასები ხასიათდება პრაქტიკული ღირებულების მქონე სირთულით?

8.3 ქვეთავი

- განსაზღვრეთ **PS**-გრამატიკა, რომელიც $\{a, b, c\}$ სიმრავლეზე განსაზღვრულ თავისუფალ მონოიდს წარმოქმნის. დაახასიათეთ. $\{a, b, c\}^+$ ენა **PS**-გრამატიკულ იერარქიაში. შეადარეთ $\{a, b, c\}^+$ და $a^k b^k c^k$ ენების გრამატიკების წარმოქმნის უნარები. რომელი მათგანია მაღალი და რატომ?
- რა მიზეზებით იხსნება **PS**-გრამატიკაში არსებობა ტერმინისა კონტექსტისგან თავისუფალი?
- რა სახის სტრუქტურები წარმოიქმნება კონტექსტისგან თავისუფალი **PS**-გრამატიკებისგან?
- დაასახელეთ ორი ხელოვნური ენა, რომლებიც არ არიან კონტექსტისგან თავისუფალი. რატომ აჭარბებენ ეს ენები კონტექსტისგან თავისუფალი გრამატიკების წარმოქმნით შესაძლებლობებს?
- განსაზღვრეთ **PS**-გრამატიკა $a^{k_1} b^{k_2} c^{k_3}$ ენისათვის. ახსენით თუ რატომ ემორჩილება ეს ენა წყვილების ინვერსირების კონტექსტისგან თავისუფალ სქემას.
- განსაზღვრეთ **PS**-გრამატიკა $ca^m dyb^n$ ენისათვის. მოიყვანეთ ამ ხელოვნური ენის დასრულებული გამოსახულებების მაგალითები. ახსენით თუ რატომ არის ეს ენა წესიერი ენა.
- რატომ არღვევს 8.3.5 კონტექსტზე-დამოკიდებული **PS**-გრამატიკის წესის განსაზღვრებას, მაშინ როცა β ნოლია?
- რა არის მტუმბავი ლემა?
- რატომ არ არსებობს კონტექსტზე დამოკიდებული ენებისათვის მტუმბავი ლემა?
- არიან თუ არა რეკურსულად გადათვლადი ენები რეკურსული?

11. დაასახელე რეკურსული ენა, რომელიც არ არის კონტექსტზე დამოკიდებული.

8.4 ქვეთავი

- ჩამოაყალიბეთ შემადგენელი სტრუქტურის განსაზღვრება.
- ახსენით კონტექსტისგან თავისუფალი **PS**-გრამატიკებსა და ფრაზის სტრუქტურულ ხეებს შორის კავშირი.
- აღწერეთ შემადგენელი სტრუქტურის ცნების ისტორიული განვითარების გზა.
- დაასახელეთ დისტრიბუციული ტესტირების ორი ტიპი და ახსენით მათი როლი მართებული, ანუ კორექტული შემადგენელი სტრუქტურის მოძიებაში.
- რატომ იყო ამერიკელი სტრუქტურალისტებისათვის მნიშვნელოვანი წინადადების მართებული სეგმენტაცია?

8.5 ქვეთავი

- აღწერეთ დისკონტინიალური ელემენტის ცნება ბუნებრივ ენაში და ახსენით, თუ რატომ იწვევს დისკონტინიალური ელემენტი შემადგენელი სტრუქტურის პარადოქსს.
- რა ხერხებით ცდილობს ტრანსფორმაციული გრამატიკა დისკონტინიალური ელემენტების მიერ წარმოქმნილი პრობლემის გადაჭრას?
- შეადარეთ ტრანსფორმაციული გრამატიკის და გამოთვლითი ლინგვისტიკის მიზნები.
- რა არის ტრანსფორმაციული გრამატიკის წარმოქმნის უნარი?
- ახსენით ბახ-პეტერსის წინადადების სტრუქტურა აღდგენისუნარიანი წაშლის პირობასთან მიმართებაში. ტრანსფორმაციული გრამატიკის რომელი მათემატიკური თვისება იქნა ნაჩვენები ამ ტიპის წინადადებით?

9 თავი ანალიზის ძირითადი ცნებები

ამ თავში გამოკვლეულია, თუ რომელი ფორმალური თვისებები ხდის წარმომქმნელ გრამატიკას ავტომატური ენობრივი ანალიზისთვის ვარგისს, და რომელი არა. ამასთან, ძირითად მაგალითებად გამოყენებულია კონტექსტისგან თავისუფალი **PS**-გრამატიკა და მისი პარსერები.

9.1. ქვეთავში აღწერილია პარსერის შემადგენელი სტრუქტურები და ახსნილია, თუ რა გავლენებს ახდენს დეკლარაციულ-პროცედურული განსხვავებები წარმომქმნელი გრამატიკებისა და პარსერის ურთიერთკავშირებზე. 9.2 ქვეთავში განხილულია კონტექსტისგან თავისუფალი **PS**-გრამატიკებისა და სტანდარტული **C**-გრამატიკების ურთიერთკავშირები. შეჯამებულია ენის სხვადასხვა განსაზღვრებების, წარმომქმნელი გრამატიკების, გრამატიკების ქვეტიპების, ენის ქვეკლასების, პარსერებისა და სირთულის ხარისხების ურთიერთმიმართებები. 9.3 ქვეთავში ახსნილია გრამატიკული ტიპის ტრანსფარენტულობის პრინციპი. განხილულია **a^kb^k** ენის ერლეის ალგორითმი, რისი მეშვეობითაც ნაჩვენებია, რომ კონტექსტისგან თავისუფალი **PS**-გრამატიკა არ არის ტრანსპარენტული. 9.4 ქვეთავში ნაჩვენებია, რომ შესაძლო ჩანაცვლებების პრინციპი, რომელსაც არსებითად ემყარება გამოყვანები **PS**-გრამატიკებში, ხელს უშლის **PS**-გრამატიკის შემავალ-გამომავალ ეკვივალენტობას მის პარსერსა და მსმენელ-მოსაუბრე სისტემებთან. 9.5 თავში წარმოდგენილია ის მათემატიკური, გამოთვლითი და ფინანსურული თვისებები, რომლებიც არ შეიძლება არ ჰქონდეს ბუნებრივი ენის ემპირიკულად აღეკვატურ წარმომქმნელ გრამატიკას.

9.1 ანალიზის დეკლარაციული და პროცედურული ასპექტები

ხელოვნური ენების ანალიზატორები, ანუ პარსერები¹ (**parsers for artificial languages**) კომპიუტერულ მეცნიერებებში გამოიყენება ერთი პროგრამული დონის მეორე პროგრამულ დონეში ტრანსფორმირებისთვის, მაგალითად, როგორც ეს ხდება კომპილატორებში (**compilation**). ბუნებრივი ენების ანალიზატორები, ანუ პარსერები (**parsers for natural languages**) გამოიყენება

¹ როგორც განმარტებულია 1.3 ქვეთავში, პარსერი არის კომპიუტერული პროგრამა, რომელიც ამჟამავებს ენობრივ გამოსახულებებს, როგორც შემაგალ (**input**) მოცემულობებს, და გამოსავალში (**output**) იძლევა ასალ, გადამუშავებულ გამოსახულებებს, რაც დასამუშავებლად შესული გამოსახულების გარკვეული სახის სტუქტურული ანალიზის შედეგა.

როგორც სიტყვების ავტომატური გამოცნობისთვის, ასევე ავტომატური სინტაქსური და სემანტიკური ანალიზისთვის. შესაბამისად, ერთმანეთისაგან უნდა განირჩენ მორფოლოგიური პარსერები (**morphological parsers**), სინტაქსური პარსერები (**syntax parsers**), და სემანტიკური პარსერები (**semantic parsers**).

მორფოლოგიური პარსერი (იხ. 13-15 თავები) მოქმედებს სიტყვა-ფორმებზე (**word form**), როგორც შემავალ მოცემულობებზე, და აანალიზებს მათ. ეს გულისხმობს: (i) მათი გარეენობრივი გამოსახულების (**word surface**) ალომორფულ სეგმენტირებას (**segmentation**), (ii) მათი სინტაქსური კომბინატორივის დახასიათებას (**categorization**) და (iii) მათი ძირეული ფორმის გამოყვანას (**lemmatization**). სინტაქსური პარსერი (იხ. 16-18 თავები) მოქმედებს წინადადებებზე, როგორც შემავალ მოცემულობებზე, და გამოსავალში იძლევა მათ უკვე გაანალიზებულ გრამატიკულ სტრუქტურას მაგ.: შემადგენელ სტრუქტურას **PS**-გრამატიკაში, ან დროში წრფივ გამოყვანას **LA**-გრამატიკაში. სემანტიკური პარსერი (იხ. 22-24 თავები) აფართოებს სინტაქსურ ანალიზს ამ ანალიზთან მიკავშირებული სემანტიკური წარმოდგენის გამოყვანით.

ბუნებრივი ენით კომპიუტერთან ურთიერთობის მექანიზმის მოდელირებისთვის საჭირო ხდება ამ სხვადასხვა ტიპის პარსერების შეჯერება ერთ მთლიან სისტემაში. — სინტაქსური პარსერი ითხოვს სიტყვების ავტომატურ ამომცნობს და, შესაბამისად, საჭიროებს გარკვეული ტიპის მორფოლოგიურ პარსერს. სემანტიკური პარსერი კი ითხოვს სინტაქსურ ანალიზს და, შესაბამისად, საჭიროებს სინტაქსურ პარსერს.

სხვადასხვა ტიპის პარსერები განსხვავდებიან ერთმანეთისაგან (i) გასაანალიზებელი გამოსახულებების სტრუქტურული აღწერით (**structural description**) და (ii) ავტომატური ანალიზის მარეალიზებელი გამოთვლითი ალგორითმით (**computational algorithm**). თანამედროვე პარსერებში ეს ორი ასპექტი სისტემურად არის განყოფილი: სტრუქტურული აღწერები კეთდება წარმომქმნელი გრამატიკის მეთოდებით. ამასთან, ეს სტრუქტურები ინტერპრეტირდება და მუშავდება სპეციალურად მათოვის შემუშავებული გამოთვლითი ალგორითმით.

სხვანაირად რომ ვთქვათ, თანამედროვე პარსერს შეუძლია გამომთვლელში (**computer**) ჩატვირთოს სხვადასხვა ფორმალიზმის მქონე **G_i**, **G_j**, **G_k** წარმომქმნელი გრამატიკებიდან ნებისმიერი (მაგ. 7.1.3, 8.3.1, 8.3.5 პუნქტებში აღწერილი კონტექსტისგან თავისუფალი **PS**-გრამატიკები, ან 10.2.2, 10.2.3, 11.5.2, 11.5.3, 11.5.5, 11.5.7, 11.5.8 პუნქტებში აღწერილი **C-LAG** გრამატიკები) და გაანალიზოს **L_j** ენა მასთან მიკავშირებულ **G_j** გრამატიკულ ინტერპრეტაციაში. **L_j** ენის გამოსახულებების ავტომატური ანალიზისას ცხადად განირჩევა ერთმანეთისაგან დატვირთვების

გადანაწილება **(i) G_j** გრამატიკაზე და **(ii) პარსერზე G_i, G_j, G_k** ფორმალური გრამატიკების მთლიანი კლასისათვის.²

გრამატიკისა და მანალიზებელი ალგორითმის განცალკავება შესაბამისობაშია კომპიუტერულ მეცნიერებებში საზოგადოდ გამოყენებადი მიზნობრივი დისტინქციის პრინციპთან, რომელიც გამოთვლითი პრობლემის გადაწყვეტისას გულისხმობს დეკლარაციული აღწერებისა (**declarative specification**) და პროცედურული რეალიზაციების (**procedural implementation**) სისტემურ დაყოფას.

9.1.1 დეკლარაციული & პროცედურული ასპექტები ენათმეცნიერებაში

- ენის გამოთვლითი ანალიზის, ანუ პარსერის დეკლარაციული ასპექტი (**declarative aspect**) მოიცემა იმ ენისათვის დაწერილი წარმომქმნელი გრამატიკით, რომელიც უნდა განალიზდეს ზოგად, მათემატიკურად მკაცრად განსაზღვრულ ფორმალიზმში.
- ენის გამოთვლითი ანალიზის, ანუ პარსერის პროცედურული ასპექტი (**procedural aspect**) მოიცემა რეალიზებული კომპიუტერული პროგრამის იმ შემადგენლებით, რომლებიც იყენებენ რა ენის ამ ზოგად, მათემატიკურად მკაცრად განსაზღვრულ გრამატიკულ ფორმალიზმს თავისებურად აინტერპრეტირებენ მას დასამუშავებელი ენის ავტომატური ანალიზისას.

განსხვავება პარსერის დეკლარაციულ და პროცედურულ ასპექტებს შორის განსაკუთრებით ცხადი ხდება მაშინ, როცა ფორმალური გრამატიკა ღიად ტოვებს რიგ ისეთ საკითხებს, რომლებიც ავტომატურმა ანალიზატორმა ამ თუ იმ ფორმით აუცილებლად უნდა გადაწყვიტოს.

სამაგალითოდ განვიხილოთ შემდეგი კონტექსტისგან თავისუფალი **PS**-გრამატიკა:

² არ არის რეკომენდებული გრამატიკული წესების სრული და პირდაპირი ფორმულირება პროგრამულ ენაში. წარმომქმნელი გრამატიკის ასეთ გამოყენებას უარყოფითი მხარეები ახლავს: ასეთ რეალიზაციებში კომპიუტერული პროგრამა ვერ აჩვენებს, თუ მისი რომელი თვისება არის თეორიულად შემთხვევითი (ე.ი. პროგრამისტის ინდივიდუალური განმასხვავებელი თავისებურების ან პროგრამული გარემოს შედეგი) და რომელია თეორიულად აუცილებელი (ე.ი. აღწერილი ენის ფორმალური ანალიზის შედეგი). ამგვარი მიღობის კიდევ ერთი უარყოფითი მხარე იმაში მდგომარეობს, რომ ასეთი პროგრამა შეძლება უკეთესადაც მუშაობდეს ერთი კონკრეტული ენისთვის, მაგრამ მთლიანი წარმომქმნელი გრამატიკისა და მისი სხვადასხვა ქვეტიპებისათვის უფრო ცუდ შედეგებს იძლევა.

წესი 1: $A \rightarrow B C$

წესი 2: $B \rightarrow c d$

წესი 3: $C \rightarrow e f$

ცვლადების ეს განაწილება თავის თავად გვარწმუნებს, რომ ამ ქვემოთგაშლად გამოყვანაში (**top-down derivation**) წესი 1 გამოიყენება უწინარესად 2 და 3 წესებისა. თუმცა, ის, ჯერ წესი 1 გამოიყენება და მერე წესი 2, თუ პირიქით, ან ორივე ერთდროულად, ღია საკითხია.

კონტექსტისგან თავისუფალი **PS**-გრამატიკის დეკლარაციული შემადგენლების დასახასიათებლად სრულიად საკმარისია წესების გამოყენების თანამიმდევრობის ასეთი ნაწილობრივი განსაზღვრა. მეორე მხრივ, კომპიუტერული პროგრამისთვის აუცილებელია წესების გამოყენების თანამიმდევრობა ცხადად და სრულად იყოს განსაზღვრული – ეს აუცილებელია თუნდაც ამან თეორიული თვალსაზრისით არაღირებული შედეგები მოგვცეს! – ასეთი მომაწესრიგებელი გადაწყვეტები, რომლებიც გადის ფორმალური გრამატიკის დეკლარაციული შემადგენლების მიღმა, ან რომლებიც – მანალიზებელი ალგორითმის მიზეზით – ეწინააღმდეგებიან გრამატიკული გამოყვანის კონცეპტუალურ დალაგებას (იხ. 9.3.4), იწოდებიან პროცედურულად.

ერთი რომელიმე მოცემული ფორმალიზმით (მაგ.: კონტექსტისგან თავისუფალი **PS**-გრამატიკით) პროგრამირების სხვადასხვა ენაში შეიძლება შემუშავდეს სხვადასხვა მანალიზებელი ალგორითმები. ეს ნიშნავს, რომ ერთი და იგივე დეკლარაციული გრამატიკა, როგორც დასამუშავებელი მოცემულობა, სხვადასხვა პროგრამულ ენებში სხვადასხვაგვარად შეიძლება იქნეს პროცედურულად რეალიზებული. მაგალითად, ორი პარსერალგორითმი (მხედველობაშია **Earley** და **CYK** ალგორითმები), რომლებიც რეალიზებულნი არიან პროგრამირების განსხვავებულ ენებში (შესაბამისად **Lisp** და **C** ენებში) ამუშავებენ **(i)** ერთი და იგივე გრამატიკას (მაგ.: 7.1.3 **a^kb^k** ენისთვის აღწერილ გრამატიკას) და **(ii)** ერთი და იგივე ენის ერთი იგივე შესაყვან **(input)** მოცემულობას (მაგ.: **aaabbb**), რაც ამ შემთხვევაში ამ ორი პროცედურულად განსხვავებულად რეალიზებული პარსერის დეკლარაციულ ინდენტურობას ადასტურებს.

გარკვეული გრამატიკული ტიპის ზოგადი პარსერი არსებობს მხოლოდ მაშინ, როცა მისი სირთულე არ არის ძალიან მაღალი. ამგვარად, ასეთი ზოგადი პარსერი არსებობს წესიერი და კონტექსტისგან თავისუფალი **PS**-გრამატიკული კლასებისთვის, მაშინ როდესაც არანაირი პრაქტიკული ღირებულების მქონე ზოგადი პარსერი არ შეიძლება შეიქმნას კონტექსტზე დამოკიდებული და შეუზღუდავი **PS**-გრამატიკული კლასებისათვის.

9.2 გრამატიკის მონიშვნა ენაზე

ის, თუ **PS**-იერარქიაში (იხ. 8.2.3) რა ადგილი შეიძლება მიეკუთვნოს ბუნებრივ ენებს, არის არა მარტო აკადემიური ინტერესის საგანი, არამედ გადამწყვეტი იმისაც, შესაძლებელია თუ არა ბუნებრივი ენის ეფექტური პარსერის აგება **PS**-გრამატიკის ფარგლებში. შედეგიანობის თვალსაზრისით ოპტიმალური იქნებოდა იმის ჩვენება, რომ ბუნებრივი ენები წესიერი ენების კლასს მიეკუთვნებიან. – ასეთ შემთხვევაში ბუნებრივი ენების **PS**-გრამატიკული ანალიზი საშუალებას მოგვცემდა აგვეგო მათი დროში წრფივი პარსერი.

თუმცა, ბუნებრივ ენებში არსებობს ისეთი სტრუქტურები, რომელთა ენობრივი ფორმები უფრო კონტექსტისგან თავისუფალი სახისანი არიან, ვიდრე წესიერი. სამაგალითოდ განვიხილოთ გერმანული ენის ცენტრზე-წყობილი რელაციური კლაუზა (**center-embedded relative clause**).

9.2.1 პონტიფისტისბან თავისუფალი სტრუქტურა გერმანულში

Der Mann,	schläft
der die Frau,	liebt,
die das Kind,	sieht,
das die Katze, füttert,	

ამ წინადადების სტრუქტურა შეესაბამება შემდეგ აბსტრაქტულ სქემას

“სახელური_ფრაზა₁ სახელური_ფრაზა₂ ... ზმნური_ფრაზა₂ ზმნური_ფრაზა₁”,

რომელიც, თავის მხრივ, შეესაბამება **abc ... cba** სტრუქტურას (იხ. 9.2.1). ეს კი კონტექსტისგან თავისუფალი სტრუქტურაა, რადგან აქ არ არის რაიმე გრამატიკული შეზღუდვა ჩალაგებული წევრების რაოდენობის თაობაზე. ეს ყველაფერი კი ამტკიცებს იმას, რომ ბუნებრივი ენის **PS**-გრამატიკული ანალიზი სულ ცოტა **n³** სირთულისაა.

ახლა უკვე კითხვა ისმის თაობაზე, თუ რამდენადა შესაძლებელი ბუნებრივი ენების მიკუთვნება კონტექსტისგან თავისუფალი ენების **PS**-გრამატიკული კლასისადმი? – ამ კითხვის პასუხი ნაკლებ ცხადია, ვიდრე ეს იყო წესიერი ენების შემთხვევაში.

1957 და 1965 წლებში ნ. ჩომსკიმ გააფართოვა კონტექსტისგან თავისუფალი **PS**-გრამატიკა ტრანსფორმაციული გრამატიკის წარმოებულ ფორმალიზმამდე იმ მიზეზით, რომ **PS**-გრამატიკა არ იყო საქმარისი იმის ფორმულირებისთვის, რასაც ის ენათმეცნიერულ განზოგადებად (**linguistics generalization**) განიხილავდა (იხ. 4.5.2, 8.5.4). ამასთან, 1985 წელს ს. შებერმა წარმოადგინა შვეიცარიული გერმანული ენის წინადადება, რომელსაც კონტექსტზე დამოკიდებული **WW** სტრუქტურა ჰქონდა (იხ. 8.3). ამით – ზემოთ უკვე მოყვანილი მტკიცების ანალოგით – ის ცდილობდა დაესაბუთებინა, რომ ბუნებრივი ენები, სულ ცოტა, კონტექსტზე დამოკიდებული არიან.

9.2.2 პოლიტიკური დამოკიდებული სტრუქტურა შვეიცარიულ გერმანულში

**mer em Hans es huus hälfed aastriiche
we the Hans the house help paint**

ამ მაგალითის ფორმალური სტრუქტურა გაანალიზებულია როგორც **a b a' b'** (სადაც **a = the Hans, b = the house, a' = help, b' = paint**). ეს სტრუქტურა არ არის კონტექსტისგან თავისუფალი, რადგან მას, ისევე როგორც კონტექსტზე დამოკიდებული **WW** ენას, არ აქვს ინვერსირებული სტრუქტურა. თუ ეს ასეა, მაშინ ბუნებრივი ენა **PS**-გრამატიკული თვალსაზრისებით სულ ცოტა კონტექსტზე დამოკიდებული მაინც არის. ეს იმას ნიშნავს, რომ იგი ექსპონენციალური სიითულისაა და რომ გრძელი წინადადების ანალიზი უარეს შემთხვევაში მიღიარდობით საუკუნეს მოგვთხოვდა (იხ. 8.2.2).

ამ წარმოუდგენელი შედეგის თავიდან ასაცილებლად პერმანენტ და გაზდარმა, შესაბამისად 1963 და 1982 წლებში წარმოადგინეს საქმარისი ზომის **PS**-გრამატიკული სისტემები, რომელებიც გამიზნული იყო იმის საჩვენებლად, რომ ბუნებრივ ენაში არ არსებობს ისეთი სტრუქტურა, რომელიც არ შეიძლება არ იქნას დაყვანილი კონტექსტისგან თავისუფალ ფორმაზე. პერმანის განაცხადი გაკეთდა იმ დროს, როდესაც სირთულეთა თეორია არ იყო ფართოდ გამოკვლეული და ტრანსფორმაციულ გრამატიკებში მოღვაწეობდა მხოლოდ ენთუზიასტთა მცირე ჯგუფი. ამავდროულად, პერმანის სისტემას აკლდა გამოკვეთილი და დეტალური ლინგვისტიკური მოტივაცია.

მეორე მხრივ, გაზდარის განაცხადის დროს სირთულეთა თეორია უკვე ფართოდ იყო შესწავლილი. ამასთან, გაზდარი უშუალოდ კი არა დაეყრდნო კონტექსტისგან თავისუფალ **PS**-გრამატიკას, არამედ ისარგებლა რა მეტაწესების (**metarules**) დამატების მექანიზმით ჯერ განსაზღვრა წარმოებული **PS**-გრამატიკული ფორმალიზმი, რომელიც განზოგადებული ფრაზების სტრუქტურულ გრამატიკად (**GPSG**) იწოდება. მისეული მეტაწესების მიზანი იყო უზარმაზარი რაოდენობის კონტექსტისგან თავისუფალი **PS**-წესების³ კომბინირება, იმისათვის, რომ შესაძლებელი ყოფილიყო იმ ლინგვისტურ განზოგადებათა ფორმალიზება, რომელთაც იმ დროს გაზდარიც და სხვებიც მნიშვნელოვნად მიიჩნევდნენ.

გარდა ამისა, იმედოვნებდნენ, რომ **GPSG** ბუნებრივ ენებს გააანალიზებდა სირთულის იმ ხარისხით, რომელიც გამოთვლითი თვალსაზრისებით პრაქტიკულად ღირებულად მიიჩნევა, ე.ო. ან კონტექსტისგან თავისუფალი, ან n^3 სირთულით. თუმცა, გაზდარის საწყისი დაშვების საწინააღმდეგოდ, ანუ იმის, რომ მეტაწესების გამოყენება არ გამოიწვევდა სირთულის გაზრდას, ამ მიმართულებით ჩატარებულმა ჩამკეტმა ფორმალურმა გამოკვლევამ⁴ დაამტკიცა, რომ სინამდვილეში **GPSG** არის რეკურსულად გადათვლადი და, აქედან გამომდინარე, არაამოხსნადი.

თუ ბუნებრივი ენა არაა კონტექსტისგან თავისუფალი – როგორც მიიჩნევს თეორეტიკოს ენათმეცნიერთა უმრავლესობა – მაშინ ფორმალურ ენათა რომელ კლასს მიეკუთვნება იგი? – ამ შეკითხვის პასუხის ძიებისას უნდა გვახსოვდეს, რომ კონტექსტისგან თავისუფალი ენების კლასი არის გარკვეული სახის ფორმალიზმის (მხედველობაშია **PS**-გრამატიკა) გარკვეული სახის წესების (მხედველობაშია ტიპი 2) გამოყენების შედეგი.

ამასთან, ერთი მხრივ, არ გვაქვს არანაირი მიზეზი იმ დასკვნის გასაკეთებლად. რომ ეს კონკრეტული ფორმალიზმი და წესთა ეს კონკრეტული ტიპი – კონტექსტისგან თავისუფალი ენების ინვერსირებადი დაწყვილების სტრუქტურის მომცემი – ბუნებრივ ენას სრულად ახასიათებს. მეორე მხრივ, კონტექსტისგან თავისუფალი ენები ქმნიან **PS**-გრამატიკის იმ უდიდეს ენობრივ კლასს, რომლის მათემატიკური სირთულის საკმარისად დაბალი ხარისხი მას პრაქტიკული თვალსაზრისებით ღირებულად აქცევს.

დაშვება, რომ ბუნებრივი ენა არაა კონტექსტისგან თავისუფალი, იწვევს ერთ-ერთს შემდეგი ორიდან:

³ "დაახლოებით ტრილიონიბით წესი", ს. შებგრი, ს. სტაკა, პ. უშკოლევაჭია, დ. კ. რობინსონი 1983

⁴ პ. უშკორეიტი & ს. პეტრისი. 1986.

1. **PS**-გრამატიკა წარმოქმნელი გრამატიკის ერთადერთი ელემენტარული ფორმალიზმია, და, შესაბამისად, უნდა ვივარაუდოთ, რომ ბუნებრივი ენები მაღალი სირთულისაა და რომ გამოთვლას არ ემორჩილებიან.

2. **PS**-გრამატიკა არაა წარმოქმნელი გრამატიკის ერთადერთი ელემენტარული ფორმალიზმი. შესაბამისად, არსებობენ სხვა ელემენტარული ფორმალიზმები, რომლებიც იძლევიან სხვა ენობრივ იერარქიებს, რომელთა შესაბამისი ენობრივი კლასები აღნიშნული **PS**-გრამატიკების ორთოგონულია.

იმ ფაქტის გათვალისწინებით, რომ ადამიანები ბუნებრივ ენას თავისუფლად, მაღალეფექტურად და შედეგიანად იყენებენ, პრეცენტი დასკვნა მიუღებელია. მეორე დასკვნას კი, თავის მხრივ, საკითხი ამ დასკვნით ხაზგასმული კონკრეტული ალტერნატივების მოძიებაზე დაჰყავს.

ისტორიული თვალსაზრისით პირველი ბუნებრივი ნაბიჯი ახალი ფორმალური ენობრივი კლასების არსებობის დადგენისა უნდა იყოს **C**-გრამატიკის წარმოქმნითი უნარისა და სირთულის გაანალიზება (7.4 თავი) და დასადგენ ენობრივ კლასებთან ამ ანალიზით გამოკვეთილი ენობრივი კლასების შედარება. ამგვარად, ამ მიზნებით **C**- და **PS**- გრამატიკების ფორმალური თვისებების შედარება აქ არსებულ პირდაპირ და უმარტივეს სტრატეგიად ისაზღვრება. ამასთან, ცხადია, რომ ასეთი შედარებისას შეძლევი სამიდან მხოლოდ ერთი შესაძლო მიმართება გამოიკვეთება.

9.2.3 შესაძლო მიმართებები რო გრამატიკულ ფორმალიზმს გვითვავთ

- არანაირი ეკვივალენცია (*no equivalence*)

ორი გრამატიკული ფორმალიზმი არ არის ეკვივალენციის მიმართებაში, თუ ისინი წარმოქმნიან/გამოიცნობენ განსხვავებულ ენობრივ კლასებს. ეს ნიშნავს, რომ ამ ფორმალიზმებს განსხვავებული წარმოქმნითი უნარი აქვთ.

- სუსტი ეკვივალენცია (*weak equivalence*)

ორი გრამატიკული ფორმალიზმი არის სუსტი ეკვივალენციის მიმართებაში, თუ ისინი წარმოქმნიან/გამოიცნობენ ერთი და იგივე ენობრივ კლასებს. ეს ნიშნავს, რომ ამ ფორმალიზმს ერთი და იგივე წარმოქმნითი უნარი აქვთ.

- ძლიერი ეკვივალენცია (*strong equivalence*)

ორი გრამატიკული ფორმალიზმი არის ძლიერი ეკვივალენციის მიმართებაში, თუ ისინი არიან **(i)** სუსტი ეკვივალენციის მიმართებაში, და ამავე დროს **(ii)** წარმოქმნიან ერთი და იგივე სტრუქტურულ აღწერებს. ეს ნიშნავს, რომ ეს ფორმალიზმები მეტი არაფერი არ არის, თუ არა ერთმანეთის ფორმალური, ანუ ნოტაციური ვარიანტები (**notational variants**).

თანამედროვე ენათმეცნიერების ისტორიული განვითარებისათვის სასურველი იქნებოდა, რომ **C** და **PS** გრამატიკის ელემენტარული ფორმალიზმები ეკვივალენტურები არ ყოფილიყვნენ, რადგან, ასეთ შემთხვევაში, ჩვენ გვექნებოდა კონტექსტისგან თავისუფლად ენების კლასის ჭეშმარიტი ალტერნატივა. თუმცა, ადრევე გაირკვა, რომ **C** გრამატიკა და **PS** გრამატიკა ერთმანეთთან **სუსტი ეკვივალენციის** მიმართებაში არიან.

საკითხი წარმოიქმნა ამ ტიპის [PS-] გრამატიკებსა და კატეგორიულ გრამატიკებს შორის ზუსტი ურთიერთკავშირების დადგენისას. 1958-ში მე მივწვდი, რომ **BCG**ები [ბილირექციული კატეგორიული გრამატიკები (იხ. 7.4.1)] არიან დაახლოებით იგივე სიძლიერის როგორც **CFPSG**ები [კონტექსტისგან თავისუფალი ფრაზების სტრუქტურული გრამატიკები]. მათი ეკვივალენტია დამტკიცა კაუჭმანმა 1959 წლის ივნისში. . . გრამატიკის ამ განსხვავებული ტიპების ეკვივალენტია არც თუ მთლად მოულოდნელი იყო. — თითოეული მათგანი გამიზნული იყო უშუალო შემადგენელთა გრამატიკის იმ იდეების ზუსტი აღწერისთვის, რომლებიც მრავალი წლის განმავლობაში ითვლებოდა ამერიკული აღწერითი ენათმეცნიერების რჩეულ იდეებად. სადემონსტრაციოდ. მაგალითად, იხილეთ პარისის [1951] და პოკეტის [1958] ცნობილ წიგნები

ი. ბარ-ჰილელი 1960 [1964. პ. 103]

ის ფაქტი, რომ **C** და **PS** გრამატიკები ეკვივალენტურები არიან მხოლოდ გარკვეულ ქვეტიპებში, კერძოდ ის, რომ ბილირექციული **C** გრამატიკა და კონტექსტისგან თავისუფალი **PS** გრამატიკა ეკვივალენტურები არიან, გააფერებთალა იმ ფაქტმა, რომ კონტექსტისგან თავისუფალი გრამატიკები წარმოქმნიან იმ ყველაზე უფრო ფართო ენობრივ კლასს, რომელიც ჯერ კიდევ ექვემდებარება გამოთვლას. ამის გათვალისწინებით, ბილირექციული **C** გრამატიკისა და კონტექსტისგან თავისუფალი **PS** გრამატიკის სუსტი ეკვივალენტია, არ იწვევს ასეთსავე ღრმა ინტერესს. არადა, ამან, თავის დროზე, საფუძველი ჩაუყარა არასწორ შეხედულებას, რომ **PS** გრამატიკა და ფორმალური ენების **PS** გრამატიკული იერარქია არის რაღაც ბუნების მიერ მოცემული საერთო საფუძველი ყველა შესაძლო ბუნებრივი და ხელოვნური ენობრივი სისტემებისა. თუმცა, როგორც უკვე დავრწმუნდით, უნდა არსებობდეს წარმომქმნელი გრამატიკის ისეთი ერთი მაინც ალტერნატიული ფორმალიზმი, რომელიც ხელოვნური ენების ერთობლიობას ყოფს სრულიად განსხვავებულ ენათა კლასებად, ვიდრე ამას **PS** გრამატიკა აკეთებს.⁵

ძირითადი დამოკიდებულებები ენის განსაზღვრებებს, წარმომქმნელ გრამატიკებს, გრამატიკის ქვეტიპებს, ენების კლასებს, პარსერებსა და სირთულეებს შორის შემდეგით შეიძლება შეჯამდეს:

- ენები (Languages)

ენები არსებობენ წარმომქმნელი გრამატიკებისგან დამოუკიდებლად. ამას ადასტურებს არა მარტო ბუნებრივი ენები, არამედ ასევე ხელოვნური ენები: **a^kb^m, a^kb^kc^k, a^kb^kc^kd^k, {a^kb^kc^k}, WW^R, WW, WWW** და ა.შ. მათი ტრადიციული სახელები ამ ენებს ისე კარგად ახასიათებენ, რომ შესაძლებელი ხდება მათი სწორად ფორმირებული გამოსახულებების როგორც ჩამოწერა, ისე გამოცნობა. რაიმე გრამატიკული ფორმალიზმის ფარგლებში რაიმე გრამატიკის ცხადი განსაზღვრა რომელიმე ბუნებრივი თუ ხელოვნური ენისათვის ითხოვს ნაბიჯს, რომელიც, როგორც წესი, არატრივიალურია: ის, რომ მოცემული ენა შეიძლება აღიწეროს სხვადასხვა გრამატიკული ფორმალიზმებითა და სხვადასხვა ფორმალური

⁵ მართლაც, კონტექსტისგან თავისუფალი **a^kb^k** (იხ. 7.1.3, 10.2.2) და კონტექსტზე დამოკიდებული **a^kb^kc^k** (იხ. 8.3.6, 10.2.3) ენები **LA**-გრამატიკაში კლასიფიცირებული არიან როგორც **C1-LAG**-ების ერთი და იგივე წრფივი კლასების წევრები. შესაბამისად, კონტექსტისგან თავისუფალი **WW^R** (იხ. 8.3.4, 11.1.5) და კონტექსტზე დამოკიდებული **WW** ენები **LA**-გრამატიკაში კლასიფიცირებული არიან როგორც **C2-LAG**-ების ერთი და იგივე პოლიომური (**n²**) კლასის წევრები.

გრამატიკებით, ნაჩვენებია 7.4.3 და 7.1.3 პუნქტებში $a^k b^k$ ენის **C** და **PS** გრამატიკული ანალიზების შედარებით.

- წარმომქმნელი გრამატიკები (Generative grammars)

ერთი მხრივ, წარმომქმნელი გრამატიკა არის საერთო, საზოგადო ფორმალური ყალიბი; მეორე მხრივ, ეს არის ამ საფრთხო ყალიბის ფარგლებში სპეციფიკური ენის აღსაწერად განსაზღვრული. წესთა სპეციფიკური სისტემა. მაგალითად, **PS** გრამატიკა, როგორც ზოგადი ფორმალური ყალიბი, განსაზღვრულია **<V, VT, S, P>** ოთხეულით და მის კომპონენტებზე მოცემული გარკვეული დამატებითი პირობებით. ამ ზოგადი ყალიბის ფარგლებში $a^k b^k$ ენის მსგავსი სპეციფიკური ენების წარმოსაქმნელად შესაძლებელი ხდება სხვადასხვა სპეციფიკური **PS** გრამატიკების განსაზღვრა.

- წარმომქმნელი გრამატიკების ქვეტიპები (Subtypes of generative grammars)

წარმომქმნელი გრამატიკის ზოგადი ფორმალური ყალიბის სხვადასხვა შეზღუდვები იძლევა სხვადასხვა გრამატიკულ ტიპებს. ამ გზით არის განსაზღვრული რეგულარული, კონტექსტისგან თავისუფალი, კონტექსტზე დამოკიდებული და შეუზღუდავი **PS** გრამატიკები, როგორც **PS** გრამატიკის ქვეტიპები, და **C1-, C2-, C3-, B-** და **A-LAG** გრამატიკები როგორც **LA**-გრამატიკის ქვეტიპები. სხვადასხვა შეზღუდვები არსებობენ არა თავისთავად და დამოუკიდებლად, არამედ არიან დამოკიდებულნი გამოყენებული გრამატიკის ფორმალურ თვისებებზე (ძირითადად წესების სტრუქტურაზე).

- ენის კლასები (Language classes)

წარმომქმნელი გრამატიკის ეს სხვადასხვა ქვეტიპები შეიძლება გამოყენებულ იქნენ შესაძლო ენათა ერთობლიობის დასაყოფად ენათა სხვადასხვა კლასებად. იმის გამო, რომ წარმომქმნელი გრამატიკის ქვეტიპები დამოკიდებულია გამოყენებულ ფორმალიზმზე, ასოცირებული ენის კლასები არ არსებობენ თავისთავად, ისინი გამოსახავენ გამოყენებული გრამატიკული ტიპის ფორმალურ თვისებებს. მაგალითად, კონტექსტისგან თავისუფალი ენების წყვილ-ინვერსიონი სპეციფიკაცია პირდაპირ გამომდინარეობს კონტექსტისგან თავისუფალი **PS** გრამატიკების წესთა სტრუქტურის სპეციფიკური შეზღუდვებისგან.

Nota bene: ენები არსებობენ იმ ფორმალური გრამატიკებისგან დამოუკიდებლად, რომლებმაც შესაძლოა წარმოქმნან ისინი. ენის კლასები, მეორე მხრივ, არ არსებობენ დამოუკიდებლად, ისინი განსაზღვრული გრამატიკული ფორმალიზმების განსაზღვრული შეზღუდვების შედეგია.

- პარსერები (Parsers)

ანალიზატორები ენის ავტომატურად მააანალიზებელი პროგრამებია, რომლებიც განსაზღვრულია წარმომქმნელი გრამატიკის მთლიან ქვეტიპებზე (მაგ.: კონტექსტისგან თავისუფალი **PS** გრამატიკა ან **C-LAG-ები**). ამგვარად, $a^k b^k c^k$ ენის მსგავსი კონტექსტზე დამოკიდებული ენების პრობლემა არ არის იმ სირთულის, რომ ვერ დაიწეროს მათი ეფექტურად მაანალიზებელი პროგრამა, მაგრამ, მიუხედავად ამისა, ვერ შეიქმნება ვერანაირი პრაქტიკულად ღირებული ანალიზატორი კონტექსტზე დამოკიდებული **PS** გრამატიკისთვის.

- სირთულე (Complexity)

წარმომქმნელი გრამატიკის ქვეტიპის სირთულე განსაზღვრულია იმ პრიმიტიული ოპერაციების რაოდენობით, რაც ყველაზე უარეს შემთხვევაში ესაჭიროება ეკვივალენტურ აბსტრაქტულ ავტომატს, ან პარსერს დასამუშავებელი გამოსახულებების გასაანალიზებლად. ინდივიდუალური... ენების სირთულე ჩვეულებრივ განისაზღვრება მათი (ენების) შესაბამისი ენობრივი კლასების სირთულით. ამასთან, რადგან ენობრივი კლასები და მათი თვისებები დამოკიდებულია ამ კლასების მომცემ ფორმალიზმზე, გასაგები ხდება, თუ რატომ არის რომ $a^k b^k c^k$ ენის ტიპის ენები **PS** გრამატიკაში ეკუთვნიან კონტექსტულ დამოკიდებულ ენათა კლასს, რომელიც ექსპონენციალური სირთულისაა, ხოლო **LA**-გრამატიკაში ასეთი ენები არიან **C1-LAG** კლასში, რომელიც წრფივი სირთულისაა.

გარდა ენის შესაბამისი ენობრივი კლასით განსაზღვრული ენის სირთულის ცნებისა შეიძლება აგრეთვე ინდივიდუალური ენის ინჰერენტული სირთულის (**inherent complexity**) გამოკვლევა. ასეთ შემთხვევაში ეყრდნობიან ენის სპეციფიკურ სტრუქტურულ თვისებებს (რამე განსაზღვრული გრამატიკული ფორმალიზმისგან დამოუკიდებლად) იმის განსასაზღვრავად თუ უარეს შემთხვევაში რამდენ ოპერაციას ითხოვს მისი ანალიზი აბსტრაქტულ მანქანაზე (მაგ.: ტიურინგის ან რეგისტრ მანქანაზე (**register machine**)). მაგალითად, **3SAT** და **Subset Sum** (შედ.: 11.4 და 11.5 ქვეთავები) მსგავსი ენები ინჰერენტულად როტული ენებია. შესაბამისად, ეს ენები ნებისმიერ შესაძლო გრამატიკულ ფორმალიზმში აუცილებლად მაღალი სირთულის კლასში იქნებიან (აյ ეს არის ექსპონენციალური კლასი).

ინდივიდუალური ენის ინჰერენტული სირთულე მნიშვნელოვანი საშუალებაა ენის კლასის მინიმალური სირთულის გამოსათვლელად. ანალიზის ეს ფორმა შემოდის ძალიან დაბალ დონეზე და, როგორც წესი, შესაბამება მანქანურ და ასემბლერულ კოდებს. ამის გამოა, რომ, როგორც წესი, ხელოვნური თუ ბუნებრივი ენების სირთულე ანალიზდება გრამატიკული ფორმალიზმის აბსტრაქტულ დონეზე, რის გამოც სირთულე ითვლება გრამატიკული ტიპის და ამ ტიპის შესაბამისი მთლიანი ენობრივი კლასისთვის.

9.3 ტიპობრივი ტრანსფარენტულობის დამოკიდებულება გრამატიკებსა და არსერებს შორის

ითვლება, რომ პარსერი გრამატიკას ყველაზე მარტივი და ტრანსფარენტული სახით იყენებს მაშინ, როდესაც იგი მხოლოდ ამ გრამატიკული მექანიზმებით ცდილობს შემავალი გამოსახულების ანალიზებას. — ასეთი ხედვა პარსერზე, როგორც გრამატიკის **ამამუშავებელსა (motor)** და **გამძლოლზე (driver)**, თავიდანვე იყო გათვალისწინებული **PS** გრამატიკაში.

