

პირითადი და დამხმარე სახელმძღვანელოები  
თანამედროვე მათემატიკურ ენათმეცნიერებაში  
2007 №2

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

გამოცემის რედაქტორი  
კ. ფხაკაძე

**რ. ჰაუსერი**

# გამოთვლითი ენათმეცნიერების საფუძვლები

ნაწილი II

ბრამატიკული  
თეორიები

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი  
ილია ვეკუას სახელობის გამოყენებითი მათემატიკის ინსტიტუტი

# გამოთვლითი ენათმეცნიერების საფუძვლები

მომავალზე ორიენტირებული გამოთვლითი ენათმეცნიერების ძირითადი ამოცანაა ისეთი კოგნიტური მანქანების კონსტრუირება, რომლებთანაც ადამიანებს შეეძლებათ თავისუფლად ისაუბრონ თავიანთივე ბუნებრივი ენის ფარგლებში. უკვე დიდი 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 მოსაუბრე რობოტების კონსტრუირებისთვის. ამასთან, წინამდებარე კურსის ძირითადი თემაა მოლაპარაკება და მსმენელს შორის არსებული ბუნებრივ-ენობრივი კომუნიკაცია.

როლანდ ჰუსერი, 1999 წელი, იენისი.

## ნაწილი II გრამატიკული თეორიები

მთარგმნელ-რედაქტორი –

– კონსტანტინე შხაპაძე

მთარგმნელები –

– ლაშა აბზიანიძე

– ნიკო ვახანია

– სანდრო მასხარაშვილი

– ნიკო შხაპაძე

– ბესო ჩიქვინიძე

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, ახსნილია ბუნებრივი კომუნიკაციის მექანიზმი, რომელიც დაფუძნებულია ენის SLIM თეორიის [2+1] დონის სტრუქტურაზე და განსხვავებული ტიპის ენობრივ ნიშნებზე. კურსის II ნაწილში ყურადღება 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 ღირებულების მქონე ფორმალური ცნებებისა და მეთოდების რაც შეიძლება მარტივი გადმოცემა და მათი სუსტი და ძლიერი მხარეების მიმოხილვა.

ქართული ტექსტის კორექტორ-რედაქტორები –

– ფილ. მმცნ. დოქტორი კახა ბაბუნიანი

– ფილ. მმცნ. დოქტორი ბადრი ცხადაძე

სამეცნიერო-საგანმანათლებლო ჟურნალი  
– ქართული ენა და ლოგიკა –  
[ggl\\_pkhakadze@caucasus.net](mailto:ggl_pkhakadze@caucasus.net)  
უნივერსიტეტის ქ. № 2  
ISSN 1512 – 2840

2006 — 2007  
№3, №4, №5, №6

რედაქტორი  
– კონსტანტინე ფსაკაძე  
რედაქტორის  
მოადგილე  
– კახა გაბუნია  
სარედაქციო  
საბჭო:  
ბადრი ცხადაძე  
გიორგი ჩიჩუა  
ლაშა აბჯიანიძე  
ნიკო ვახანია  
სანდრო მასხარაშვილი  
ნიკო ფსაკაძე  
ნინო ლაბაძე  
ბესო ჩიქვინიძე

SCIENTIFIC-EDUCATIONAL JOURNAL  
- THE GEORGIAN LANGUAGE AND LOGIC -  
[ggl\\_pkhakadze@caucasus.net](mailto:ggl_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

გამოცემა ქართულ-შვეიცარული საწარმოს  
„ვისოლის“ ძველმოქმედებითი მხარდაჭერით

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

სარჩევი

7 თავი – წარმომქმნელი გრამატიკა	4
7.1 ენა როგორც თავისუფალი მონოდიის ქვესიმრავლე	4
7.2. წარმომქმნელი გრამატიკის უპირატესობათა განმსაზღვრელი მეთოდოლოგიური მიზეზები	9
7.3 წარმომქმნელი გრამატიკის ადეკვატურობა	11
7.4 C-გრამატიკის ფორმალიზმი	13
7.5 C-გრამატიკა ბუნებრივი ენისათვის	17
სავარჯიშოები	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. წარმოქმნითი უნარი და ფორმალურ ენათა კლასები	29
8.4. PS-გრამატიკა ბუნებრივი ენისათვის	35
8.5. შემადგენელი სტრუქტურის პარადოქსი	40
სავარჯიშოები	45
9 თავი – ანალიზის ძირითადი ცნებები	48
9.1 ანალიზის დეკლარაციული და პროცედურული ასპექტები	48
9.2 გრამატიკის მორგება ენაზე	51
9.3 ტიპობრივი ტრანსფარენტულობის დამოკიდებულება გრამატიკებსა და პარსერებს შორის	56
9.4 შემავალ-გამომავალი ეკვივალენცია მსმენელ-მოლაპარაკე სისტემასთან	63
9.5 გრამატიკის უკმარისობა კონვერგენციის მისაღწევად	66
სავარჯიშოები	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	85
სავარჯიშოები	92
11 თავი – LA გრამატიკის იერარქია	94
11.1 შეუზღუდავი LA -ების წარმოქმნითი უნარი	94
11.2 A-, B- და C-LAG-ების LA-იერარქია	94
11.3 ორპარამეტრული LA-გრამატიკაში	97
11.4 გრამატიკებისა და ავტომატების სირთულე	100
11.5 C1-, C2-, და C3-LAG-ების ქვეიერარქია	107
სავარჯიშოები	114
12 თავი – LA- და PS- იერარქიების ურთიერთმიმართება	116
12.1 LA- და PS- გრამატიკების ენობრივი კლასები	116
12.2 ქვესიმრავლეობის 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

### წარმომქმნელი გრამატიკა

7.1 ქვეთავში ენა ფორმალური მიდგომებით ისაზღვრება როგორც 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 იმდენად 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 ენა 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 ენის განსაზღვრება

ენა არის სიტყვითი მიმდევრობების სიმრავლე.

თავად ეს სიმრავლე (სიმრავლეთა თეორიის გაგებით) დაულაგებელია. მიუხედავად 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<sub>1</sub> (meaning<sub>1</sub>)**.

მაგალითად,  $a$  და  $b$  სიტყვებით წარმოქმნილი  $LX=\{a,b\}$  სიტყვათა სიმრავლე არის ქვემოთ განსაზღვრული ფორმალური (ე.ი. აბსტრაქტული) ენის ლექსიკონი, ანუ ალფაბეტი<sup>1</sup>.  $LX$  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

$\varepsilon$   
 $a, b,$   
 $aa, ab, ba, bb$   
 $aaa, aab, aba, abb, baa, bab, bba, bbb$   
 $aaaa, aaab, aaba, aabb, abaa, abba, abbb...$   
 $...$

$LX$  სიმრავლეზე განსაზღვრული თავისუფალი მონოიდი  $LX$  სიმრავლის **კლინის ჩაკეტვად (Kleene closure)** იწოდება და მას  $LX^*$  გამოსახულებით აღნიშნავენ.  $LX^*$  თავისუფალი მონოიდი  $\varepsilon$  **ნეიტრალური წევრის (neutral element)** გარეშე  $LX^+$  გამოსახულებით აღინიშნება და იგი  $LX$  სიმრავლის **დადებით ჩაკეტვად (positive closure)** იწოდება.<sup>2</sup> თავისუფალი მონოიდის  $\varepsilon$  ნეიტრალურ წევრს ზოგჯერ **ცარიელ (empty)**, ზოგჯერ კი **ნულოვან (zero)** მიმდევრობასაც (**