მიღერის და ჩომსკის ძირითადი შემოთავაზება (1963) იყო ის, რომ გრამატიკები უნდა ყოფილიყვნენ რეალიზებულნი როგორც მაანალიზებელი ალგორითმები. ჩვენ შეგვიძლია ავიდოთ ეს როგორც მეთოდოლოგიური პრინციპი. ამ შემთხვევაში, გასაგები ხდება, რომ გრამატიკაში არსებული წესებისა და სტრუქტურების მკაცრი ლოგიკური ორგანიზების პირდაპირი და ზუსტი სარკელი ასახვით უნდა მოხდეს შესაბამისი პარსერის მექანიზმის ორგანიზება. ამას ჩვენ ტიპობრივი ტრანსფარენტულობა (**type transparency**) ვუწოდეთ.

რ. კ. ბერვიკი & ა. ს. ვაინბერგი, 1984, გვ. 39.

ქვემოთ მოყვანილ განსაზღვრებას შემოყავს **სრული ტიპობრივი ტრანსფარენტულობის (absolute type transparency)** ცნება (იხ. **R. C. Berwick & A. S. Weinberg** 1984, გვ. 41.) ზუსტი და ინტუიციურად ცხადი ზოგადი გზით.

9.3.1 სრული ტიპობრივი ტრანსპარენტულობის განსაზღვრება

- ნებისმიერი ენისათვის პარსერი და წარმომქმნელი იყენებს ერთსა და იმავე ფორმალურ გრამატიკას,
- შესაბამისად, პარსერიც და წარმომქმნელიც მიმართავს გრამატიკის წესებს პირდაპირ.
- ეს ნიშნავს, რომ პარსერიც და წარმომქმნელიც წესებს იმავე თანამიმდევრობით ამუშავებს, რა თანმიმდევრობითაც ისინი მუშავდება გრამატიკული დერივაციისას.
- ამგვარად, თითოეული ამ წესის გამოყენებისას პარსერიც და წარმომქმნელიც ამუშავებს იმავე შემავალ გამოსახულებას, რასაც გრამატიკა და
- ამასთან, თითოეული წესის გამოყენებისას პარსერიც და წარმომქმნელიც იძლევა იმავე გამომავალ გამოსახულებას რასაც გრამატიკა.

PS გრამატიკაში ადვილად დგინდება, რომ პარსერის მიერ გრამატიკის წესების პირდაპირი გამოყენება შეუძლებელია. ამის ისტორიული მიზეზი არის ის, რომ 1936 წელს პოსტმა მისი მარტარმოებელი, ანუ გადამწერი სისტემა ააგო რეკურსის თეორიაში ეფექტური გამოთვლადობის (**effective computability**) მათემატიკური ცნების განსასაზღვრავად.⁶ პოსტის სისტემაში დერივაციული რიგი ეწყორებოდა ერთი სახის ნიშნების სხვა სახის ნიშნებით ჩანაცვლებას, რაც მისი მიზნებისთვის სრულიად ბუნებრივი იყო.

როცა ჩომსკიმ 1957 წელს ბუნებრივი ენის გასაანალიზებლად **PS** გრამატიკის სახელით გამოიყენა პოსტის მარტარმოებელი სისტემა, იგი იძულებული გახდა გამოეყენებინა აგრეთვე პოსტის სისტემის ჩანაცვლებაზე დამყარებული გამოყვანის რიგი. იმის გამო, რომ პარსერი დასამუშავებლად იღებს ტერმინალური სიმბოლოების სტრიქონს, **PS** გრამატიკები და მათი პარსერები არ არიან შემავალ-გამომავალ ეკვივალენციაში (**input-output equivalent**). – ეს იმას ნიშნავს, რომ ვერ იარსებებს ისეთი პარსერი, რომელსაც შეეძლება **PS**-გრამატიკის წესების პირდაპირი გამოყენება.

პარსერში კონტექსტისგან თავისუფალი **PS**-გრამატიკის გამოყენების სტრუქტურული პრობლემების სადემონსტრაციოდ ქვემოთ წარმოდგენილია **aaabbbb** გამოსახულების ნაბიჯ-ნაბიჯითი დერივაცია (**derivation**) 7.1.3 განაწერში $a^k b^k$ ენისათვის განსაზღვრულ **PS** გრამატიკი.

9.3.2 **aaabbbb** გამოსახულების ქვემოთ მიმართული გამოყვანა

⁶ მაგალითისთვის იხილეთ **A. Church**, გვ. 52, სქოლით № 119

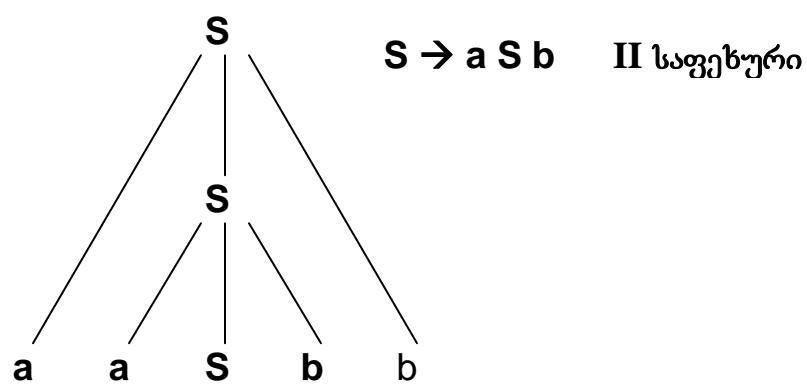
S

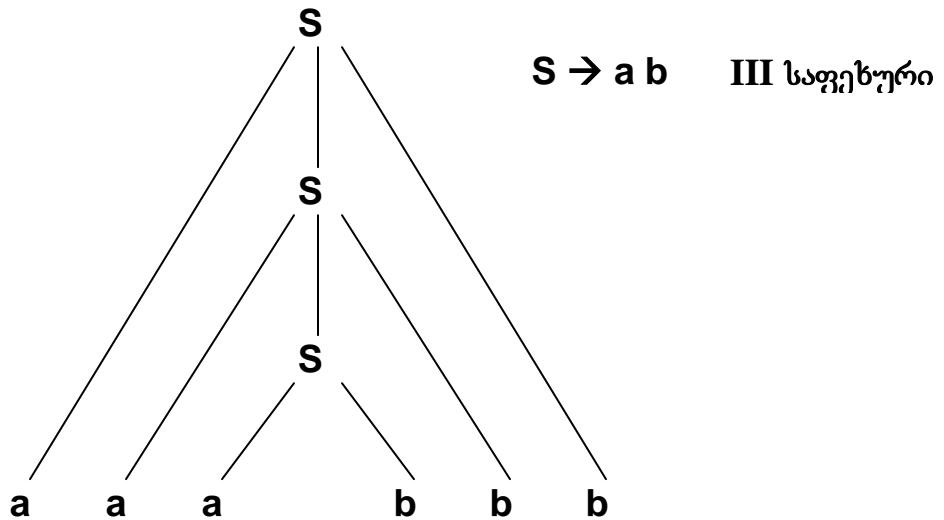
S → a S b

I საფუნქცი



a S b





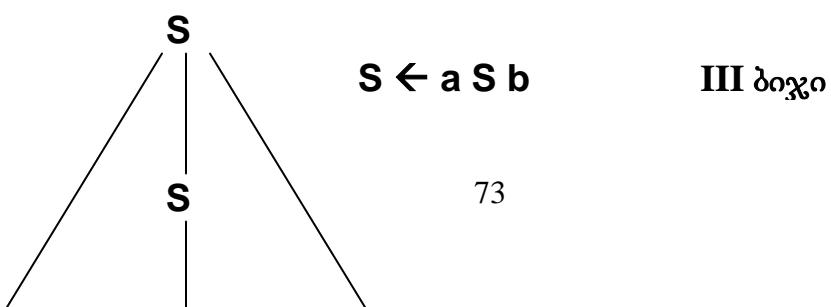
ამ ჩანაცვლებაზე დაფუძნებული გამოყვანის (**substitution-based derivation**) წესების მარცხნა მხარის **S** ცვლადი მრავალჯერაა ჩანაცვლებული წესების მარჯვენა მხარეს განთავსებული ნიშნების სტრიქონით. გამოყვანა იწყება **S** ცვლადით და მთავრდება ყველა ცვლადის ტერმინალური ნიშნებით ჩანაცვლების შემდეგ. ცვლადების ასეთი დამუშავება ქვემოთ მიმართულ გამოყვანად (**top-down derivation**) იწოდება.

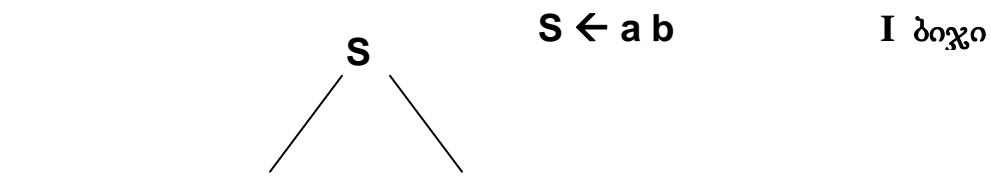
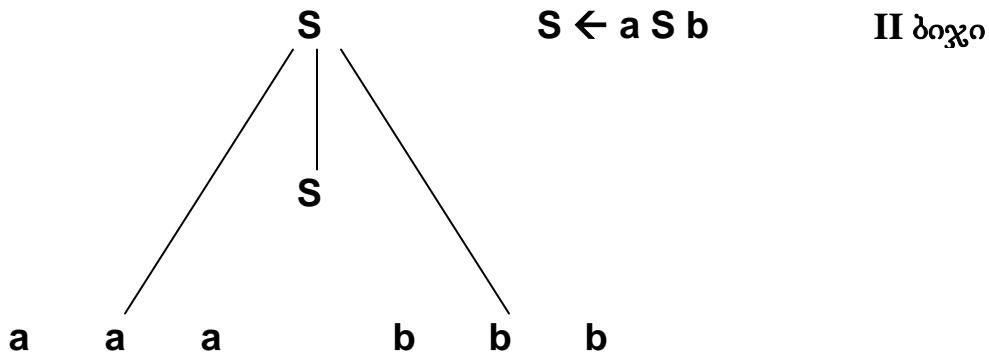
გამოთვლითი პროცედურა **aaabbbb** შემავალი გამოსახულების ანალიზისას პირდაპირ ვერ იყენებს **PS** გრამატიკის დერივაციულ მექანიზმს, რადგანაც **PS**-გრამატიკა გამომავალ სტრიქონში ტერმინალურ სიმბოლოებს ყოველთვის ორ სხვადასხვა ადგილას სვამს. **PS** გრამატიკისგან განსხვავებით კომპიუტერულ პროგრამას დასამუშავებელი აქვს გაუანალიზებელი ტერმინალური სტრიქონი და, შესაბამისად, სტრუქტურული კავშირი ნებისმიერი მანძილით დაშორებულ ორ ადგილს შორის სრულიად არაცხადია.

შემავალი გამოსახულების ავტომატური ანალიზისას ქვემოთ მიმართული გრამატიკული გამოყვანის პირდაპირი გამოყენების ერთადერთი შესაძლებლობა იქნებოდა ყველა გამომავალი გამოსახულების თანდათანობითი გამოყვანა, იმ იმედით, რომ გასაანალიზებელი სტრიქონი რაღაც მომენტში გამოჩნდება დამუშავებულ გამოსახულებებს შორის. **a^kb^k** ენის შემთხვევაში თავისუფლად შესაძლებელია, რომ ეს ასე მოხდეს. ასეთი მიდგომა არ მოგვცემს ამოხსნას **WW^R** ენისათვის, რადგან ამ შემთხვევაში შესაძლო გამოსავალ გამოსახულებათა რაოდენობა შეიძლება გაიზარდოს მათი სიგრძის ექსპონენციალურად. — ეს არის მიზეზი იმისა, რომ ყველა შესაძლო სტრიქონის წარმოქმნის მეთოდი მაღლევე შეფასდა არავარგის და არაეფექტურ მეთოდად.

როგორც ალტერნატივა, შესაძლოა იცადოს **PS**-გრამატიკის გამოყენება ზემოთ მიმართული (**bottom up**) გამოყვანის ასაგებად.

9.3.3 aaabbbb გამოსახულების ზემოთ მიმართული გამოყვანა





ზემოთ მიმართული გამოყვანა (bottom-up derivation) იწყება ერთი ან რამოდენიმე წესის მარჯვნა მხარეს განთავსებული გამოსახულებით და ცვლის მათ წესის მარცხენა მხარეს განთავსებული ცვლადით. ეს ცვლადი და ასე წარმოქმნილი გამოსახულება, ან მისი ნაწილი წყვილება სხვა წესის მარჯვნა მხარესთან და შესაბამისად ამ წესისა ეს დაწყვილება იცვლება წესის მარცხენა მხარეს განთავსებული ცვლადით. ამ ტიპის ზემოთ მიმართული გამოყვანა ასრულებს მუშაობას, როცა მხოლოდ ერთი წესია აქტიური და როცა მისი მარჯვნა მხარე იცვლება საწყისი **S** სიმბოლოთი.

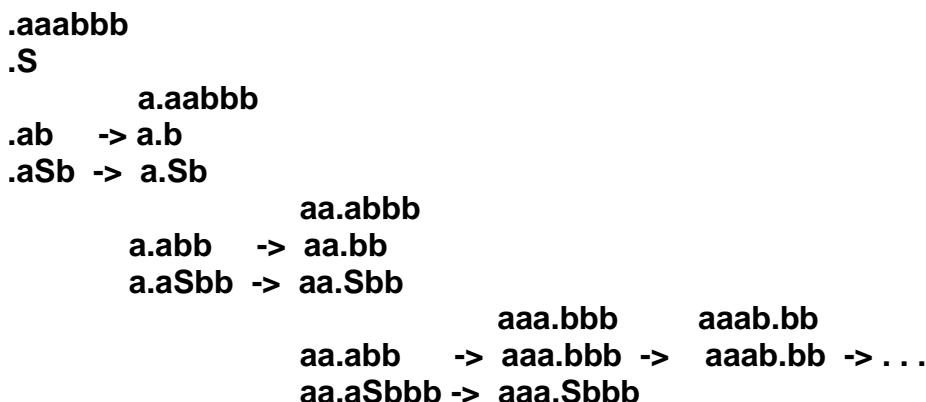
კონტექსტისგან თავისუფალ **PS**-გრამატიკაში ზემოთ მიმართული გამოყვანის სისუსტე არის ის, რომ ის პროგრამისგან ითხოვს დაწყვილებული ინვერსირებადობის სტრუქტურის თეორიული ცენტრის მოძებნას. ეს პრობლემაა, რადგანაც პროგრამას შეიძლება შემთხვევით შესასვლელიდან გასაანალიზებლად შეუვიდეს ისეთი სტრიქონი, რომელშიც აღნიშნული ცენტრი საერთოდ არც არსებობს.

a^kb^k ენის შემთხვევაში პროგრამას ცენტრის განსასაზღვრად შეუძლია გამოიყენოს საზღვარი **a** და **b** სიტყვებს შორის, მაგრამ ასეთი მიღეობა არ იძლევა ზოგად გადაწყვეტას კონტექსტისგან თავისუფალი **PS**-გრამატიკისთვის. მაგალითისთვის **WW^R** ენა შეიცავს **aaaaaa** და **aaaaaaaa** სახის სტრიქონებს, რომელთა ცენტრის გამოთვლა შეიძლება მხოლოდ მათი სიგრძის გამოთვლით. თუმცა, არც სიგრძე იძლევა ზოგად კრიტერიუმს, რადგან **a^kb^{3k}** (იხ. 8.3.4) სახის კონტექსტისგან

თავისუფალი ენებში გამოყვანის ცენტრი არ არის განთავსებული შესავალი გამოსახულების სიგრძით ცენტრში.

იმის გამო, რომ ვაჩვენოთ, თუ როგორ შეიძლება გადალახოს პარსერმა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ლოგიკური წარმოების წესრიგთან შექმნილი სირთულეები ზოგადად მოქმედი მეთოდით, სჯობს განვიხილოთ ანალიზი ერლეის ალგორითმის ფარგლებში, რომელიც **Hopcroft & Ullman, 1979**, გვ. 145-ის მიხედვით „არის ყველაზე პრაქტიკული და ზოგადი კონტექსტისგან თავისუფალი გამოცნობისა და პარსირების ალგორითმი.“ შემდეგი, 9.3.4 განაწერის ძირითადი მიზანია კონტექსტისაგან თავისუფალი ენების ზოგადი პარსირებისათვის საჭირო **PS**-გრამატიკის წესების რესტრუქტურიზაციის მიმოხილვა.

9.3.4 $a^k b^k$ ენის მრავალი მარალიზებელი ალგორითმი



ერლეის⁷ ალგორითმი იყენებს სამ ოპერაციას, რომლებიც არ არიან **PS**-გრამატიკის ალგებრული განსაზღვრების შემდგენლები (იხ. 8.1.1). ესენია პრედიკატორ- (**predictor**), სკანერ- (**scanner**) და კომპლეტორ- (**completer**) ოპერაციები (**operation**) (იხ. ერლეი 1970). მათი მეშვეობით, 9.3.4 ალგორითმის შემავალი გამოსახულება ანალიზდება მარჯვნიდან მარცხნივ (**left to right**). ამასთან, წერტილი გვიჩვენებს, თუ სადამდე უნდა გაგრძელდეს ანალიზი.

ანალიზის დასაწყისში წერტილი დასმულია შემავალი

.a a a b b b

სტრიქონის წინ, რომელიც ინტერპრეტირებულია როგორც

.S

მდგომარეობა (**state**), ანუ როგორც **S** ცვლადი. ცვლადის ასლი ვერ იქნება შესატან სტრიქონში, რის გამოც ვერ გამოვიყენებთ სკან-ოპერაციას. თუმცა, $a^k b^k$ ენისთვის განსაზღვრული **PS**-გრამატიკის შემდეგი ორი წესის **S** → **a b** და **S** → **a S b** გამოყენებით პრედიკატორ-ოპერაცია წარმოქმნის ორ ახალ, სახელდობრ

.a b და .a S b

მდგომარეობას, რომლებიც ემატებიან ალგორითმის მდგომარეობათა სიმრავლეს (**state set**).

ალგორითმის შემდეგ ციკლში როგორც დასამუშავებელ, ე.რ.

⁷ ჯერ ერლეიმ ააგო თავისი მანალიზებელი ალგორითმი **PS**-გრამატიკებისთვის თავის დისერტაციაში კარნეგ-მელონის უნივერსიტეტის (პიტ საურგი, აშშ) კომპიუტერულ მეცნიერებათა დეპარტამენტში. ამ მიღწევის შემდეგ ის გაუჩინარდა სამეცნიერო სცენიდან და ამჟამად დაკარგულად ითვლება.

a. a a b b b

სტრიქონში, ისე ორივე ახალ, ე.ი.

a. b და a. S b

მდგომარეობებში წერტილი წაძრულია ერთი პოზიციით მარჯვნივ (კომპლექტორ-ოპერაცია). სკან-ოპერაცია ამოწმებს შეესაბამება თუ არა მდგომარეობაში წერტილამდე არსებული ნიშანი დასამუშავებელი სტრიქონის წერტილამდე განთავსებულ ნიშანს. ეს ასეა ორივე შემთხვევაში.

შეძლებ ციკლში წერტილი კვლავინდებურად წაიძრა კიდევ ერთი პოზიციით მარჯვნივ. შეძლებ დასამუშავებელ სტრიქონად დაჯდა

a a. a b b b

და მდგომარეობებად:

a b. და a S. b

პირველ მდგომარეობაში წერტილის წინ არის **b** ტერმინალური სიმბოლო, ხოლო დასამუშავებელ სტრიქონში წერტილამდე არის **a** ტერმინალური სიმბოლო, რაც **სკან-ოპერაციის** შესრულების მცდელობას წარუმატებლობით ასრულებს. მეორე მდგომარეობაში, მეორე მხრივ, წერტილამდე გვაქვს **S**, რომელზეც, გრამატიკული წესების მიხედვით, დაშვებულია პრედიკატორ-ოპერაციის გამოყენება. ეს შეძლებად იძლევა მდგომარეობებს

a a. b b და a a. S b b

რომელთაგანაც ორივეს **სკან-ოპერაცია** უსაბამებს დასამუშავებელ სტრიქონს. შეძლებ კომპლექტორ-ოპერაცია წერტილს კვლავ წაძრავს ერთი პოზიციით მარჯვნივ. შეძლებად ვლებულობთ

a a a. b b b

სტრიქონსა და

a a b. b და a a S. b b

მდგომარეობებს.

პირველი მდგომარეობისათვის **სკან-ოპერაციის** გამოყენება წარუმატებლად სრულდება. მიუხედავად ამისა, პრედიკატორ-ოპერაციის გამოყენება მეორე მდგომარეობაზე შეძლებად იძლევა ორ ახალ მდგომარეობას, კერძოდ:

a a a. b b b და a a a. S b b b

რომელთაგანაც ორივეს **სკან-ოპერაცია** წარმატებით აწყვილებს დასამუშავებელ სტრიქონთან.

წერტილი კიდევ ერთხელ წაიძრა მარჯვნივ ერთი პოზიციით და წარმოქმნა დასამუშავებელი სტრიქონი

a a a b. b b

და ახალი

a a a b. b b და a a a S. b b b

მდგომარეობები. ამჯერად **სკან-ოპერაცია** წარმატებულია პირველ მდგომარეობაზე.⁸ წერტილი კიდევ ერთხელ წაიძრა მარჯვნივ და შეძლებად მოგვცა

a a a b b. b

⁸ ამ სტადიაში მეორე გამოსახულებაც ასევე ჯერ კიდევ აქტიურია, რადგან ის ნებას რთავს მომდევნო პრედიკატორ-ოპერაციას, რომელიც შეძლებად გვაძლევს ახალ მდგომარეობებს

a a a a. b b b b b

a a a a. S b b b b b

თუმცა ეს მდგომარეობები მიმყოლი სკან-ოპერაციით წარუმატებლად სრულდებიან დასამუშავებელ

a a a b. b b

სტრიქონზე.

დასამუშავებელ სტრიქონი და

a a a b b. b

მდგომარეობა.

ის, რომ ახალი მდგომარეობა შეიცავს მხოლოდ ტერმინალურ სიმბოლოებს, განაპირობებს იმას, რომ მხოლოდ სკან-ოპერაცია შეიძლება იქნეს გამოყენებული. ამასთან, ეს ოპერაცია ამ შემთხვევაში რეზულტატურია და შედეგად იძლევა

a a a b b. b

დასამუშავებელ სტრიქონისა და

a a a b b. b.

მდგომარეობას. დასამუშავებელი სტრიქონის ბოლოს განთავსებული წერტილი გვანიშნებს ანალიზის წარმატებულ დასასრულზე.

ჩვენ ვნახეთ, რომ **ერლეის** ალგორითმი იყენებს გრამატიკის წესებს, მაგრამ არა პირდაპირ. თუმცა, ალგორითმი წარამატებით შლის გრამატიკის წესებს მათ შემადგენელ ძირითად ელემენტებად, როთაც რიგითობა დგინდება დასამუშავებელ სტრიქონში ტერმინალური სიმბოლოების თანამიმდევრობით - და არა გრამატიკული წესების სისტემის ლოგიკური დერივაციული რიგით. ამგვარად, **ერლეის** ალგორითმსა და მის მიერ გამოყენებულ კონტექსტისგან თავისუფალ **PS**-გრამატიკას შორის კავშირი არ არის ტრანსფორმული ტიპის.

კონტექსტისგან თავისუფალი **PS**-გრამატიკისთვის, **ერლეის** ალგორითმის გარდა, არსებობს პარსერების საკმაოდ დიდი ოდენობა, მაგ.: **CYK** ალგორითმი (იხ. ჰოფკროფტი & ულმანი, 1979, გვ. 139-141) და **chart** პარსერი (იხ. კეი, 1980). ყველა ეს პარსერი აინტერპრეტირებს ნებისმიერ კონტექსტისგან თავისუფალ **PS**-გრამატიკას და აანალიზებს მათთან მიკავშირებულ ნებისმიერ ენას. თუმცა, არცერთი ეს პარსერი არ არის ტრანსპარენტული ტიპის.

პარსერები მსგავსად ერლეის და **CYK** ალგორითმებისა, ანდა ჩარტ პარსერისა კონტექსტისგან თავისუფალი **PS**-გრამატიკის წესებს უშუალოდ ვერ ამუშავებენ და ითხოვენ მასიურ შუალედურ სტრუქტურებს (მაგ. როგორებიცაა: მდგომარეობათა სიმრავლე, ცხრილები, დიაგრამები და. ა. შ.) იმ მიზნით, რომ შესაბამისობაში მოიყვანოს განსხვავებები გრამატიკული სისტემისა და პარსერის ალგორითმის მიერ წარმოებულ დერივაციებს შორის: რაც, თავის მხრივ, სირთულეებს უქმნის მომხმარებელს, მაგალითად გრამატიკის დამწერს, რომელსაც წარმოდგენაც კი არა აქვს პარსერის პროცედურულ რეტრინაზე. მიუხედავად ამისა, თუ ამ მიღეობით ხერხდება არსებული განსხვავებების სრული კორელაცია, ანუ თუ **PS**-გრამატიკა აღეკვატურად არის აღწერილი, ასეთი გადაწყვეტები გამართლებულად მიიჩნევა.

ამის საპირისპიროდ, თუ პარსერი გამოყენებულია აღწერითად აღეკვატურ განვითარებად გრამატიკაში, ტიპობრივი ტრანსპარენტულობის ნაკლი ძლიერ აფერხებს გრამატიკის შემდგომი დახვეწისა და მაღალ საფეხურზე აყვანის საქმეს. მაგალითად, თუ მოხდა შეცდომა იმის გამო, რომ ანალიზატორმა ვერ იპოვა კანონიერი (სწორი) გრამატიკული სტრუქტურა სწორად ფორმირებული დასამუშავებელი გამოსახულებისათვის, მაშინ ეს შეცდომა მოძიებული უნდა იქნას **PS**-გრამატიკის წესთა სისტემაში.

ხდება, რომ პრაქტიკული მიზნით გამოყენებადი კონტექსტისგან თავისუფალი **PS**-გრამატიკის წესების რაოდენობა ხშირად რამდენიმე ათასს აჭარბებს. ამ შემთხვევაში უდიდესი დახმარება

იქნებოდა, თუ მოხერხდებოდა შეცდომის ლოკალიზება პარსერის კვალის დახმარებით. თუმცა, როცა პარსერი პირდაპირ ვერ იყენებს **PS**-გრამატიკის წესებს, მისი კვალი (მდგომარეობათა განაწერი) ისევე არაწაკითხვადია, როგორც ასემბლერის კოდის (იხ. 9.3.4) და, შესაბამისად, მისი (ე.ი. პარსერის ტრასირების) ევრისტიკული მნიშვნელობა მეტად მცირდება.

რა თქმა უნდა, შეცდომების ლოკალიზების მიზნით, შესაძლოა დაიწეროს მეორადი პარსერი ძირითადი პარსერის არაწაკითხვადი სტრუქტურის წაკითხვად ფორმაში სათარგმნად: ტრანსპარენტული ტიპის **LA**-გრამატიკის მსგავსი გრამატიკებისთვის როგორც შუალედური სტრუქტურების, ისე მათი ადამიანთათვის წაკითხვად ფორმაში რეინტერპრეტატორების კონსტრუირება დაკავშირებულია დიდ ძალის ხმელებთან და დანახარჯებთან. ამასთან, ირკვევა, რომ სასურველი შედეგის მისაღწევად არასაკმარისია პროგრამისტული დამუშავება და რომ რიგ შემთხვევებში საჭირო ხდება მეხსიერების გაზრდა.

9.4 შემავალ-გამომავალი ეპიგვალენცია მსხველ-მოლაპარაკე სისტემასთან

ნეიტივისტების მხარდამჭერებს ჰქონდათ მცდელობა მიეჩქმალათ **PS**-გრამატიკის სტრუქტურული სისუსტეები, როგორც პროგრამირების პრობლემა, რომელთა გადაწყვეტა კომპიუტერულ მეცნიერებათა ამოცნაა. ისინი ამბობდნენ, რომ მათი ნამდვილი მიზნი მოლაპარაკე-მსმენელის ბუნებრივი ენობრივი ცოდნის ზოგადი გაანალიზებაა და რომ ამ მიზნით განსაზღვრულ კვლევებზე იმ ვიწრო პროგრამისტული ხასიათის პრობლემებს, რომლებზედაც საუბარია, არ აქვთ არანაირი უარყოფითი გავლენა.

თუმცა, ზემომოყვანილი მსჯელობა არამართუბულია. რადგანაც, ცხადია, რომ თანდაყოლილი სტრუქტურების ფორმები განაპირობებენ მათ ფუნქციებს (იხ. 4.5.3). ამასთან, ცხადია ისიც, რომ თანდაყოლილი ენობრივი შესაძლებლობების აღწერა აუცილებლად ითხოვს აღწერაში გამოყენებული ფორმალიზმის შემავალ-გამომავალ ეკვივალენციას აღსაწერი ენის მსმენელ-მოლაპარაკე სისტემასთან. არადა, მოლაპარაკე-მსმენელი სისტემის შემავალ-გამომავალ პირობებს ვერც **PS**-გრამატიკა აკმაყოფილებს და ვერც **PS**-გრამატიკის ის პარსერები, რომლებიც ქვემოთაა მოყვანილი.

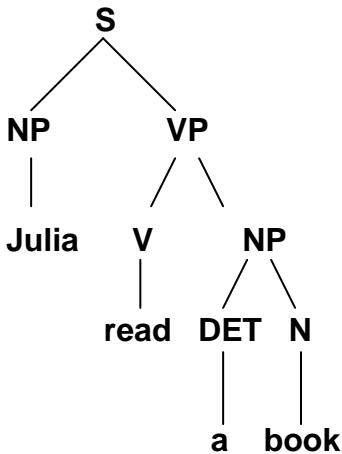
9.4.1 კონტექსტისგან თავისუფალი **PS**-გრამატიკა ინგლისური ენის მარტივი ფინალებისათვის

1. **S** → **NP VP**
2. **NP** → **DET N**
3. **VP** → **V NP**
4. **NP** → **Julia**
5. **V** → **read**
6. **DET** → **a**
7. **N** → **book**

9.4.2 განაწერით აღწერილ ანალიზის ხეს მიკავშირებულ გამოყვანაში 9.4.1 ჯერ წესი 1 ანაცვლებს (**Substitutes**) საწყის **S** სიმბოლოს **NP** (**noun**) და **VP** (**verb**) ფრაზებით (**phrase**). შემდეგ წესი 4 **NP** კვანძს ანაცვლებს სიტყვით **Julia**, და წესი 3 **VP** კვანძს ცვლის **V** (**verb**) და **NP** კვანძებით. შემდეგ წესი 5 **V** კვანძს ცვლის სიტყვით **read**, და წესი 2 **NP** კვანძს

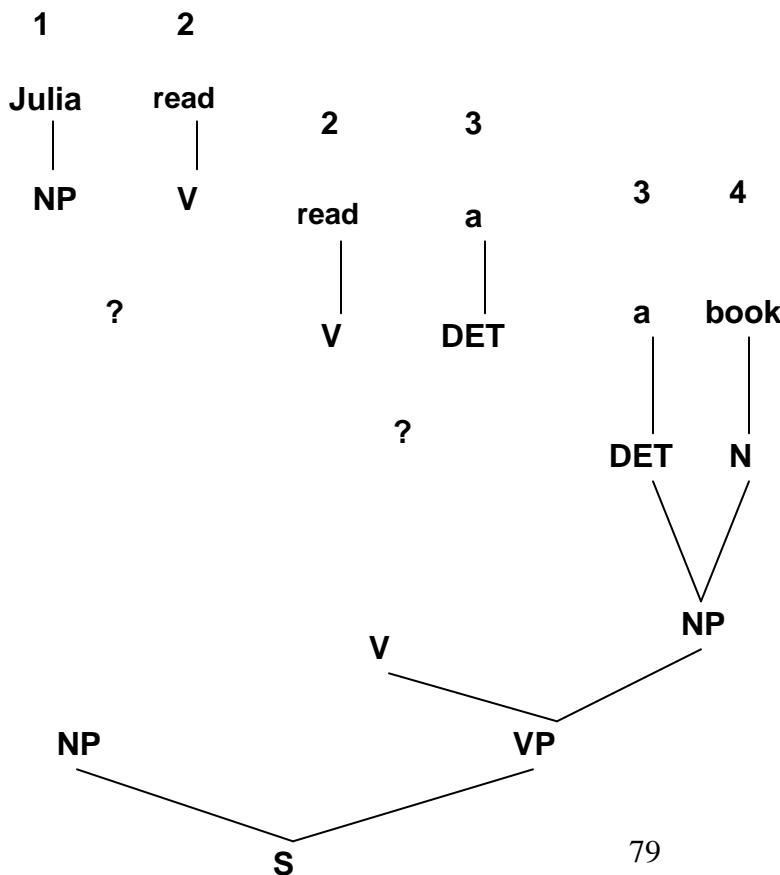
ცვლის **DET** (**determiner**) და **N** (**noun**) კვანძებით. საბოლოოდ, წესი 6 ცვლის **DET** კვანძს სიტყვით **a**, და წესი 7 ცვლის **N** კვანძს სიტყვით **book**.

9.4.2 PS-გრამატიკული ანალიზი (ქვემოთ მიმართული გამოყვანა)



ლოგიკური თვალსაზრისებით ეს გამოყვანა ბუნებრივად იმზირება. მიუხედავად ამისა, ის არის ცხად წინააღმდეგობაში ბუნებრივი ენის დროში-წრფივ სტრუქტურასთან. შემავალ-გამომავალი პირობების თვალსაზრისით **PS**-გრამატიკებსა და მოლაპარეკე-მსმენელის ენობრივ ფუნქციონირებას შორის არსებული განსხვავებები დემონსტრირებულია ქვემოთ მოყვანილი მცირე მონახაზით, რომელიც 9.4.1 **PS**-გრამატიკაზე დამყარებული დროში-წრფივი ანალიზის გაკეთების მცდელობაა.

9.4.3 დროში-წრფივი ანალიზის ჩატარების მცდელობა **PS**-გრამატიკაში



წინადადების - **Julia read a book** წაკითხვის ბუნებრივი თანამიმდევრობა დროში-წრფივია, ე.გ. სიტყვები იკითხება მარცხნიდან მარჯვნივ ერთი მეორეს მიყოლებით.⁹ მას შემდეგ, რაც პირველი ორი სიტყვა **Julia** და **read** ჩანაცვლდა **NP** და **V** კვანძებით (9.4.1 სისტემის 4 და 5 წესების შესაბამისად), მკითხველის გრამატიკული ალგორითმი კვანძების **NP** და **V** მიმდევრობის უფრო მაღალი საფეხურის კვანძით ჩასახაცვლებლად გთქას **PS**-წესს. თუმცა, რადგან ასეთი სახის წესის არსებობას ვერ უზრუნველოფე 9.4.1 გრამატიკა, ქვეხე, რომელიც შედგება პირველი სიტყვისა და მისი კატეგორიისგან, დროებით გვერდზე უნდა გადაიდოს მოგვიანებითი გამოყენებისთვის.

შემდეგ, მკითხველის გრამატიკული ალგორითმი 6-ე წესის გამოყენებით ანაცვლებს **a** მესამე სიტყვას მისი **DET** კატეგორიით, და ცდილობს მეორე სიტყვის **V** კატეგორის კომბინირებას ამ მესამე სიტყვის **DET** კატეგორიასთან უფრო მაღალი კვანძში. კვლავაც, 9.4.1 გრამატიკაში შესაბამისი წესის არ არსებობის გამო, მეორე სიტყვის შესაბამისი ქვეხე დროებით გვერდზე უნდა გადაიოდს შემდგომი მოგვიანებითი გამოყენებისთვის.

ბოლოს, მკითხველის გრამატიკული ალგორითმი ცდილობს მესამე სიტყვის კატეგორიის კომბინირებას შესატანი წინადადების ბოლო, ანუ მეოთხე სიტყვის კატეგორიასთან. ამ შემთხვევაში 9.4.1 გრამატიკა უზრუნველყოფს შესაბამისი წესის არსებობას, სახელდობრ, ეს არის წესი 2, რომელიც **DET N**-კვანძების მიმდევრობას უფრო მაღალ **NP** კვანძში აერთიანებს.

ამის შემდეგ მკითხველის გრამატიკული ალგორითმი კიდევ ერთხელ ცდილობს იმ მეორე **read** სიტყვის **V** კატეგორიის კომბინირებას, რომელიც მოგვიანებითი გამოყენებისთვის დროებით გერერდზე იყო გადადებული: ამჯერად 9.4.1 გრამატიკის 3 წესით იგი კომბინირდება ახლად გამოყვანილ **NP** კვანძთან და შედეგად უფრო მაღალ **VP** კვანძს იძლევა. საბოლოოდ, მკითხველის გრამატიკული ალგორითმი კიდევ ერთხელ უცდება **NP** კატეგორიის პირველი სიტყვა **Julia**, რომელიც იმ დროიდან მოყოლებული, რაც ამ ოთხსიტყვიანი წინადადების ანალიზი დაიწყო, გვერდზე იყო გადადებული, დააკომბინიროს ამ დროისათვის ახლად გამოყვანილ **VP** კვანძთან. ამ შემთხვევაში წესი 1-ის ძალით მცდელობა წარმატებულია და ანალიზიც სრულდება.

9.4.3 ანალიზის მონახაზი გვიჩვენებს, რომ შემადგენელ-სტრუქტურებზე დაფუძნებულ კონტექსტისგან თავისუფალ **PS**-გრამატიკებში დროში-წრფივი ინტერპრეტაცია შეუძლებელია: წინადადების პირველმა სიტყვამ სრული ანალიზის აგებისას უკანასკნელად გასაანალიზებელი მოცემულობის ადგილი დაიკავა. **PS**-გრამატიკების ტიპობრივი არატრანსფორმულობა შემაფერხებელია არა მხოლოდ ბუნებრივი და ხელოვნური ენების ავტომატურად მაანალიზებელი სისტემების აგების თვალსაზრისებით. ეს ამავე დროს შეუთავსებლობაშია ეჭვგარეშე ძირეულ ბუნებრივ პრინციპთან: ფორმები განაპირობებენ ფუნქციებს – ეს განსაკუთრებით ძალისმიერია ბუნებრივი ფორმებისათვის.

ახლა რაც შეეხება იმ სტანდარტულ პასუხს, რომ ნეიტრივისტური გრამატიკა **მიზნად არ ისახავდა**¹⁰ მოლაპარაკე-მსმენელის საკომუნიკაციო ურთიერთობების აღწერას და რომ ის

⁹ ტრადიციული ბერძნულ-რომაული წერის შეთანხმების მიხედვით.

¹⁰ ნ. ჩოშკი დაუღალავად აკონტარებას, რომ მისი ნეიტრივისტური პროგრამის მიზანი არ იყო მოლაპარაკე-მსმენელს შორის კომუნიკაციის პროცედურის მოდელირება. მაგალითისთვის იხილეთ ჩოშკი 1965, გვ.9.

გამიზნული იყო მოლაპარაკე-მსმენელის თანდაყოლილი ცოდნის აბსტრაქტულ, ანუ მხოლოდ იმ დონეზე აღსაწერად, რომელიც არ არის დამოკიდებული ენის კონკრეტულ კომუნიკაციაში გამოყენებისგან – ეს არადამაჯერებელია, რადგანაც ნეიტივისტური თეორიული მოსაზრებების დამაჯერებლობისთვის აუცილებელია, რომ მისი ძირითადი სტრუქტურები თავსიებადნი (**compatible**) იყვნენ ბუნებრივი ენის კომუნიკაციის მექანიზმთან. და ეს კი, თუ ნეიტივისტური თეორია **PS**-გრამატიკულ ფორმალიზმს დაეყრდნობა, შეუძლებელია.

9.5 გრამატიკის უპარისობა კონვენციის მისაღწევად

მეცნიერული თვალსაზრისების ისტორიულ ჭრილში **PS**-გრამატიკაზე დამყარებული ნეიტივიზმი შეიძლება შეფასდეს სხვადასხვაგვარად. – დადებითი მხარე: განუწყვეტელი რევიზირების გზით ნეიტივიზმა შეძლო ის... რომ უკვე ორმოც წელზე მეტია რაც ინარჩუნებს ინტერესს, არ კარგავს გავლენას, იერთებს ახალ მიმღებრებს და შედა დაგას იყენებს არა დაცემისთვის, არამედ განვითარებისთვის. უარყოფითი მხარე: ეს სახელმძღვანელო წიგნი დასტურია ამ საკითხის ირგვლივ არასაკმარისი კონვერგენციისა.¹¹

ეს არაკონვერგენციული დამოკიდებულებები თავს იჩენს სულ ცოტა ოთხი მიმართულებით მაინც. პირველი: ნაცვლად შედეგების გაერთიანებისა და გაძლიერებისა განუწყვეტლივ იგება ახალ-ახალი წარმოებული სისტემები. (იხ. ჯ. დ. მაკოლის (**J.D. McCawley**) მიერ გაკეთებული აღწერა, რომელიც 1982 წელს გამოქვეყნდა წიგნში სათაურით „ოცდაათი მილიონი გრამატიკული თეორიის დასახელება“¹². მეორე: ყოველ ჯერზე, როცა დამატებითი მექანიზმები (მაგ.: ტრანსფორმაციები, მეტაწესები და ა.შ., განხილული როგორც აღწერითად აუცილებელი) შემოჰქონდათ, მათემატიკური და გამოთვლითი მახასიათებლები საგრძნობ დეგრადაციას განიცდიდნენ კონტექსტისგან თავისუფალი **PS**-გრამატიკის ელემენტარული ქვეფორმალიზმის შესაბამის მახასიათებლებთან შედარებით. მესამე: **PS**-გრამატიკის ფარგლებში ბუნებრივი ენის აღწერის ყოველი მცდელობა საბოლოო ჯამში პრობლემების სახით იძლეოდა აღწერითი ტიპების სიუხვითა და მრავალფეროვნებით განპირობებულ სირთულეებს.¹³ მეოთხე: ბუნებრივი ენის დამამუშავებელი

¹¹ მეცნიერების ისტორიაში კვლევათა ველი განხილულია როგორც დადებითიად განვითარებადი თუ მისი სხვა და სხვა არები იკრიბება. ე.რ. თუ გაუმჯობესება ერთ არეში იწვევს გაუმჯობესებას სხვა არებში. ამის საპირისარიოდ, კვლევათა ველი განხილულია როგორც უარყოფითად განვითარებადი თუ გაუმჯობესება ერთ არეში იწვევს გაუარესებას სხვა არებში.

¹² ეს რაოდენობა გამოთვლილია ენის ნეიტივისტური თეორიის ფარგლებში ღია ალტერნატივების საფუძველზე. მაკოლის გამოთვლით დღეში საშუალოდ 2055 განსხვავებული გრამატიკული თეორია იქმნებოდა, ანუ თითო ახალი თეორია ყოველ 42 წაში, რაც მთლიანობაში 40 წელი გრძელდებოდა (1957 წლიდან - ანუ, სინტაქტიკური სრტუქტურების გამოჩენიდან - 1997 წლამდე).

¹³ **PS**-გრამატიკა შემადგენელი სტრუქტურის ანალიზთან კომბინაციაში აჩვენებს აღწერით ეიფორიას, მაგ., ამას გვიჩვენებს გერმანულის ენის ისეთი ძირითადი კლაუზა, როგორიცაა **Peter hat die Tür geschlossen**, რომელიც მისი დისკონტინუალური ელემენტების გამო კანონიერი სტრუქტურული ანალიზით ვერ ჯდება კონტექსტისგან თავისუფალი **PS**-გრამატიკის ფარგლებში (იხ.: 8.5.1 და 8.5.2). ამ და სხვა პრობლემების გადასაწყვეტად კონტექსტისგან თავისუფალ **PS**-გრამატიკას დამატა ტრანსფორმაციები (იხ. 8.5.3). ამან შედეგად მოგვცა სიძღვირის გართულების ტიპის ბევრი პრობლემა. მაგალითისთვის, წარმოიშვა კითხვა, გერმანულის ძირითადი კლაუზები უნდა იწარმოოს ტრანსფორმაციულად დამოკიდებული წინადადების (მაგ.: **weil Peter hat die Tür geschlossen hat**) ღრმა სტრუქტურისგან თუ პირიქით. ძირითადი კლაუზების ტრანსფორმაციული წარმოება დამოკიდებული კლაუზისაგან მოტივირებულია დამოკიდებული კლაუზის შემადგენელი სრტუქტურის შესაბამისობით, როცა დამოკიდებული კლაუზის ტრანსფორმაციული წარმოება ძირითადი კლაუზისგან მოტივირებულია იმ გრძნობით, რომ ძირითადი კლაუზა უფრო მნიშვნელოვანია (იხ. ე. ბახი (**E. Bach**), 1962 და მ. ბერვიში (**M. Bierwisch**), 1963). გერმანული ენის ძირითადი და დამოკიდებული წინადადების განხილვა ტრანსფორმაციის გარეშე იხილეთ 18-ე თავში, კურძოდ 18.5 ქვეთავში.

პრაქტიკულად გამოყენებადი სისტემები ან მხოლოდ ზედაპირულად იყენებენ ნეიტივიზმის თეორიულ კონსტრუქციებს, ან საერთოდ უარყოფენ მათ.

ამასთან, ამ კონვერგენციული დამოკიდებულებების ნაკლულობას ორი ძირითადი მიზეზი აქვს: პირველი: ნეიტივიზმი არ არის ემპირიულად სრულად განსაზღვრული, რადგან ის არ შეიცავს კომუნიკაციის ფუნქციონალურ თეორიას. მეორე: ნეიტივიზმის მიერ არჩეული **PS**-გრამატიკული ფორმალიზმი არაა თავსებადი მოლაპარაკე-მსმენელი სისტემის შემავალ-გამომავალ პირობებთან.

ქვემოთ დამწყები ლინგვისტებისათვის მოცემულია სხვადასხვა ფორმალიზმებზე ხანგრძლივი მუშაობის შედეგად მიღებული გამოცდილების საფუძველზე გაკეთებული რჩევები¹⁴: (i) თქვენ უნდა გახსოვდეთ, რომ თქვენი მიზანია ბუნებრივი კომუნიკაციის მექანიზმის მოღელირება, ამასთან, საჭიროა (ii) თქვენს მიერ შემუშავებული მოღელების გამოთვლითი ვერიფიცირება, და (iii) მათი უტილიზება ისეთ პრაქტიკულ გამოყენებებში, როგორიცაა ადამიანი-მანქანის კომუნიკაციები. – თქვენი მოღელების სასაფუძვლო გრამატიკული ფორმალიზმი არ უნდა აირჩიოთ ნაჩქარევად და გაუაზრებლად, არამედ თქვენი არჩევანი უნდა იყოს ცოდნასა და განათლებაზე დამყარებული ღრმად გაცნობიერებული არჩევანის შედეგი. – **PS**-გრამატიკის თვისებები შეჯამებულია ქვემოთ.

9.5.1. **PS**-გრამატიკის თვისებები

- მათემატიკური:

პრაქტიკული ღირებულების მქონე პარსერ ალგორითმები არსებობს მხოლოდ კონტექსტისგან თავისუფალი **PS**-გრამატიკისთვის. მათ აქვთ საკმაოდ დაბალი სირთულის დონე (**n³**), მაგრამ მათი წარმოქმნითი უნარი ბუნებრივი ენების დასამუშავებლად არასაკმარისია. ამასთან, როგორც წესი, ბუნებრივი ენის აღწერის მიზნით ფორმალიზმის წარმოქმნითი უნარის გაფართოება იძლევა სირთულის ისეთ მაღალ ხარისხს (არაამოხსნადს, ან ექსპონენციალურს), რომ არ შეიძლება მისთვის არსებობდეს რაიმე პრაქტიკული ღირებულების მქონე პარსერი.

- გამოთვლითი:

PS-გრამატიკა არ არის ტრანსფარენტული ტიპის. ეს ხელს უშლის გრამატიკაში შეცდომების გამოსწორებისა და სისტემის უფრო მაღალ დონეზე აყვანის მიზნით პარსერების ავტომატური კვალის გამოყენებას. გარდა ამისა, გრამატიკასა და პარსერ ალგორითმს შორის არაპირდაპირი კავშირი მოითხოვს ძვირადღირებულ რუტინულ შრომასა და მოცულობითი შუალედური სტრუქტურების გამოყენებას.

- ემპირიკული:

PS-გრამატიკის ჩანაცვლებაზე-დაფუძნებული გამოყვანის რიგი შეუთავსებელია ბუნებრივი ენის დროში-წრფივ სტრუქტურასთან.

¹⁴ დარწმუნებით გამოწვეული ერთი ფორმალიზმის ცვლა მეორით შესაძლოა ძვირადღირებული აღმოჩნდეს. ეს ნიშნავს, რომ განსაზღვრულ არეში ექსპერტის ფრთხილად გაზრდილი სტატუსი ზოგჯერ იწვევს წინა სამუშაოების გამოსწორებას, და შესაძლოა იქინიოს დიდი სოციალური გამოძახვილი როგორც მის ჯგუფში, ისე მის ქვეყანაშიც და საზღვარგარეთაც. თუმცა, მეცნიერებათა ველები, როგორც ცნობილია, ზოგჯერ მოულოდნელად იცვლება სხვადასხვა მიმართულებით. – შექმნილი კვლევითი ჯგუფები მოულოდნელად აწყდებიან მათსავე რადიკალურად შემცირებულ ფონდებს, ქმედებებსა და წევრებს. როცა ეს ნდება, ერთი ფორმალიზმების ცვლა შესაძლოა გადაიქცეს გადარჩენის აუცილებელ საშუალებად. ამ შესაძლებლობების გაშუქებისას ველებაზე ხანგრძლივი სტრატეგია სამუცნიერო არჩევანთა შეფასება და გრამატიკული ფორმალიზმის რაცონალური არჩევაა.. როგორც ყველა კაზგ მეცნიერებაში, ლინგვისტური კვლევაც, უნდა იმართებოდეს კვლევის ველის ფართო და ღრმა ცოდნაზე დამყარებული რწმენით.

რადგან ბიდირექციული **C**-გრამატიკა სუსტ ეკვივალენციაშია კონტექსტისაგან თავისუფალ **PS**-გრამატიკასთან (იხ. 9.2), **C**-გრამატიკა ვერ გამოდგება **PS**-გრამატიკის აღტერნატივად. ამგვარად, ჯერ კიდევ მიმდინარეობს ინტენსიური სამეცნიერო კვლევები შემდეგი თვისებების მქონე ძირული ფორმალიზმების მოძიების მიზნით.

9.5.2 მოთხოვები დარმომამხელ გრამატიკულ ფორმალიზმები

1. გრამატიკული ფორმალიზმი უნდა იყოს მათემატიკურად მკაცრად განსაზღვრული, ისე რომ
2. იძლეოდეს ხელოვნური და ბუნებრივი ენების ცხადი, დეკლარაციული აღწერის საშუალებას.
3. ეს ფორმალიზმი უნდა იყოს რეაურსიული (და ამდენად ამოხსნადი), ასევე
4. ის უნდა იყოს ტიპობრივად ტრანსპარენტული მის პარსერსა და წარმომქმნელთან მიმართებაში.
5. ამ ფორმალიზმით უნდა ისაზღვრებოდეს თვისობრივად განსხვავებული ენათა კლასები და ეს უნდა ხდებოდეს მის წესთა სისტემაზე განსაზღვრული სტრუქტურულად ცხადი აკრძალვებით (ანალოგიურად - მაგრამ ორთოგონალურად **PS**-გრამატიკული იერარქიისა),
6. რის შედეგადაც იერარქია უნდა შეიცავდეს დაბალი - სავარაუდოდ წრფივი სირთულის წარმოქმნითი შესაძლებლობის ისეთ ენათა კლასს, რომელიც საკმარისი იქნება ბუნებრივი ენის სრული აღწერისთვის.
7. ფორმალიზმი უნდა იყოს შემავალ-გამომავალ ეკვივალენციაში მოლაპარაკე-შსმენელის სისტემასთან (ანუ უნდა იყენებდეს დროში წრფივ გამოყვანებს).
8. ფორმალიზმს უნდა ჰქონდეს წარმოებისა (შინაარსთა ენობრივ გამოსახულებებზე გადასახვის აზრით) და ინტერპრეტირების (ენობრივ გამოსახულებათა ფორმების მნიშვნელობებში გადასახვის აზრით) უნარები.

შემდეგი თავები გვიჩვენებენ, რომ **LA**-გრამატიკა თავს ართმევს ამ მოთხონებს. საპირისპიროა ვითარებაა **C**- და **PS**- გრამატიკულ ფორმალიზმებში, რომლებშიც გამოყვანები დაფუძნებულია შესაძლო ჩასმების (**possible substitutions**) პრინციპზე. **LA**-გრამატიკაში გამოყვანები დაფუძნებულია შესაძლო გაგრძელებათა (**possible continuations**) პრინციპზე.

სავარჯიშოები

9.1 ქვეთავი

1. რის საფუძველზე განვითარდა ტერმთა პარსერები და, საზოგადოდ, რა არის პარსერის ფუნქციები?
2. როგორ არიან მორფოლოგიური, სინტაქსური და სემანტიკური პარსერები ერთმანეთთან დაკავშირებულნი და რა განასხვავებს მათ?

3. აღწერეთ წარმომქმნელი გრამატიკის პარსერში გამოყენების ორი სხვადასხვა გზა და შეაფასეთ ეს აღტერნატივები.
4. განსაზღვრეთ ცნებები: დეკლარაციული, პროცედურული. როგორ აისახებიან ისინი პარსერებში?
5. შესაძლებელია თუ არა სხვადასხვა პარსერის დაწერა ერთი და იგივე გრამატიკული ფორმალიზმისათვის?

9.2 ქვეთავი

1. აღწერეთ კონტექსტისგან თავისუფალი სტრუქტურა ბუნებრივ ენებში.
2. არის თუ არა კონტექსტისგან თავისუფალი სტრუქტურები ბუნებრივ ენაში?
3. რას ვიღებთ იმის დაშვებით, რომ ბუნებრივი ენა არ არის კონტექსტისგან თავისუფალი? არის თუ არა პასუხი დამოკიდებული იმაზე, თუ რომელ გრამატიკულ ფორმალიზმს ვიყენებთ? იქნებოდა თუ არა ბუნებრივი ენის წრფივ დროში გაანალიზება შესაძლებელი, ის რომ კონტექსტზე დამოკიდებული ყოფილიყო?
4. ახსენით შესაძლო ეკვივალენტური დამოკიდებულებები გრამატიკის ორ ფორმალიზმს შორის.
5. რა კავშირშია ბიდირექციული **C**-გრამატიკა **PS**-გრამატიკულ ტიპებთან?
6. არიან თუ არა ხელოვნური ენები მათ ფორმალურ გრამატიკაზე დამოკიდებულინი?
7. არიან თუ არა ენობრივი კლასები გრამატიკის ფორმალიზმზე დამოკიდებულინი?
8. რა გავლენას ახდენს ენობრივი კლასის სირთულის დონე მისთვის პრაქტიკული ანალიზატორის შექმნის შესაძლებლობაზე?
9. რა არის ენის ინდივიდუალური სირთულის ხარისხი და როგორ არის ის განსაზღვრული?

9.3 ქვეთავი

1. ახსენით ტიპობრივი ტრანსფორმტულობის ცნება.
2. რა მიზნით ააგო პოსტმა 1936 წელს თავისი მაწარმოებელი სისტემა?
3. როდის არიან გრამატიკული ფორმალიზმი და პარსერი შემავალ-გამომავალ ეკვივალენციაში?
4. რა განსხვავებაა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ქვემოთ მიმართულ და ზემოთ მიმართულ გამოყვანებს შორის?
5. რატომ არ არის კონტექსტისგან თავისუფალი **PS**-გრამატიკა შემავალ-გამომავალ ეკვივალენციაში მისსავე პარსერებთან? თქვენი ახსნები დააფუქნეთ ქვემოთ მიმართულ და ზემოთ მიმართულ გამოყვანებს.
6. ახსენით ერლეის ალგორითმის ფუნქციონირება **a^kb^k** ენის გამოსახულებების გამოყენებით. როგორ აკომპესირებს ერლეის ალგორითმი **PS**-გრამატიკის ჩანაცვლებაზე დამყარებული გამოყვანითი რიგის სირთულეებს?
7. მიმოიხილეთ ერლეის ალგორითმისათვის კონტექსტისგან თავისუფალი **PS**-გრამატიკის წყვილ-ინვერსირებადი სტრუქტურის გადამწყვეტი მნიშვნელობები.
8. არის თუ არა შესაძლებელი **a^kb^kc^k** ენის გაანალიზება ერლეის ალგორითმით?
9. არსებობს თუ არა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ტრანსპარენტული პარსერები?
10. დაასახელეთ ორი პრაქტიკული მოუხერხებლობა იმ ავტომატური ანალიზისა, რომელიც არაა ტრანსფორმტული ტიპის.

9.4 ქვეთავი

- შემავალ-გამომავალი ეკვივალენტურობის ცნების გამოყენებით ახსენით, რატომაა **PS**-გრამატიკაზე დამყარებული ენის ნეიტივისტური თეორია შეუთავსებელი პრინციპთან – ფორმები განაპირობებენ ფუნქციებს (**form follows function**).
- წარმოაჩინეთ მაგალითით, რომ **PS**-გრამატიკის წარმოების რიგი შეუთავსებელია ბუნებრივი ენის დროით-ხაზოვან სტრუქტურასთან.
- ახდენს თუ არა დამატებითი ტრანსფორმაციული კომპონენტი (იხ. 2.4 და 8.5) **PS**-გრამატიკის წარმოების რიგსა და ბუნებრივი ენის დროში-წრფივ რიგს შორის შეუთავსებლობის შემცირებას?

9.5 ქვეთავი

- ახსენით კონვერგენციის ცნება და ამ ცნების მნიშვნელობა მეცნიერების ისტორიაში.
- როგორ იჩენს თავს ნეიტივიზმის ისტორიულ განვითარებაში კონვერგენციის ნაკლი და რა არის ამის მიზეზები?
- როგორ გამოთვალა მაკოლიმ თავისი წიგნის სათაურში „გრამატიკის ოცდაათი მილიონი თეორია“ ასახული რიცხვი? არის თუ არა ეს რაოდენობა ლინგვისტიკაში არსებული დადებითი ძვრების მანიშვნებელი?
- რატომ შეიძლება აღმოჩნდეს გრამატიკული ფორმალიზმის შეცვლა ძვირადლირებული?
- რა კავშირს ხედავთ, ერთი მხრივ, ‘დესკრიფიული ეიფორიებსა’ და ‘რაოდენობრივ შეფერხებებს’ და, მეორე მხრივ, ახალ-ახალი წარმოებული ფორმალიზმების განუწყვეტელი წარმოების მაღალ მატერიალურ სირთულეებს შორის?
- აღწერეთ **PS**-გრამატიკის მათემატიკური, გამოთვლითი და ემპირიკული თვისებები.
- რა ნაკლები უნდა აღმოითხვრას წარმომქმნელ გრამატიკებში რომ იგი ბუნებრივი ენის გამოთვლითი ანალიზისთვის თავსებადი გახდეს?

10 თავი მარცხნივ ასრულებადი გრამატიკა

წინა 7-ე – 9-ე თავებში **C**-გრამატიკისა (1929) და **PS**-გრამატიკის (1935) ისტორიული ფორმალიზმების ფარგლებში აღიწერა ხელოვნური და ბუნებრივი ენების ძირითადი საანალიზო ცნებები. მომდევნო 10-ე – 12-ე თავებში ეს ცნებები გამოყენებული იქნება **LA**-გრამატიკად წოდებულ (1985)¹ მესამე ელემენტალურ ფორმალიზმში. ბუნებრივი ენის ანალიზის ეს შედარებით ახალი ფორმალიზმი არ არის ნასესხები სამეცნიერო კვლევის სხვა დარგებიდან, არამედ თავიდანვე იყო ჩაფიქრებული როგორც ისეთი დროში წრფივი ტრანსფარენტული ტიპის ალგორითმი შემავალ-გამომავალ ეკვივალენციაში იქნებოდა მოლაპარაკე-მსმენელის სისტემასთან.

10.1 ქვეთავში ახსნილია, თუ როგორ მოდელირდება ბუნებრივი ენის დროში წრფივი სტრუქტურა მარცხნივ ჯგუფდებადი დერივაციული რიგით, განსაზღვრულია შესაძლო გავრძელებების პრინციპი და **C**-, **PS**- და **LA**- გრამატიკებისთვის გაშუქებულია ზოგადი კავშირი წესის ფორმატსა და

¹ ოფიციალური პუბლიკაცია არის **Hausser**, ჟურნალში თეორიული კომპიუტერული მეცნიერება (**TCS**).

კონტექტუალურ გამოყვანით დალაგებას შორის. 10.2 ქვეთავში მოცემულია **LA**-გრამატიკის ალგებრული განსაზღვრება. 10.3 ქვეთავში წარმოდგენილია დროში წრფივი დერივაციის ფორმატები. 10.4 ქვეთავში კონტექსტზე დამოკიდებული $a^k b^k c^k$ ენის გამოყენებით გაშუქებულია ავტომატური ანილიზისა და ავტომატური გენერაციის ურთირთკავშირები **LA**-გრამატიკაში და ნაჩვენებია **LA**-გრამატიკის ტიპობრივი ტრანსფორმატულობა. 10.5 ქვეთავში წარმოდგენილია ბუნებრივი ენის **LA**-გრამატიკული ანალიზის ენათმეცნიერული მოტივაციები.

10.1 დესტა ტიპები და დერივაციული რიგი

LA-გრამატიკის სახელწოდება მოტივირებულია იმ მარცხნივ ჯგუფდებადი გამოყვანითი, ანუ დერივაციული რიგით (**left-associative derivation order**), რასაც ეს ელემენტალური ფორმალიზმი ეფუძნება. მარცხნივ ჯგუფდებადობის ცნება ლოგიკიდანაა ცნობილი.

როდესაც ჩვენ ოპერატორების კომბინირებით ვაყალიბებთ გამოსახულებებს, რიგი, რომლის მიხედვითაც ოპერატორები გამოიყენებიან, შეიძლება არ იყოს ცხადი. მაგალითად: **a+b+c** შეიძლება ინტერაციული რიგორუც (**(a+b)+c**), ან როგორუც (**a+(b+c)**). ჩვენ ვაბობთ, რომ + არის მარცხნივ ჯგუფდებადი (**left-associative**), თუ ის ოპერანდებს აჯგუფებს მარცხნიდან მარჯვნივ, ისე როგორუც (**(a+b)+c**). ჩვენ ვაბობთ, რომ + არის მარჯვნივ ჯგუფდებადი (**right-associative**), თუ ის ოპერანდებს საპირისპირ მიმართულებით აჯგუფებს, როგორუც (**a+(b+c)**).

ა. ვ. ახო & ჯ. დ. ულმანი (A. V. Aho & J. D. Ullman) 1977. გვ. 47

მარცხნივ და მარჯვნივ ჯგუფდებად ფრჩხილებიან სტრუქტურებს აქვთ ერთი განსაკუთრებული თვისება, რაც ამ სტრუქტურების რეგულარული მზარდობის უნარით არის განპირობებული.

10.1.1 მარცხნივ და მარჯვნივ მზარდად ჯგუფდებადი გამოყვანა

მარცხნივ ჯგუფდებადი:

$$\begin{array}{c} a \\ (a + b) \\ ((a + b) + c) \\ (((a + b) + c) + d) \\ \dots \Rightarrow \end{array}$$

მარჯვნივ ჯგუფდებადი:

$$\begin{array}{c} a \\ (b + a) \\ (c + (b + a)) \\ (d + (c + (b + a))) \\ \Leftarrow \dots \end{array}$$

ამ ორი რეგულარულად მზარდი სტრუქტურიდან მარცხნივ ჯგუფდებადი შეესაბამება დასავლეთის, ანუ ბერძნულ-რომაული დამწერლობის ტრადიციულ მიმართულებას.

როდესაც ენის ანალიზისთვის მარცხნივ ასოცირებად დერივაციულ რიგს ვიყენებთ, ბუნებრივია, **a**, **b**, **c** ... სიმბოლოებით სიტყვების, ხოლო + სიმბოლოთი მიწერის, ანუ კონკატენაციის (**concatenation**) ოპერაციის აღნიშვნა. წინადადების პირველი **a** სიტყვა ამავდროულად იწოდება წინადადების საწყისად. ამასთან, ეს **a**, ანუ, წინადადების ეს საწყისი კომბინირდება წინადადების მომდევნო **b** სიტყვასთან და შედეგად იძლევა (**a + b**) გამოსახულებას, რაც ასევე წინადადების საწყისად გაიგება. შემდეგ ეს ახალი წინადადების საწყისი კომბინირდება მომდევნო **c** სიტყვასთან და (**(a + b) + c**) გამოსახულების სახით შედეგად კვლავ ახალ წინადადების საწყისს იძლევა.

მოკლედ, ამ ალგორითმის მიხედვით წინადადების საწყისი ყოველ ბიჯზე მთლიანდება მომდევნო სიტყვასთან და შედეგად წინადადების ახალ საწყისს იძლევა. ამასთან, ეს გრძელდება მანამ, სანამ ალგორითმში შემავალი მოცემულობის სახით არსებობს წინადადების უკვე განსაზღვრული საწყისი ნაწილის მომდევნო სიტყვა.

მარცხნივ ასოცირებადი დერივაციული რიგი წრფივია იმ აზრით, რომ რეგულარულად უმატებს მომდევნო სიტყვას წინა სიტყვას, და არის დროში წრფივი, რადგან ზრდის მიმართულება შეესაბამება დროის მიმართულებას (იხ. 5.4.3). ამგვარად, მარცხნივ ასოცირებადი დერივაციული რიგი შეესაბამება ბუნებრივი ენის დე სოსიურის (**de Saussure**) მეორე კანონისეულ დროში წრფივ სტრუქტურას.

რეგულარული მარცხნივ და მარჯვნივ ჯგუფდებადი ფრჩხილებიანი სტრუქტურების გარდა, არსებობს ამავე ტიპის უამრავი არარეგულარული სტრუქტურა. მაგალითად:

$$\begin{aligned} & (((a + b) + (c + d)) + e) \\ & ((a + b) + ((c + d)) + e) \\ & (a + ((b + c)) + (d + e)) \\ & ((a + (b + c)) + (d + e)) \\ & (((a + b) + c) + (d + e)) \end{aligned}$$

...

ამ არარეგულარული ფრჩხილებიანი სტრუქტურების რიცხვი გამოსახულების სიგრძის ზრდასთან ერთად ექსპონენციალურად იზრდება და თუ **(a)**, **((a))**, **((((a)))**, ..., და ა.შ. სახის გამოსახულებებსაც ჩავთვლით, უსასრულოა.

ეს რეგულარული და არარეგულარული ფრჩხილებიანი სტრუქტურები (და მათი შესაბამისი ხეები), ის სტრუქტურებია, რომელთაც **C-** და **PS-** გრამატიკები წარმოშობები შესაძლო ჩასმების (**possible substitution**) პრინციპის საფუძველზე. ტერმინალური ჯაჭვების შესაბამისი შესაძლო ხეების ამ საკმარისად დიდი რიცხვიდან ‘კორექტული’ ფრჩხილებიანი სტრუქტურებისა და ‘კორექტული’ ფრაზული სტრუქტურების ხეების გამორჩევა შემადგენელ სტრუქტურებზე დამყარებული ანალიზით **C-** და **PS-** გრამატიკული ლინგვისტური კვლევების ცენტრალური ამოცანაა.

ამასთან, **LA**-გრამატიკა დაფუძნებულია შესაძლო გაგრძელებათა (**possible continuations**) პრინციპზე, რომელიც ფორმალურად იძლევა რეგულარულ მარცხნივ ჯგუფდებად ფრჩხილებიან სტრუქტურებსა (იხ. 10.1.1) და მათ შესაბამის ხეებს (იხ. 10.1.6).

10.1.2 შესაძლო გაგრძელებათა პრიცეპი

დაწყებული პირველი სიტყვით, გრამატიკა აღწერს შესაძლო გაგრძელებებს დასამუშავებელი წინადადების თითოეული საწყისისათვის იმ წესების ცხადი მინიშნებით, რომელთა მიხედვითაც შეიძლება განხორციელდეს შემდეგი გრამატიკული კომპოზიცია (ე.ი. წინადადების საწყისზე მისი მომდევნო სიტყვის დამატება).

დროში წრფივი დერივაციული რიგი და შესაძლო გაგრძელებების სტრუქტურული მახასიათებელი ფორმალიზებულია **LA**-გრამატიკის სპეციფიკური სქემატური წესით.

10.1.3 LA-გრამატიკის მარცხენი ჯგუფდებადი სქემატური ფესი

$$r_1: \text{cat}_1 \text{ cat}_2 \Rightarrow \text{cat}_2 \text{ rp}_i$$

წესი შედგება წესის r_i სახელისაგან, cat_1 , cat_2 და cat_2 კატეგორიათა ნიმუშებისაგან (**category patterns**), და rp_i წესის დამაპაკეტებელისაგან (**rule package**). კატეგორიების ნიმუშები განსაზღვრავენ იმ კატეგორიალურ ოპერაციას, რომელიც წინადადების ss საწყისს (დაწყვილებულია cat_1 -თან) და nw მომდევნო სიტყვას (დაწყვილებულია cat_2 -თან) ასახავს წინადადების ახალ ss' საწყისში (აღნიშნულია cat_3 -ით). r_1 წესის წარმატებული გამომავალი არის მდგომარეობა (**state**), რომელიც ($\text{ss}' \text{ rp}_i$) დალაგებული წყვილით მოიცემა.

შემდეგ კომბინაციაში, rp_i წესების დამაპაკეტებლის წესები ამუშავებენ წინადადების ss' ახალ საწყისსა და ახალ მომდევნო სიტყვას. ამგვარად, **LA**-გრამატიკული ანალიზი იწყება შემავალი წინადადების საწყისი სიტყვით, ხოლო მისი კონცეპტუალური დერივაციული რიგი არის **ზემოთ მიმართული მარცხნივ ჯგუფდებადი (bottom-up left-associative)**.

შედარების გასაადვილებლად, ქვემოთ კიდევ ერთხელ წარმოვადგენთ **C-** და **PS-** გრამატიკების სქემატურ წესებს.

10.1.4 C-გრამატიკის გაპათილების ფესის სქემა

$$\alpha(x|y) \circ \beta(y) \Rightarrow \alpha\beta(x)$$

სქემატური წესი α და β გამოსახულებებს აკომბინირებს $\alpha\beta$ გამოსახულებაში. ამასთან, α გამოსახულების **Y** კატეგორიული მახასიათებელი ბათილდება β გამოსახულების შესაბამისი კატეგორიული მახასიათებლით. შედეგი არის ხეობრივი სტრუქტურა, რომელშიც (**X**) კატეგორიის $\alpha\beta$ გამოსახულება დომინირებს α და β კატეგორიებზე. კატეგორიალური გაბათილების წესის კონცეპტუალური დერივაციული რიგი არის **ზემოთ მიმართული შევიწროებადი (bottom-up amalgamating)**.

10.1.5 PS-გრამატიკის გადაფერის ფესის სქემა

$$A \rightarrow BC$$

A სიმბოლოს ჩანაცვლება **B** და **C** სიმბოლოთი შეისაბამებს ხეობრივ სტრუქტურას, რომელშიც **B** და **C** სიმბოლოები უშეალოდ მიიღებიან **A** სიმბოლოსაგან. კონცეპტუალური დერივაციული რიგი არის ქვემოთ მიმართული გაფართოებადი (**top-down expanding**).

ამ განსხვავებული სქემატური წესების შესაბამის დერივაციულ რიგებს ქვემოთ აღწერილი სქემატური წარმოდგენები შეესაბამება

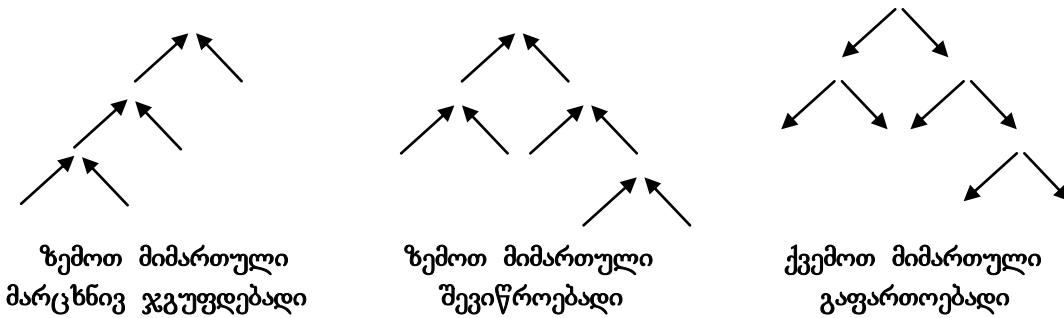
10.1.6 სამი აბსტრაქტული ფარმობული დალაგება

LA-გრამატიკა

C-გრამატიკა

PS-გრამატიკა





C- და **PS-** გრამატიკები, რომლებიც შესაძლო ჩანაცვლებათა პრინციპს ეფუძნებიან, განსხვავდებიან ერთმანეთისაგან მათი კონცეპტუალური² დერივაციული რიგის მიმართულებებით. **C-**გრამატიკაში ორი კატეგორიზებული გამოსახულება შეცვლილია ერთი კატეგორიზებული გამოსახულებით (ზემოთ მიმართულობა). **PS-**გრამატიკაში ერთი ნიშანი შეცვლილია ერთი, ორი ან მეტი ნიშანით (ქვემოთ მიმართულობა).

შესაძლო ჩანაცვლებათა პრინციპი წარმოშობს არარეგულარულ დერივაციულ სტრუქტურებს, რომლებიც იძლევან არარეგულარულ ფრაზათა სტრუქტურულ ხებს. ენის **SLIM** თეორიის თვალსაზრისით შესაძლო ჩანაცვლებათა პრინციპზე დაფუძნებული დერივაციული დალაგება წინააღმდეგობაში მოდის ენის საურთიერთობო ფუნქციასთან (ირღვევა შემავალ-გამომავალი ეკვივალენცია გრამატიკულ და მოლაპარაკე-მსმენელ სისტემებს შორის) და შეფერხებებს იძლევა ენის ავტომატურ ანალიზში (არ იძლევა არანაირ ტიპობრივ ტრანსფარენტულობას გრამატიკასა და შესაბამის პარსერს შორის).

შესაძლო გაგრძელებათა პრინციპი წარმოშობს რეგულარულ დერივაციულ დალაგებებს. მათი შესაბამისი **LA-**გრამატიკის ხეობრივი სტრუქტურები შეიძლება შემადგენელ სტრუქტურებზე დაფუძნებული **C-** და **PS-**გრამატიკების ლინგვისტური თვალსაზრისებით არც კი იყოს საინტერესო. ენის **SLIM** თეორიის თვალსაზრისით ის (i) ამოდელირებს ბუნებრივი ენის ფუნდამენტურ დროში წრფივ სტრუქტურას, (ii) იძლევა შემავალ-გამომავალ ეკვივალენციას გრამატიკულ და მოლაპარაკე-მსმენელ სისტემებს შორის და (iii) უზრუნველყოფს ტიპობრივ ტრანსფარენტულობას გრამატიკასა და პარსერს შორის.

ამგვარად, გამოდის, რომ შემადგენელი სტრუქტურის ფორმალური ცნება, განსაზღვრული ფრაზათა სტრუქტურული ხების ტერმინებით (იხ. 8.4.3), არ გამოიყენება **LA-**გრამატიკაში. თუმცა, ბუნებრივი ენის მაგალითების ინტუიციური ანალიზისას ეს ცნება არაფორმალურად შეიძლება გამოვიყენოთ **შედეგნილი გრამატიკული ერთეულის (complex grammatical unit)** ცნების უფრო ზოგადი გააზრებისათვის. ამ აზრით, იგი არის კლასიკური გრამატიკის უფრო ძველი ცნების, კერძოდ სინტაგმის ცნების განზოგადებული გააზრება. მაგალითად, როდესაც ვამბობთ, რომ ინგლისურ დეკლარაციულ წინადადებაში

² **C-** და **PS-** გრამატიკებში აბსტრაქტული წარმოებული დალაგება განსაზღვრულია როგორც პროცედურული წარმოებული დალაგება. მაგალითად, შინაარსისგან თავისუფალი **PS-**გრამატიკის განსხვავებული სინტაქსური პარსერები იყენებენ ბევრ განსხვავებულ წარმოებულ დალაგებას: ზოგიერთი მოქმედებს მარცხნიდან მარჯვნივ; ზოგიერთი – მარჯვნიდან მარცხნივ; ზოგიერთი იწყება შეუ ნაწილიდან (კუნძულოვანი სინტაქსური პარსერები (**island parsers**)); ზოგიერთი იყენებს მარცხნა კუთხის, სხვები – მარჯვენა კუთხის დალაგებას და ა.შ. ეს ალტერნატიული დალაგებები, როგორც წესი, სისტემებს პროცედურული ნაწილის შემადგენლებას და ემსაზურებას ენის განსაკუთრებული თვისებების გამოყენებით პარსერის ეფექტურობის გაუმჯობესების მიზნებს.

The little dog found a bone

ზმნა **found** (იპოვნა) არის ‘მეორე ადგილზე’, ჩვენ არ ვგულისხმობთ წინადადების მეორე სიტყვის ადგილს, არამედ ვგულისხმობთ წინადადების პირველი გრამატიკული ერთეულის შემდეგ განთავსებულ მეორე სინტაქტურ, ანუ შემადგენლურ ადგილს. საქმე ისაა, რომ **LA**-გრამატიკაში ეს პოზიციურობა, ანუ წინადადებაში სიტყვის კონკრეტული ადგილი, არ არის დამოკიდებული რაიმე სახის ჩანაცვლებაზე და/ან გადანაცვლებებზე, არამედ დამოკიდებულია შესაძლო გაგრძელებათა გაანგარიშებაზე დროით წრფივ დერივაციაში.

10.2 LA-გრამატიკის ფორმალიზაცია

LA-გრამატიკის ქვემოთ მოყვანილ ალგებრულ განსაზღვრებაში, ჩვენ, მოხერხებულობისათვის, დადებით მთელ რიცხვებს გავაიგივებთ სიმრავლეებთან, ანუ. $n = \{ i \mid 0 \leq i < n \}$

10.2.1 LA-გრამატიკის ალგებრული განსაზღვრება

მარცხნივ ჯგუფდებადი გრამატიკა (ანუ **LA**-გრამატიკა) არის $\langle W, C, LX, CO, RP, ST_s, ST_f \rangle$ შვიდეული, სადაც

1. **W** არის სიტყვითი ენობრივი გამოსახულებების (**word surface**) სასრული სიმრავლე.
2. **C** არის კატეგორიათა სეგმენტების (**category segments**) სასრული სიმრავლე.
3. $LX \subset W \times C^*$ არის ლექსიკონის (**lexicon**) მომცველი სასრული სიმრავლე.
4. $CO = (co_0 \cdots co_{n-1})$ არის $(C^* \times C^*)$ სიმრავლიდან $C^* \cup \{\perp\}$,³ სიმრავლეში განსაზღვრული სრულად რეკურსული ფუნქციების (**total recursive function**) სასრული მიმდევრობა. ამ ფუნქციებს კატეგორიულ ოპერაციებს (**categorial operations**) უწოდებენ.
5. $RP = (rp_0 \cdots rp_{n-1})$ არის n სიმრავლის ქვესიმრავლების მიმდევრობა და მას წესების პაკეტის (**Rule packages**) უწოდებენ.
6. $ST_s = \{(cat_s, rp_s), \dots\}$ არის საწყისი მდგომარეობების (**initial state**) სასრული სიმრავლე, სადაც თითოეული rp_s არის საწყის წესების პაკეტად (**start rule package**) წოდებული n სიმრავლის ქვესიმრავლე, და თითოეული $cat_s \in C^*$.
7. $ST_f = \{(cat_f, rp_f), \dots\}$ არის დამსრულებელი მდგომარეობების (**final state**) სასრული სიმრავლე, სადაც თითოეული $cat_f \in C^*$ და თითოეული $rp_f \in RP$.

კონკრეტული **LA**-გრამატიკა ისაზღვრება შემდეგი მონაცემებით:

- (i) ლექსიკონი **LX** (იხ. 3),
- (ii) პირველადი მდგომარეობების სიმრავლე **ST_s** (იხ. 6),

³ თეორიული მიზნები განაპირობებენ იმას, რომ ეს კატეგორიალური ოპერაციები არიან სრულად განსაზღვრული (ტოტალური) ფუნქციები. პრაქტიკულად კი, ისინი გაიგებან როგორც $(C^* \times C^*)$ სიმრავლის მარტივად გამოცნობადი ქვესიმრავლები, რომელთა გარეთ დარჩენილი განსაზღვრის არის ნებისმიერი წერი "don't care" პრინციპის საფუძველზე აისახება (\perp) სიმბოლოთ აღნიშნულ მნიშვნელობაზე, რითაც ეს კატეგორიალური ოპერაციები გარდაიქმნებან სრულად განსაზღვრულ ფუნქციებად.

- (iii) r_i წესის მიმდევრობა, რომელთაგან თითოეული განსაზღვრულია როგორც დალაგებული წყვილი (co_i, rp_i) (იხ. 4 და 5), და
- (iv) დამასრულებელი მდგომარეობების სიმრავლე ST_F (იხ. 7).

მარცხნივ ასოცირებადი r_i წესი დასამუშავებლად იღებს წინადადების ss საწყისს და nw მომდევნო სიტყვას და ცდილობს co_i კატეგორიალური ოპერაციის გამოყენებას. თუ მიღებული დასამუშავებელი გამოსახულებების კატეგორიები შესაბამისად უწყვილდება (უთანხმდება) cat_1 და cat_2 ნიმუშებს, მაშინ r_i წესის გამოყენება წარმატებით სრულდება გამოსავალი მოცემულობის წარმოქმნით. გამოსავალი შედგება წყვილისგან ($ss \ rp_i$), სადაც ss არის დასამუშავებელი წინადადების ახლად წარმოქმნილი საწყისი და rp_i არის r_i წესის მაპაკეტებელი. თუ დასამუშავებელი გამოსახულებების კატეგორიები არ წყვილდება cat_1 და cat_2 ნიმუშებთან, მაშინ r_i წესის გამოყენება არ სრულდება წარმატებით და, შესაბამისად, არც გამოსავალი არ წარმოიქმნება.

წესის მაპაკეტებელი rp_i შეიცავს ყველა იმ წესს, რომლებიც შეიძლება წარმატებით იქნება გამოყენებულნი r_i წესის შემდეგ. წესის მაპაკეტებელი არის წესის სახელების სიმრავლე, სადაც წესის სახელი არის CO მიმდევრობაში მისი co_g კატეგორიალური ოპერაციის ადგილის გამომსახველი g რიცხვი. პრაქტიკაში წესების უმრავლესობას მნემონიკური სახელები აქვთ, მაგ.: rule-g ან Fverb+main.

წესის წარმატებული გამოყენების შემდეგ ალგორითმი შესასვლელში შემოსული დასამუშავებელი წინადადებიდან იღებს ახალ მომდევნო სიტყვას (თუკი ის არსებობს) და იყენებს წესის მაპაკეტებლის წესებს წინადადების ახლად წარმოქმნილ საწყისზე და ამ ახლად აღებულ მომდევნო სიტყვაზე. ამგვარი მიღომით **LA**-გრამატიკა შესასვლელის გავლით შემოსულ დასამუშავებელ წინადადებას ამუშავებს მარცხნიდან მარჯვნივ და პარალელურად ამოწმებს შესაძლო ალტერნატიულ გაგრძელებებს. დერივაცია, ანუ გამოყვანა ჩერდება იმ შემთხვევაში თუ არ არის არცერთი შესაძლო გრამატიკული გაგრძელება (არაგრამატიკული შესასვლელი, მაგ. 10.5.5), ან თუ აღარ არის შემდეგი მომდევნო სიტყვა (სრული ანალიზი, მაგ. 10.5.3).

LA-გრამატიკების ზოგადი ფორმატი **C**-გრამატიკისა (იხ. 7.4.3) და **PS**-გრამატიკის (იხ.. 7.1.3) ფარგლებში აღწერილი კონტექსტისგან თავისუფალ $a^k b^k$ ენისათვის წარმოდგენილია 10.2.2 პუნქტში.

10.2.2 LA-გრამატიკა $a^k b^k$ ენისათვის

$$\begin{aligned} LX &=_{\text{def}} \{ [a (a) |, b (b)] \} \\ ST_S &=_{\text{def}} \{ [(a) \{r_1, r_2\}] \} \\ r_1: (X) (a) &\Rightarrow (aX) \{r_1, r_2\} \\ r_2: (aX) (b) &\Rightarrow (X) \{r_2\} \\ ST_F &=_{\text{def}} \{ [\varepsilon \ rp_2] \} \end{aligned}$$

LX ლექსიკონი შეიცავს ორ **a** და **b** სიტყვას. ამასთან, სიტყვა არის დალაგებული წყვილი, რომელიც შედგენილია მისი ენობრივი გამოსახულებისა და მისივე კატეგორიისაგან. კატეგორიები,

როგორც კატეგორიული სეგმენტების სიები, აქ შეიცავენ მხოლოდ განცალკავებულ სეგმენტებს⁴. ეს განპირობდა სიტყვებისა და მათი ენობრივი გამოსახულებების გაიგივებით.

ST_S საწყისი მდგომარეობა გვანიშნებს იმას, რომ პირველი სიტყვა უნდა იყოს (a) კატეგორიის, ანუ იმას, რომ იგი უნდა იყოს **a** და რომ თავდაპირველად **r₁**, ან **r₂** წესი უნდა იქნეს გამოყენებული. ამგვარად, ალგორითმის პირველ ბიჯში შეიძლება იქნეს გამოყენებული ნებისმიერი წესი⁵, მაგრამ არა ნებისმიერი სიტყვა.

წესების კატეგორიალური ნიმუშები, ანუ მოდელები იყენებენ **X** ცვლადს (ის წარმოგვიდგენს ნულ, ან მეტ კატეგორიულ სეგმენტს), და **a** და **b** სეგმენტურ კონსტანტებს. **r₁** წესი დასამუშავებლად იღებს წინადადების ნებისმიერი კატეგორიის საწყის (წარმოდგენილს (**X**) ნიმუშით) და (a) კატეგორიის მომდევნო სიტყვას, ე.ი. **a** სიტყვას. კატეგორიალური ოპერაციის შედეგი გამოსახულია (**a X**) ნიმუშით: **a**-სეგმენტი წინიდან მიეწერება წინადადების საწყისის კატეგორიას.

r₂ წესი დასამუშავებლად იღებს წინადადების საწყისს (რომლის კატეგორია იწყება **a**-სეგმენტით და წარმოდგენილი (**aX**) ნიმუშით) და (b) კატეგორიის მომდევნო სიტყვას, ე.ი. **b** სიტყვას. ამასთან, ამ კატეგორიალური ოპერაციის შედეგი გამოსახულია (**X**) სქემით: **a**-სეგმენტი წინიდან აკლდება წინადადების საწყისის გამომხატველ კატეგორიას.

r₁ წესის მაპაკეტებელი, სახელად **rp₁**, შეიცავს **r₁** და **r₂** წესებს. ამასთან, სანამ მომდევნო სიტყვა არის **a**, **r₁** არის წარმატებული, ხოლო **r₂** წარუმატებელი, რადგან ის მომდევნო სიტყვად მოითხოვს **b** სიტყვას. როგორც კი შესასვლელში პირველი **b** გამოჩნდება, **r₂** წარმატებული ხდება, ხოლო **r₁** წარუმატებელი. მაპაკეტებელი წესი **rp₂** შეიცავს მხოლოდ ერთ წესს, სახელად **r₂**. მაშასადამე, ყოველი შემდეგი (ე.ი. არაპირველი) **b** მოსალოდნელია ახალ მომდევნო სიტყვად მხოლოდ იმ შემთხვევაში, როცა **b** უკვე იყო წინა მომდევნო სიტყვად.

ანალიზი სრულდება, როდესაც წინადადების საწყისის კატეგორიის ყველა **a**-სეგმენტი ბათილდება **b**-სეგმენტით. ანუ, ანალიზი მთავრდება წინადადების ცარიელ საწყისზე **r₂** წესის გამოყენების შემდეგ. ამას განაპირობებს 10.2.2 პუნქტში მოცემული **ST_F** დამასრულებელი მდგომარეობა, რომელიც მოითხოვს **ε** სიმბოლოს (ე.ი. ცარიელ მიმდევრობას), როგორც დამასრულებელ რეზულტატურ კატეგორიას.

შემდეგი **LA**-გრამატიკა წარმოშობს კონტექსტზე დამოკიდებულ **a^kb^kc^k** ენისათვის.

10.2.3 **LA**-გრამატიკა **a^kb^kc^k** ენისათვის

LX =_{def} { [a (a)] , [b (b)], [ε (ε)] }

ST_S =_{def} { [(a) {r₁, r₂}] }

r₁: (X) (a) ⇒ (aX) {r₁, r₂}

r₂: (aX) (b) ⇒ (Xb) {r₂, r₃}

r₃: (bX) (c) ⇒ (X) {r₃}

ST_F =_{def} { [ε rp₂] }

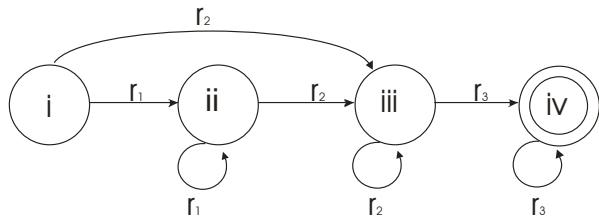
⁴ უფრო რთული ენების **LA**-გრამატიკებში ლექსიკური კატეგორიები შედგება რამდენიმე სეგმენტისაგან.

⁵ უფრო რთულ ენებში საწყისი მდგომარეობა დაწკრილებით აღწერს გრამატიკული წესების საკუთრივ ქვესიმრავლეს.

8.3.7 განაწერში აღწერილი **PS**-გრამატიკასთან შედარებით აქ აღწერილი **LA**-გრამატიკა ბევრად უფრო მარტივია. ამასთანავე, 10.2.2 განაწერში კონტექსტისგან თავისუფალი $a^k b^k$ ენისთვის აღწერილი **LA**-გრამატიკა ძალიან ჰგავს 10.2.3 განაწერში კონტექსტზე დამოკიდებული $a^k b^k c^k$ ენისთვის აღწერილ **LA**-გრამატიკას.

LA-გრამატიკაში წესებსა და მათ პაკეტებს შორის დამოკიდებულება განსაზღვრავს სასრულ მდგომარეობიან გარდაქმნელ ქსელს (**finite state transition network (FSN)**). მაგალითად, 10.2.3 პუნქტში $a^k b^k c^k$ ენისათვის აღწერილი **LA**-გრამატიკის **FSN** შემდეგი ფორმა აქვს.⁶

10.2.4 **LA**-გრამატიკის სასრულ მდგომარეობისი ქსელის ჩარჩო $a^k b^k c^k$ ენისათვის



ეს **FSN** შედგება **i-iv** წრებით წარმოდგენილი ოთხი მდგომარეობისგან. თითოეული მდგომარეობა განსაზღვრულია, როგორც დალაგებული წყვილი, რომელიც შედგება კატეგორიის ნიმუშისა და წესის პაკეტისგან. **i** მდგომარეობა შეესაბამება **ST_s** სქემის საწყის მდგომარეობას, მაშინ როდესაც **ii**, **iii** და **iv** მდგომარეობები შეესაბამისად შეესაბამებიან **r₁**, **r₂** და **r₃** წესების გამოსავალს. **iv** მდგომარეობას აქვს მეორე წრეც, რომელიც აღნიშნავს იმას, რომ **iv** მდგომარეობა შეიძლება იყოს დამასრულებელი (იხ. **ST_F** სქემის განსაზღვრება 10.2.3 განაწერში).

მარცხნივ ასოცირებადი წესის გამოყენება შემავალ წყვილზე, რომელიც შედგენილია წინადადების საწყისისა და მომდევნო სიტყვისგან, შედეგად იძლევა გარდაქმნას (**transition**). 10.2.4 სქემაზე ასეთი გარდაქმნები გრაფიკულად გამოსახულია ისრებით, რომლებიც მონიშნულია მიკავშირებული წესების სახელებით. მდგომარეობაში შემავალი გარდაქმნები მოიცემა მდგომარეობასთან მიკავშირებული წესების კატეგორიალური ოპერაციებით. მდგომარეობიდან გამომავალი გარდაქმნები მოიცემა ამ მდგომარეობაში შემავალი გარდაქმნელი წესების პაკეტში გაერთიანებული წესების კატეგორიალური ოპერაციების გაერთიანებით.

მაგალითად, **i** მდგომარეობიდან გამომავალი გარდაქმნები განსხვავდებიან ერთმანეთისაგან (კერძოდ ესენია, **r₁** და **r₂**) და შეესაბამება **ST_s** სქემის საწყისი მდგომარეობების წესის პაკეტს. **ii** მდგომარეობაში შემავალი გარდაქმნები არიან ერთი და იგივე და გამოხატავენ **r₁** წესის გამოყენებებს განსხვავებულ წინა მდგომარეობებზე. შესაბამისად, **ii** მდგომარეობიდან გამომავალი გარდაქმნები განსხვავდებიან ერთმანეთისაგან და შეესაბამებიან **r₁** წესის პაკეტს. გარდაქმნები, რომლებსაც მივყავართ **iii** მდგომარეობამდე, ერთი და იგივენი არიან და განისაზღვრებიან **r₂** წესის გამოყენებით სამ განსხვავებულ წინა მდგომარეობაზე, და ასე შეძლევ.

⁶ უფრო დეტალური აღწერა ნახეთ **CoL**, ქვეთავი 8.2.

შემავლის მანალიზებელი და გამოსავლის წარმოქმნელი **LA**-გრამატიკა გარდაქმნის ისრების შესაბამისად მოძრაობს თავის **FSN**-ის გასწვრივ ერთი მდგომარეობიდან მომდევნო მდგომარეობაში. ამასთან, **LA**-გრამატიკა, რომლის წარმოქმნითი უნარი რეალიზდება მისი წესების კატეგორიული ოპერაციებით, წარმოქმნის ზუსტად რეკურსიული ენების კლასს, მაშინ როდესაც **FSN** ალგორითმი მხოლოდ წესიერ ენებს წარმოქმნის (იხ. ქვეთავი 11.1).

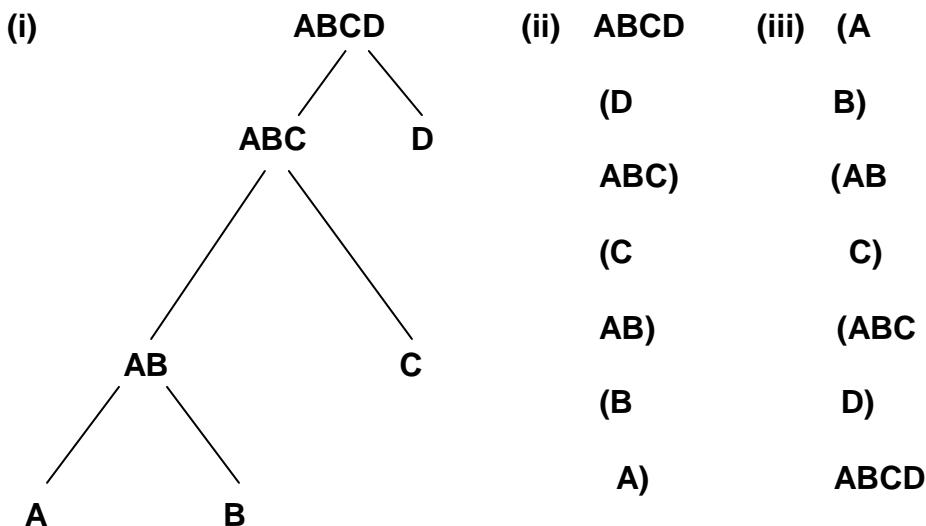
ეს უზარმაზარი წარმოქმნითი უნარი შეიძლება ძუნწად იქნეს გამოყენებული. მაგალითად, $\mathbf{a^k b^k c^k}$ (იხ. 10.2.2) და $\mathbf{a^k b^k c^k}$ (იხ. 10.2.3) ენების **LA**-გრამატიკები არიან უმდაბლესი სირთულის: ისინი ორივე აპარსერებენ წრფივ დროში.⁷

ამგვარად, **LA**-გრამატიკაში კონტექსტისგან თავისუფალი და კონტექსტე დამოკიდებულ ენებს შორის არ არის გამყოფი ხაზი. ეს არის პირველი ნიშანი იმისა, რომ ფორმალური ენებისათვის **LA**-გრამატიკული ფორმალიზმი **PS**-გრამატიკული ფორმალიზმისგან განსხვავებულ იერარქიას წარმოქმნის.

10.3 დროში შროვივი ანალიზი

LA-გრამატიკის ხისებრი სტრუქტურა შესაძლებელია წარმოვადგინოთ ეკვივალენტური სტრუქტურული ცხრილის სახით.

10.3.1 **LA**-ხეები როგორც სტრუქტურირებული ცხრილები



(i)-ში მარცხივ ჯგუფდებადი დერივაცია ნაჩვენებია ხისებრი სახით (იხ. 10.1.6). (ii)-ში იგივე დერივაცია ნაჩვენებია ეკვივალენტური სტრუქტურირებული ცხრილის სახით, სადაც წინადადების

⁷ იმის გამო, რომ მათ წესებს აქვთ შეუთავსებელი შესავლის მდგომარეობა (მათგან თითოეული იღებს განსხვავებულ მომდევნო სიტყვას), ეს ორივე **LA**-გრამატიკა არის არაორაზროვანი. ასეთი სახის არაორაზროვანი **LA**-გრამატიკები იწოდებიან როგორც **C1-LAG-ები** და ისინი წრფივ დროში აპარსერებენ (იხ. თავი 11).

საწყისს, მომდევნო სიტყვას და შედეგად მიღებულ წინადადების საწყისს აქვს საკუთარი სტრიქონი. (ii) და (iii) განსხვავდებიან მხოლოდ კითხვის დროში წრფივი მიმართულებით: (ii)-ში კითხვა იწყება ქვემოდან, ხოლო (i) და (iii)-ში კითხვა იწყება ზემოდან.

დერივაციის ეკრანზე საჩვენებლად უფრო მეტად მოსახერხებელია (iii) სტრუქტურა. ეს ფორმა წარმოდგენილია 10.3.2 ამონაბეჭდში, სადაც 10.2.2 პუნქტში აღწერილი **LA**-გრამატიკით ანალიზდება **aaabbb** გამოსახულება.

10.3.2 **LA**-გრამატიკული დერივაცია $a^k b^k$ ენისათვის ($k=3$)

NEWCAT > a a a b b b

***START – 0**

1

(A) A
(A) A

***RULE - 1**

2

(A A) A A
(A) A

***RULE - 1**

3

(A A A) A A A
(B) B

***RULE - 2**

4

(A A) A A A B
(B) B

***RULE - 2**

5

(A) A A A B B
(B) B

***RULE - 2**

6

(N | L) A A A B B B

ეს **LA**-გრამატიკული ანალიზი არის **NEWCAT** **LA**-პარსერის ფორმატირებული პროტოკოლი (ტრასირებული კვალი, ამონაბეჭდი). **LA**-გრამატიკის სრული ტიპობრივი ტრანსფარენტულობა განაპირობებს იმას, რომ ეს ლინგვისტური დეკლარაციული ანალიზი და კომპიუტერული დერივაციული პროცედურა მეტი არაფერი არ არიან თუ არა სხვადასხვა რეალიზაციები ერთი და იგივე მარცხნივ ჯგუფდებადი ალგორითმის.

10.3.2 პროტოკოლში გასაანალიზებელი შემავალი (**input**) დაბეჭდილია **NEWCAT >** ნიშნის შემდეგ. შემავალი კატეგორიების უკეთესად შედარების მიზნით, გამოსახულება და კატეგორია დაბეჭდილნი არიან შებრუნებული რიგით, მაგალითად [A (A)] გამოსახულების ნაცვლად დაბეჭდილია (A) A გამოსახულება.

10.3.2 პროტოკოლის პირველი ნაწილი იწყება მოქმედი წესის პაკეტის სახელით ***START-0**. შემდეგ მას მოჰყება მაკომპლექტებელი რიცხვი 1, წინადადების საწყისი (შედგენილი პირველი სიტყვისგან), მომდევნო (ანუ მეორე) სიტყვა, და ბოლოს იმ წესის სახელი **(RULE-1)**, რომელიც აკომბინირებს წინადადების ამ საწყისსა და ამ მომდევნო სიტყვას. მომდევნო მაკომპლექტებელ რიცხვს მოჰყება ჯერ კიდევ დამუშავების პროცესში მყოფი შედეგი:

მოქმედი წესის პაკეტი:	*START-0
მაკომპლექტებელი რიცხვი:	1
წინადადების საწყისი:	(A) A
მომდევნო სიტყვა:	(A) A
წარმატებული წესი:	*RULE-1
მომდევნო მაკომპლექტებელი რიცხვი:	2
შედეგი:	(A A) A A

როგორც წესის სახელს (იხ. “წარმატებული წესი”), ისე შედეგად მიღებულ წინადადების საწყისს (იხ. “შედეგი”) ორმაგი ფუნქციები აქვთ. წესის სახელი ერთდროულად გვანიშნებს (i) იმ წესზე, რომელიც წარმატებული იყო n კომპოზიციისას, და (ii) იმ წესის პაკეტზე, რომლის შემადგენელი წესებიც გამოიყენება მომდევნო $n+1$ კომპოზიციაში, შესაბამისად, შედეგი ერთდროულად წარმოგვიდგენს (i) n კომპოზიციის გამოსავალს და (ii) $n+1$ კომპოზიციის წინადადების საწყისს.

მარცხნივ ჯგუფდებად დერივაციაში ეს დაშვებული ორმაგი ფუნქციონირება ცხადად ჩანს 10.3 2 პროტოკოლის მეორე კომპლექტაციაში:

მოქმედი წესის პაკეტი:	*RULE-1
კომპლექტაციის რიცხვი:	2
წინადადების საწყისი:	(A A) A A
მომდევნო სიტყვა:	(A) A
წარმატებული წესი:	*RULE-1
მომდევნო კომპლექტაციის რიცხვი:	3
შედეგი:	(A A A) A A A

10.3.2 პროტოკოლში აღწერილი დერივაციის ფორმა შედგენილია წესების კატეგორიალური ოპერაციების დასახასიათებლად. მიიჩნევა, რომ ეს დერივაცია საზღვრავს **LA**-გრამატიკის პირველი სილრმის ფორმატს (**depth first format**), რადგან იგი იძლევა მხოლოდ წარმატებული გაგრძელებების ნიმუშებს. გარდა ამისა, აქ ისაზღვრება აგრეთვე პირველი სიგანის ფორმატი (**breadth first format**), რომელიც გამოიყენება სიტყვათა ფორმების მორფოლოგიურ ანალიზში (იხ. 14.4.2 და 14.4.3). ამ გამოსავალ ფორმატებს საერთო აქვთ ის, რომ ისინი ერთდროულად არიან პარსერის პროტოკოლიც და გრამატიკული ანალიზიც.

10.4 LA-გრამატიკის სრული ტიპობრივი ტრანსფორმაციები

LA-გრამატიკული ალგორითმი თანაბრად ვარგისია პარსირებისთვის. ამის სტრუქტურული მიზეზი არის შესაძლო გაგრძელებათა პრინციპი. პარსირებისას მომდევნო სიტყვა ირჩევა შემავალი მოცემულობიდან, გენერაციისას კი მომდევნო სიტყვას ლექსიკონიდან ვირჩევთ.

პარსირებასა და გენერაციას შორის მჭიდრო ურთიერთკავშირები ილუსტრირებულია 10.4.1 და 10.4.2 პუნქტებში. ეს ანალიზიცა და გენერაციაც ეფუძნება კონტექსტზე დამოკიდებული $a^k b^k c^k$ ენისთვის 10.2.3 პუნქტში აღწერილ **LA**-გრამატიკას.

LA-პარსერი ტვირთავს **C-LAG**-ებიდან ნებისმიერ **LA**-გრამატიკას. ამის შემდეგ გასაანალიზებლად შესაძლებელია ნებისმიერი გამოსახულების შეტანა. **NEWCAT** პარსერი შეიცავს წესების ავტომატურ მთვლელს, რომელიც ყოველი დერივაციული ბიჯის დაწყებისას გვიჩვენებს ამ მარცხნივ ჯგუფდებად კომპლექტაციაში გამოყენებად წესებს (ქვემოთ გამოყენებადი წესების რიცხვი უფრო ნაკლებია ვიდრე 2n).

10.4.1 aaabbbccc გამოსახულების 1 პარსირება შესთა აქტიური მთვლელით

```
NEWCAT> a a a b b b c c c
; 1: გამოყენებული წესები (RULE-1 RULE-2)
; 2: გამოყენებული წესები (RULE-1 RULE-2)
; 3: გამოყენებული წესები (RULE-1 RULE-2)
; 4: გამოყენებული წესები (RULE-2 RULE-3)
; 5: გამოყენებული წესები (RULE-2 RULE-3)
; 6: გამოყენებული წესები (RULE-2 RULE-3)
; 7: გამოყენებული წესები (RULE-3)
; 8: გამოყენებული წესები (RULE-3)
; წესის გამოყენებების რიცხვი: 14.
```

```
*START-0
1
(A) A
(A) A
*RULE-1
2
(A A) A A
(A) A
*RULE-1
3
(A A A) A A A
(B) B
*RULE-2
4
(A A B) A A A B
(B) B
*RULE-2
5
(A B B) A A A B B
(B) B
*RULE-2
6
(B B B) A A A B B B
(C) C
*RULE-3
```

7

(C C) A A A B B B C

(C) C

*RULE-3

8

(C) A A A B B B C C

(C) C

*RULE-3

9

(NIL) A A A B B B C C

LA-გენერატორი, მსგავსად **LA**-პარსერისა, ტვირთავს **C-LAG**-ებიდან ნებისმიერ **LA**-გრამატიკას. ამის შემდეგ ფუნქცია **gram-gen**, ოპერირებს რა გრამატიკის რეკურსულ ფაქტორად (**recursion factor**)⁸ და/ან გენერაციაში გამოყენებულ სიტყვათა სიად (**list of words to be used generation**) წოდებულ არგუმენტებზე, უზრუნველყოფს გენერაციული პროცედურის ლიმიტირებას სიტყვათა გარკვეული სიმრავლითა და სიტყვების გარკვეული სიგრძით. ამჯერადაც, გენერაციის გრამატიკული ანალიზი ამავდროულად არის **LA**-გენერატორის პირდაპირი კვალი, ანუ პროტოკოლი.

10.4.2 გენერაციული ფარმომაღალგენელი ნიმუში $a^k b^k c^k - \text{ში}$

NEWCAT> (gram-gen 3 '(a b c))

Parses of length 2:

A B

2 (B)

A A

1 (A A)

Parses of length 3:

A B C

2 3 (NIL)

A A B

1 2 (A B)

A A A

1 1 (A A A)

Parses of length 4:

A A B B

1 2 2 (B B)

A A A B

1 1 2 (A A B)

A A A A

1 1 1 (A A A A)

Parses of length 5:

A A B B C

1 2 2 3 (B)

⁸ Col, p. 193 ff. სხვა გერსიით ‘gram-gen’ მოხსენიებულია მაქსიმალურ ზედაპირის სიგრძესთნ რეკურსული ფაქტორის ნაცვლად.

A A A B B
1 1 2 2 (A B B)

A A A A B
1 1 1 2 (A A A B)

Parses of length 6:

A A B B C C
1 2 2 3 3 (NIL)

A A A B B B
1 1 2 2 2 (B B B)

A A A A B B
1 1 1 2 2 (A A B B)

Parses of length 7:

A A A B B B C
1 1 2 2 2 3 (B B)

A A A A B B B
1 1 1 2 2 2 (A B B B)

Parses of length 8:

A A A B B B C C
1 1 2 2 2 3 3 (C)

A A A A B B B B
1 1 1 2 2 2 2 (B B B B)

Parses of length 9:

A A A B B B C C C
1 1 2 2 2 3 3 3 (NIL)

A A A A B B B B C
1 1 1 2 2 2 2 3 (B B B)

Parses of length 10:

A A A A B B B B C C
1 1 1 2 2 2 2 3 3 (B B)

Parses of length 11:

A A A A B B B B C C C
1 1 1 2 2 2 2 2 3 3 3 (B)

Parses of length 12:

A A A A B B B B C C C C
1 1 1 2 2 2 2 3 3 3 3 (NIL)

ეს სისტემური გენერაცია იწყება 2 სიგრძის მართებულად აგებული, მაგრამ დაუსრულებელი გამოსახულებით და წარმოგვიდგენს ყველა იმ მართებულად აგებულ შეალებურ გამოსახულებას, რომელთა სიგრძე არ აღემატება 12-ს. ენის დასრულებული გამოსახულები გამოიცნობა მათი რეზულტირებული (NIL) კატეგორიით.

თითოეული დერივაციული ბიჯი მოიცემა ენობრივი გამოსახულების, რიცხვებით წარმოდგენილი წესების სახელების მიმდევრობისა და რეზულტირებული კატეგორიის მინიშნებით. 10.4.3 განაწერში წარმოდგენილია ერთი ცალკე აღებული გამოყვანითი, ანუ დერივაციული ბიჯი.

10.4.3 $a^k b^k c^k$ ენის დასრულებული სტრად-ფორმირებული გამოსახულება

A A A B B B C C C

1 1 2 2 2 3 3 3 (NIL)

ენობრივი გამოსახულებისა და წესის სახელების მიმდევრობა ისეა რანგირებული, რომ ცხადი ხდება თუ რომელ სიტყვას, რომელი წესი მიემართება. 10.4.3 ბიჯი ახასიათებს მართებულად აგებულ გამოსახულებას, რადგან ის შეისაბამებს ფინალურ მდგომარეობას (ε rp_3), რომელიც აქ მოცემულია '**3 (NIL)**' ჩანაწერით, რაც იმას ნიშნავს, რომ იგი $a^k b^k c^k$ ენისთვის 10.2.3 განაწერში აღწერილი **LA**-გრამატიკის **ST_F** სიმრავლის წევრია.

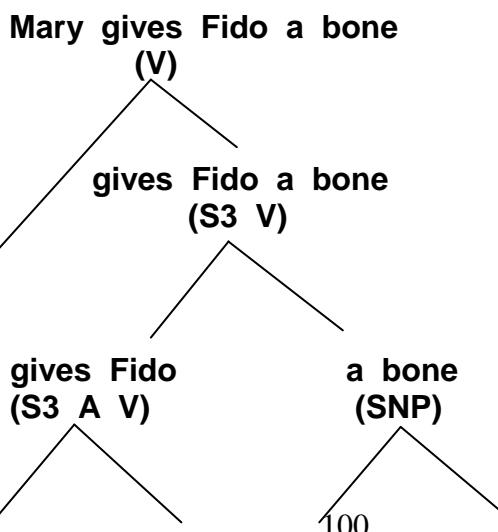
10.2.3-ში განსაზღვრულ **LA**-გრამატიკას, 10.4.1-ში ილუსტრირებულ **LA**-პარსერსა და 10.4.2-ში ილუსტრირებულ **LA**-გენერატორს შორის ურთიერთკავშირები ცხად დემონსტრირებას უკეთებს გრამატიკულ ფორმალიზმს, პარსერსა და გენერატორს შორის არსებული სრული ტიპობრივი ტრანსფარენტულობის მიმართებას. ამგვარად, **LA**-გრამატიკა არის ერთადერთი ელემენტარული გრამატიკული ფორმალიზმი, რომელიც ამომწურავად აკმაყოფილებს სრული ტიპობრივი ტრანსფარენტულობის თაობაზე ბერვიკისა და ვეინბერგის (**Berwick & Weinberg (1984. p. 41)**) მიერ დაყენებულ მოთხოვნებს.

10.5 **LA**-გრამატიკა ბუნებრივი ენისათვის

სანამ მომდევნო თავში მივუბრუნდებით **LA**-გრამატიკის ფორმალურ თვისებებს (როგორიცაა ენის კლასები, წარმოქმნის უნარი, და სირთულე), ვნახოთ როგორ მუშაობს **LA**-გრამატიკა ბუნებრივ ენაში. ამ ქვეთავის მიზანია **LA**-გრამატიკული ანალიზის ენათმეცნიერული მოტივაცია ისეთი ტრადიციული ცნებების საფუძველზე, როგორებიცაა, ერთის მხრივ, ვალენტობა (**valency**), შეთანხმება (**agreement**), და სიტყვების წყობა (**word order**) და, მეორეს მხრივ, დროში წრფივი გამოყვანა (**time-linear derivation**).

LA-გრამატიკის ფორმალიზმი წარმოიშვა იმ მცდელობათა საფუძველზე, რომლის მიზანიც იყო ის, რომ **SCG**-ის ფარგლებში განსაზღვრული **C**-გრამატიკა დანერგილიყო როგორც პარსერი (იხ.. **NEWCAT**, გვ. 7). ინტუიციური დამოკიდებულებები **C**- და **LA**- გრამატიკებს შორის იხსნება 10.5.1 და 10.5.2 მაგალითების შედარებით.

10.5.1 შემადგენელ სტრუქტურებიანი ანალიზი **C**-გრამატიკაში



Mary (SNP)	gives (S3 D A V)	Fido (SNP)	a (SN SNP)	bone (SN)
---------------	---------------------	---------------	---------------	--------------

ეს ხე აკმაყოფილებს შემადგენელი სტრუქტურების 8.4.3 პირობებს და შეიძლება იქნას გამოყენებული როგორც **PS**-გრამატიკაში, ასევე, **C**-გრამატიკაში. თუმცა, მისი შედგენილი კატეგორიებისა და ზემოთ მიმართული დერივაციის თანხლები გარემოებებით იგი უფრო ახლოსაა **C**-გრამატიკულ ანალიზთან.

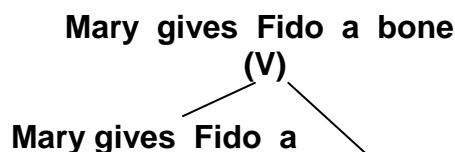
C-გრამატიკის ალგებრულ განსაზღვრებასთან შედარებით (იხ. 7.4.1) აქ (იხ. 10.5.1) გამოყენებული კატეგორიები განსაკუთრებულად მარტივი სახისანი არიან: ისინი წარმოგვიდგებიან კატეგორიული სეგმენტების სიების სახით – და არ შეიცავენ ხაზზე (ე.ო. / სიმბოლოზე) დამყარებულ კატეგორიალურ ფუნქტორ-არგუმენტებს სტრუქტურებს და მრავალრიცხოვან ფრჩხილებს, რაც ტრადიციულად დამახასიათებელია **C**-გრამატიკებისათვის.

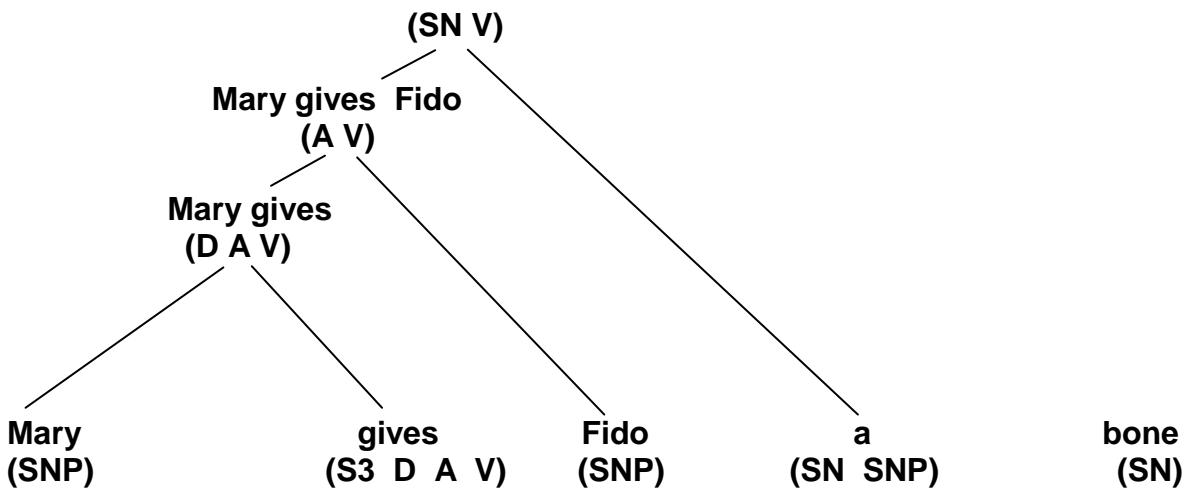
ეს გამრტივებული კატეგორიები სრულიად საკმარისია ვალენტობის ენათმეცნიერულად მეტად ღირებული პრინციპის მნიშვნელობის წარმოსაჩენად. ასე მაგალითად, **gives** სიტყვის კატეგორია (**S3 D A V**) აღნიშნავს ზმნას (**V**), რომელიც ილებს მესამე პირის მხოლობითი რიცხვის ნომინატივის (**nominative**) (**S3**), დატივის (**dative**) (**D**) და აკუზატივის (**accusative**) (**A**) ბრუნვებს არგუმენტებად. შესაბამისად, **a** დეტერმინირის კატეგორია (**SN SNP**) აღნიშნავს ფუნქტორს, რომელიც მოქმედებს მხოლობით ნაუნზე (**singular noun**) (**SN**) და მოქმედების შედეგად იძლევა მხოლობით ნაუნურ ფრაზას (**singular noun phrase**) (**SNP**).

10.5.1 პუნქტში წარმოდგენილი **C**-გრამატიკული დერივაციის შედგენა შეიძლება დავიწყოთ **gives** და **Fido** სიტყვების გაერთიანებით, სადაც **Fido** სიტყვის კატეგორიის **SNP** სეგმენტი აბათილებს **gives** სიტყვის კატეგორიაში **D** სეგმენტს, საიდანაც წარმოიშვება (**S3 A V**) კატეგორიის შუალედური გამოსახულება (**intermediate expression**) **gives Fido**. შემდეგ **a** დეტერმინატორი (**determiner**) და **bone** ნაუნი (**noun**) უნდა გაერთიანდნენ, საიდანაც წარმოიშვება (**SNP**) კატეგორიის შუალედური გამოსახულება **a bone**, რომელიც **gives Fido** შუალედური გამოსახულების კატეგორიაში აბათილებს **A** სეგმენტს, საიდანაც წარმოიშვება (**S3 V**) კატეგორიის შუალედური გამოსახულება **gives Fido a bone**. საბოლოოდ, **Mary** სიტყვის კატეგორიის **SNP** სეგმენტი აბათილებს **gives Fido a bone** გამოსახულების კატეგორიის **S3** სეგმენტს, საიდანაც წარმოიშვება დასრულებული წინადადება (**complete sentence**).

10.5.1 განაწერში გაკეთებული **C**-გრამატიკული ანალიზის ლექსიკური კატეგორიები ხელახლა გამოვიყენოთ ქვემოთ 10.5.2 განაწერში განხილულ შესაბამის მარცხნივ ჯგუფდებად ანალიზში. ამ კატეგორიების ასეთი გამოყენების უფლებას გვაძლევს ის, რომ მათი სიის სტრუქტურა შეთანხმებულია 10.2.1 განაწერში გაკეთებულ **LA**-გრამატიკის ალგებრულ განსაზღვრებასთან.

10.5.2 ლოგი საზოვანი ანალიზი **LA**-გრამატიკაში





ეს ანალიზი ეფუძნება 17.4.1 განაწერში განსაზღვრულ **LA**-გრამატიკას, რომლის აბრევიატურული დასახელებაა **LA-E2**.

10.5.2 გამოყვანაში, ანუ დერივაციაში, მსგავსად 10.5.1 გამოყვანისა, იგივე ვალენტური პოზიციები ბათილდება იგივე შემავსებლებით. ამგვარად, მარცხნივ ჯგუფდებადი დერივაცია იწყება წინადადების პირველი სიტყვით. ამასთან, ეს დერივაცია ყოველ დერივაციულ ბიჯზე წინადადების უკვე წარმოქმნილი საწყისისა მისი მომდევნო სიტყვის გაერთიანებით წინადადების ახალ საწყისს წარმოქმნის.

Mary და **gives** სიტყვების საწყისი კომბინაცია ეფუძნება შემდეგ კატეგორიალურ ოპერაციას:

$$(\text{SNP}) (\text{N D A V}) \Rightarrow (\text{D A V}).$$

SNP კატეგორიალური სეგმენტი (შეესაბამება მხოლობითი რიცხვის სახელურ ფრაზას), რომლითაც ხასიათდება სიტყვა **Mary** აბათილებს **gives** სიტყვის კატეგორიის პირველ **N** სეგმენტს. ენათმეცნიერული ენით ეს იმას ნიშნავს, რომ ზმნის ნომინატიური ვალენტობა შევსებულია პირველი სიტყვის კატეგორიით. **SNP** და **N** სეგმენტების შორის არსებული შეთანხმება განსაზღვრულია 17.4.1 პუნქტში მოცემული ცვლადებიანი განსაზღვრებით.

ამ პირველი კომბინაციის შედეგი არის **(D A V)** კატეგორიის წინადადების საწყისი. შესაბამისად, წინადადების ეს საწყისი წარმოდგენილია იმ შუალედური გამოსახულებით, რომელსაც ჯერ კიდევ სჭირდება **D** (დატივური) და **A** (აკუზატიური) ბრუნვა რათა იქცეს დასრულებულ წინადადებად. ეს ახალი წინადადების საწყისი ერთიანდება **(SNP)** კატეგორიის **Fido** მომდევნო სიტყვასთან, რომლის კატეგორიული სეგმენტი **SNP** აბათილებს **(D A V)** კატეგორიის წინადადების საწყისის პირველ სეგმენტს.

ამ მეორე გაერთიანების შედეგი არის **(A V)** კატეგორიის წინადადების საწყისი. შესაბამისად, წინადადების ეს საწყისი წარმოდგენილია შუალედური გამოსახულებით, რომელსაც ჯერ კიდევ სჭირდება **A** (აკუზატიური) ბრუნვა რათა იქცეს დასრულებულ წინადადებად. ეს ახალი წინადადების საწყისი ერთიანდება მომდევნო სიტყვასთან, ანუ **(SN SNP)** კატეგორიის **a** დეტერმინერთან. წესის კატეგორიალურ ოპერაციას, რომელიც გამოყენებულია ამ გაერთიანებისას, აქვს შემდეგი სახე:

(A V) (SN SNP) ⇒ (SN V).

ამგვარად, დეტერმინერის შედეგის **SNP** სეგმენტი ავსებს აკუზატივის სავალენტო პოზიციას წინადადების საწყისის კატეგორიაში, მაშინ როდესაც დეტერმინერის **SN** არგუმენტული პოზიცია უმატება წინადადების საწყისის კატეგორიას და ხსნის ახალ ვალენტურ ადგილს მხოლობითი რიცხვის ნაუნისთვის.

Mary gives Fido a წინადადების საწყისისა და **bone** მოძევნო სიტყვის კომპოზირებისას ეს ახალი ვალენტური ადგილი ბათილდება შემდეგი სიტყვის **(SN)** კატეგორიით. შედეგი არის **(V)** კატეგორიის მქონე წინადადების საწყისი. ზმნასთან შეუკებელი ადგილის არარსებობა გვანიშნებს, რომ გამოსახულება დასრულებულია. ეს დასრულება მხოლოდ პოტენციური დასრულებულობაა, რადგან ბუნებრივი ენის გამოსახულებები დაუსრულებლად შეიძლება გაგრძელდეს – სამაგალითოდ განვიხილოთ შემდეგი ინგლისური ტექსტი: **regularly or because she is so fond of this cute little dog which she picked up in Denver visiting her mother who told her while driving to the cleaners that the Millers had recently moved to detroit because...** და ასე შემდეგ.

10.5.2-ში აღწერილი დროში წრფივი ანალიზი სრულად იჭერს ცნობილ ენათმეცნიერულ ინტუიციებს ვალენტობაზე, შეთანხმებაზე და სიტყვათა წყობაზე, ისევე როგორც 10.5.1-ში აღწერილი ანალიზი იჭერს შემადგენელ სტრუქტურებზე დამყარებულ ინტუიციებს – ეს ნაჩვენებია 10.5.1 და 10.5.2 განაწერებით, რომლებიც ერთი და იგივე ლექსიკურ კატეგორიებს იყენებენ. მათ განასხვავებთ მხოლოდ ის, რომ 10.5.1 ალგორითმი აგებულია შესაძლო ჩანაცვლებების პრინციპის, ხოლო 10.5.2 ალგორითმი კი შესაძლო გაგრძელებათა პრინციპის საფუძველზე.

10.5.1-ის მსგავსი შემადგენელ სტრუქტურათა ანალიზით ისეთი შუალედური გამოსახულებები, როგორებიცა **Mary gives** და **Mary gives Fido a**, უნდა განვიხილოთ როგორც არაკანონიერები, რადგანაც ისინი არ დასტურდებიან იმ ჩანაცვლებებისა და გადანაცვლებების ტესტით, რაზეც სტრუქტურული ანალიზია დაფუძნებული. 10.5.2-ში აღწერილი მარცხნივ ჯგუფდებადი გამოყვანით კი ისინი კანონიერი შუალედური გამოსახულებებია, რადგან ისინი შეიძლება გაგრძელდნენ მართებულად აგებულ წინადადებამდე.

ამავდროულად, ისეთი გამოსახულებები, მაგალითად როგორებიცაა **gives Fido** და **gives Fido a bone**, არ არიან **LA**-გრამატიკის შუალედური გამოსახულებები, რადგან მათი გაგრძელება დასრულებულ, მართებულად აგებულ წინადადებამდე შეუძლებელია. არადა, **C-** და **PS-** გრამატიკებში ასეთი გამოსახულებები კანონიერ შუალედურ გამოსახულებად გაიგებიან, რადგანაც ისინი დასტურდებიან იმ ჩანაცვლებებისა და გადანაცვლებების ტესტებით, რითაც ეს **C-** და **PS-** გრამატიკებია მოტივირებული.

ხაზგასასმელია, რომ 10.5.2 აღწერილი **LA**-გრამატიკული დერივაცია პირდაპირ შეიძლება იქნას გამოყენებული პარსერში: პარსერი კითხულობს წინადადებას სიტყვაზე სიტყვის მოყოლებით; ამასთან, თითოეული ასეთი ახალი სიტყვის წაკითხვისას იყენებს **LA**-გრამატიკის წესებს და, ყოველთვის, წინადადების საწყისს ამ წესების მიხედვით აერთიანებს მომდევნო სიტყვასთან წინადადების ახალ საწყისში. ამგვარად, გასაგები ხდება, რომ პარსირების პროცესში პარსერის პროტოკოლი, ანუ განვითარების კვალი (**evolving trace**) შეიძლება განხილულ იქნეს როგორც გრამატიკული ანალიზის განაწერიც.

შემდეგი ავტომატური ანალიზი წარმოგვიდგენს **NEW.CAT** და **Col.** სისტემებში გამოყენებული სტუქტურირებული სის ფორმატს 10.3.2 და 10.4.1 განაწერებში აღწერილი ხელოვნური ენების უკვე ცნობილი გრამატიკული ანალიზისთვის.

10.5.3 მარცხნივ ჯგუფდებადი პარსერი 10.5.2 მაგალითისათვის

```
NEWCAT> Mary gives Fido a bone .
*START
1
    (SNP) MARY
    (S3 D A V) GIVES
*NOM + FVERB
2
    (D A V) MARY GIVES
    (SNP) FIDO
*FVERB + MAIN
3
    (A V) MARY GIVES FIDO
    (SN SNP) A
*FVERB + MAIN
4
    (SN V) MARY GIVES FIDO A
    (SN) BONE
*DET + NOUN
5
    (V) MARY GIVES FIDO A BONE
    (V DECL)
*CMPLT
6
    (DECL) MARY GIVES FIDO A BONE .
```

საბოლოო კომბინაციას ემატება სრული სდექი (**full stop**), რომელიც ახასიათებს გამოსახულებას როგორც დეკლარაციულ (**declarative**) წინადადებას (იხ. **LA-E3** (იხ. 17.5.5)).

10.5.3 პარსერანალიზი (**parsing analysis**) შეიცავს არამარტო სრულ ინფორმაციას 10.5.2 დერივაციის დროში წრფივ ხისებრ ფორმატზე, არამედ დერივაციისას გამოყენებადი თითოეული მარცხნივ ჯგუფდებადი კომბინაციისათვის დამატებითი შეტყობინების სახით გვაწვდის ამ კომბინირებისას გამოყენებული წესის სახელსაც, რომელიც ამავდროულად გვმანიშნებლობს აგრეთვე მოქმედების პროცესში მყოფი წესის პაკეტის სახელზეც (იხ. 10.3 ქვეთავი). სტრუქტურული განსხვავება 10.5.2 წრფივ ხესა და 10.5.3 ფორმატირებულ ცხრილს შორის შეესაბამება 10.3.1 განაწერში (i) და (iii) ეკვივალენტური ფორმატებისათვის უკვე მონიშნულ განსხვავებებს.

მომდევნო მაგალითი წარმოგვიდგენს დისკონტინიალური ელემენტის **LA**-გრამატიკულ დამუშავებას.

10. 5. 4 დისკონტინიალური ელემენტის ანალიზი

NEWCAT> Fido dug the bone up \.

*START

1

(SNP) FIDO

(N A UP V) DUG

*NOM + FVERB

2

(A UP V) FIDO DUG

(SN SNP) THE

*FVERB + MAIN

3

(SN UP V) FIDO DUG THE

(SN) BONE

*DET-NOUN

4

(UP V) FIDO DUG THE BONE

(UP) UP

*FVERB + MAIN

5

(V) FIDO DUG THE BONE UP

(V DECL) .

*CMPLT

6

(DECL) FIDO DUG THE BONE UP .

dug და **up** დისკონტინიალურ ელემენტებს შორის კავშირი წარმოდგენილია **dug** სიტყვის (**N A UP V**) კატეგორიის **UP** სეგმენტით. **up** სიტყვის ბოლო ადგილი სპეციფირებულია **dug** ფუნქტორის კატეგორიაში შესავსები ადგილების რიგით (აյ **N, A** და **UP**).

დისკონტინიალური ელემენტების დამუშავება პრობლემაა მხოლოდ შემადგენელ სტრუქტურებზე დამყარებული ანალიზისათვის. მართლაც, ბუნებრივი ენის დროში წრფივ სტრუქტურებზე დამყარებული გრამატიკული თეორია და ვალენტობის, შეთანხმებისა და სიტყვათა რიგის ამ თეორიისეული გაგებებით დისკონტინიალური ელემენტები მუშავდება სტანდარტული მიდგომებით: რელევანტურ ფუნქტორ კატეგორიაში ხდება შესავსებელი ადგილის კოდირება და ამის საფუძველზე მისი შემდგომი განათელება.

დასასრულს განვიხილოთ არაგრამატიკული წინადადების დამუშავება.

10.5.5 არაგრამატიკული შესატანის LA-ანალიზი

NEWCAT> the young girl give Fido the bone \.

ERROR

Ungrammatical continuation at: "GIVE"

*START

1

(SN SNP) THE

(ADJ) YOUNG

*DET + ADJ

2

(SN SNP) THE YOUNG

(SN) GIRL

***DET + NOUN**

3

(SNP) THE YOUNG GIRL

დერივაცია იწყება ჩვეულებრივ, მაგრამ მესამე სიტყვის შემდეგ წყდება იმიტომ, რომ გრამატიკა არ ითვალისწინებს შესაძლო გაგრძელებას წინადადების საწყისისათვის **the young girl** და მომდევნო სიტყვისათვის **give**. მიზეზი ის არის, რომ **SNP** კატეგორიული სეგმენტი ვერ თანხმდება **give** სიტყვის **NOM** სახელობით სეგმენტთან. ეს დაწვრილებით აღწერილია 17.4.1 პუნქტში **LA-E3** ენის ცვლადებიან განსაზღვრებაში.

არავრამატიკული შემავალის გრამატიკულად გამართული საწყისის **LA-პარსერ** ანალიზი, რამდენადაც ეს შესაძლებელია, გვაწვდის წინადადების ამ საწყისის ზუსტ გრამატიკულ აღწერას. ამას აქვს პრაქტიკული ღირებულება **LA**-გრამატიკის როგორც გასწორების, ისე გაუმჯობესების თვალსაზრისებით. ანალიზის წყვეტის მიზეზისა და წინადადების საწყისის გრამატიკული სტრუქტურა გვკარნახობს გრამატიკის რომელი წესია დარღვეული, როგორი უნდა ყოფილიყო შემავალი კატეგორიები და ა. შ.. პირიქით, თუ სისტემა ამუშავებს არაგრამატიკულ შემავალს (ნეგატიური ტესტირებისას გამოკვეთილი შეცდომა), მაშინ **LA**-ანალიზი გვაწვდის იმ ადგილის ზუსტ გრამატიკულ აღწერას, სადაც არაგრამართული გრამატიკული აღნაგობის გამო წყვეტა უნდა ყოფილიყო. ასეთ შემთხვევაში ახალი კორექტული კონსტრუქცია შეიძლება აიგოს ნაბიჯ-ნაბიჯ წინადადების კორექტული საწყისიდან და შემდეგ ისევ იმ წინადადების კორექტულ საწყისთან დაბრუნებით. **LA**-გრამატიკის ეს დამახასიათებელი თვისებები განპირობებულია მისი დროში წრფივი დერივაციული სტრუქტურით და ფართოდ იქნება გამოყენებული III ნაწილში, ინგლისური (**LA-E1 – LA-E3** 17-ე თავში, **LA-E3** 23.4 ქვეთავში) და გერმანული (**LA-D1 – LA-D4** 18-ე თავში) ენების სულ უფრო და უფრო ფართო ფრაგმენტების დასამუშავებლად.

სავარჯიშოები

10.1 პრეტაზი

- რა იგულისხმება ოპერაციების მარცხნივ ჯგუფდებადობაში? ოპერაციების კიდევ სხვა რა დაჯგუფებია შესაძლებელი?
- ბუნებრივი ენის რომელი თვისება არის ფორმალურად მოდელირებული მარცხნივ ჯგუფდებადი დერივაციული გამოყვანით?
- შეადარეთ ერთმანეთს შესაძლო გაგრძელებათა და შესაძლო ჩანაცვლებათა პრინციპები. როგორ უკავშირდებიან ეს განსხვავებული პრინციპები დროში წრფივობისა და შემადგენელი სტრუქტურების ინტუიციურ ცნებებს? რომელი ელემენტარული ფორმალიზმი რომელ პრინციპზეა დაფუძნებული?
- ახსენით დამოკიდებულება განსხვავებული წესის ფორმატებსა და **C-**, **PS-** და **LA**-გრამატიკების კონცეპტუალურ დერივაციულ დალაგებებს შორის.
- რა არის სინტაგმა?

10.2 ქვეთავი

1. ახსენით **LA**-გრამატიკის ალგებრული განსაზღვრება.
2. მარცხნივ ჯგუფდებადი წესის რომელი ნაწილია გამოყენებული ნიმუშების დაწყვილებაში?
3. მარცხნივ ჯგუფდებადი წესის რომელი ნაწილი იღებს მონაწილეობას სასრულ მდგომარეობისანი გარდამქმნელი სქემის განსაზღვრებაში?
4. რატომ შედის გარდაქმნა მდგომარეობაში ერთი და იმავე წესებით და რატომ გამოდის იგი იქიდან განსხვავებული წესებით?
5. ჩამოაყალიბეთ $a^k b^k$ ენის **LA**-გრამატიკა, ახსენით თუ როგორ მუშაობს იგი და შეადარეთ შესაბამის **C**- (იხ. 7.4.2) და **PS**- გრამატიკებს (იხ. 7.1.3).
6. განსაზღვრეთ $a^{2k} b^{2k}$ ენის **LA**-გრამატიკა.
7. განსაზღვრეთ $a^{k_1} b^{k_2}$ ენის **LA**-გრამატიკა.
8. რა არის მარცხნივ-ასოციაციური დერივაციის დამთავრების პირობები?

10.3 ქვეთავი

1. რა დამოკიდებულებაა მარცხნივ-ასოციაციურ ხისებრ სტრუქტურებსა და სტრუქტურულ სიებს (ცხრილებს) შორის? რომელი ფორმატია საუკეთესოდ მორგებული გამოთვლითი ანალიზისათვის და რატომ?
2. აღწერეთ **NEWCAT** პარსერის გამოსავალის დანომრილი კომპლექტები და ახსენით მათი ორმაგი ფუნქცია.
3. რატომ არის ხაზგასასმელი ის, რომ $a^k b^k c^k$ ენის **LA**-გრამატიკა ასე ემსგავსება $a^k b^k$ ენის **LA**-გრამატიკას. თქვენი პასუხი დააფუძნეთ შემდეგ ცნებებზე: **PS**-იერარქია, ენობრივი კლასი, სირთულე და პარსერის ალგორითმი.
4. განსაზღვრეთ **LA**-გრამატიკა $\{a^k b^k c^k\}^+$, $a^k b^k c^k d^k$ და $a^k b^k c^k d^k e^k f^k$ ენებისათვის. რა არის ამ ენების **PS**-გრამატიკული ენობრივი კლასი? რა არის მათი **LA**-გრამატიკული სირთულე?

10.4 ქვეთავი

1. რა დამოკიდებულებაა **LA**-პარსერს, **LA**-გრამატიკასა და **LA**-გენერატორს შორის?
2. რა მნიშვნელობა აქვს, რომ მაქსიმალური სიგრძისა და განსაკუთრებული სიტყვების ზუსტ განსაზღვრებას **LA**-გენერატორისათვის? ამ შეზღუდვებიდან რომელია განსაკუთრებით სასარგებლო ხელოვნური ენებისათვის და რომელი ბუნებრივი ენებისათვის?
3. ახსენით თუ რით განსხვავდება **PS**-გრამატიკული დერივაციის დასაწყისი **LA**-გრამატიკული დერივაციის დასაწყისისაგან. არის თუ არა **PS**-გრამატიკული დერივაციის დასაწყისის სიმბოლოს ანალოგი **LA**-გრამატიკულ დერივაციაში?
4. არის თუ არა კავშირი ფორმალურ **LA**-ანალიზსა (იხ. 10.4.1) და ენის გაგების სქემას შორის (იხ. 5.4.1)?
5. ხედავთ თუ არა კავშირს ფორმალურ **LA**-გენერაციასა (იხ. 10.4.2) და ენის წარმოების სქემას შორის (იხ. 5.4.2)?
6. ახსენით ტიპობრივი ტრანსპარენტულობის ზუსტი თვისებები **LA**-გრამატიკაში.

10.5 ქვეთავი

1. რა მსგავსება და განსხვავებაა ბუნებრივი ენის **C**-გრამატიკულ და **LA**-გრამატიკულ ანალიზებს შორის?

2. რით განსხვავდებიან ბუნებრივი ენის შემადგენელ სტრუქტურალური ანალიზისა და მარცხნივ ჯგუფდებადი ანალიზის შუალედური გამოსახულებები?
3. რა არის მიზეზი იმისა, რომ დისკონტინიალური ელემენტების დამუშავება არის პრობლემური ბუნებრივი ენის შემადგენელ სტრუქტურაზე დაფუძნებული ანალიზისათვის, მაგრამ არა ბუნებრივი ენის დროში წრფივ სტრუქტურებზე დაფუძნებული ანალიზისათვის?
4. რომელ წერტილში წყვეტს **LA**-პარსერი არაგრამატიკული გამოსახულების ანალიზებას?
5. დაასახელეთ სამი განსხვავებული მიზეზი იმისა, თუ რატომ შეუძლია **LA**-პარსერს ანალიზის შეწყვეტა მანამ, სანამ მიაღწევს შესავლის წარმატებულ დასასრულს.

11 თავი **LA** გრამატიკის იმრაობია

ამ თავში **LA**-გრამატიკის წესების ბუნებრივი შეზღუდვების გზით განსაზღვრულია სხვადასხვა ტიპის **LA**-გრამატიკები და თითოეული მათგანი დახასიათებულია მათი წარმოქმნის უნარისა და გამოთვლითი სირთულის მიხედვით.

- 11.1 ქვეთავი გვიჩვენებს, რომ ძირეული **LA**-გრამატიკა, ანუ **LA**-გრამატიკა, მისი ალგებრული განსაზღვრების შეუზღუდვავი ფორმით, წარმოქმნის რეკურსული ენების კლასს. 11.2 ქვეთავში აღწერილია **LA**-გრამატიკის სხვადასხვა შესაძლო შეზღუდვები და განსაზღვრულია **A**-, **B**- და **C**-ტიპის **LAG**-ები. 11.3 ქვეთავში ახსნილია რეკურსული და არარეკურსული ორაზროვნების მიზეზები და გაშუქებულია მათი გავლენა დერივაციებში გამოყენებულ წესთა ოდენობაზე. 11.4 ქვეთავი საზღვრავს და ადარებს იმ ელემენტარულ ოპერაციებს, რომლებიც გამოიყენება სირთულის გამოსათვლელად კონტექსტისგან თავისუფალ **PS**-გრამატიკებსა და **C-LAG**-ებში და აშუქებს **C-LAG**-ებისა და ავტომატთა თეორიიდან ცნობილი დეტერმინისტული და არადეტერმინისტული ავტომატების ურთიერთდამოკიდებულებებს. 11.5 ქვეთავში ისაზღვრება **C-LAG** კლასის წრფივი,

ანუ **C1**-, პოლინომური, ანუ **C2**-, და ექსპონენციალური, ანუ **C3**- ქვეირარქიები, რომლებიც ერთმანეთისგან განსხვავდებიან მხოლოდ მათი ორაზროვნების_ხარისხით.

11.1 შეუზღუდავი LAG-ების ფაროქმითი უნარი

გრამატიკული ფორმალიზმის წარმოქმნითი უნარი პირდაპირ კავშირშია მის ალგებრულ განსაზღვრებასთან და დგინდება იმ სპეციფიკური ფორმის წესებით, რომლებიც, თავის მხრივ, განსაზღვრავენ იმას, თუ რა სახის სტრუქტურები წამოიქმნება ამ ფორმალიზმით, და რა სახის არა.

LA გრამატიკის ალგებრული განსაზღვრების თანახმად (იხ. 10.2.1), წესი r_i შედგება ორი ნაწილისგან, ესენია: **co_i** კატეგორიული ოპერაცია და მასთან მიკავშირებული **rp_i** წესთა პაკეტი. ამასთან, **co_i** კატეგორიული ოპერაცია განსაზღვრულია, როგორც სრულად რეკურსული ფუნქცია.

ფუნქცია რეკურსულია თუ მისი შესაბამისობის წესი შეიძლება აღიწეროს მექანიკური ლოგიკური ალგორითმის სახით. ამასთან, ფუნქცია არის სრულად რეკურსული თუ ამ ალგორითმით განსაზღვრის არის ნებისმიერ წევრს ეთანადება მნიშვნელობათა არის ერთი რომელიღაც წევრი. ნაწილობრივ რეკურსული ფუნქციებისთვის, მეორე მხრივ, ალგორითმული შესაბამისობა არ არის განსაზღვრული განსაზღვრის არის რაღაც ნაწილზე.

ინტუიციური თვალსაზრისებით, სრულად რეკურსული ფუნქციებით იგება ის სტრუქტურები, რომელთა გამოთვლა შესაძლებელია ცხადი და ამომწურავი გზით. ფორმალური თვალსაზრისებით კი, სრულად რეკურსული ფუნქციებით მოიცემა რეკურსულად წოდებული ენების კლასი.

ენა რეკურსულია მაშინ, და მხოლოდ მაშინ, როდესაც არსებობს ალგორითმი (ე.ი. სრულად რეკურსული ფუნქცია), რომელიც ნებისმიერი შემავალი მოცემულობისთვის სასრული რაოდენობა ბიჯების მეშვეობით საზღვრავს ეკუთვნის თუ არა იგი მოცემულ ენას. რეკურსულ ენათა კლასი აერთიანებს ყველა ამონსნად (**decidable**) ენას. ამგვარად, რეკურსულ ენათა კლასი წარმოადგენს ენათა იმ ყველაზე ფართო კლასს, რომლის წევრებსაც აქვთ სრულად სპეციფირებადი სტრუქტურა (**completely specifiable structure**).¹

LA-გრამატიკის ძირეული, ანუ შეუზღუდავი ფორმალიზმის (იხ. 10.2.1) წარმოქმნითი უნარი შემდეგნაირად ხასიათდება.

11.1.1 შეუზღუდავი LA-გრამატიკის ფაროქმითი უნარი

შეუზღუდავი **LA**-გრამატიკა წარმოქმნისა და ამუშავებს_რეკურსულ ენებს და მხოლოდ მათ.

¹ ავტომატთა თეორიაში რეკურსული ენები განსაზღვრულია როგორც ისეთი ენები, რომელთა ნებისმიერი გამოსახულება შეიძლება გამოცნობილ იქნას ერთი რომელიმე ტიურინგის მანქანის მიერ სასრული რაოდენობა ბიჯების მეშვეობით („ყველა შემაგალზე გაჩერებადი“ ჯ. ე. ჰოკროტ-ულმანი & ჯ. დ. ულმანი (“Halt on all inputs”, J.E. Hopcroft & J.D. Ullman), 1970. გვ. 150). **PS**-გრამატიკული იერარქია ვერ უზრუნველყოფს რეკურსული ენების ფორმალურ დახსასიათებას. – განსხვავებით რეგულარული, კონტექსტუსისგან თავისუფალი, კონტექსტზე დამოკიდებული და რეკურსულად გადათვლადი ენებისაგან, რომელებიც განსაზღვრულია არიან როგორც ავტომატურ-თეორიული, ისე **PS** გრამატიკული თვალსაზრისებით.

ამ მოსაზრების შემადგენელი ნაწილებიდან ორივე ჩამოყალიბებულია და დამტკიცებულია თეორემების სახით.²

11.1.2 თეორემა 1

შეუზღუდავი \mathbf{LA} -გრამატიკა წარმოქმნის და ამუშავებს მხოლოდ რეკურსულ ენებს.

დამტკიცება: დავუშვათ, რომ შემავალ სტრიქონს აქვს რაღაც სასრული n სიგრძე და რომ შემავალი სტრიქონის ყოველ სიტყვას აქვს სასრული რაოდენობა წაკითხვები (>0).

კომბინაციური ბიჯი n : **STs** საწყისი მდგომარეობების სასრული სიმრავლე და პირველი W_1 სიტყვის სხვადასხვა შესაძლო წაკითხვები შედეგად იძლევა მართებულად აგებული გამოსახულებების $WE_1=\{(ss' rp_s) / ss' \in (W^* \times C^*)\}$ სასრულ სიმრავლეს.

კომბინაციური ბიჯი n : კომბინაციური ბიჯი **k-1 (k>1)**, წარმოქმნის მართებულად აგებული გამოსახულებების $WE_k=\{(ss' rp_i) / i \in n, ss' \in (W^* \times C^*)\}$ და თითოეული ss' გამოსახულებების სიგრძე არის $k\}$ სასრულ სიმრავლეს. ამასთან, მომდევნო W_{k+1} სიტყვა ხასიათდება წაკითხვების სასრული ოდენობით. აქედან გამომდინარე, WE_k სასრული სიმრავლის და W_{k+1} მომდევნო სიტყვის სხვადასხვა შესაძლო წაკითხვების ასევე სასრული სიმრავლის დეკარტული ნამრავლი იძლევა წყვილთა სასრულ სიმრავლეს. ეს წყვილები მიკავშირებულნი არიან წესთა პაკეტთან, რომელიც სასრული ოდენობა წესებისგან შესდგება. ამგვარად, k კომბინაციური ბიჯი წარმოქმნის წინადადების ახალი საწყისების მხოლოდ სასრულ ოდენობას. ამასთან, ამ სასრული ოდენობის ახალი საწყისების წარმოქმნა, ანუ დერივაცია ამოხსნადია, რადგან მათი წარმომქმნელი კატეგორიული ოპერაციები სრულად განსაზღვრული რეკურსული ფუნქციებია.

რ.დ.გ.

რადგან ნებისმიერი სასრული სიგრძის შემავალი გამოსახულების ნებისმიერი შესაძლო მარცხნივ ჯგუფდებადი ანალიზი კეთდება სასრული რაოდენობა ბიჯების გამოყენებით, რომელთაგან თითოეული ამოხსნადია, უფლება გვაქვს დავასკვნათ, რომ არც \mathbf{LA} -გრამატიკასა და არც მის პარსერს არანაირი შემაფერხებელი ვითარება არ ექნება. ეს იმას ნიშნავს, რომ \mathbf{LA} -გრამატიკა აკმაყოფილებს 3-ე პირობას წარმოქმნითი გრამატიკის 9.5.2 პირობებიდან.

11.1.3 თეორემა 2

შეუზღუდავი \mathbf{LA} გრამატიკა წარმოქმნის და ამუშავებს ყველა რეკურსულ ენას.

დამტკიცება:³ ვთქვათ W არის ნებისმიერად აღებული L რეკურსული ენის ალფაბეტი. რადგან L რეკურსულია, არსებობს L ენის მახასიათებელი სრულად განსაზღვრული $\rho: W^* \rightarrow \{0,1\}$ რეკურსული ფუნქცია. ვთქვათ \mathbf{LAG}^L არის შემდეგნაირად განსაზღვრული \mathbf{LA} -გრამატიკა:

\mathbf{LAG}^L სიტყვითი გამოსახულებების სიმრავლე არის W .
კატეგორიული სეგმენტების სიმრავლე არის $C =_{\text{def}} W \cup \{0,1\}$.

² **Col**, თეორემები 1 და 2 გვ. 134.

³ ეს დამტკიცება ეპუთვნის დიანა სკოტს.

ნებისმიერი $e, f, \epsilon W^+$ გამოსახულებებისათვის, $[e (f)] \in LX$ მხოლოდ მაშინ, როცა $e = f$.
 $LX =_{\text{def}} \{[a (a)], [b (b)], [c (c)], [d (d)], \dots\}$
 $ST_s =_{\text{def}} \{[(seg_c) \{r_1, r_2\}]\},$ სადაც $seg_c \in \{a, b, c, d, \dots\}$
 $r_1:(X) (seg_c) \Rightarrow (X seg_c) \quad \{r_1, r_2\}$
 $r_2:(X) (seg_c) \Rightarrow \rho (X seg_c) \quad \{\}$
 $ST_F =_{\text{def}} \{[(1) rp_2]\}$

ნებისმიერი კომბინაციური ბიჯის შემდეგ წესთა rp_1 პაკეტი გვთავაზობს ორ არჩევანს: გამოვიყენოთ r_1 წესი შემავალი სტრიქონის კითხვის გასაგრძელებლად, ან გამოვიყენოთ r_2 წესი იმის შესაძლებლად, უკვე წაკითხული არის თუ არა L ენის მართვულად აგებული გამოსახულება. ამ უკანასკნელ შემთხვევაში ρ ფუნქცია გამოიყენება იმ შემავალი კატეგორიების კონკეტუნაციაზე, რომლებიც შემავალი გამოსახულებების იდენტურები არიან. ამასთან, თუ r_2 წესის ამ გამოყენების შედეგი არის $[(1) rp_2]$ ⁴, მაშინ შემავალი გამოსახულება დაშვებულია (**accepted**), ხოლო თუ ეს შედეგი არის $[(0) rp_2]$, მაშინ შედეგი უარყოფილია (**rejected**).

რადგანაც **LAG^L** გრამატიკის კატეგორიული ოპერაცია შეიძლება იყოს ნებისმიერი სრულად რეკურსული ფუნქცია, **LAG^L** შესაძლებელია დაფუძნდეს L ენის ρ მახასიათებელ ფუნქციაზეც. ეს იმას ნიშნავს, რომ ამგვარად განსაზღვრული **LAG^L** ამუშავებს და წარმოქმნის თავდაპირველად სრულიად ნებისმიერად აღებულ L რეკურსულ ენას.

რ.დ.გ.

LAG^L-ის არადამასრულებელი კომბინაციური ბიჯები ემსახურება შემავალი გამოსახულებების შემადგენლებების კატეგორიზებულ წაკითხვას. დამასრულებელ კომბინაციურ ბიჯში უკვე სრულად კატეგორიზებული გამოსახულება ანალიზდება ერთი ძალიან რთული კატეგორიალური ოპერაციით, რომელიც დასამუშავებელი ენის ρ მახასიათებელ ფუნქციას ემთხვევა.

შეუზღუდავი **LA**-გრამატიკები აღინიშნება, როგორც **A-LAG**-ები, და ისინი წარმოქმნიან ყველა რეკურსულ ენას.

11.1.4 შეუზღუდავი **LA**-გრამატიკების (**A-LAG**-ების) კლასის განსაზღვრება.

A-LAG-ების, ანუ შეუზღუდავი **LA**-გრამატიკების კლასი შესდგება შეუზღუდავი **LA**-გრამატიკებისაგან და წარმოქმნის ყველა რეკურსულ ენას.

ამის საპირისპიროდ, შეუზღუდავი **PS** გრამატიკები (იხ. 8.1.1) წამოქმნიან რეკურსულად გადათვლად ენებს. ამგვარად, **LA** გრამატიკისა და **PS** გრამატიკის ძირები, ანუ შეუზღუდავი ფორმალიზმები და მათი შესაბამისი ალგებრული განსაზღვრებები არ არიან ეკვივალენტურნი.

11.2 A-, B- და C-LAG-ების LA-იმპარატივი

გრამატიკული ფორმალიზმის სირთულე – გაგებული როგორც ნებისმიერი შემავალი მოცემულობის ანალიზისთვის საჭირო ოპერაციების ზედა საზღვარი – დამოკიდებულია შემდეგ ორ პარამეტრზე:

⁴ ე. ი. თუ ρ შემავალი გამოსახულების $(X seg_c)$ კატეგორიას ასახავს **(1)** კატეგორიაში.

11.2.1 სირთულის პარამეტრები

- ყველაზე ცუდ შემთხვევაში გამოყენებული ალგორითმული ბიჯების ოდენობა (**amount**).
- ყველაზე ცუდ შემთხვევაში მოცემული სიგრძის შემავალი მოცემულობის დამუშავებისთვის გამოყენებული წესების რიცხვი (**number**).

ეს პარამეტრები ერთმანეთისგან დამოუკიდებელნი არიან და, როგორც წესი, მიყენებადნი არიან წესებზე დაფუძნებულ ნებისმიერ გრამატიკულ ფორმალიზმზე.

LA-გრამატიკაში ალგორითმული ოდენობითი პარამეტრი დამოკიდებულია კატეგორიალურ **CoI** ოპერაციებზე, მაშინ როცა წესთა რიცხვითი პარამეტრი სრულად ისაზღვრება ორაზროვნების ხარისხით. ამგვარად, ბუნებრივად იხაზება **LA** გრამატიკის შეზღუდვის ორი ძირითადი მიმართულება.

11.2.2 LA გრამატიკის შეზღუდვის მთავარი მიმართულებები

R1: შეზღუდვები კატეგორიალურ ოპერაციებზე იმ მიზნით, რომ შემოგვაზღვროთ ამ წესების გამოყენებისას გამოყენებადი ალგორითმული ბიჯების მაქსიმალური ოდენობა.

R2: შეზღუდვები ორაზროვნების ხარისხზე იმ მიზნით, რომ შემოგვაზღვროთ წესთა შესაძლო გამოყენების მაქსიმალური რიცხვი.

R1 მიმართულება, თავის მხრივ, გვთავაზობს ორი განსხვავებული სახის შეზღუდვებს.

11.2.3 კატეგორიალური ოპერაციების შესაძლო შეზღუდვები

R1.1: კატეგორიების სიგრძეების ზედა საზღვრების სპეციფირება;

R1.2: კატეგორიალური ოპერაციების განსაზღვრებებში გამოყენებული მოდელების სპეციფირება.

კატეგორიალური ოპერაციების შეზღუდვის ეს ორი გზა შედეგად გვაძლევს **A-LAG**-ების კლასის ორ ქვეკლასს, რომლებიც შესაბამისად **B-LAG** და **C-LAG** ქვეკლასებად იწოდებიან.

11.2.4 B-LAG-ების კლასის განსაზღვრება

შემოსაზღვრული **LA**-გრამატიკების კლასი (**class of bounded LA-grammar**), ანუ **B-LAG**-ები შედგება იმ **LA**-გრამატიკებისგან, რომლებშიც ნებისმიერი დასრულებული მართებულად აგებული **E** გამოსახულების ნებისმიერი შუალედური წინადადების საწყისის კატეგორიის სიგრძე შემოსაზღვრულია **k·n** –ნამრავლით (**n** არის **E** გამოსახულების სიგრძე, **k** არის კონსტანტი).

B-LAG-ების მიერ წარმოქმნილი ენების კლასი ემთხვევა **PS**-გრამატიკულ იერარქიაში კონტექსტზე დამოკიდებული ენების კლასს. დამტკიცება⁵ ანალოგიურია 11.1.3 პუნქტში არსებული დამტკიცებისა და ემყარება შესაბამის შეზღუდვებს იმ წრფივად შემოსაზღვრულ ავტომატზე (**linearly bound automata**), რომელიც უკანასკნელი კომბინაციური ბიჯის **LA**-კატეგორიებს იყენებს, როგორც ლენტას. **B-LAG**-ები არის **A-LAG**-ების საკუთრივი ქვესიმრავლე. რაც

⁵ **CoL**, თეორემა 5, გვ. 142.

განპირობებულია იმით, რომ **CS ⊂ REC**.⁶

C-LAG-ით აღნიშნულია **B-LAG**-ების ის ქვეკლასი, რომელიც მოიცემა **R1.2** სახის შემზღვდავი პირობით. კერძოდ, **C-LAG**-ებში **co** კატეგორიალური ოპერაციების ალგორითმული ბიჯების მაქსიმალური ოდენობა გარკვეული მუდმივი, ანუ კონსტანტური რიცხვით უნდა ისაზღვრებოდეს.

თუ კატეგორიალურ ოპერაციებს გავიაზრებთ სრულიად ნებისმიერ სრულად განსაზღვრულ რეკურსულ ფუნქციებად, მაშინ არ გვექნება იმ მაქსიმალური ოდენობის დადგენის საშუალება, რომლითაც შემოიზღვდებოდა მათი მეშვეობითვე ორგანიზებული გამოთვლები. მეორე მხრივ, თუ კატეგორიალურ ოპერაციებს განვსაზღვრავთ ფორმალური ნიმუშების, ანუ ფორმალური მოდელების ტერმინებში, მაშინ შესაძლებელი ხდება იმ წესების ერთმანეთისაგან განრჩევა, რომლებიც გამოთვლების კონსტანტურ და არაკონსტანტურ ოდენობებს ითხოვენ.

შემდეგ სქემატურ წესებში კატეგორიალური ოპერაციებისთვის საჭირო გამოთვლების ოდენობა მუდმივია და დამოკიდებულია შეყვანილი კატეგორიის სიგრძეზე.

11.2.5 სემატური ფესხი კონსტანტური პატეგორიალური რაერაციებით

r_i:(seg₁...seg_k X) cat₂⇒ cat₃ rp_i
r_i:(X seg₁...seg_k) cat₂⇒ cat₃ rp_i
r_i:(seg₁...seg_m X seg_{m-1}...seg_k) cat₂ ⇒ cat₃ rp_i

ამ სქემებს აქვთ საერთო ის, რომ მათი მეშვეობით მოცემულ კატეგორიალურ ოპერაციებს **ss** კატეგორიაში აქვთ შესამოწმებელი ზუსტად **k** სეგმენტი. კატეგორიალური ოპერაციების ეს მოდელები არიან კონსტანტური, რადგან ისინი **seg₁...seg_k** სეგმენტების მიხედვით კატეგორიებს ყოველთვის გარე მხრიდან ამოწმებენ, და ეს მიუხედავად იმისა, თუ რა სიგრძის მიმდევრობაა კატეგორიის შუაში.

ის კატეგორიალური ოპერაციები, რომლებსაც ეს თვისება არა აქვთ, ვერ უზრუნველყოფენ მუდმივ ზედა საზღვარს გამოთვლითი საჭიროების ოდენობაზე. სქემატური წესი არაკონსტანტური კატეგორიული ოპერაციებით განხილულია ქვემოთ.

11.2.6 სემატური ფესი პატეგორიულ რაერაციათა არაკონსტანტური ღდენობით

r_i:(X seg₁...seg_k Y) cat₂ ⇒ cat₃ rp_i

ამ სახის წესებში საჭირო ხდება უცნობი ოდენობის კატეგორიული სეგმენტების გადამოწმება (წარმოდგენილი ან **X**-ით, ან **Y**-ით, რაც დამოკიდებულია იმაზე, თუ რომელ მხრიდან იწყება ძებნა). რადგან **X**-ს და **Y**-ს შესაძლებელია პქნდეთ ნებისმიერი სიგრძე, ამ წესების გამოყენებისას კატეგორიული ოპერაციის შესასრულებლად საჭირო გამოთვლების ოდენობა დამოკიდებული ხდება **ss**-კატეგორიის მთლიან სიგრძეზე.

⁶ თეორემა კ. ე. პოპკროფტი და კ. დ. ულიმანი 1979, თეორემა 9.8 გვ. 228. **CS** აღნიშნავს კონტექსტზე დამოკიდებული ენების კლასს, ხოლო **REC** აღნიშნავს რეგულურ ენათა კლასს.

LA გრამატიკები, რომელთა წესებიც იყენებენ მხოლოდ კონსტანტურ კატეგორიულ ოპერაციებს, ადგენენ **C-LAG**-ების კლასს.

11.2.7 C-LAG-ების კლასის განსაზღვრება

კონსტანტური (constant) **LA** გრამატიკების კლასი, ანუ **C-LAG**-ების კლასი შედგება იმ გრამატიკებისაგან რომლებშიც არცერთი **CO_i** კატეგორიული ოპერაცია არ უყურებს რაიმე სასრულ **k** რიცხვზე მეტ სეგმენტს წინადადების საწყისის კატეგორიაში.⁷

LA გრამატიკული კლასები დღემდე მიიჩნევა, რომ შესდგება შემდეგი იერარქიული კლასებისაგან:

11.2.8 A-LAG-ების, B-LAG-ებისა და C-LAG-ების იერარქიები

A-LAG-ების კლასი იღებს და წარმოქმნის ყველა რეკურსულ ენას, **B-LAG**-ების კლასი იღებს და წარმოქმნის ყველა კონტექსტზე დამოკიდებულ ენას, ხოლო **C-LAG**-ების კლასი იღებს და წარმოქმნის მრავალ კონტექსტზე დამოკიდებულ ენას და ყველა კონტექსტისგან თავისუფალსა და რეგულარულ ენას.

ის, რომ ყველა კონტექსტისგან თავისუფალი ენა გამოცნობადია და წარმოქმნადია **C-LAG**-ების მიერ, დამტკიცებულია⁸ იმის საფუძველზე, რომ **C-LAG**-ების კატეგორიალური ოპერაციები მოდიფიცირებას უკეთებენ მხოლოდ **cat_i** კატეგორიის ნიმუშის საწყისს, რაც ბუნებრივ შესაბამისობაშია ძირსმბიძგი (**pushdown**) ავტომატების შეზღუდვებთან. კონტექსტისგან თავისუფალი ენების კლასი არის საკუთრივი ქვესიმრავლე **C**-ენების, რადგან **C**-ენები მოიცავენ აგრეთვე კონტექსტზე დამოკიდებულ ენებსაც (მაგ. 10.3.3). რეგულარული ენების კლასი დამტკიცებადია და წარმოქმნადია იმ **C-LAG**-ების მიერ, რომელთა კატეგორიების სიგრძე შემოსაზღვრულია აბსოლიტური **k** მუდმივით.⁹

ამგვარად, საბუთდება, რომ **LA**-გრამატიკაში იგება **PS** გრამატიკული იერარქიის რეგულარული, კონტექსტისგან თავისუფალი და კონტექსტზე დამოკიდებული ენათა კლასები.¹⁰

⁷ ამ გრამატიკებს მხოლოდ ეს სასრული მუდმივი განასხვავებს ერთმანეთისაგან.

⁸ 8 **CoL**, თეორემა 4, გვ. 138. კონტექსტისგან თავისუფალი **C-LAG**-ები, მოკლეთ **CF-LAG**-ები შედგება მხოლოდ შემდეგი ფორმის წესებისგან: $r_i:(aX) \rightarrow (\alpha X) \ r_{pi} \ (a, \ a,b \in C \text{ და } \alpha \in C^+)$. ამგვარად, შეზღუდვები **ss** და **ss'** კატეგორიებზე შესაბამება **PDA**-ს მუშაობის პრინციპს, რომელსაც შეუძლია ერთდროულად ამობეჭდოს არა მარტო ერთი სიმბოლო, არამედ სიმბოლოთა მთელი მიმდევრობა. (იხ. ჯ. ე. ჰოპკროფტი და ჯ. დ. ულმანი, 1979, თავი 5.2). შემდეგი მოსაზრებები ითხოვენ დამტკიცებას:

1. ყოველი **PDA M**-ისთვის, **CF-LAGσ** შეიძლება აიგოს ისე, რომ $L(M)=L(\sigma)$.
აქედან გამომდინარეობს, რომ $CF \subseteq C_{cf}$.

2. ყოველი **CF-LAG σ**-ისთვის, **PDA M** შეიძლება აიგოს ისე, რომ $L(\sigma)=L(M)$.
აქედან გამომდინარეობს, რომ $C_{cf} \subseteq CF$.

1-ის და 2-ის დამტკიცებისას ყურადღება უნდა მიექცეს იმას, რომ **CF-LAG** იყენებს წესებს, ხოლო **PDA** იყენებს გამონათქვამებს, რაც მთლად ერთი და იგივე არაა. ამგვარად, აუცილებელია გვქონდეს კონსტრუქცია იმ პროცედურის რომელიც გარდაქმნის გამონათქვამებს წესებად და წესებს გამონათქვამებად. ეს დიდი, მაგრამ არც ძალიან რთული ამოცანა. – შევნიშნოთ, რომ ϵ -გადაადგილებები არ არის დაშვებული **CF-LAG**-ებში, მაგრამ იგი დაშვებულია **PDA**-ებში (ჯ. ე. ჰოპკროფტი და ჯ. დ. ულმანი 1979, გვ. 24). მეორეს მხრივ, ყოველი კონტექსტისგან თავისუფალი ენისათვის არსებობს ϵ -გადაადგილებების გარეშე მომუშავე **PDA** (პარისონი 1978, თეორემა 5.5.1) და ყოველი ϵ -თავისუფალი **PDA** -ისთვის შესაძლებელია აიგოს **CF-LAG**.

⁹ **CoL**, თეორემა 3, გვ. 138.

¹⁰ განსხვავებული გზა **LA**-გრამატიკის წარმოქმნითი უნარის განსაზღვრის შედგება – უკიდურეს შემთხვევაში

11.3 ማጠቃሚነት የሚከተሉት LA-ስነዱ አገልግሎት

LA-გრამატიკის შეზღუდვის მეორე მთავარი მიმართულება (იხ. **R2** 11.2.2) ეხება წესთა გამოყენებების რიცხვს. ეს რიცხვი განისაზღვრება შემდეგი ფაქტორებით.

11.3.1 მასთა გამოყენების რეცეპტ განვითარების ფაზი

წესთა გამოყენებების რიცხვი **LA**-გამოყვანაში დამკიდებულია:

1. შემავალი მოცემულობის სიგრძეზე;
 2. წესთა პაკეტის იმ წესების ოდენობაზე, რომლებიც გამოიყენება ამ წესის კომბინაციურ ბიჯეტი შემავალი წყვილის საანალიზოდ (წასაკითხად);
 3. თითოეულ კომბინაციურ ბიჯეტი არსებულ წაკითხვათა¹¹ ოდენობაზე.

ფაქტორი 1 გრამატიკისგან დამოუკიდებელია და სირთულის განსასაზღვრავ ფორმულებში აისახება სიგრძეზე მანიშნებელი **n** სიმბოლოთი.

ვაქტორი 2 გრამატიკაზე დამოკიდებული კონსტანტაა. მაგალითად: თუ **LAG**-ების ყველაზე დიდ წესთა პაკეტი შეიცავს 5 წესს, მაშინ მაქსიმალური რიცხვი წესთა გამოყენებების ცალსახა წარმოებაში იქნება არაუგრძელეს 5-ი ნამრავლისა, სადაც 1 შემავალი გამოსახულების სიგრძეა.

մեռլուծ Յ-յ ջայտկորս Շեշմլուս գամոցանու Եւստա գամոցենքին մտլանան ըստեզ Մրգով պայմանական մուգայուղ մարզենու չափութեած կոմիտենու օրու դամադրեած կոտեա օդպահ մեռլուծ մաժն, Ռուսասաւ մոմունարյ Եւստա Ազգայի արու շրտե մեժու օւղու Եւսո, Ռուսասաւ Շեշմլուս Վարմագրեալ սլավոն Շեմազալ մուգայուղակաչյ. օս, Շեմազալ մուգայուղակաչյ շրտե մեժու Եւսու Շեշմլուս արա օպու Վարմագրեալ, ամուսնութեած ամ Եւսու Շեմազալ Արմագանի.

თეორიულად – 10.2.1 ალგებრული განსაზღვრების მე-4 პუნქტის შეცვლით. მაგალითად, თუ კატეგორიალურ თეორიაციებად დაშვებული იქნებოდა ნაწილობრივ რეკურსული ფუნქციები, მაშინ **LA**-გრამატიკა წარმოქმნიდა რეკურსულად გადათვლად ენებს, რაც **LA**-გრამატიკას გახდიდა **PS** გრამატიკის სუსტად ეკვივალენტურს. მეორეს მხრივ, თუ კატეგორიალურ თეორიაციებად დაშვებული იქნებოდა პრიმიტიული რეკურსული ფუნქციები, მაშინ თეორება 2-ის ანალოგიურად შეგვეძლო გვეჩენებია, რომ ასე გაგებული **LA**-გრამატიკა წარმოქმნიდა ზუსტად პრიმიტიულ რეკურსულ ენებს, რაც სტანდარტულ განსაზღვრებასთან შედარებით უფრო დაბალი ხარისხის წარმოქმნითი უნარის თაობაზე განიშნებს. – თუმცა მე-4 პუნქტის ალტერნატიული განსაზღვრების გამოყენება **LA**-გრამატიკის ქვეკლასების მისაღებად ნაკლებ გამართლებულად: კატეგორიალური თეორიაციების გადაზრება ხან ნაწილობრივ რეკურსულ, ხან სრულად რეკურსულ, ხანაც პრიმიტიულად რეკურსულ ფუნქციებად ძალიან უხეში მეორდა. მეორე მიზეზი ისაა, რომ პრაქტიკული ოფელსაზრისებიდან გამომდინარე ასე მიღებული ენობრივი კლასები ძალიან დიდია: პრიმიტიული რეკურსული ფუნქციები საკუთრივი ქვესიმრავლეა სრულად რეკურსული ფუნქციების. ამასთან, პრიმიტიული რეკურსული ფუნქციები საკუთრივი ქვესიმრავლის სახით შეიცავენ კონტექსტზე დამოკიდებული ენების მთელ კლასს. მესამე მიზეზი: 10.2.1 განაწერში კატეგორიალური თეორიაციები განსაზღვრულა როგორც სრულად რეკურსული ფუნქციები და ეს მართლდება იმ მნიშვნელოვანი საფუძვლის გათვალისწინებით, რომ **LA** გრამატიკას აქვს მაქსიმალური წარმნითი უნარი, იქამდე სანამ ამოქსნათა.

¹¹ სიმარტივისთვის აქ ორაზროვნების მხოლოდ სინტაქსური მიზეზებია განხილული. ლექსიკური ორაზროვნება, რომელიც ჩნდება სიტყვის მრავალნაირი გაგების გამო, ფორმალურ წევებში დადი ნაწილით აღმოვტვრილ იქნა, მაგრამ ამას ვერ ავიცილებთ ბუნებრივი ენის **LA**-გრამატიკულ ანალიზში. გამოთვლით სირთულეებზე ლექსიკური ორაზროვნების შესძლო ზეგავლენება განხილულია **CoL**-ში გვ. 157. და გვ.248.

ორი განსხვავებული წესის შემავალ პირობებს შორის შემდეგი ორი დამოკიდებულებიდან ერთ-ერთი ყოველთვის ძალისმიერია:

11.3.2 შესაძლო დამოკიდებულებანი შემავალ პირობებს შორის

1. **არათავსებადი შემავალი პირობები (incompatible input conditions):** ორ წესს აქვს არათავსებადი შემავალი პირობები, თუ არ არსებობს ისეთი შემავალი წყვილი, რომელიც მიღებადია, ანუ დამუშავებადია ორივე წესის მიერ.

2. **თავსებადი შემავალი პირობები (compatible input conditions):** ორ წესს აქვს თავსებადი შემავალი პირობები თუ არსებობს ერთი მაინც ისეთი შემავალი წყვილი, რომელიც მიღებადია ორივე წესის მიერ და ერთი მაინც ისეთი, რომელიც მიღებადია ერთი წესის მიერ, მაგრამ არა მეორის მიერ.

3. **იდენტური შემავალი პირობები (identical input conditions):** ორ წესს აქვთ იდენტური შემავალი პირობები თუ ყველა შემავალი წყვილი ან ორივე წესის მიერ მიღებადია, ან ორივე წესი უარყოფს მას.

არათავსებადი შემავალი პირობების მაგალითებია **(a X) (b)** და **(c X) (b)**, ასევე **(a X) (b)** და **(a X) (c)**. თუ წესთა პაკეტის ყველა წესს აქვს არათავსებადი შემავალი პირობები, მაშინ ამ წესთა სისტემის გამოყენება ვერ გახდება სინტაქსური ორაზროვნების მიზეზი.

11.3.3 არაორაზროვანი LA გრამატიკების განსაზღვრება

გრამატიკა არაორაზროვანია მხოლოდ მაშინ, როცა (i) წესთა თითოეული პაკეტისთვის ძალაშია ის, რომ მათ წესებს აქვთ არათავსებადი შემავალი პირობები და (ii) არ არსებობს არანაირი ლექსიკური ორაზროვნება.¹²

არაორაზროვანი **C-LAG**-ების მაგალითებია 10.2.2-ში $a^k b^k$ და 10.3.3-ში $a^k b^k c^k$ ენებისთვის აღწერილი გრამატიკები.

თავსებადი შემავალი_პირობის მაგალითია **(a X) (b)** და **(X a) (b)**. თავსებადი შემავალი პირობები **LA**-გრამატიკის სინტაქსური ორაზროვნების ფორმალური წინაპირობაა.

11.3.4 სინტაქსურად ორაზროვანი LA-გრამატიკების განსაზღვრება

¹² **LA**-გრამატიკა ლექსიკურად ორაზროვანია თუ მისი ლექსიკონი შეიცავს მინიმუმ ორ ისეთ ანალიზებად სიტყვას, რომელთა ზედაპირები, ანუ ენობრივი გამოსახულებები ერთი და იგივე. არალინგვისტური მაგალითი ლექსიკური ორაზროვნების არის პროპროზიტურულ წინადადებათა აღგებრა. მაგალითად **(xVyVz)&(...),** სახის გამოსახულებებში ცვლადები **x, y, z** და ა.შ. შეიძლება ლექსიკურად გაანალიზებულ იქნან როგორც **[x (1)]** და **[x (0)], [y (1)]** და **[y (0)],** და ა.შ., სადაც **[x (1)]** ანალიზით **x** გაიგება როგორც ჭეშმარიტი, ხოლო **[x (0)]** ანალიზით როგორც მცდარი. – თუ წესებზე-დაფუძნებულ გამოყვანაში სინტაქსური ორაზროვნება ჩნდება ერთზე მეტი წინადადების ახალი საწყისის წარმოქნით, ლექსიკურ ორაზროვნებას იწვევს მომდევნო სიტყვის ერთზე მეტი წაკითხვები. – **LA**-გრამატიკა შეძლება ერთდროულად ხასიათდებოდეს სინტაქსური და ლექსიკური ორაზროვნებებით. ამასთან, სინტაქსური ორაზროვნება შესაძლებელია გააზრდეს როგორც ლექსიკური ორაზროვნება, და პირიქით. (იხ. პაუზერი 1992, გვ. 303/4.)

LA გრამატიკა სინტაქსურად ორაზროვანია მხოლოდ მაშინ, როცა: (i) მასში არის ერთი მაინც წესთა ისეთი პაკეტი, რომელიც შეიცავს ერთ მაინც ისეთ წესთა წევილს, რომელთაც თავსებადი შემავალი_პირობები აქვთ, და (ii) არ არსებობს არანაირი ლექსიკური ორაზროვნება.

სინტაქსურად ორაზროვანი **LA**-გრამატიკით წარმოქმნილი შესაძლო წაკითხვების რიცხვი დამოკიდებულია მის ორაზროვან სტრუქტურაზე (**ambiguity structure**), რომელიც შესაძლებელია დახასიათდეს შედეგი ორი ნიშნით: **±გლობალური** (**±global**) და **±რეკურსული** (**±recursive**). დავიწყოთ გლობალური ნიშნით.

ენათმეცნიერებაში წინადადებას ეწოდება **სინტაქსურად ორაზროვანი (syntactically ambiguous)**, თუ მას აქვს ერთზე მეტი წაკითხვა, ანუ სტრუქტურული ანალიზი, როგორც შემდეგ მაგალითს.

11.3.5 +გლობალური სინტაქსური ორაზროვნება

flying airplane can be dangerous

ამ წინადადების ერთი წაკითხვა ეხება საპარო თვითმფრინავებს, მეორე კი მათ პილოტირებას. ეს ორაზროვნება არის +გლობალური რადგან ეს მთელი წინადადების თვისებაა.

+გლობალური ორაზროვნების გარდა ენის დროში წრფივი ანალიზისას წამოიქმნება აგრეთვე -გლობალური (ან ლოკალური) ორაზროვნება. მაგალითად 11.3.6 წინადადების მარცნიდან მარჯვნივ კითხვისას განსხვავებულ გაგებებს იძლევა წაკითხვა **barn** სიტყვამდე და სრული წაკითხვა, რაც ხსნის მანაძლე არსებული ორი განსხვავებული კითხვის პრობლემას.

11.3.6. –გლობალური სინტაქსური ორაზროვნება

The horse raced by the barn fell

პირველი კითხვით **raced by the barn** ინტერპრეტირებულია როგორც მთავარი კლაუზას პრედიკატი. ეს პირველჯერობი კითხვა არის –გლობალური, რადგან ის ელიმინირდება **fell** გაგრძელებით. კერძოდ, ეს მეორე, ანუ სრული კითხვა **raced by the barn** გამოსახულებას აინტერპრეტირებს რედუცირებულ რელაციური კლაუზად და იძლევა მთლიანი წინადადების სრულიად ერთსახა კითხვას. ასეთი მაგალითები იწოდება **ბარის გზის წინადადებად (garden path sentence)**, რადგან ისინი პირველსაწყისად გვთავაზობენ კითხვის ერთ გზას, რომელიც შემდეგ უქმდება.

LA-გრამატიკაში განსხვავება +გლობალურ და –გლობალურ ორაზროვნებებს შორის ისაზღვრება იმისდა მიხედვით ბოლოს ერთზე მეტი კითხვა რჩება (იხ. 11.3.5), თუ არა (იხ. 11.3.6). ამგვარად, ორივე, +გლობალური და –გლობალური წაკითხვები მუშავდება როგორც პარალელური დროში წრფივი დერივაციული შტოები. ეს სრულიად განსხვავდება **C-** და **PS-** გრამატიკებში ჩასმებზე დამყარებული მიდგომებისაგან, რომლებიც განარჩევენ მხოლოდ იმ ორაზროვნებებს, რომლებიც არიან +გლობალური სახის.

±გლობალური ორაზროვნებების განრჩევა არ ახდენს არსებით გავლენას **LA**-გრამატიკის სირთულზე და ეს ძირითადად კეთდება ლინგვისტური ინტერესების გათვალისწინებით. ±რეკურსული ორაზროვნებების განრჩევა, მეორე მხრივ, გადამწყვეტ გავლენას ახდენს ანალიზის სირთულეზე, რადგან მტკიცდება, რომ არარეკურსულ ორაზროვნებიან **LA**-გრამატიკაში ყოველ კომბინაციურ ბიჯზე ერთი რომელიმე წესის გამოყენების მაქსიმალური რიცხვი ლიმიტირდება გრამატიკაზე დამოკიდებული კონსტანტით. (იხ. 11.3.7, თეორემა 3).

ორაზროვნება არის რეკურსული, თუ ის წარმოიქმნება წესთა გამოყენების რეკურსულ კვანძში. სხვა სიტყვებით: გარკვეულ (**cat**, **rp_i**) მდგომარეობას გარკვეულად განსაზღვრული შემდეგი სიტყვისთვის აქვს რამოდენიმე გაგრძელება, თანაც ისეთი, რომ ერთი, ან რამოდენიმე მათგანი ამ გაგრძელებათა შედეგად კვლავ ბრუნდება თავდაპირველ მდგომარებაში, რითაც წარმოიქმნება ორაზროვნანი გახლების კვლავგამეორების იძულება. +რეკურსული ორაზროვნების მაგალითებია **SubsetSum** (იხ. 11.5.2) და **WW^R** და **WW** ენებისთვის 11.5.4 და 11.5.6 განაწერებში აღწერილი **C-LAG**-ები, რომელთაგან პირველი ამავდროულად +გლობალურია, ხოლო ბოლო – –გლობალური.

ორაზროვნება –რეკურსულია, თუ არცერთი ორაზროვნული გაგრძლება არ ბრუნდება იმ მდგომარეობაში, რომელმაც წარმოშვა იგი. –რეკურსული ორაზროვნების მაგალითებია **a^kb^kc^mdⁿ უa^kb^mcⁿd^k** ენისთვის 11.5.2 განაწერში აღწერილი **C-LAG**, რომელიც ორაზროვნანია აგრეთვე +გლობალური თვალსაზრისით, და ბუნებრივი ენის **C-LAG**-ები (იხ. თავი 17, 18), რომლებიც არიან ორაზროვნები როგორც +გლობალური, ისე –გლობალური თვალსაზრისებით.

11.3.7 თეორემა 3

მხოლოდ –რეკურსულ ორაზროვნებიან **LA** გრამატიკაში წესთა გამოყენებების მაქსიმალური ოდენობა ტოლია

$$(n-(R-2)) 2^{(R-2)}.$$

ეს იმ შემთხვევაში, თუ **n > (R - 2)**, სადაც **n** არის შემავალი გამოსახულების სიგრძე და **R** არის წესთა რაოდენობა გრამატიკაში.

თეორემა 3 ითვალისწინებს, რომ იმ **LA** გრამატიკებში, რომლებიც არ არიან რეკურსულად ორაზროვნები, წესთა გამოყენებათა რიცხვი მხოლოდ წრფივად იზრდება შემავალი გამოსახულების სიგრძესთან ერთად.

11.4 გრამატიკებისა და ავტომატების სირთულე

გრამატიკული ტიპისა და მისი ენობრივი კლასის სირთულე დგინდება იმ პრიმიტიული ოპერაციების ოდენობით, რომელიც ყველაზე ცუდ შემთხვევაში საჭიროა შემავალი გამოსახულების დასამუშავებლად. რომელი ოპერაცია უნდა იქნეს განხილული პრიმიტიულ ოპერაციად და როგორ უნდა იქნეს განსაზღვრული საჭირო პრიმიტიული ოპერაციების ოდენობა ყოველთვის არაა ადვილად გადასაწყვეტი საკითხი.

ეს ასეა განსაკუთრებით იმ გრამატიკულ ფორმალიზმებში, რომლებიც არ არიან ტრანსფორმატული ტიპის. ისტორიულად, **PS**-გრამატიკული ტიპებისა და მათი ენობრივი კლასების სირთულეები ისაზღვრებოდა არა უშუალოდ გრამატიკული თვისებებით, არამედ ეკვივალენტური აბსტრაქტული

ავტომატების მეშვეობით.¹³

ამ ავტომატებიდან თითოეულს თავისი პრიმიტიული ოპერაცია ჰქონდა. მაგალითად, იმისათვის, რომ განესაზღვრათ ერლეის ალგორითმის სირთულე (იხ. 9.3.4), პრიმიტიული ოპერაცია თავდაპირველად არჩეული და მოტივირებული იქნა შემდეგნაირად:

გრიფიტის და პატრიკის მონაცემები არ არის აღწერილი ფაქტობრივი დროის ტერმინებში, არამედ ისინი აღწერებან "პრიმიტიული ოპერაციების" ტერმინებში. მათ გამოსახული აღმოჩენითი ალგორითმები ტიურინგის მანქანის მსგავსი მანქანისათვის გაწერილი არადეტერმინირებული გადამწერი წესების სიმრავლეების სახით. თითოეული ამ მანქანის თითოეული გამოყენება პრიმიტიულ ოპერაციად არის გაგებული. ჩვენ ჩვენს პრიმიტიულ ოპერაციად ამორჩეული გვაქვს მოქმედება, რომელიც მდგომარეობათა სიმრავლეს უმატებს ახალ მდგომარეობას (ან მცდელობა ასეთი მიმატების, მაშინ როცა მისამატებელი მდგომარეობა უკვე სიმრავლეშია). ჩვენ ვხვდებით, რომ ეს რაღაცით ემსგავსება მათ პრიმიტიულ ოპერაციას, რადგანაც ყველა ესენი არიან რაღაც აზრით იმ ალგორითმით ჩატარებული ყველაზე როგორი ინიციატივის დამოუკიდებელია გრამატიკის და შემავალი სტრიქნის ზომაზე.

XX. ერლეი 1970, გვ. 100

ამგვარად, მტკიცებულება "კონტექსტისგან თავისუფალი **PS**-გრამატიკები ააარსერებს **n³-ში**" პირდაპირ არ ახასიათებს კონტექსტისგან თავისუფალ **PS**-გრამატიკებს, არამედ გარკვეულ პარსერ ალგორითმს, რომელიც ამუშავებს **PS**-გრამატიკულ კლასებს, როგორც შემავალ მოცემულობებს. შესაბამისად, როცა ვალიანტმა (**Valiant**) 1975 წელს შეძლო კონტექსტისგან თავისუფალი ენების სირთულე **n³**-დან **n^{2.8}**-მდე შექმცირებინა, ეს არ იყო **PS**-გრამატიკის გაუმჯობესების შედეგი, არამედ ეს იყო შედეგი ვალიანტის იმ საკმაოდ გაუმჯობესებული მაანალიზებელი ალგორითმის, რომელსაც იგი აბსტრაქტულ ავტომატად განიხილავდა.

ამ მიდგომებისგან განსხვავებით, **C-LAG**-ები გვაძლევენ საშუალებას (i) განვსაზღვროთ პრიმიტიული ოპერაცია უშუალოდ გრამატიკული ფორმალიზმისთვის, და (ii) უცვლელად გადავიტანოთ იგი მის პარსერზე.

11.4.1 C-LAG-ების კლასის პრიმიტიული ოპერაცია არის წესის გამოყენება (აგრეთვე წარუმატებელი მცდელობების გადათვლა).

C-LAG-ების კლასის პრიმიტიული ოპერაცია არის წესის გამოყენება (აგრეთვე წარუმატებელი მცდელობების გადათვლა).

C-LAG-ების კლასის პრიმიტიულ ოპერაციებად წესის გამოყენების განხილვა მოსახერხებელია, რადგან მათი კატეგორიალური ოპერაციების გამოთვლითი მოთხოვნები კონსტანტით არის შემოსაზღვრული. ამასთან, ის, რომ **C-LAG**-ების კლასის პრიმიტიული ოპერაცია ამავდროულად

¹³ აბსტრაქტული ავტომატები შედგებან ისეთი კომპონენტებისაგან როგორებიცაა წამკითხავ/დამწერი-თავი, ჩაწერისგან-დაცული შემავალი ლენტი, მომუშავე ლენტები, ლენტის გასწროვ უჯრდან უჯრაზე გადაადგილებადი წამკითხავ/დამწერი-თავის, წაკითხვა ან წაშლა უჯრაში მყოფი მოცემულობის და ა.შ.. აბსტრაქტული ავტომატების კლასიფირი მაგალითება ტიურინგის მანქანები (**TM**), წრფივად შემოსაზღვრული ავტომატები (**LBA**), და ძირის მბიძვი ავტომატები (**PDA**), და სასრულ მდგომარეობანი ავტომატები (**FSA**). არსებობს კიდევ მრავალი სხვა აბსტრაქტული ავტომატები, თითოეული განსაზღვრულია იმ მიზნით, რომ გადაიჭრას სხვადასხვა განსაკუთრებული სასიათის სირთულეებთან დაკავშირებული პრობლემები, ეკვივალენტობის საკითხები, და გამოთვლადობის თვისებები.

გამოიყენება მისი პარსერის პრიმიტიულ ოპერაციად, განპირობებულია **LA**-გრამატიკის ტიპობრივი ტრანსფარენტულობით.

ეს, თავის მხრივ, განაპირობებს იმას, რომ **C-LAG**-ების სირთულე დამოკიდებულია მხოლოდ მათი ორაზროვნების ხარისხზე. ორაზროვნების ლინგვისტური ცნება მჭიდრო კავშირშია ავტომატების თეორიიდან კარგად ცნობილ არადეტერმინირებულობის ცნებასთან. ამიტომ, **C-LAG**-ების სირთულის თვისებების გამოკვლევამდე განვიხილოთ დეტერმინირებულ და არადეტერმინირებულ ავტომატებს (**deterministic and nondeterministic automata**) შორის არსებული განსხვავებები **LA**-გრამატიკასთან მიმართებაში.

ავტომატი იწოდება დეტერმინირებულად (**deterministic**), თუ მისი ალგორითმის თითოეული დერივაციული ბიჯი გრძელდება მაქსიმუმ ერთი მომდევნო ბიჯით. ამ განმარტების თანახმად ყველა არაორაზროვანი **LA**-გრამატიკა დეტერმინირებულია.

ავტომატი იწოდება არადეტერმინირებულად (**nondeterministic**), თუ მის ალგორითმს არ აქვს ცალსახად არჩეული ერთი რომელიმე რამოდენიმე შესაძლო მომდევნო ბიჯიდან (ამ რამოდენიმე შესაძლო მომდევნო ბიჯების რაოდენობა სასრულია), ანუ, რაც იგივეა, თუ ეს ალგორითმი ამ რამოდენიმე მომდევნო ბიჯს ერთდროულად ასრულებს. ამ განსაზღვრების თანახმად ყველა ორაზროვანი **LA**-გრამატიკა (იხ. 11.3.4) არადეტერმინირებულია.¹⁴

ითვლება, რომ არადეტერმინირებული ავტომატი (**nondeterministic automaton**) ამუშავებს შემავალ მოცემულობას, თუ ერთი მაინც შესაძლო გზა მთავრდება დამუშავებული მდგომარეობით. ამასთან, დღესაც არადეტერმინირებული ავტომატებისთვის პრობლემატურია როგორც არადეტერმინირებული დროითი სირთულის (**nondeterministic time complexity**) **NTIME(f(n))** (აქ n -ით აღნიშნულია შემავალი მოცემულობის სიგრძე, ხოლო $f(n)$ -ათ იმ ოპერაციების რაოდენობა, რასაც ყველაზე გრძელი დასაშვები გზა საჭიროებს), ისე არადეტერმინირებული მოცულობითი სირთულის (**nondeterministic space complexity**) **NSPACE(f(n))** (აქ n -ით აღნიშნულია შემავალი მოცემულობის სიგრძე, ხოლო $f(n)$ -ათ მეხსიერების უკრების ის რაოდენობა, რასაც ყველაზე გრძელი დასაშვები გზა საჭიროებს) საკითხი.

NTIME და **NSPACE** სირთულეები დამყარებულია იმ დაშვებაზე, რომ არადეტერმინირებულ ავტომატებს შეუძლიათ გამოიცნონ (**guess**) ყველაზე გრძელი ალგორითმული გზებიც კი. ამგვარად, **NTIME** და **NSPACE** ცხადად უნდა არჩევდნენ ამოხსნების მომცემ ისეთ სხვადასხვა ალტერნატიულ გზებსაც, რომლებიც ვერც კი ეტევიან სირთულის მოცემულ ფარგლებში.

დეტერმინირებული ავტომატებისთვის, მეორე მხრივ, ასევე პრობლემატურია **DTIME(f(n))** დროითი სირთულისა და **DSPACE(f(n))** მოცულობითი სირთულის დახასიათება. ამასთან, რადგან განსაზღვრების თანახმად დეტერმინირებულ ავტომატებში შეიძლება არსებოდეს მხოლოდ ერთი გზა, **DTIME** და **DSPACE** სირთულეები უნდა ისაზღვრებოდეს ამ გზისთვის საჭირო დროისა და მეხსიერების მოცულობის ფაქტობრივი ოდენობებით.

ავტომატების დეტერმინირებული და არადეტერმინირებული ვარიანტების ურთიერთგანრჩევით წამოიჭრა საკითხი არის თუ არა მათი წარმოქმნითი უნარები ეკვივალენტური. სასრულ მდგომარეობიანი ავტომატების (**finite state automata (FSA)**) შემთხვევაში ეს ასეა ნებისმიერი **L**

¹⁴ **PS**-გრამატიკებში, დეტერმინირებულობასა და არადეტერმინირებულობას შესაბამისები არ გააჩნია.

ენისთვის: თუ **L** ენა მუშავდება რომელიმე არადეტერმინირებული **FSA** ავტომატით, მაშინ არსებობს მისი ეკვივალენტური დეტერმინირებული **FSA** ავტომატიც, რომელიც ასევე ამუშავებს **L** ენას.¹⁵ ანალოგიური შედეგია ძალაში ტიურინგის მანქანებისთვისაც (**Turing machine (TM)**).¹⁶

ძირსმბიძე ავტომატების (**pushdown automata (PDA)**), მეორე მხრივ, დეტერმინირებულ და არადეტერმინირებულ ვერსიებს არ აქვთ ეკვივალენტური წარმოქმნითი უნარი.¹⁷ მაგალითად, **WW** ენა, სადაც **{n=R}** მუშავდება არადეტერმინირებული **PDA** ავტომატით, მაგრამ არ მუშავდება დეტერმინირებულით. და პოლის, წრფივად შემოსაზღვრული ავტომატების შემთხვევაში (**linear bounded automata (LBA)**) ჯერ კიდევ უცნობია არის თუ არა იმ ენების სიმრავლე, რომლებიც მუშავდება დეტერმინირებულ **LBA** ავტომატის მიერ საკუთრივი ქვესიმრავლე იმ ენების სიმრავლის, რომლებიც მუშავდება არადეტერმინირებულ **LBA** ავტომატის მიერ (**LBA** პრობლემა).

ამგვარად, ძალაშია შემდეგი დამოკიდებულებები **FSA**, **PDA**, **LBA**, და **TM** ავტომატების დეტერმინირებულ და არადეტერმინირებულ ვარიანტებს შორის:

DFSA = NFSA

DPDA ⊂ NPDA

DLBA ? NLBA

DTM = NTM

ავტომატები, რომელთა დეტერმინირებულ და არადეტერმინირებულ ვერსიებს აქვთ ტოლი წარმოქმნითი უნარი, წარმოშობები დამატებით შეკითხვას: რა ხარისხით იზრდება მეხსიერებითი მოთხოვნები არადეტერმინირებული ვერსიიდან დეტერმინირებულ ვერსიაზე გადასვლისას.

დეტერმინირებულ (არაორაზროვან) და არადეტერმინირებულ (ორაზროვან) **LA**-გრამატიკებს შორის განსაზღვრული განსხვავებები, სირთულის ავტომატურ-თეორიული განსაკუთრებით კი **NTIME** და **DTIME** სირთულეები, შეიძლება გამოყენებულ იქნას **LA**-გრამატიკაზე. კერძოდ, ძალაშია შემდეგი (**C** მუდმივია და **n** შემავალი გამოსახულების სიგრძეა):

არაორაზროვანი **C-LAG**-ების **DTIME** სირთულე = **C·(n-1)**

ორაზროვანი **C-LAG**-ების **NTIME** სირთულე = **C·(n-1)**

ეს ძალაშია რადგან – ვეურდნობით რა მარცხნივ ჯგუფდებად დერივაციულ სტრუქტურებს – ნებისმიერი **n** სიგრძის შემავალი ანალიზდება **n-1** კომბინაციურ ბიჯში. ამასთან, წესთა გამოყენება ყოველ კომბინაციური ბიჯში საჭიროებს დროს, რომელიც გამოთვლისთვის გამოყოფილი **C** კონსტანტის ტოლია. სხვა სიტყვებით: არაორაზროვანი **C-LAG**-ებს აქვთ დეტერმინირებული წრფივი დროითი_სირთულე (**deterministic linear time complexity**),¹⁸ მაშინ როდესაც ორაზროვანი **C-LAG**-ებს აქვთ არადეტერმინირებული წრფივი დროითი სირთულე (**nondeterministic linear time complexity**).

¹⁵ მ. ო. რაბინი და დ. სკოტი (**M. O. Rabin & D. Scott**) 1959

¹⁶ ჯ. ე. ჰოპროფტი და ჯ. დ. ულიმანი 1979, გვ. 164, თეორემა 7.3

¹⁷ ჯ. ე. ჰოპროფტი და ჯ. დ. ულიმანი 1979, გვ. 113.

¹⁸ მეცრად თუ ვიტყვით, არაორაზროვანი **C-LAG**-ები (**unambiguous C-LAGs**) უფრო კარგი სირთულით ხსიათდებან, ვიღრე ეს არის დეტერმინირებული წრფივი დრო (**deterministic linear time**), კერძოდ რეალური დრო (**real time**).

თეორიული კუთხით, **NTIME** და **DTIME** სირთულეების **LA**-გრამატიკაში გამოყენებადობა ფართო ინტერესებით იტვირთება, რადგან იგი აადვილებს შედეგებისა და გადაუჭრელი პრობლემების გადატანას ავტომატთა თეორიიდან **LA**-გრამატიკაში, და პირიქით. მაგალითად, შესაძლებელია ნაჩვენებ იქნას, რომ შეკითხვა არის თუ არა **C-LAG**-ების კლასი საკუთრივი ქვესიმრავლე **B-LAG**-ების კლასის, ეკვივალენტურია **LBA**-პრობლემის, ანუ იმის არის თუ არა **DLBA** კლასი საკუთრივი ნაწილი (**properly contained**) **NLBA** კლასის (იხ. 12.2).

თუ შევხედავთ პრაქტიკული კუთხით, ერთი მხრივ, ავტომატურ-თეორიული სირთულის გაზომვისას ხელს გვიშლის ის, რომ **DTIME** მიუწებადია მხოლოდ არაორაზროვან **LA**-გრამატიკებზე, ხოლო **NTIME** მხოლოდ ორაზროვან **LA**-გრამატიკებზე. ამასთან, **NTIME** არ იძლევა ნებისმიერად შემოსული შემავალი გამოსახულების გასააღიზებლად საჭირო ოპერაციების ფაქტობრივ ოდენობას, არამედ იმ ოდენობას რომელიც საჭიროა გამოთვლილი რეზულტატის შესამოწმებლად.

იმისათვის, რომ სირთულის გასაზომად გამოვიყენოთ ისეთი რეალური მეთოდი, რომელიც შესაძლებელი იქნება მსგავსად და ერთდროულად გამოყენებულ იქნას როგორც ორაზროვან, ისე არაორაზროვან გრამატიკებში, საჭიროა **C-LAG**-ების სირთულის ანალიზი დახასიათებულ იქნეს იმ გამოყენებულ წესთა მთლიანი ოდენობით, რომელიც საჭიროა ნებისმიერად აღებული შემავალი მოცემულობის ანალიზისთვის ყველაზე ცუდ შემთხვევაში (შემავალი მოცემულობის სიგრძესთან კავშირში). – **LA**-გრამატიკის სირთულის გაზომვის ეს გრამატიკულად წოდებული მეთოდი მარტივიცაა და ბუნებრივიცა.

11.5 C1-, C2-, და C3-LAG-ების ქვემოთარაობია

A- და **B-** LAG-ებთან შედარებით **C-LAG**-ები წარმოადგენენ **LAG**-ების ყველაზე მეტად შეზღუდულ კლასს, რომლებიც წარმოქმნიან ენების ყველაზე პატარა **LA**-კლასს. კონტექსტუსიგან თავისუფალ ენებთან შედარებით (რომლებიც საკუთრივ შედიან **C**-ენებში) **C**-ენების კლასი საკმაოდ ფართოა. აქედან გამომდინარე, თეორიული და პრაქტიკული თვალსაზრისებით მნიშვნელოვანია **C-LAG**-ების ქვეირარქიის განსაზღვრა და ამ გზით მისი სხვადასხვა ქვეკლასების გამოყოფა..

C-LAG-ების სირთულე იზომება გამოყენებული წესების რიცხვით. ამასთან, გამოყენებული წესების რიცხვი დამოკიდებულია მხოლოდ სტრუქტურათა ორაზროვნებაზე. ამდენად, ბუნებრივია რომ **C-LAG**-ების ქვეკლასები განვსაზღვროთ ორაზროვნების განსხვავებული ხარისხების ტერმინებში. ეს ქვეკლასები იწოდებიან, როგორც **C1**, **C2**, და **C3**- **LAG**-ები, რაც იმით იხსნება, რომ ორაზროვნების ხარისხის ზრდა პირდაპირ აისახება სირთულის ხარისხის ზრდაზე.

C1-LAG-ები ყველაზე დაბალი სირთულისა და ყველაზე დაბალი წარმოქმნითი უნარის ქვეკლასია. **C-LAG** არის **C1-LAG**, თუ კი ის არ არის რეკურსულად ორაზროვანი. **C1** ენების კლასი პარსირდება წრფივ დროში და შეიცავს ყველა იმ დეტერმინირებულ კონტექსტუსიგან თავისუფალ ენებს, რომლებიც გამოცნობადია **E**-გადანაცვლების ოპერაციის არმქონე **DPDA** ავტომატის მიერ, აგრეთვე –რეკურსული ორაზროვნებით დახასიათებულ კონტექსტისგან თავისუფალ ენებს, მაგ: $a^k b^k c^m d^m \cup a^k b^m c^m d^k$ ენას, და აგრეთვე $a^{2i} b^m c^{km}$ და a^{il} ენებს. ამასთან, ხაზგასასმელია, რომ ამ უკანასკნელთაგან ბოლო ინდექსირებული ენაც კი არა არის.¹⁹

¹⁹ $a^k b^k c^m d^m \cup a^k b^m c^m d^k$ ენის **C1-LAG** განსაზღვრულია 11.5.2-ში; L^k_{square} და L^k_{hast} ენების – 1993 წლის ბ.

კონტექსტზე დამოკიდებული არაორაზროვანი **C1-LAG**-ების მაგალითია 10.3.3 პუნქტში განსაზღვრული $a^k b^k c^k$ ენა და $a^i =_{\text{def}} \{a^i / i \text{ არის } 2\text{-ის რომელიმე დადებითი ხარისხი}\}$, რომელიც შეძლება ისაზღვრება

11.5.1 a^{2^i} ენის პროფესიული დამოკიდებული C1-LAG

$$\begin{aligned} LX &=_{\text{def}} \{[a(a)]\} \\ ST_S &=_{\text{def}} \{[(a)\{r_1\}]\} \\ r_1:(a) &\quad (a) \Rightarrow (aa) \quad \{r_2\} \\ r_2:(a X) &\quad (a) \Rightarrow (Xbb) \quad \{r_2, r_3\} \\ r_3:(b X) &\quad (a) \Rightarrow (Xaa) \quad \{r_2, r_3\} \\ ST_F &=_{\text{def}} \{[(aa) rp_1], [(bXb) rp_2], [(aXa) rp_3]\}. \end{aligned}$$

ეს **C1-LAG** არაორაზროვანია, რადგან r_2 და r_3 წესებს აქვთ არათავსებადი შემავალი პირობები. a^{2^i} ენის **PS** გრამატიკისა და ამავე ენის აქ აღწერილი **LA** გრამატიკის შედარება აშკარად გვიჩვენებს **LA** გრამატიკის ფორმალურ და კონცეპტუალურ სიმარტივეს.²⁰

11.5.2 განაწერში აღწერილია კონტექსტისაგან თავისუფალი $a^k b^k c^m d^m \cup a^k b^m c^m d^k$ ენის +გლობალური და $-$ რეკურსული ორაზროვანი **C1-LAG**. **PS**-გრამატიკაში ამ ენას ქვია შინაგანად ორაზროვანი (**inherently ambiguous**), რადგან არ არსებობს მისი არაორაზროვანი **PS**-გრამატიკა.²¹

11.5.2 ორაზროვანი $a^k b^k c^m d^m \cup a^k b^m c^m d^k$ ენის C1-LAG

$$\begin{aligned} LX &=_{\text{def}} \{[a(a)], [b(b)], [c(c)], [d(d)]\} \\ ST_S &=_{\text{def}} \{[(a)\{r_1, r_2, r_5\}]\} \\ r_1:(X) &\quad (a) \Rightarrow (a X) \quad \{r_1, r_2, r_5\} \\ r_2:(a X) &\quad (b) \Rightarrow (X) \quad \{r_2, r_3\} \\ r_3:(X) &\quad (c) \Rightarrow (c X) \quad \{r_3, r_4\} \\ r_4:(c X) &\quad (d) \Rightarrow (X) \quad \{r_4\} \\ r_5:(X) &\quad (b) \Rightarrow (b X) \quad \{r_5, r_6\} \\ r_6:(b X) &\quad (c) \Rightarrow (X) \quad \{r_6, r_7\} \\ r_7:(a X) &\quad (d) \Rightarrow (X) \quad \{r_7\} \\ ST_F &=_{\text{def}} \{[\varepsilon rp_4], [\varepsilon rp_7]\} \end{aligned}$$

11.3.4 განსაზღვრების თანახმად ეს **C1-LAG** სინტაქსური ორაზროვნების მატარებელია: rp_1 წესის პაკეტი შეიცავს თავსებად შემავალ პირობებიან r_2 და r_5 წესებს. მიუხედავად ამისა, გრამატიკული პარსინება ხდება წრფივ დროში, რადგან ორაზროვანი გაგრძელებები არ ხდებიან რეკურსული ციკლის ნაწილები: rp_2 და rp_5 არ შეიცავნ r_1 წესს. ცუდ შემთხვევაში, მაგალითად $aabbccdd$

შტუბერტის (**B. Stubert**) წიგნში გვ.16 და გვ.12; $a^k b^k c^k d^k e^k$ ენის – **Col** გვ. 223; $a^k b^m c^m$ ენი – ჰაუსერის 1992 წელს გამოშვებულ წიგნში გვ. 296; a^i ენის – 11.5.1-ში; a^{ii} ენის – ჰაუსერის 1992 წელს გამოშვებულ წიგნში (იხ. მინიშნება გვ. 296, შენიშვნა 13).

²⁰ ჯ. ე. პოპკოროვტმა და ჯ. დ. ულმანმა 1979-ში წარმოადგინეს კანონიკური კონტექსტზე დამოკიდებული a^{2^i} ენის **PS**-გრამატიკა (იხ. გვ. 224) და მისი შეუზღუდული **PS**-გრამატიკული გერსია (იხ. გვ. 220).

²¹ იხ. ჯ. ე. პოპკოროვტმა და ჯ. დ. ულმანი 1979 გვ. 99-103.

სიტყვის შემთხვევაში, გრამატიკა გენერირებას უკეთებს (წარმოქმნის) ორ პარალელურ დროში წრფივ გამოყვანაზე დაფუძნებულ ანალიზს, რომელთაგან თითოეული იწყება საწყისი **a**-მიმდევრობის შემდეგ.²²

C2-LAG-ები არაან **C-LAG**-ების მეორე ყველაზე დაბალი სირთულის და მეორე ყველაზე დაბალი წარმოქმნითი უნარის კლასი. **C-LAG** არის **C2-LAG** თუ (i) ის წარმოქმნის +რეკურსულ ორაზროვნებას და თუ (ii) მისი ორაზროვნებები შეზღუდულია ერთის დაბრუნების პრინციპით.

11.5.3 ერთის დაბრუნების პრინციპი_(THE SINGLE RETURN PRINCIPLE (SRP))

+რეკურსული ორაზროვნება არის ერთხელ დაბრუნებადი, თუ პარალელური გზებიდან მხოლოდ ერთი ბრუნდება ამ ორაზროვნების მომცემ მდგომარეობაში.

C2 ენების კლასი აპარსერებს პოლინომიალურ დროში და მოიცავს ისეთი სახის არადეტერმინირებულ კონტექსტისგან თავისუფალ ენებს, როგორებიცაა WW^R და L^k_{hast} , და, ამის გარდა, ისეთ კონტექსტზე დამოკიდებულ ენებს, როგორებიცაა: WW , $W^{k\geq 3}$, $\{WWW\}^*$, და $W_1W_2W_1^RW_2^R$.²³

SR-რეკურსული ორაზროვნება შუქდება ქვემოთ WW^R ენისთვის განხილულ **C2-LAG** მაგალითზე. როგორც 8.3 ქვეთავშია ახსნილი, ეს ენა შეიცავს სრულიად ნებისმიერი სიტყვების სრულიად ნებისმიერი სიგრძის W მიმდევრობას მიღენებულს მისი W^R ინვერსირებული მიმდევრობით.

11.5.4 პონტექსტისგან თავისუფალი WW^R ენის **C2-LAG**

$$\begin{aligned} LX &= \text{def}\{[a(a)], [b(b)], [c(c)], [d(d)]...\} \\ ST_S &= \text{def}\{[(seg_c)\{r_1, r_2\}]\}, \text{სადაც } seg_c \in \{a, b, c, d, \dots\} \\ r_1:(X) (seg_c) &\Rightarrow (seg_c X)\{r_1, r_2\} \\ r_2:(seg_c X) (seg_c) &\Rightarrow (X)\{r_2\} \\ ST_F &= \text{def}\{[\varepsilon \text{ rp}_2]\} \end{aligned}$$

ყოველთვის, როცა r_1 წესის გამოყენება წარმატებულია, შემდეგ კომპოზიციაზე გამოიყენება r_1 და r_2 წესები. ამასთან, r_2 წესის წარმატებული გამოყენების შემდეგ გამოყვანა აღარ უბრუნდება r_1 წესს, რადგან r_1 არაა მითითებული r_2 წესის პაკეტში. აქედან გამომდინარე, ორაზროვნების მხოლოდ ერთი ტოტი (ე.გ. ის, რომელიც r_1 წესის განმეორებითი გამოყენებით მიიღება) შეიძლება დაბრუნდეს რეკურსიაში.

WW^R პარსირების ყველაზე ცუდ შემთხვევას იძლევა ის შემავალი მოცემულობა, რომელიც შედგება ერთი და იგივე სიტყვების ლუწი ოდენობით. ეს ნაჩვენებია 11.5.5-ში **aaaaaa** შემავალი მოცემულობის მაგალითზე.

11.5.5 გამოყვანის სტრუქტურა ყველაზე ცუდი შემთხვევისთვის WW^R -ში.

²² ცხადი გამოყვანა მოცემულია **CoL**-ში გვ. 154 ფფ.

²³ **C2-LAG** განსაზღვრულია WW^R ენისათვის 11.5.4-ში, L^∞_{hast} ენისათვის – შტუპერტი 1993, გვ. 16, WW ენისათვის – 11.5.6-ში, WWW ენისათვის – **CoL**-ში, გვ. 216-, და $W_1W_2W_1^RW_2^R$ ენისათვის – განმარტიბულია 11.5.7-ში.

წესები:

ანალიზები:

2	a \$ a
1 2 2	a a \$ a a
1 1 2 2 2	a a a \$ a a a
1 1 1 2 2	a a a a \$ a a
1 1 1 1 2	a a a a a \$ a
1 1 1 1 1	a a a a a a \$

11.5.5 გამოყვანის მსვლელობისას წარმოქმნილი შუალედური სტრიქონის არამარკირებული ნაწილი აღნიშნულია **\$** სიმბოლოთი. ამ მარცხნივ ჯგუფდებადი ანალიზის მსვლელობისას მიღებული ექვსი ჰიპოთეზიდან პირველი ორი დაწესებულია (უარყოფილია), იმ ფაქტის გამო, რომ შემავალი სტრიქონები გასაგრძელებელია, მესამე ჰიპოთეზა ზუსტად ეთანადება **a a a a a a** შემავალს, მაშინ როდესაც დარჩენილი სამი ჰიპოთეზა უარყოფილია იმის გამო, რომ შემავალი არ შეიცავს მეტ სიტყვებს.

ქვემოთ განსაზღვრულია **C2-LAG** კონტექსტზე დამოკიდებული **WW** ენისათვის.

11.5.6 პონტიფისტზე დამოკიდებული **WW**-ენის **C2-LAG**

$$\begin{aligned} LX &= \text{def}\{[a(a)], [b(b)], [c(c)], [d(d)]\ldots\} \\ ST_S &= \text{def}\{[(seg_c)\{r_1, r_2\}]\}, \text{სადაც } seg_c \in \{a, b, c, d, \ldots\} \\ r_1:(X) &\quad (seg_c) \Rightarrow (X \ seg_c) \quad \{r_1, r_2\} \\ r_2:(seg_c \ X) &\quad (seg_c) \Rightarrow (X) \quad \{r_2\} \\ ST_F &= \text{def}\{[\varepsilon \ rp_2]\} \end{aligned}$$

ეს გრამატიკა 11.5.4 გრამატიკისაგან განსხვავდება მხოლოდ **r₁** წესის შედეგის კატეგორიით: კონტექსტზე დამოკიდებულ **WW** ენაში **r₁** წესი შედეგ კატეგორიაში იძლევა (**X segc**) კატეგორიის ნიმუშს, მაშინ როდესაც კონტექსტისგან თავისუფალი **WW^R** ენაში **r₁** წესი შედეგ კატეგორიაში იძლევა (**segc X**) ნიმუშს. კონტექსტზე დამოკიდებულ **WW** და კონტექსტისგან თავისუფალი **WW^R** ენებში ყველაზე ცუდ შემთხვევებს ერთი და იგივე შემავალი გამოსახულებები იძლევა: ორივეს შესაბამისი **C2-LAG** არის **SR**-ორაზროვანი, და ორივე არის **n²** სირთულის. **WW** და **WW^R** ენებში ანალიზის ყველაზე ცუდ შემთხვევებში წაკითხვების **n²**-ული ზრდა თვალსაჩინოდ არის წარმოდგენილი 11.5.5 პუნქტში.

არადეტერმინირებული კონტექსტზე დამოკიდებული **W₁W₂W^R₁W^R₂** ენა **n³** სირთულის **C2-LAG** ტიპისაა (შედეგი ეკუთვნის ბ. შტუბერტს).

11.5.7 პონტიფისტზე დამოკიდებული **W₁W₂W^R₁W^R₂** ენის **C2-LAG**

$$\begin{aligned} LX &= \text{def}\{[a(a)], [b(b)]\} \\ ST_S &= \text{def}\{[(seg_c)\{r_{1a}\}], [(seg_c)\{r_{1b}\}]\} \text{ სადაც } seg_c, seg_d \in \{a, b\} \\ r_{1a}:(seg_c) &\quad (seg_d) \Rightarrow (\#seg_c \ seg_d) \quad \{r_2, r_3\} \\ r_{1b}:(seg_c) &\quad (seg_d) \Rightarrow (seg_d \ # \ seg_c) \quad \{r_3, r_4\} \\ r_2:(X) &\quad (seg_c) \Rightarrow (X \ seg_c) \quad \{r_2, r_3\} \\ r_3:(X) &\quad (seg_c) \Rightarrow (seg_c \ X) \quad \{r_3, r_4\} \end{aligned}$$

$r_4:(X \text{ seg}_c)$	$(\text{seg}_c) \Rightarrow (X)$	$\{r_4, r_5\}$
$r_5: (\text{seg}_c X \#)$	$(\text{seg}_c) \Rightarrow (X)$	$\{r_6\}$
$r_6: (\text{seg}_c X)$	$(\text{seg}_c) \Rightarrow (X)$	$\{r_6\}$
$\text{ST}_F = \text{def}\{[\varepsilon \text{ rp}_5], [\varepsilon \text{ rp}_6]\}$		

$$\frac{1}{8}n^3 + \frac{1}{4}n^2 + \frac{1}{2}n$$

ამ ენის ზუსტი სირთულე არის $\frac{1}{8}n^3 + \frac{1}{4}n^2 + \frac{1}{2}n$. პოლინომიალური სირთულის ხარისხი დამოკიდებულია გამოყვანაში **SR**-რეკურსიული ორაზროვნების მაქსიმალურ ღდენობაზე. ამგვარად, **C2-LAG** ერთი **SR**-რეკურსიული ორაზროვნებით შემძლეა n^2 ღროში პარსირების, ხოლო **C2-LAG** ორი **SR** რეკურსიული ორაზროვნებით შემძლეა n^3 ღროში პარსირების და ა.შ..

C-LAG-ების უმაღლესი სირთულისა და წარმოქმნითი უნარის ქვეკლასია **C3-LAG**-ები. **C-LAG** არის **C3-LAG**, თუ ის წარმოქმნის შემოუსაზღვრელ +რეკურსულ ორაზროვნებას. **C3** ენების კლასი პარსირდება ექსპონენციალურ ღროში და შეიცავს დეტერმინირებულ კონტექსტისგან თავისუფალ **L_{no}** ენას, ურთულეს კონტექსტისგან თავისუფალ **HCFL** ენას, და, აგრეთვე, **SubsetSum** და **SAT** ენების მსგავს კონტექსტზე დამოკიდებულ ენებს, რომლებიც **NP**-სრულია.²⁴

SubsetSum ენა მიიჩნევა თვისობრივად რთულ ენად, ანუ ისეთ ენად, რომელიც დასამუშავებლად გამოყენებული ალგორითმის მიუხედავად დასამუშავებლად მოითხოვს ექსპონენციალურ ღროს. მისი გამოსახულებები $y \# a_1 \# a_2 \# a_3 \dots \# a_n \#$ განსაზღვრული როგორც y, a_1, a_2, a_3, \dots არიან ციფრების ერთი და იგივე რაოდენობის შემცველი ორობითი სტრიქონები. უფრო მეტიც, როდესაც ისინი განიხილება როგორც ორობითი რიცხვები, ყველაზე ნაკლებ მნიშვნელოვანია პირველი ციფრი, და, ამასთან, y ტოლია a_i სიმრავლეების ქვესიმრავლეების გაერთიანების.

11.5.8 SubsetSum მნიშვნელი C3-LAG

$LX = \text{def}\{[0(0)], [1(1)], [\#(\#)]\}$		
$\text{ST}_S = \text{def}\{\{\text{seg}_c\} \{r_1, r_2\}\},$ სადაც $\text{seg}_c \in \{0, 1\}$		
$\text{seg}_c \in \{0, 1\}$		
$r_1:(X)$	(seg_c)	$\Rightarrow (\text{seg}_c X) \{r_1, r_2\}$
$r_2:(X)$	$(\#)$	$\Rightarrow (\# X) \{r_3, r_4, r_6, r_7, r_{12}, r_{14}\}$
$r_3:(X \text{ seg}_c)$	(seg_c)	$\Rightarrow 0 X) \{r_3, r_4, r_6, r_7\}$
$r_4:(X \#)$	$(\#)$	$\Rightarrow (\# X) \{r_3, r_4, r_6, r_7, r_{12}, r_{14}\}$
$r_5:(X \text{ seg}_c)$	(seg_c)	$\Rightarrow (0 X) \{r_5, r_6, r_7, r_{11}\}$
$r_6:(X 1)$	(0)	$\Rightarrow (1 X) \{r_5, r_6, r_7, r_{11}\}$
$r_7:(X 0)$	(1)	$\Rightarrow (1 X) \{r_8, r_9, r_{10}\}$
$r_8:(X \text{ seg}_c)$	(seg_c)	$\Rightarrow (1 X) \{r_8, r_9, r_{10}\}$
$r_9:(X 1)$	(0)	$\Rightarrow (0 X) \{r_5, r_6, r_7, r_{11}\}$
$r_{10}:(X 0)$	(1)	$\Rightarrow (0 X) \{r_8, r_9, r_{10}\}$
$r_{11}:(X \#)$	$(\#)$	$\Rightarrow (\# X) \{r_3, r_4, r_6, r_7, r_{12}, r_{14}\}$
$r_{12}:(X 0)$	(seg_c)	$\Rightarrow (0 X) \{r_4, r_{12}, r_{14}\}$
$r_{13}:(X 0)$	(seg_c)	$\Rightarrow (0 X) \{r_{11}, r_{13}, r_{14}\}$
$r_{14}:(X 1)$	(seg_c)	$\Rightarrow (1 X) \{r_{11}, r_{13}, r_{14}\}$
$\text{ST}_F = \text{def}\{[(X) \text{ rp}_4]\}$		

²⁴ **L_{no}** ენის **C3-LAG** განსაზღვრულია 12.3.3-ში; **HCFL** ენის განსაზღვრულია ბ. შტუბერტთან 1993, გვ. 16; **SubsetSum** ენისათვის განსაზღვრულია 11.5.8-ში, და **SAT** ენის განსაზღვრულია პაუსერთან 1992, მინიშნება 19.

ეს **C3-LAG** (აგებულია დ. ეპლიგეიტის **(D. Applegate)** მიერ) გადაწერს **y**-ს კატეგორიაში. ამის შემდეგ +რეკურსული ორაზროვნება წარმოიქმნება **y**-ისგან თითოეული **a_i**-ის არადეტერმინირებულად განსაზღვრული ან გამოკლებით, ან არ გამოკლებით. გრამატიკა მხოლოდ იმ შემთხვევაში შედის დასაშვებ მდგომარეობაში, თუ სხვაობის შედეგი ნულია.

და ბოლოს, შევაჯამოთ **LA**-გრამატიკის სხვადასხვა შეზღუდვები, ამ შეზღუდვების შესაბამისი **LA**-გრამატიკული კლასები, ამ გრამატიკული კლასების შესაბამისი ენობრივი კლასები, და მათი სირთულეები. **LA**-იერარქია მსგავსია **PS**-იერარქიის (იხ. 8.1.2), იმით, რომ **LAG**-ების ქვეკლასები და მათი ენები განსაზღვრება წესთა სისტემის მზარდი შეზღუდულობის ტერმინებში.

C1-LAG ენათა კლასი **C2**-ების კლასის ქვესიმრავლეა, რადგან **C1-LAG**-ების –რეკურსული ორაზროვნული სტრუქტურა უფრო შეზღუდულია ვიდრე **C2-LAG**-ების **SR**-რეკურსული ორაზროვნული სტრუქტურა (11.5.4). **C2-LAG**-ის ენათა კლასი არის **C3**-ენათა კლასის ქვესიმრავლე, რადგან **C2-LAG**-ის **SR**-რეკურსული ორაზროვნული სტრუქტურა უფრო შეზღუდულია ვიდრე **C3-LAG**-ის +რეკურსული ორაზროვნული სტრუქტურა. **C3-LAG**-ის ენათა კლასი **B**-ების ქვესიმრავლეა, რადგან **C3-LAG**-ის კატეგორიული ნიმუშები უფრო შეზღუდულია ვიდრე **B-LAG**-ის. **B-LAG**-ში გამოყენებადი კატეგორიების სიგრძე იზღუდება, მაშინ როდესაც **A-LAG**ში გამოყენებადი კატეგორიები სიგრძის თვალსაზრისებით არ იზღუდებიან.

LA-გრამატიკაზე გამოყენებადი შეზღუდვები განეკუთვნებიან სირთულის შემდეგ ორ პარამეტრს: წესთა გამოყენების **მოცულობას (amount)** და **ოდენობას (number)**. (იხ. 11.2.1)

11.5.9 შეზღუდვების ფიპები **LA**-გრამატიკაში

0. **A LA-ტიპი:** არავითარი შეზღუდვები

1. **B LA-ტიპი:** შუალედური გამოსახულებების კატეგორიების სიგრძე შემოსაზღვრულია **k·n**-ით, სადაც **k** არის კონსტანტა და **n** არის შემავალი გამოსახულების სიგრძე (**R1.1, მოცულობა (amount)**).
2. **C3 LA-ტიპი:** კატეგორიის ნიმუშების ფორმა შედეგად ითხოვს კატეგორიული ოპერაციებით მოთხოვნილი ოპერაციების კონსტანტურ ლიმიტირებას (**R1.2. მოცულობა (amount)**).
3. **C2 LA-ტიპი:** **C3 LA-ტიპი** და ყველაზე მეტი **SR**-რეკურსულად ორაზროვანი გრამატიკა (**R2, ოდენობა (number)**).
4. **C1 LA-ტიპი:** **C3 LA-ტიპი** და ყველაზე მეტი –რეკურსულად ორაზროვანი გრამატიკა (**R2, ოდენობა (number)**).

პარამეტრების სიდიდე და რიცხვი კონტროლდება (1) შუალედური გამოსახულებების კატეგორიების სიგრძით, (2) წესებში კატეგორიის ნიმუშების (მოდელების) ფორმით, (3) წესების შესაძლო კვლავ გამოყენებით ისე როგორც ეს განსაზღვრულია წესთა პაკეტით, და (4) ერთი და იგივე წესთა სისტემის წესების შემავალი თავსებადობით.

LA-გრამატიკული იერარქია შეჯამებულია 11.5.10 ცხრილში. 8.1.2 განაწერში აღწერილი **PS**-გრამატიკული იერარქიის მსგავსად, ეს იერარქიაც ემყარება სტრუქტურული წესთა სისტემის შეზღუდვების, გრამატიკის ტიპების, ენათა კლასების, და სირთულის ხარისხების ურთიერთკავშირებს.

11.5.10 LA-გრამატიკული იმრაობია ფორმალური მნების

ტიპი	LAG-ტიპი	ენები	სირთულე
ტიპი	C1-LAG	C1-ენები	წრფივი
ტიპი	C2-LAG	C2-ენები	პოლინომიალური
ტიპი	C3-LAG	C3-ენები	ექსპონენციალური
ტიპი	B-LAG	B-ენები	ექსპონენციალური
ტიპი	A-LAG	A-ენები	ექსპონენციალური

LA-გრამატიკის ხუთი სხვადასხვა კლასიდან მხოლოდ **B**-ენების კლასი გვხვდება **PS**-გრამატიკის იერარქიაში, კერძოდ, როგორც კონტექსტისგან თავისუფალი ენების კლასი. **A**-ენების კლასი, მეორე მხრივ, საკუთრივ შედის შეუზღუდავი **PS**-გრამატიკის მიერ წარმოქმნილ რეკურსულად გადათვლად ენების კლასში. **C1**, **C2**, და **C3**- ენების კლასებს **PS**-გრამატიკულ იერარქიაში ანალოგი არა აქვთ. ამასთან, რეგულარული ენების **PS**-გრამატიკული კლასი საკუთრივ შედის **C1**-ენების კლასში, ხოლო კონტექსტისგან თავისუფალი ენების კლასი საკუთრივ შედის **C3**-ენების კლასში.

ამის გარდა, თუ შევადარებთ **PS**-გრამატიკის შეზღუდვებს, რომლებიც ეხება მხოლოდ წესების ფორმას, **LA**-გრამატიკის შეზღუდვები ეხება წესთა სისტემას მთლიანად. ამის მიზეზია ის, რომ **LA**-გრამატიკაში წესების გამოყენება ექსპლიციტურადაა აღწერილი კატეგორიის ნიმუშებისა და წესების პაკეტის ტერმინებში, მაშინ როდესაც **PS**-გრამატიკაში ისინი მხოლოდ იმპლიციტურად აღიწერება ცვლადების შემცველი გადამწერი წესების ტერმინებში.

სავარჯიშოები

ქვეთავი 11.1

- ახსენით სრულად რეკურსული და ნაწილობრივ რეკურსული ფუნქციების ცნებები.
- როგორ ხასიათდება ფორმალურად **PS**-გრამატიკულ იერარქიაში რეკურსული ენების კლასი?
- რა არის **LA**-გრამატიკის ძირეული შეუზღუდავი ფორმის წარმოქმნითი უნარი?
- ახსენით 11.1.2 და 11.1.3 განაწერში მოცემული დამტკიცებები.
- აღწერეთ შეუზღუდავი **LA**-გრამატიკისა და შეუზღუდავი **PS**-გრამატიკის წარმოქმნით უნარებს შორის არსებული განსხვავებები.

ქვეთავი 11.2

- რა არის **LA**-გრამატიკის შესაძლო სტრუქტურული შეზღუდვები?
- ახსენით **A-LAG**, **B-LAG**, და **C-LAG** გრამატიკული კლასები.
- რა განსხვავებაა კონსტანტურ და არაკონსტანტურ კატეგორიულ ოპერაციებს შორის?
- აღწერეთ **PS**-გრამატიკული იერარქიის რეკონსტრუქცია **LA**-გრამატიკაში.
- არსებობს თუ არა **LA**-გრამატიკის კლასი, რომელიც წარმოქმნის რეკურსულად გადათვლად ენებს?
- არსებობს თუ არა **PS**-გრამატიკის კლასი, რომელიც წარმოქმნის რეკურსულ ენებს?
- არსებობს თუ არა **PS**-გრამატიკის კლასი, რომელიც წარმოქმნის **C**-ენებს?

ქვეთავი 11.3

- ახსენით \pm გლობალური ორაზროვნების ცნებები. რა კავშირშია ისინი ავტომატთა თეორიის \pm დეტერმინირებული დერივაციის ცნებებთან და ორაზროვნების ცნებებთან **PS**-გრამატიკაში.
- რას განსაზღვრავს წესთა გამოყენებების რიცხვი **LA**-დერივაციაში?
- აღწერეთ შემავალი მოცემულობის სამი განსხვავებული პირობა.
- განსაზღვრეთ ორაზროვანი და არაორაზროვანი **LA**-გრამატიკები.
- განსაზღვრეთ რეკურსული ორაზროვნების ცნებები **LA**-გრამატიკაში.
- განსაზღვრეთ წესთა პაკეტი $-$ რეკურსული ორაზროვანი **LA**-გრამატიკის 7 წესით, რომელიც მაქსიმალურად ორაზროვანია. რამდენი წაკითხვა წარმოიქმნება ამ გრამატიკის მიერ 4, 5, 6, 7, 8, და 9 კომბინაციური ბიჯის შემდეგ?

ქვეთავი 11.4

- ახსენით **LA**-გრამატიკის სირთულის პარამეტრები.
- როგორ არის განსაზღვრული **C-LAG**-ების ელემენტარული ოპერაცია?
- შეადარეთ ერლეიის პრიმიტიული ოპერაცია, რომელიც გამოიყენება კონტექსტისგან თავისუფალი **PS**-გრამატიკის მისეული ალგორითმის სირთულის გამოსათვლელად, **C-LAG**-ების პრიმიტიულ ოპერაციას.
- როგორ ითვლის ერლეი კონტექსტისგან თავისუფალი **PS**-გრამატიკის მოცულობით და რიცხვით პარამეტრებს?
- დაასახელეთ სირთულის ოთხი ცნება ავტომატთა თეორიაში და ახსენით ისინი.
- ახსენით დამოკიდებულება **FSA**, **PDA**, **LBA**, და **TM** დეტერმინირებულ და არადეტერმინირებულ ვერსიებს შორის.
- ახსენით **NTIME** და **DTIME** ცნებების გამოყენება **C-LAG**-ებში. რა ალტერნატივებია?

ქვეთავი 11.5

- აღწერეთ **C1-**, **C2-**, და **C3-** **LAG**-ების ქვეირარქიები. რა კავშირებია სირთულესა და ორაზროვნებას შორის?
- განსაზღვრეთ $a^k b^k c^m d^m \cup a^k b^m c^m d^k$ ენის **C1-LAG** და ახსენით როგორ მუშაობს ეს გრამატიკა. არის თუ არა იგი ორაზროვანი? რა არის მისი სირთულე?
- ახსენით ერთხელ დაბრუნების პრინციპი.

4. შეადარეთ **WW** და **WW^R** ენების **PS**-გრამატიკული და **LA**-გრამატიკული დამუშავებები.
5. აღწერეთ **SubsetSum** ენის სწორად ფორმირებული გამოსახულებები და ახსენით რატომ არის ეს ენა თვისობრივად რთული.
6. ახსენით **LA**-გრამატიკული იერარქია.
7. აღწერეთ დამოკიდებულება **PS**-გრამატიკასა და ენის ნეიტივისტურ თეორიას შორის, ერთი მხრივ, და **LA**-გრამატიკასა და **SLIM** თეორიას შორის, მეორე მხივ. შესაძლებელია თუ არა, რომ **PS**-გრამატიკა გამოყენებულ იქნას როგორც **SLIM** თეორიის სინტაქსური კომპონენტი? შესაძლებელია თუ არა, რომ **LA**-გრამატიკა გამოყენებულ იქნას როგორც ნეიტივისტური თეორიის სინტაქსური კომპონენტი?

12 თავი

LA- და PS- იერარქიების ურთიერთობისართება

გენერაციული გრამატიკის¹ ორი არაეკვივალენტური ფორმალიზმის აღწერის შემდეგ, რომელთაგან თითოეული ენობრივი კლასებისა და სირთულეების თავისი სრულად განვითარებული იერარქიით ხასიათდება, ამ თავში ჩვენ ვუბრუნდებით ბუნებრივი ენების ფორმალურ თვისებებს. 12.1 ქვეთავში შედარებულია **LA-** და **PS-** გრამატიკული ენობრივი კლასების სირთულეები. 12.2. ქვეთავში შუქლება **PS-** და **LA-** იერარქიების ენობრივ კლასებს შორის არსებული ქვესიმრავლეობის მიმართება. 12.3 ქვეთავში ისაზღვრება კონტექსტისგან თავისუფალი ენა, რომელიც ერთ-ერთი კერძო სახის **C3**-ენაა. 12.4 ქვეთავში შუქლება ორთოგონალური მიმართება კონტექსტისგან თავისუფალ ენებსა და **C3**-ენობრივ კლასებს შორის. 12.5 ქვეთავში გამოკვლეულია ორაზროვნება ბუნებრივ ენებში და დასაბუთებულია, რომ ბუნებრივი ენა არის **C1-LAG**-ენების კლასში და რომ იგი წრფივ დროში პარსირდება.

12.1 LA- და PS- გრამატიკების ენობრივი კლასები

¹ სახელდობრ, 8-ე და 9-ე თავში **PS**-გრამატიკა, და 10-ე და 11-ე თავში **LA**-გრამატიკა.

ბუნებრივი ენისთვის არჩეული გრამატიკული ფორმალიზმი ვარგისად მიიჩნევა, თუ ის ძლიერ მცირეა – მაგალითად როგორიც არის წესიერი (3 ტიპის) **PS**-გრამატიკა – ეს ფენომენი შეუძლებელია აიხსნას მხოლოდ მისივე საშუალებებით. თუმცა, თუ არჩეული ფორმალიზმი ძლიერ დიდია – მაგალითად როგორიცაა კონტექსტზე დამოკიდებული (1 ტიპის) **PS**-გრამატიკა –, მაშინ ბუნებრივი ენის მახასიათებელი სტრუქტურები უჩინარდებიან მასში, რომელიც, თავის მხრივ, მისი ძლიერ დიდობის გამო, ისევე მარტივად და დაუმახინჯებლად აღწერს ყველაზე უფრო რთულ ხელოვნურ ენებს, როგორც ბუნებრივ ენებს.

ამასთან, რაც უფრო ‘დიდია’ ფორმალიზმი, მით უფრო ‘ძვირი’ ჯდება იგი მათემატიკური სირთულის თვალსაზრისებით. კომპიუტერული თვალსაზრისებით კი, ეს სიძვირე ისახება ენობრივი გამოსახულების პარსიტებისთვის საჭირო დროზე და მეხსიერებაზე.

ბუნებრივი ენის აღსაწერად ესა თუ ის ფორმალიზმი შეიძლება იყოს არა უბრალოდ ძლიერ დიდი, ან ძლიერ პატარა, არამედ ერთდღოულად ძლიერ დიდიც და ძლიერ პატარაც – მსგავსად იმ შარვალისა, რომელიც არის ძლიერ მოკლე და ამავდროულად ძლიერ განიერი. მაგალითად, კონტექსტისგან თავისუფალი **PS**-გრამატიკული ფორმალიზმი არის ძლიერ პატარა ბუნებრივი ენის აღსაწერად, რაც ტრანსფორმაციებისა და სიმილარული გამოსახულებების შემოტანის აუცილებლობითაც დასტურდება. თუმცა, ამავე დროს, ის არის ძლიერ დიდი, რასაც აშკარად გვიდასტურებს კონტექსტისგან თავისუფალი **HCFL**² ენების მსგავსი ენები, რომელთა საკმაოდ რთულ სტრუქტურებს არ გააჩნიათ თავიანთი შესატყვისები ბუნებრივ ენებში.

პრინციპში, არ არსებობს არანაირი თეორიული და პრაქტიკული მიზეზი იმისა, რომ შეუძლებლად მივიჩნიოთ ისეთი დაბალი წარმოქმნითი უნარისა და სირთულის (ქვე-)ფორმალიზმის (**sub-formalism**) მოძიება, რომელიც სრულად მოერგებოდა ბუნებრივ ენას. თუმცა, უკვე შეიძლება ითქვას, რომ მიუხედავად დიდი ძალისხმეულისა **PS**-გრამატიკის ფარგლებში ვერანაირი ამგვარი ქვეფორმალიზმი ვერ იქნა მოძიებული.

PS- ძირეული გრამატიკული ფორმალიზმის ალტერნატიული ძირეული ფორმალიზმებია **C**- და **LA**- გრამატიკები. ამასთან, **C**-გრამატიკას ჯერ კიდევ არა აქვს სხვადასხვა ენობრივი კლასებისა და სირთულის ხარისხების სრულად განვითარებული იერარქიული სისტემა. მოუხედავად ამისა, ის, რომ ბიდირექციული **C**-გრამატიკა სუსტად ეკვივალენტურია კონტექსტისგან თავისუფალი **PS**- გრამატიკის (იხ. 9.2 ქვეთავი), უკვე დიდი ხანა ცნობილია.

ამგვარად, **PS**-გრამატიკული ენობრივი იერარქიის ჯერჯერობით ერთადერთ ორთოგონალურ ენობრივ იერარქიას **LA**-გრამატიკა წარმოშობს. 12.1.1 ცხრილში **LA**- და **PS**- გრამატიკების ენობრივი კლასები დახასიათებულია სირთულის ოთხი ძირითადი ხარისხის მიხედვით.

12.1.1 **LA**- და **PS**- გრამატიკების სირთულის ხარისხები

LA-გრამატიკა

არაამოხსნადი

PS- გრამატიკა

რეკურსულად გადათვლადი ენები

² ურთულესი კონტექსტისგან თავისუფალი ენა (**S. Greibach** 1973).

ექსპონენციალური	A-ენები B-ენები C3-ენები	კონტექსტზე დამოკიდებული ენები
პოლინომიალური	C2-ენები	კონტექსტისგან თავისუფალი ენები
წრფივი	C1-ენები	რეგულარული ენები

LA- და **PS-** გრამატიკების ელემენტარული ფორმალიზმების არაეკვივალენტურობას ადასტურებს ისეთი ენების არსებობა, რომლებიც ერთსა და იმავე **PS-** გრამატიკულ კლასში არიან და, რომლებიც, ამავდროულად, არიან განსხვავებულ **LA-** გრამატიკულ კლასებში, და პირიქით. მაგალითად, $a^k b^K$ და WW^R ენები **PS-** გრამატიკის ერთი და იგივე კლასში არიან (კონტექსტისაგან თავისუფალ ენათა კლასში), მაგრამ ისინი არიან **LA-** გრამატიკის სხვადასხვა ენობრივ კლასებში: $a^k b^K$ არის წრფივ დროში პარსირებადი **C1-LAG** კლასის ენა, ხოლო WW^R არის n^2 დროში პარსირებადი **C2-LAG** კლასის ენა. პირიქით, $a^k b^K$ და $a^k b^K c^k$ ენები არიან **LA-** გრამატიკის ერთსა და იმავე კლასში (ორივენი არიან **C1-LAG** კლასის ენები), მაგრამ ისინი არიან **PS-** გრამატიკის სხვადასხვა ენობრივ კლასებში: $a^k b^K$ კონტექსტისგან თავისუფალ ენათა კლასშია, ხოლო $a^k b^K c^k$ კი – კონტექსტზე დამოკიდებულ ენათა კლასში.

ის, რომ $a^k b^K c^k$ ენის მსგავსი ენები ორი განსხვავებული ენათა კლასითა (კერძოდ, კონტექსტზე დამოკიდებული და **C1**) და ორი განსხვავებული სირთულის ხარისხით (კერძოდ, ექსპონენციალური და წრფივი) კლასიფიცირდება, ადასტურებს ინდივიდუალურად აღებული ენის შინაგანი სირთულის და ამ ენის ენობრივი კლასის სირთულის ცნებებს შორის არსებულ თვისობრივ განსხვავებულობას.

$a^k b^K c^k$ ენის დაკლასიფიცირება, **PS**-გრამატიკის კონტექსტზე დამოკიდებულ ენად, არ არის განპირობებული იმით, რომ ეს ენა თვისობრივადაა როული, არმედ იმით, რომ **PS**-გრამატიკის ჯერ-ჯერობით ცნობილი არცერთი სხვა უფრო დაბალი საფეხურის ქვეკლასი არ არის სტრუქტურულად მისი შესატყვისი. აქედან გამომდინარე, არც ის უნდა გამოვრიცხოთ, რომ დაიმებნოს ისეთი ახალი ძირებული ფორმალიზმი, რომელიც $a^k b^K c^k$ ენას წრფივი სირთულის ენად დააკლასიფიცირებს.

ამასთან, უფრო დაბალი ენობრივი კლასები განსაზღვრულია როგორც უფრო მაღალი ენობრივი კლასების ქვეკლასები. მაგალითად, $a^k b^K$ ენების მსგავსი ენები **PS**-გრამატიკაში კონტექსტისგან დამოუკიდებლად იწოდება, რადგან ეს ის უმცირესი ენობრივი კლასია, რომელიც ამ ტიპის ენებს შეიცავს. თუმცა, სინამდვილეში, $a^k b^K$ ამავდროულად არის კონტექსტზე დამოკიდებულიც, რადგან კონტექსტზე დამოუკიდებელი ენათა კლასი შედის კონტექსტზე დამოკიდებულ ენათა კლასში. შესაბამისად, განაცხადი იმის თაობაზე, რომ „ $a^k b^K c^k$ არის კონტექსტზე დამოკიდებული ენა, რომელიც **LA**-გრამატიკაში წრფივ დროში **C1-LAG** პარსირებადია“ არ არის წინააღმდეგობრივი: მართლაც, **C1**-ენების კლასი ქვესიმრავლეა **B**-ენების კლასის და, შესაბამისად, **C1-LAG** ტიპის ენა ამავდროულად არის **B-LAG** ტიპის, ანუ კონტექსტზე დამოკიდებული ტიპის ენაც.

12.2 ქვესიმრავლების მიმართებაზე ამ რო იმრაობიაზი

X ენათა კლასი არის Y ენათა კლასის ქვესიმრავლე, ანუ ქვეკლასი (ფორმალურად, $X \subseteq Y$), თუ X კლასის ნებისმიერი ენა ამავდროულად არის Y კლასის ენა. X ენათა კლასი არის Y ენათა კლასის საკუთრივი ქვესიმრავლე, ანუ ქვეკლასი (ფორმალურად, $X \subset Y$), თუ X ქვესიმრავლეა Y -ის და, ამასთან, თუ Y კლასში არსებობს ერთი მაინც ისეთი ენა, რომელიც არ არის X კლასის ენა.

PS-იერარქიის ენობრივ კლასებს შორის ძალაშია შემდეგი სახის ქვესიმრავლეობის მიმართება.

12.2.1 ქვესიმრავლეობის მიმართება PS -იერარქიაში

წესიერი ენები \subset კონტექსტისაგან თავის. ენები \subset კონტექსტზე დამოკ. ენები \subset რეპ. გადათვლაზე ენები ასეთი ქვესიმრავლეობის მიმართება განპირობებულია **PS**-გრამატიკის გადამწერი წესების მზარდი შეზღუდვებით (იხ. 8.2), და იგი საკუთრივი ქვესიმრავლეობის მიმართებად იწოდება.³

12.2.2 ქვესიმრავლეობის მიმართება LA -იერარქიაში

$$C_1 \text{ ენები} \subseteq C_2 \text{ ენები} \subseteq C_3 \text{ ენები} \subseteq B \text{ ენები} \subset A \text{ ენები}$$

ეს ქვესიმრავლეობის მიმართებაც მზარდი შეზღუდვებითაა განპირობებული. თუმცა, თუ **B**-ენების კლასი **A**-ენების⁴ კლასის საკუთრივი ქვესიმრავლეა, სხვა კლასებს შორის საკუთრივი ქვესიმრავლეობის მიმართების არსებობა შეიძლება მხოლოდ ვიგარაუდოთ. საკითხი იმის თაობაზე არის თუ არა **C₂** ენების კლასი **C₃** ენების კლასის საკუთრივი ქვესიმრავლე, დაიყვანება კლასიგური ავტომატების თეორიის ჯერ კიდევ გადაუჭრელ პრობლემაზე, კერძოდ კი იმაზე, თუ რომელია ძალისმიერი: ის, რომ $P \subset NP$, თუ ის, რომ $P = NP$.

NP ენობრივი კლასი შეიცავს ყველა იმ ენას, რომლებიც შეიძლება გამოცნობილ იქნას არადეტერმინირებულ პოლინომურ დროში, მაშინ როდესაც P შეიცავს ყველა იმ ენას, რომლებიც გამოცნობადნი არიან დეტერმინირებულ პოლინომურ დროში.

დეტერმინირებულ პოლინომურ დროში ამოცნობადი ენები აყალიბებენ ენების ბუნებრივ და მნიშვნელოვან კლასს. ეს კლასია **DTIME(nⁱ)**, რომელიც ჩვენთან P -თი აღინიშნება. ინტუიციურად მეტად მიმზიდველია ის, რომ P არის იმ პრობლემათა კლასი, რომლებიც შეიძლება იქნენ უფექტურად გადაჭრილნი. თუმცა ვინმე შეიძლება არ დაგვეთანხმოს იმაში, რომ n^{57} ბიჯიანი ალგორითმი უფექტურია. პრაქტიკა გვარწმუნებს, რომ P სირთულის პრობლემებს, როგორც წესი, აქვთ პოლინომურ დროით სირთულეზე უფრო დაბალი დროითი სირთულის ხარისხი.

J. E. Hopcroft & J. D. Ullman 1979, გვ., 320

³ იერარქიული ლემა (**Hopcroft & Ullman 1979**, გვ.. 228).

⁴ (იხ. **Hopcroft & Ullman 1979**, გვ.. 228, თეორემა 9.8.) **A-LAG** წარმოქნის რეკურსულ ენებს, მაშინ როდესაც, როგორც ეს ნაჩვენებია 11.1 და 11.2-ში, **A-LAG** წარმოქნის კონტექსტზე დამოკიდებულ ენებს.

L ენა იწოდება **NP -სრულად (NP-complete)**, თუ (i) NP კლასის ნებისმიერი ენა დეტერმინირებულ პოლინომურ დროში შეიძლება დაყვანილ იქნეს L ენაზე და თუ (ii) L ენა NP კლასის წევრია. NP -სრული ენები იგება არადეტერმინირებული პოლინომური სირთულის ყველაზე ცუდი შემთხვევების წარმოსადგენად.

ისტორიულად, პირველი NP -სრული ენის კლასიკური მაგალითია **SAT**-ენა, ანუ ბულისეული შესრულებადობის⁵ (**Boolean SATisifiability**) პრობლემა. განვიხილოთ შემდეგი ბულისეული გამოსახულება.

12.2.3 3SAT ენის სფორად ფორმირებული გამოსახულები

$$(x \vee \neg y \vee \neg z) \& (y \vee z \vee u) \& (x \vee z \vee \neg u) \& (\neg x \vee y \vee u)$$

✓ ნიშნით აღნიშნულია ლოგიკური ან (ანუ, დიზუნქცია (**disjunction**)), და ნიშნით აღნიშნულია ლოგიკური და (ანუ, კონუნქცია (**conjunction**)), ასოები გაგებულია პროპოზიციულ, ანუ წინადაღებრივ ცვლადებად, და \neg ნიშნით აღნიშნულია ლოგიკური არ (ანუ, უარყოფა (**negation**)). **3SAT** ენა არის **SAT** ენის ოდნავი გამარტივების შედეგი: კერძოდ, **3SAT** არის შეზღუდული კონუნქტირებაში იმით, რომ ნებისმიერი კონუნქტი უნდა წარმოადგენდეს ზუსტად სამი ცვლადისგან აგებულ დიზუნქციას.

12.2.3 სახის გამოსახულების შესრულებადობის პრობლემა გაიგება როგორც მისი ცვლადების მნიშვნელობათა ისეთი დამნიშვნელების მოძიებაზე, რომელიც გამოსახულებას ჭეშმარიტად აქცევს – თუ კი ასეთი დამნიშვნელება საერთოდ არსებობს. ეს პრობლემა თვისობრივად რთულია, რადგან ანალიზი ითხოვს 2^n ოდენობის სხვადასხვა დანიშვნების გადამოწმებას. მაგალითად, x ცვლადს შეიძლება მნიშვნელობად დაწინშნოს 1 (**ჭეშმარიტი (true)**), ან 0 (**მცდარი (false)**). როცა გამოთვლებში მეორე ცვლადიც ერთვება ჩვენ იძულებულნი ვხდებით განვიხილოთ შემდეგი ოთხი განსხვავებული დანიშვნა: $(x=1, y=1)$, $(x=1, y=0)$, $(x=0, y=1)$ და $(x=0, y=0)$. ანუ, სხვა სიტყვებით: ყოველთვის, როცა თვლისას შემოდის ახალი ცვლადი, შესაძლო დანიშვნების რიცხვი ორმაგდება.

NP-სრული ენის კიდევ ერთი მაგალითია ჩვენთვის უკვე ნაცნობი **SubsetSum** ენა. **3SAT** ენის მსგავსად **SubsetSum** ენაც **C3**-ენების კლასშია (იხ. **C3-LAG**-ის 11.5.9 განსაზღვრება). ამგვარად, **C3**-ენების კლასი მოიცავს NP -სრულ ენებს. უფრო მეტიც, **C3**-ენების კლასი ცხადია შედის NP კლასში, რადგან განსაზღვრების ძალით **C-LAG** მოწმდება არადეტერმინირებულ წრფივ დროში – და, ამგვარად, იგი მით უფრო შემოწმდება არადეტერმინირებულ პოლინომურ დროში.

C2-ენები, მეორე მხრივ, აგებულნი არიან დეტერმინირებულ პოლინომურ დროში პარსირებისთვის. ეს განაპირობებს იმას, რომ ისინი P კლასს მიეკუთვნებიან. იმის დაშვება, რომ **C2**-ენათა კლასი არ არის **C3**-ენათა კლასის საკუთრივი ქვესიმრავლე, გამოიწვევდა იმას, რომ

$$\text{C2-ენათა } \text{კლასი} = \text{C3-ენათა } \text{კლასი}.$$

⁵ იხ. Hopcroft & Ullman 1979, გვ. 324.

ეს კი დაადასტურებდა იმას, რომ არსებობს **SubsetSum** ენის **C2-LAG** გრამატიკა. ამასთან, რადგან NP ენა დეტერმინირებულ პოლინომურ დროში შეიძლება დაყვანილ იქნას **SubsetSum** ენაზე, ამით ისიც დადასტურდებოდა, რომ

$$P = NP.$$

მაგრამ, რადგან P და NP ენების ეკვივალენტობა დაუჯერებლად მიიჩნევა, რწმენა იმის თაობაზე, რომ **C2** ენათა კლასი არის **C3** ენათა კლასის საკუთრივი ქვესიმრავლე, მყარდება.⁶

ის, არის თუ არა **C3** ენათა კლასი **B** ენათა კლასის საკუთრივი ქვესიმრავლე, ასევე ღია საკითხია. აქაც ვინმებ შეიძლება სავარაუდო საყრდენად აიღოს **C3-LAG** ენების NP სისრულე. ცნობილია, რომ კონტექსტური დამოკიდებული ენების გამოცნობა (**CS**-გამოცნობა) არის **PSPACE**-სრული (იხ. **Hopcroft & Ullman 1979**, გვ. 346,7). რადგან დაუჯერებელია⁷ ის, რომ NP კლასში იყოს **PSPACE**-სისრულის პრობლემა, ასევე დაუჯერებელია, ისიც რომ **C**-ენათა კლასი არ იყოს საკუთრივი ქვესიმრავლე **B**-ენათა კლასის.

12.3 LA- და PS- იმრაობიების არაეპიზალუნტობა

დ. ეპლიგეიტის (**D. Applegate**) მიხედვით **L_{no}** ენა, რომელიც **PS**-გრამატიკაში კონტექსტისგან თავისუფალია, **LA**-გრამატიკაში კი – **C3**-ტიპის, ხმაურიან-ენად (**noise-language**) იწოდება. **L_{no}** ენა წარმოქმნის გამოსახულებებს, რომლებიც შეიცავენ 0 და 1 სიმბოლოებს, და რომლებსაც აქვთ **W'#W^R** სტრუქტურა. # სიმბოლო ერთმანეთისგან გამოყოფს **W'** და **W^R** გამოსახულებებს, სადაც **W^R** არის სარკისებული სახე **W**-სი, ხოლო **W'** განსხვადება **W**-სგან იმით, რომ დამატებით შეიცავს 0 და 1 სიმბოლოების გარკვეულ ოდენობას. **W'**-ის ეს დამატებითი სიტყვები არ განსხვავდება მათგან, რომელთაც აქვთ მეტყვილები **W^R**-ში და ისინი ფუნქციონირებენ როგორც ხმაური. ამ ენის კონტექსტისგან თავისუფალი **PS**-გრამატიკა განსაზღვრულია 12.3.1-ში.

12.3.1 L_{no} ენის **PS**-გრამატიკა

$$\begin{array}{lll} S \rightarrow 1S1 & S \rightarrow 1S & S \rightarrow \# \\ S \rightarrow 0S0 & S \rightarrow 0S & \end{array}$$

მარცხენა სვეტში განთავსებული წესები აწარმოებენ # სიმბოლოს მარცხნივ და მარჯვნივ განთავსებულ სარკულად შესაბამის სიტყვებს, შეუა სვეტში განთავსებული წესები აწარმოებენ მხოლოდ **W'**-ის სიტყვებს.

კონტექსტისგან თავისუფალი ენის ერლეისა და **CYK** ალგორითმების მსგავს ჭრადიციულ პარსერებს არ უჭირთ გააკეთონ **L_{no}** ენის n^3 ანალიზი. ეს ასეა რადგან პარსერი იყენებს

⁶ ის, რომ **C2 ⊂ C3** არ არის გამორიცხული დასაბუთდეს პირდაპირ (მაგ. **C2** ენის მტუმბავი ლემის გამოყენებით), რაც იმ შემთხვევაში, თუ ნაჩვენები იქნება, რომ **C2=P** (რაც, თავის მხრივ, მეტად არასარწმუნოა), დაადასტურებდა იმასაც, რომ $P \subset NP$.

⁷ **PSPACE**-სისრულის პრობლემა მოსალოდნელია, რომ არ იყოს არა მხოლოდ **P** კლასში, არამედ ის არც NP კლასში არ უნდა იყოს. უფრო მეტიც, თვისება, რომლის არსებობაც განაპირობებს **PSPACE**-სისრულეს სავარაუდოა არ იყოს ვერიფიცირებადი არც პილონომურ დროში და არც პოლინომურ სიგრძეში

კონტექსტისგან თავისუფალი ენის ძირეულ ინვერსირებულად დაწყვილებულ სტრუქტურას. ქვემოთ აღწერილი 12.3.2 მაგალითი წარმოგვიდგენს L_{no} ენის **10010#101** გამოსახულების **PS**-გრამატიკულ დერივაციას. შესაბამისი მდგომარეობები ნაწარმოებია ერლეის ალგორითმით.

12.3.2 L_{no} ენის **10010#101** გამოსახულების **PS**-გრამატიკული დერივაცია

დერივაციული ზე:	გენერაციული ჯაჭვი:	მდგომარეობები:
	1S1	1. S1 1S1.
	10S01	1.S 0.S0 0S0.
	100S01	0.S 0.S0
	1001S101	0.S OS. 1.S1 1S1.
	10010S101	1.S 0.S0
	10010#101	0.S OS. #

აյ გამოყენებული ერლეის ალგორითმი წარმოქმნის მხოლოდ ორ მდგომარეობას **#** სიმბოლოს წინ მჯდომი თითოეული ტერმინალური სიმბოლოსთვის. მაგალითად, ‘**1.S1**’ და ‘**1.S**’. ამგვარად, თუ შემავალ ჯაჭვში **#** სიმბოლოს წინ ზის **k** ტერმინალური სიმბოლო, მაშინ ალგორითმი **#** სიმბოლოს მიღწევამდე **2k** მდგომარეობას აწარმოებს.⁸

ამის საპირისპიროდ, **C-LAG**-ენებში კატეგორიული ოპერაციები შეეხება დაწყვილებული კუდით დამთავრებულ სტრუქტურებს. ეს სტრუქტურები კარგად ერგება განმეორებათა ნებისმიერ რიცხვს, რის გამოც შეიძლება ამ განმეორებების მოდიფიცირება, მაგ.: ინვერსირება, დაწყვილება, განახევრება და ა.შ..

საზოგადოდ პარსერი და, მათ შორის, **C-LAG** პარსერიც არაეფექტურია, თუ შემავალი მოცემულობა შეიცავს უცნობ რაოდენობას ისეთი სიტყვებისა, რომლებიც შეიძლება განისაზღვრონ მხოლოდ ანალიზის ბოლოს, იმის მიხედვით, უკანასკნელი სიტყვა შეესაბამება წინა სიტყვებს, თუ არა. ეს არის მახასიათებელი თვისება როგორი **NP** ენებისა, ანუ ისეთი ენებისა, რომლებიც მოითხოვენ არადეტერმინირებულ პოლინომიალურ დროს ვერიფიკაციისთვის და ექსპონენციალურ დროს ანალიზისთვის.

LA-გრამატიკისთვის, ერთი მხრივ, **HCFL**⁹ და L_{no} ენებისნაირი კონტექსტისგან თავისუფალი ენები და, მეორე მხრივ, **3SAT** და **SubsetSum** ენებისნაირი **NP**-სრული კონტექსტები დამოკიდებული

⁸ რადგან L_{no} არის დეტერმინირებული კონტექსტისგან თავისუფალი ენა, ის **PS**-გრამატიკაში შეიძლება დაპარსერდეს წრფივ დროში. (იხ. **B. Stuber**, 1993, გვ. 71, ლემა 5.1).

⁹ ის, რომ **PS**-გრამატიკაში **HCFL** პარსირდება პოლინომურ დროში განპირობებულია რამოდენიმე მიზეზით:

1. კონტექსტისგან თავისუფალი **PS**-გრამატიკა იყენებს სირთულის დადგენის **C-LAG**-სგან განსხვავებულ მეთოდებს. უფრო კონკრეტულად, n^3 დროითი სირთულე კონტექსტისაგან თავისუფალი ენებისთვის

ენები სტრუქტურული თვალსაზრისებით ერთი და იგივენი არიან. ამ ოთხ ენას საერთო აქვთ ის, რომ შემავალი მოცემულობა პირველ ნახევარში შეიძლება შეიცავდეს ისეთი სიტყვების ნებისმიერ ოდენობას, რომლებისთვისაც უცნობია საჭირონი არიან თუ არა ისინი, როგორც მეწყვილეები მეორე ნახევარში განთავსებული სიტყვებისთვის.¹⁰

L_{no} ენის **C-LAG** განალიზები ერთადერთი გზა არის # სიმბოლოს წინ მჯდომი სიტყვის ორგვარი ინტერპრეტირება, ერთი, როგორც ‘ნამდვილი’ სიტყვა, და მეორე, როგორც ‘ხმაურიანი’ სიტყვა. ეს შედეგად იძლევა # სიმბოლოს წინ მჯდომი შემავალი ჯაჭვის წაკითხვის ექსპონენციალურ ოდენობას, თთოვეულ წაკითხვას მისი საკუთარი კატეგორიით. მაგალითად, თუ შემავალი ჯაჭვი არის **10010#**. . . მაშინ ერთ-ერთ კითხვას აქვს კატეგორია (**10010**), რაც იმ ჰიპოტეზას წარმოადგენს, რომ # სიმბოლოს წინ მჯდომ ყველა სიტყვა ‘ნამდვილი’. ამ შემავალი ჯაჭვის სხვა წაკითხვა იძლევა კატეგორიას (**1001**), რაც იმ ჰიპოტეზას წარმოადგენს, რომ # სიმბოლოს წინ მჯდომი უკანასკნელი სიტყვა ‘ხმაურია’, და ა.შ..

12.3.3 L_{no} მნის C3-LAG

$$LX =_{\text{def}} \{[0(0)], [1(1)], [\#(\#)]\}$$

$$ST_S =_{\text{def}} \{ [(\text{seg}_c), \{r_1, r_2, r_3, r_4, r_5\}] \} \text{ სადაც } \text{seg}_c, \text{seg}_d \in \{0,1\}.$$

$$r_1: (\text{seg}_c)(\text{seg}_d) \Rightarrow \varepsilon \quad \{r_1, r_2, r_3, r_4, r_5\}$$

$$r_2: (\text{seg}_c)(\text{seg}_d) \Rightarrow (\text{seg}_d) \quad \{r_1, r_2, r_3, r_4, r_5\}$$

$$r_3: (X)(\text{seg}_c) \Rightarrow (X) \quad \{r_1, r_2, r_3, r_4, r_5\}$$

$$r_4: (X)(\text{seg}_c) \Rightarrow (\text{seg}_c X) \quad \{r_1, r_2, r_3, r_4, r_5\}$$

$$r_5: (X)(\#) \Rightarrow (X) \quad \{r_6\}$$

$$r_6: (\text{seg}_c X)(\text{seg}_c) \Rightarrow (X) \quad \{r_6\}$$

$$ST_F =_{\text{def}} \{ [\varepsilon \text{ rp}_6] \}$$

ამ **C3-LAG**-ში არსებული და **r₃**-ისა და **r₄**-ის მიზეზით წარმოქმნილი +რეკურსული ორაზროვნება, მაგალითად, ხასათდება (i) შემავალ პირობასთან თავსებადობით (ფაქტიური იდენტურობით), (ii) თანაშემოსვლებით წესთა პაკეტში, და (iii) ერთსა და იმავე საანალიზო გზაზე კვლავ გამოყენებადობით. **r₃** წესი იგორირებას უკეთებს შემდეგ სიტყვას და ამუშავებს მას როგორც ხმაურს. **r₄** წესი მარცხნიდან, ანუ დასაწყისიდან აბამს ახალი შემდეგი სიტყვის კატეგორიას სიტყვის საწყისის კატეგორიას, რათა უზრუნველყოს ის, რომ მას ჰყავდეს მეწყვილე შემავლის მეორე ნახევარში.

L_{no} ენისთვის 12.3.3-ში აღწერილი ეს **C3-LAG** ემთხვევა იმ **C3-LAG**-ს, რომელიც 11.5.8-ში აღიწერა **NP**-სრული **SubsetSum** ენისთვის, სადაც თთოვეული **a_i** შეიძლება ინტერპრეტირებულ იქნეს, ან როგორც ხმაური, ან როგორც ‘ნამდვილი’ ქვესიმრავლე. **SubsetSum** ენის ეს **C3-LAG** ანალიზი კონტექსტზე დამოკიდებულია, რადგან ზოგიერთი წესი (მაგ. **r₅**) ამოწმებს წინადადების

საზოგადოდ მეტწილ დამოკიდებულია იმაზე, რომ გამოიყენება მრავალლენტიანი ტიურინგის მანქანა. სირთულე **C-LAG**-ში კი ისაზღვრება უშუალოდ გრამატიკის საფუძვლზე.

2. ამსტრაქტული ავტომატისაგან განსხვავებით არანაირი **E**-მოძრაობა არ დაიშვება **C-LAG**-ში.

¹⁰ კონტექსტისგან თავისუფალი **L_{no}** ენის **C-LAG**-სირთულე იგივეა, რაც კონტექსტისგან თავისუფალი **L³_{no}** ენის სირთულე, რომელიც, თავის მხრივ, წარმოქმნის **W#W#W**” სახის გამოსახულებებს, სადაც **W'** და **W''** არიან ხმაურიანი ვერსიები **W** სიტყვის. **L_{no}** ენის **LA**-გრამატიკა არის უფრო მაღალი სირთულის კლასში, ვიდრე შესაბამისი **PS**-გრამატიკა, რადგან **C-LAG** გრამატიკები არ არიან აგბული კონტექსტისგან თავისუფალი ენების ფიქსირებული ინვერსირებადად დაწყვილებული სტრუქტურების დასამუშავებლად და **E**-მოძრაობა არ არის დაშვებული.

საწყისის კატეგორიის დასაწყისს, სხვა წესები (მაგ. 14) კი ბოლო ნაწილს. ამასთან, **L_{no}** ენის ეს **C3-LAG** კონტექსტისგან თავისუფალია, რადგან მისი კატეგორიალური ოპერაციები ამოწმებენ მხოლოდ წინადადების საწყისის კატეგორიის დასაწყისს.

12.4 უმცირესი **LA-** და **PS-** კლასების შედარება

კონტექსტისგან თავისუფალი **PS**-გრამატიკა ფართოდ გამოიყენება, რადგან იგი **PS**-გრამატიკული იერარქიის შიგნით უზრუნველყოფს გამოთვლადობას დაქვემდებარებული სხვადასხვა წარმოქმნითი უნარის მქონე ენების დიდ რიცხვს. არსებობს ზოგადი ენათმეცნიერული შეთანხმება, რომ კონტექსტისაგან თავისუფალი **PS**-გრამატიკა არ ესადაგება ბუნებრივი ენის სტრუქტურულ მახასიათებლებს. იგივე სიტუაცია კომპიუტერულ მეცნიერებებში, სადაც, ბუნებრივი ენების მსგავსად, კონტექსტისგან თავისუფალი **PS**-გრამატიკა არასაკმარისად მიიჩნევა პროგრამული ენების სტრუქტურული აღწერისათვის.

საიდუმლო არ არის, რომ კონტექსტისაგან თავისუფალი **PS**-გრამატიკები იძლევიან მხოლოდ პირველი რიგის აპროქსიმაციას იმ მრავალი მექანიზმისა, რაც პროგრამული ენების სინტაქსის გამოიყენება.¹¹

ს. გინზბურგი, 1980, გვ 7.

ამგვარად, დიდი ხანია მოთხოვნაშია ალტერნატივა ბუნებრივი და პროგრამული ენების უკეთ აღსაწერად. ასეთი ახალი ენობრივი კლასის მოძიების მცდელობების უმეტესობა მიმდინარეობდა ისეთი კონსერვატორული გაფართოებების აგების გზით, რომლებიც მეტად მჭიდროდ მიჰყვებოდნენ კონტექსტისგან თავისუფალ **PS**-გრამატიკებს. ისინი იგებოდნენ გარკვეული მექანიზმების დამატებით, რაც შედეგად იძლეოდა ახალ-ახალ დამატებით ენობრივ კლასებს. ამასთან, როგორც წესი, ეს ახალი ენობრივი კლასები ქვესიმრავლეობის მიმართებაში იყვნენ **PS**-გრამატიკული იერარქიის ენობრივ კლასებთან. მაგ. კონტექსტისგან თავისუფალი ენები აყალიბებენ ხეებთან მიკავშირებული ენებად (**TAL**)¹² წოდებულ საკუთრივ ქვესიმრავლეს, რომელიც, თავის მხრივ, ქმნის ინდექსირებულ ენებად¹³ წოდებულ ენებს, რაც უკვე კონტექსტზე დამოკიდებული ენების საკუთრივი ქვესიმრავლეა.

ორმოცდათ წელზე მეტი ხნის **PS**-გრამატიკული ტრადიცია და მრავალ ხნიანი არარსებობა საფუძვლიანი ალტერნატივისა არასაკმარისი მიზეზია იმისათვის, რომ ფორმალური ენების **PS**-გრამატიკული იერარქია და მათი გაფართოებები განვიხილოთ როგორც განსაკუთრებით ‘ბუნებრივი’: ყველაფერს რომ თავი დავანებოთ ეს სხვადასხვა ენობრივი კლასები მეტი არაფერია, თუ არა ერთი გარკვეული ფორმალიზმის სხვადასხვაგვარი შეზღუდვები.

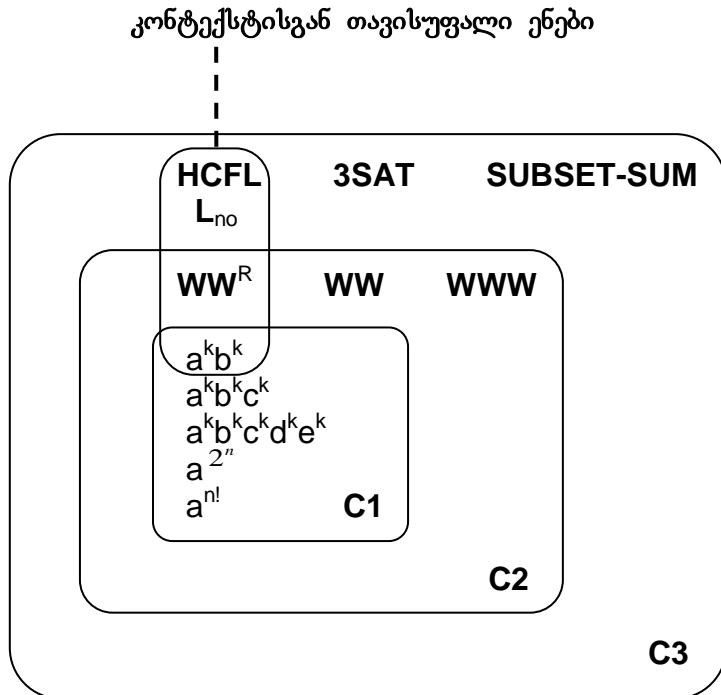
მაგალითად, კონტექსტისგან თავისუფალი ენები განსაზღვრულია იმ შეზღუდვებით, რომლებიც ეხება მხოლოდ გადაწერის წესებს (იხ. 8.1.2). ანალოგიურად **C1**, **C2** და **C3** ენები განსაზღვრულია მხოლოდ იმ შეზღუდვებით, რომლებიც ეხება მხოლოდ **LA**-გრამატიკულ ფორმალიზმს (იხ. 11.5.9) ეს ორი განსხვავებული ტიპის შეზღუდვები შედეგად იძლევა **PS**- და **LA**- გრამატიკულ იერარქიებს, რომლებიც ერთმანეთის ორთოგონალურებია:

¹¹ იხ. აგრეთვე მ. ჰარისონი (**M. Harrison**) 1978, გვ. 219, სადაც იგივენაირი განწყობაა ასახული.

¹² ა.კ. ჯოში, 1975.

¹³ იხ. **Hopcroft & Ullman**, 1979, გვ. 389. მტუმბავი ლემა ინდექსირებული ენებისთვის (დამტკიცება ეპუთვნის ტ. ჰაიაში) 1973.

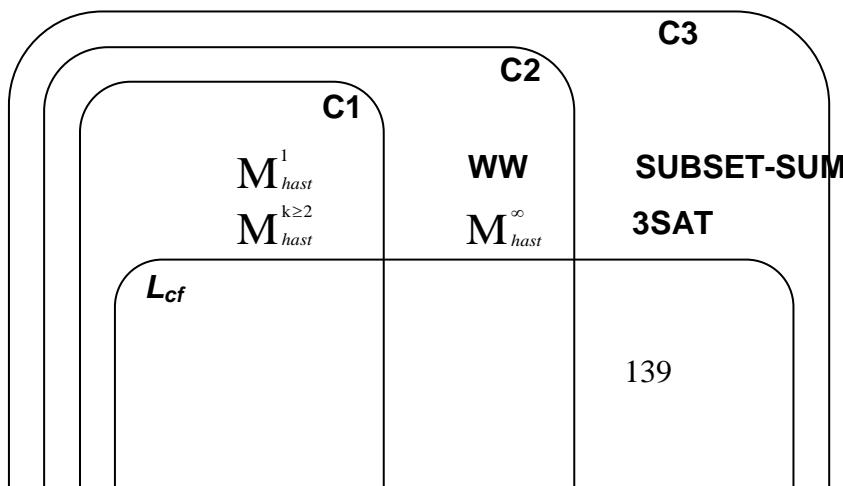
12.4.1 ორთოგონალური მიმართება C- და CF- ენებს შორის



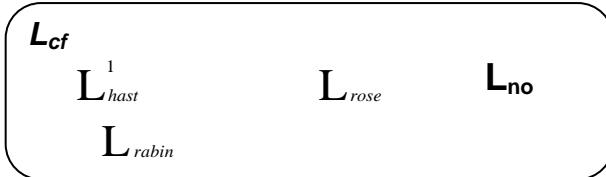
წინათეორიული ხედვებით ვინმე შეიძლება განეწყოს $a^k b^k$ ენის $a^k b^k c^k$, $a^k b^k c^k d^k$, და ა.შ. ენებთან ერთად დაკლასიფიცირებლად, ისევე როგორც WW^R და WW ენებისა. – ის, რომ **PS**-იერარქია დებს $a^k b^k$ და WW^R ენებს ერთ (კონტექსტუალური თავისუფალ ენათა) კლასში, ხოლო $a^k b^k c^k$, $a^k b^k c^k d^k$, და ა.შ. ენებს **WW** ენასთან ერთად მეორე (კონტექსტუალური თავისუფალ ენათა) კლასში, გაუთვითცნობირებული მკითხველისთვის გასაკვირვიც კია. ამ თვალსაზრისებით **LA**-იერარქია ინტუიციურად უფრო ბუნებრივად არის მოწესრიგებული, რადგან აյ a^k , $a^k b^k$, $a^k b^k c^k$, $a^k b^k c^k d^k$ და ა.შ.. კლასიფიცირებულნია წრფივ **C1**-ენათა კლასში. ხოლო, WW^R და WW , $WW^R W$ და WWW და ა.შ. ენები დაკლასიფიცირებული არიან, როგორც **C2**-ენები.

თუ L_{dcf} დეტერმინირებულ კონტექსტუალური თავისუფალ ენებსა და L_{cf} არადეტერმინირებულ კონტექსტუალური თავისუფალ ენებს შორის განსხვავებას განვიხილავთ **PS**-გრამატიკის შიგნით, ორთოგონალური მიმართება **PS**- და **LA**- იერარქიებს შორის უფრო ცხადად გამოიკვეთება

12.4.2 ორთოგონალური L_{dcf} , L_{cf} , **C1**, **C2**, და **C3** პლასივიპანია



$$\begin{array}{c} L_{hast}^{k \geq 2} \\ 0^k 1^k U 0^k 1^{2k} \\ \hline \end{array} \quad \begin{array}{c} L_{hast}^{\infty} \\ WW^R \end{array} \quad \text{HCFL}$$



L_{acf} კლასი, ისევე როგორც L_{cf} კლასი, გასწვრივ კვეთს **C1**, **C2** და **C3** კლასებს.

LA- გრამატიკის ეს ალტერნატიული კლასიფიკაცია უზრუნველყოფს ახალ პერსპექტივებს ფორმალურ ენათა თეორიაში. უფრო მეტიც, რადგან **PS**-გრამატიკული იერარქია რეკონსტრუქტირებადია **LA-** გრამატიკულ იერარქიაში (იხ. 11.2) კლასიკური ავტომატების თეორიის ღია პრობლემები შეიძლება პირდაპირ იქნენ გადატანილნი **LA-** გრამატიკაში. (იხ. 12.2).

12.5 ბუნებრივი ენების ფრაგი სირთულე

ისეთი კონტექსტზე დამოკიდებული ენის არსებობა, რომელიც არ არის **C**- ენა, დაადასტურებდა **C**- ენების საკუთრივ ქვესიმრავლების **B**-ენებთან მიმართებაში (იხ. 12.2.2). ასეთ ენაში კატეგორიის სიგრძე უნდა იზრდებოდეს კონტექსტზე დამოკიდებული ენების **LBA**-განსაზღვრების მიხედვით, მაგრამ ეს ზრდა უნდა იყოს უფრო სწრაფი ვიდრე **C-LAG**-ების წესებზე დაფუძნებული კატეგორიული ოპერაციები ამას უშვებენ. თუმცა, ის, რომ ბუნებრივი ენის სტრუქტურები ამ სახის ენით აღიწერებოდეს, ძლიერ არადამაჯერებელია.

თუ ბუნებრივი ენები აღიწერებიან **C-LAG**-ებით, მაშინ შემდეგი ორი ეკვივალენტური შეკითხვებია:

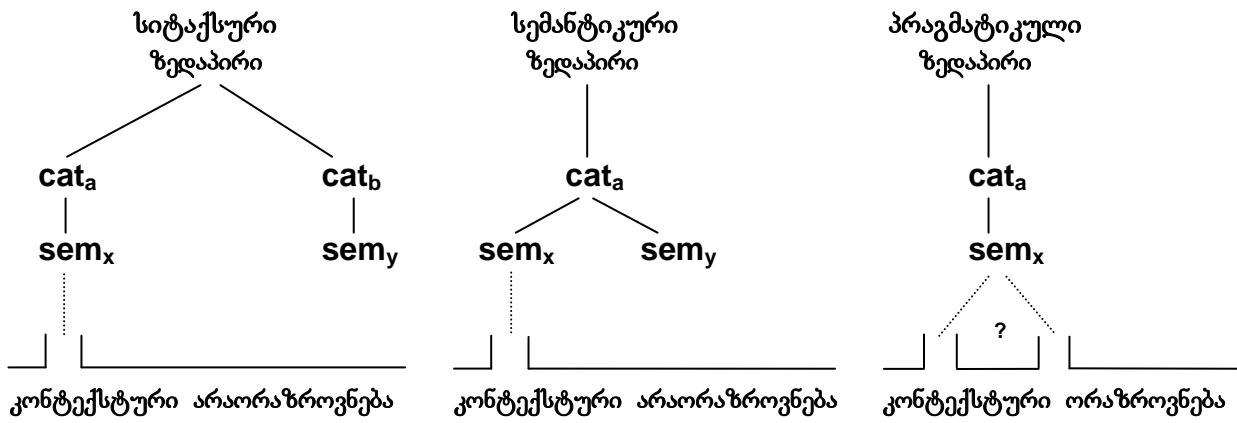
- (i) რამდენად როგორია ბუნებრივი ენა?
- (ii) რამდენად ორაზროვანია ბუნებრივი ენა?

ეს ასეა, რადგან **C-LAG**-ების სხვადასხვა ქვეკლასები ერთმანეთისგან განსხვავდებიან მხოლოდ ორაზროვნების ხარისხით.

ბუნებრივი ენის გამონათქვამი იწოდება ორაზროვანად, თუ მის მსმენელს მისგან შეუძლია გამოიყვანოს ერთზე მეტი მნიშვნელობა. ეს შეიძლება განპირობდეს სინტაქსური, სემანტიკური, ანდა პრაგმატული ორაზროვნებით. თუმცა, გამოსახულების კომბინაციურ სტრუქტურაზე დაფუძნებული ანალიზის სირთულეზე გავლენას მხოლოდ სინტაქსური ორაზროვნება ახდენს.

ენის **SLIM** თეორიის მიხედვით: გამოსახულება სინტაქსურად ორაზროვანია, თუ იგი იღებს ერთზე მეტ სტრუქტურულ ანალიზს. გამოსახულება სემანტიკურად ორაზროვანია, როცა სინტაქსურად არაორაზროვანია და როცა აქვს სულ ცოტა ორი მნიშვნელობა. პრაგმატულ ორაზროვნებას განაპირობებს ისეთი მნიშვნელობა, რომელსაც მოცემულ კონტაქსტში ერთზე მეტი გამოყენება (**use**) აქვს.

12.5.1 ორაზროვნების SLIM თეორიული ანალიზი



სინტაქსური და სემანტიკური ორაზროვნებები განპირობებულია და, შესაბამისად, ხასიათდება თავად გამოსახულების ტიპით, პრაგმატიკული ორაზროვნება კი – იმ მოღიანი გამონათქვამით (**utterance**), რომელშიც ეს ორაზროვნება ვლინდება.

ენობრივი გამოსახულება სინტაქსურად ორაზროვანია, თუ ის ხასიათდება ისეთი ალტერნატიული კატეგორიებით (აյ **cat_x** და **cat_y**), რომელთაგან თოთოეულის კატეგორიული წაკითხვა იძლევა თავის საკუთარ მნიშვნელობას (აյ **sem_x** და **sem_y**). მაგალითად, **won** წარმოქნის სინტაქსურ ორაზროვნებას კორეული ვალუტის აღმნიშვნელ არსებით სახელსა და **to win** ზმნის წარსულ დროს შორის. ამასთან, თუ კონტექსტი არ ხსნის **sem_x** და **sem_y** კატეგორიების ორაზროვნულ გავებას, მაშინ ეს სინტაქსური ორაზროვნება წარმოქმნის ისეთ გამონათქვამს, რომელსაც აქვს ერთზე მეტი მნიშვნელობა.

ენობრივი გამოსახულება სემანტიკურად ორაზროვანია, თუ ის ხასიათდება მხოლოდ ერთი სინტაქსური ანალიზით, მაგრამ ერთზე მეტი მნიშვნელობას ითვლის. მაგალითად, ენობრივი გამოსახულება **perch** სემანტიკურად ორაზროვანია: მისი ერთი მნიშვნელობა გადის თევზის სახეობაზე, მეორე კი – ჩამოსაჯდომ ადგილზე. ეს განაპირობებს იმას, რომ სინტაქსურად **perch** არ არის ორაზროვანი, რადგან ორივე წაკითხვის შემთხვევაში იგი ნაუნის, ანუ არსებითი სახელის კატეგორიით ხასიათდება.¹⁴ შინაგანი დაწყვილებისთვის, სემანტიკური ორაზროვნება ემსგავსება სინტაქსურს, რადგან ორივე შემთხვევაში გამოსახულებას აქვს ერთზე მეტი მნიშვნელობას.

პრაგმატიკული ორაზროვნება განპირობებულია მოცემულ კონტექსტში ერთი მნიშვნელობას ითვლის ალტერნატიული გამოყენებით. მაგალითად, კონტექსტში, სადაც ორი მაგიდა დგას ერთმანეთის გვერდი გვერდ, ატერანი, ანუ ენობრივი გამონათქვამი **Put the coffee on the table** იქნებოდა პრაგმატიკულად ორაზროვანი, რადგან არ არის ცხადი ამ ორი მაგიდიდან რომელს გულისხმობს მეტყველი (იხ. 5.2 ქვეთავი). სინტაქსური და სემანტიკური ორაზროვნებისგან განსხვავებით პრაგმატიკული ორაზროვნების ბუნება არ იხსნება კონტექსტით.

სინტაქსური ორაზროვნება განაპირობებს იმას, რომ გამოსახულებას ერთზე მეტი მნიშვნელობა აქვს. ამასთან, გამოსახულება სემანტიკურად ორაზროვანია მხოლოდ მაშინ, თუ ის არ არის

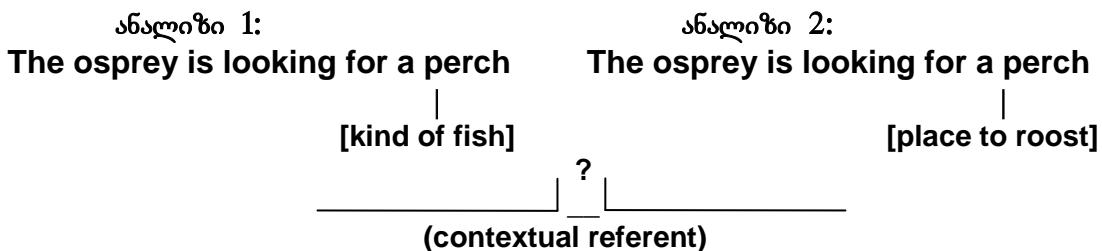
¹⁴ სინტაქსური კატეგორიზაციისას იგნორირება **to perch**-ის შინაარსულ დატვირთვებს.

სინტაქსურად ორაზროვანი (სხვადასხვა სინტაქსური წაკითხვების არსებობის თვალსაზრისით). ანალოგიურად, გამონათქვამი პრაგმატულად ორაზროვანია მხოლოდ მაშინ, თუ იგი აგებულია სინტაქსურად და სემანტიკურად არაორაზროვანი გამოსახულებებით. ცალკე აღებული სიტყვის სინტაქსურ, და სემანტიკურ ორაზროვნებას ლექსიკურ ორაზროვნებასაც უწოდებენ.

ბუნებრივი ენის გამოსახულების სირთულის განსასაზღვრავად პრაგმატული ორაზროვნება არაძალისმიერია. ეს ასეა, რადგან პრაგმატული ორაზროვნება არ არის დამოკიდებული გამოსახულების ტიპზე, არამედ იგი წარმოიქმნება გამოსახულების სემანტიკური ინტერპრეტაციისა და კონტექსტის ურთიერთმიმართებისას (იხ. 4.4).

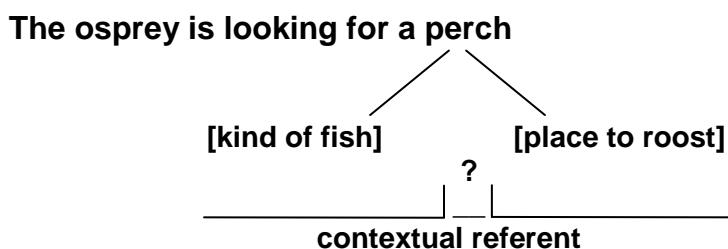
მსგავსად ამისა, არც სემანტიკური ორაზროვნებაა ძალისმიერი ბუნებრივი ენის სირთულის განსაზღვრისას, რადგან სემანტიკური ორაზროვნებები განსაზღვრების ძალით უკავშირდება მხოლოდ სინტაქსურად არაორაზროვან გამოსახულებებს. მაგალითად, იმ ფაქტის გათვალისწინებით, რომ **perch** სიტყვის ორივე წაკითხვას აქვს ერთი და იგივე კატეგორია, გასაგები ხდება, რომ **The osprey is looking for a perch** წინადადებას არ შეიძლება ქონდეს ორი განსხვავებული სინტაქსური ანალიზი:

12.5.2 სემანტიკური ორაზროვნების არაპრეპერატული ანალიზი



ლინგვისტურად კორექტული ანალიზი განიხილავს საანალიზო წინადადებას როგორც სინტაქსურად არაორაზროვანს და ორაზროვნებას ამუშავებს სემანტიკურად **perch** სიტყვისთვის ორი განსხვავებული მნიშვნელობის მინიჭებით.

12.5.3 სემანტიკური ორაზროვნების პრეპერატული ანალიზი



მაშინაც კი, როცა სინტაქსური დერივაცია არაორაზროვანია, სემანტიკური ინტერპრეტაცია უზრუნველყოფს ორ განსხვავებულ მნიშვნელობას, რომელთაგან თითოეული, თავის მხრივ, კონტექსტურად იძლევა შესაძლებლობას უზრუნველყოს ორი განსხვავებული მნიშვნელობა₂. 12.5.3 ანალიზში გამოყენებული მეთოდი იწოდება სემანტიკურ გაორმაგებად¹⁵ (**semantic doubling**).

¹⁵ პირველად შემოთავაზებულ იქნა **Col.-ში**, გვ. 219-232 და 239-247.

სემანტიკური გაორმაგების მეთოდი ეფუძნება ბუნებრივი კომუნიკაციის [2+1] დონის სტრუქტურას. გაანალიზებული გამოსახულებისთვის ერთზე მეტი მნიშვნელობას მინიჭება გვარწმუნებს, რომ აუცილებელი არ არის სემანტიკური განსხვავებების სინტაქსურ დონეზე განრჩევა.

ორაზროვანი გამოსახულება შეიძლება გაანალიზდეს სემანტიკური გაორმაგების ტერმინებში, მაშინ როცა განსხვავება სემანტიკურ დონეზე სინტაქსურ საფეხურთან კავშირშია **არა (no)** მიმართებით (იხ. 12.5.3), ან სისტემატიკური განსხვავების მიმართებით (იხ. 12.5.4). ამგვარად, სინტაქსური ორაზროვნება შეიძლება შეიზღუდოს იმ შემთხვევებით, რომლებშიც განსხვავებული სემანტიკური წაკითხვები დაკავშირებულია არასისტემურ – და, ამგვარად, წინასწარგანუზღვრელ – სინტაქსურ ალტერნატივებთან.

სისტემური სინტაქსური ალტერნატივების არსებობის შემთხვევაში სემანტიკური გაორმაგების მაგალითებად განვიხილოთ პრეპოზიციული ფრაზები (**prepositional phrases**). ისინი, როგორც ეს ნაჩვენებია შემდეგი მაგალითით, საზოგადოდ უშვებენ პოსტნომინალურ და ადვერბიალურ ინტერპრეტაციებს.

12.5.4 პირველი ფილი ფრაზების მრავალჯერადი ინტერპრეტაციები

The man saw the girl with the telescop.

Julia ate the apple on the table behind the tree in the garden.

პირველი მაგალითის ინტერპრეტაცია ორ განსხვავებულ მნიშვნელობას იძლევა. ადვერბიალური კითხვისას პრეპოზიციული ფრაზა **with the telescop** მოდიფიცირებას უკეთებს **saw** ზმნას. პოსტნომინალური კითხვისას იგივე პრეპოზიციული ფრაზა მოდიფიცირებას უკეთებს **the girl** კონსტრუქციას.

12.5.4.-ის მეორე მაგალითი ერთის ნაცვლად შეიცავს სამ პრეპოზიციულ ფრაზას და ილუსტრირებას უკეთებს თეორიულ შესაძლებლობას იმისა, რომ ასეთი ფრაზების დამატება შეიძლება შეუზღუდავი ოდენობით. ეს წამოჭრის საკითხს იმის თაობაზე, ორმაგდება თუ არა სინტაქსური წაკითხვების რაოდენობა ყოველი ასეთი დამატებისას. ქვემოთ მოყვანილი ანალიზი წარმოგვიდგენს ალტერნატიული პრეპოზიციური ფრაზების შეცდომით სინტაქსურ დამუშავებას.

12.5.5 რეპროსული ფსევდო-მრაზროვნება

წაკითხვების ოდენობა:

2¹

2²

2³

Julia ate the apple

behind the tree
postnominal

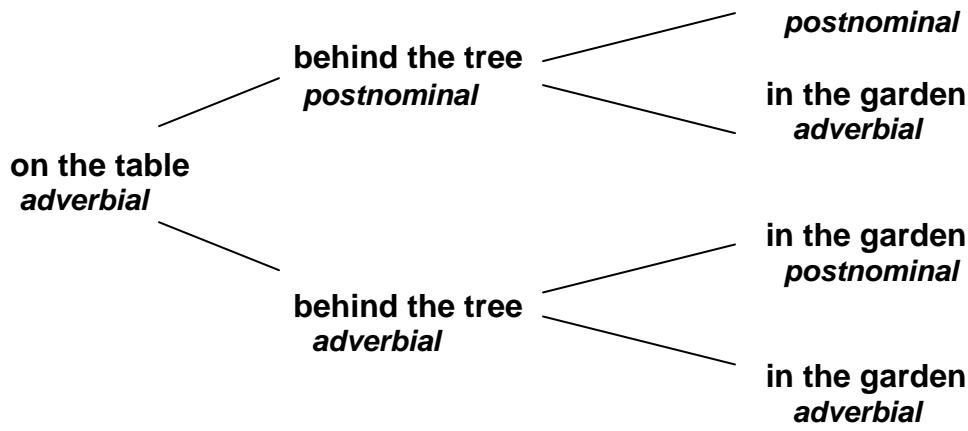
behind the tree
adverbial

in the garden
postnominal

in the garden
adverbial

in the garden
postnominal

in the garden
adverbial
in the garden



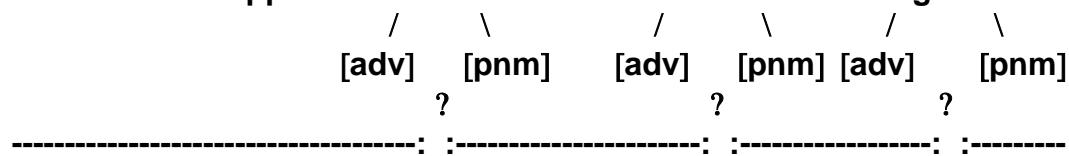
ამგვარი ანალიზების საფუძველზე იყო ის, რომ ბუნებრივი ენები განიხილებოდა სულ ცოტა ექსპონენციალური სირთულის მქონედ.

ბუნებრივი კომუნიკაციის მექანიზმებისთვის, სხვათაშორის, სემანტიკური წაკითხვების გარეთ მყოფი სინტაქსური წაკითხვების ასეთი გამრავლება არამიზნობრივია. 12.5.5 ანალიზის მსგავსი ანალიზი დასაშვებია, თუმცა მხოლოდ იმის საჩვენებლად, რომ ზოგჯერ არაეფექტური გრამატიკული ანალიზიც შეიძლება დაიწეროს.¹⁶ – კარგი სინტაქსური ანალიზი უნდა იმიზნებდეს ლიტერალური მნიშვნელობების ისეთი აბსოლიტური მინიმუმის მოძიებას, რომელიც საკმარისი იქნებოდა ყველა შესაძლო გამოყენებების დასამუშავებლად.

კომუნიკაციური მიზნებისთვის პრეპოზიციული ფრაზების ადეკვატური დამუშავება ითხოვს მხოლოდ სემანტიკურ დონეზე გაკეთებულ ადვერბიალურ და პოსტნომინალურ წაკითხვებს. ზემოაღნიშნულის შესაბამისად, 12.5.6 წარმოგვიდგენს სემანტიკური გაორმაგების პრინციპზე დამყარებულ ალტერნატიულ ანალიზს.

12.5.6 სემანტიკურ გამორჩებები და მყარებული კორექტული ანალიზი

Julia ate the apple on the table behind the tree in the garden



წინადადების ენობრივი გამოსახულება გაანალიზებულია როგორც სინტაქსურად არაორაზროვანი.¹⁷ ამასთან, სემანტიკური ინტერპრეტაცია სემანტიკურ დონეზე სისტემურად უნიშნავს თითოეულ პრეპოზიციულ ფრაზას ორ განსხვავებულ ლიტერალურ მნიშვნელობას.

¹⁶ უფრო მეტიც, მაგ., არაორაზროვანი აბ^k ენისთვის თავისუფლად შეიძლება დაიწეროს მრავალი ისეთი ორაზროვანი გრამატიკა, რომლებიც სირთულის ხარისხს წრფივიდან ექსპონენციალურად და რიგ შემთხვევებში ამოუხსნადობადეც კი ზრდიან.

¹⁷ სიტაქსურად, პრეპოზიციონალური ფრაზები კატეგირიზდებიან მრავალგაზისადი ალნიშვნებით (**multicat notation**), მაგ., როგორც (**adv&pnm**) (იხ. 15.2).

12.5.3 ანალიზის მსგავსად 12.5.6 ეფუძნება ბუნებრივი კომინიკაციების [2+1] დონის სტუქტურებს.
12.5.6 ანალიზი საკმარისია განსხვავებული მნიშვნელობაზების მოდელირებისათვის, რომლებიც შეიძლება წარმოიქმნან განსხვავებულ კონტექსტებზე მიკავშირებულ ინტერპრეტაციებში.

12.5.6-ის სემანტიკურ გაორმაგებაზე დამყარებული ანალიზი უფრო ეფექტურიცაა და უფრო კორექტულიც, ვიდრე 12.5.5-ის სემანტიკური წაკითხვების გარეთ მყოფი სინტაქსური წაკითხვების გამრავლებაზე დამყარებული ანალიზი. მაშინ როდესაც 12.5.5 არის ექსპონენციალური სირთულის, 12.5.6-ის ალტერნატიული ანალიზი არის უმდაბლესი შესაძლო, კერძოდ კი წრფივი სირთულის. ეს ძალაშია ორივე – როგორც სინტაქსური, ისე სემანტიკური დონეებისათვის. 12.5.6 ანალიზით განსაზღვრული და ამ ანალიზის საინტერპრეტაციო კონტექსტებთან დამატებული მნიშვნელობაზების ალტერნატიული ოდენობათა რიცხვი მხოლოდ **2n**-ის ტოლია.

თუ შესაძლებელი იქნება როგორც ჩანს ბუნებრივ ენებში მართლაც არსებული +რეკურსული სინტაქსური ორაზროვნებების (მაგ. 12.5.5) დამუშავება, ისევე როგორც მუშავდება სემანტიკური ორაზროვნებები (მაგ. 12.5.6), მაშინ მხოლოდ 11.3.5-ისა და 11.3.6-ის მსგავსი –რეკურსული ორაზროვნებები რჩება როგორც კანდიდატები სიტაქსური დამუშავებისა. ამასთან, +რეკურსული ორაზროვნების გარეშე დარჩენილი ბუნებრივი ენათა კლასი **C1**-ენათა კლასის ქვესიმრავლედ აზრდება. ჩვენ ვაყალიბებთ ამ დასკვნას როგორც ბუნებრივი ენების სინტაქსის ემპირიული სირთულის ჰიპოტეზის სახით, მოკლედ მას **CoNSyx** ჰიპოტეზას ვუწოდებთ.

12.5.7 CoNSyx ჰიპოტეზა (ბუნებრივი ენების სინტაქსის სირთულე)

ბუნებრივი ენები ქვესიმრავლედ შედიან **C1**-ენების კლასში და პარსირდებიან წრფივ დროში.

წრფივი სირთულე, რომელიც ბუნებრივ ენებს **CoNSyx** ჰიპოტეზით ესაზღვრება თანხმობაშია იმ ფაქტთან, რომ მეტყველსა და მსმენელ ადამიანებს ჩვეულებრივ არა აქვთ არანაირი სირთულე იმისა, რომ მშობლიური ენით აწარმოონ და გაიგონ¹⁸ რთული ტექსტებიც კი. ეს **CoNSyx** ჰიპოტეზა 21 თავში გასრულდება სემანტიკის სირთულის თაობაზე ჩამოყალიბებული ჰიპოტეზით, რომელიც **CoNSem** ჰიპოტეზად იწოდება (იხ. 21.5.2).¹⁹

სავარჯიშოები

ქვეთავი 12.1

¹⁸ სულ ცოტა სინტაქსურ დონეზე მაინც.

¹⁹ **CoNSyx** ჰიპოტეზით ბუნებრივი ენის სინტაქსისთვის განსაზღვრული სტუქტურული თვისებები შეიძლება და უნდა იქნეს ემპირიულად ტესტირებული მრავალი სხვადასხვა ენის გაფართოებადი ანალიზით. 12.5.7-ის შესაძლო კონტრაგალითო თუ არის უნდა იყოს +რეკურსული სინტაქსური ორაზროვნების მქონე ბუნებრივ ენაში, რაც, ბუნებრივია, სემანტიკური გაორმაგების პრინციპზე დამყარებული ალტერნატიული ანალიზის გაკეთების საშუალებას არ მოგეხმდა. ბუნებრივი ენების **SLIM** თეორიის ფარგლებში მოცემული ბუნებრივენობრივი კომუნიკაციის მექანიზმები გავუიქრებინებს ამგარი ბუნებრივი ენების არ არსებობას.

1. რატომ არიან ენობრივი იერარქიის ქვედა კლასები (ე.ი. შედარებით დაბალი წარმოქმნითი უნარის მქონები) განსაკუთრებით საინტერესონი ემპირიული ენათმეცნიერული კვლევებისთვის?
2. აღწერეთ **LA**- და **PS**- გრამატიკული იერარქიის ქვეკლასების სირთულის ხარისხები.
3. როგორ მტკიცდება ფორმალურად ორი ენობრივი კლასის არაეკვივალენტურობა?
4. შეადარეთ **$a^k b^k c^k$** და **SubsetSum** ენების შინაგანი სირთულე.
5. რომელი თვისებით ისაზღვრება ენის ენობრივ კლასთან კუთვნილება?
6. ახსენით რა აზრით არიან **LA**- და **PS**- გრამატიკული იერარქიები ორთოგონალური.

ქვეთავი 12.2

1. შეადარეთ ქვესიმრავლეობის მიმართებები **LA**- და **PS**- იერარქიებში.
2. რა ფორმალური მეოთხით მტკიცდება ის, რომ 3 ტიპის, შესაბამისად 2 ტიპის ენათა კლასი ქვესიმრავლეა 2 ტიპის, შესაბამისად 1 ტიპის ენათა კლასის?
3. განსაზღვრეთ **P** და **NP** ენობრივი კლასები?
4. რატომ არის **3SAT** ენა შინაგანად რთული?
5. ავტომატთა თეორიის რომელი გადაუჭრელი პრობლემა არის დაკავშირებული **LA**- გრამატიკის ღია საკითხებთან იმის თაობაზე, არის თუ არა **C2 ⊂ C3** და **C3 ⊂ B**?

ქვეთავი 12.3

1. **C**-ენების რომელ ქვეკლასშია **L_{no}**-ენა და რატომ?
2. დაწერეთ **PS**- და **LA**- გრამატიკები **L³_{no}**-ენისათვის, რომელიც განსაზღვრულია როგორც **W#W#W'**, სადაც **W'** და **W''** არიან **W**-ს ხმაურიანი ვერსიები.
3. რომელ **PS**-გრამატიკულ კლასში არიან **L³_{no}** და **L³_{no}** ენები?
4. რომელ **LA**-გრამატიკულ კლასში არიან **L³_{no}** და **L³_{no}** ენები?
5. **PS** გრამატიკული კლასიფიკაციის გათვალისწინებით შეადარეთ **L³_{no}** და **L³_{no}** ენების შინაგანი სირთულე?

ქვეთავი 12.4

1. რა არის კონტექსტისგან თავისუფალი **PS**-გრამატიკის კონსერვატორული გაფართოება? მოიყვანეთ ორი მაგალითი და შეადარეთ მათი ენობრივი კლასები.
2. **a^kb^kc^kd^ke^kf^k** არის თუ არა **TAL** ტიპის ენა? რა არის ამ ენის **LA**-კლასი?
3. რა არის სტრუქტურული მიზეზი იმისა, რომ **a^{n!}** არ არის ინდექსირებული ენა? ჰოპკოფფისა & ულმანის (1979) მიხედვით რა არის ამ ენის **LA**-კლასი?
4. აღწერეთ კონტექსტისგან თავისუფალი ენათა კლასისა და **C**-ენათა კლასის ურთიერთკავშირი.
5. რატომ მოიცავს **C-LAG**-ენათა კლასის ქვეკლასებს **LA**-გრამატიკა უფრო ბუნებრივად ვიდრე **PS**-გრამატიკა.
6. ახსენით **PS**- და **LA**- იერარქიების ორთოგონალური მიმართება რატომ იჩენს თავს დეტერმინირებულ და არადეტერმინირებულ კონტექსტისგან თავისუფალ ენებს შორის?

ქვეთავი 12.5

1. რატომ არ ესადაგება ბუნებრივ ენებს კონტექსტისგან თავისუფალ ენათა კლასი?
2. რატომ ვვარაუდობთ, რომ ბუნებრივი ენები არიან **C**-ენების ქვესიმრავლე?

3. თუ ბუნებრივი ენები არიან **C**-ენათა კლასში, მაშინ რა არის ბუნებრივი ენის შესაძლო მაღალი (ე.ი. არაწრიფივი) სირთულის ხარისხი?
4. რა ტიპის ორაზროვნებებია ბუნებრივ ენებში?
5. რა ტიპის ორაზროვნება არის არაძალისმიერი ბუნებრივი ენის სირთულისთვის და რატომ?
6. ახსენით სემანტიკური გაორმაგების მეთოდი და მისი შედეგები ბუნებრივი ენების სირთულის ხარისხის განსაზღვრაში?
7. არსებობს თუ არა +რეკურსული ორაზროვნება ბუნებრივ ენებში?
8. ახსენით რატომ არის ბუნებრივი ენის ანალიზის სირთულე დამოკიდებული გამოყენებულ გრამატიკულ თეორიაზე. ისარგებლეთ 9.2.1 და 9.2.2 მაგალითები.
9. ახსენით რატომ არის ბუნებრივი ენის ანალიზის სირთულე დამოკიდებული საანალიზოდ გამოყენებულ ენობრივ თეორიაზე. ისარგებლეთ 12.5.4-ის მეორე მაგალითით.
10. ახსენით **CoNSyx** პიპოტეზა. როგორ შეიძლება მისი უარყოფა ემპირიულად.
11. რა პრაქტიკული ლირებულება აქვს **CoNSyx** პიპოტეზას?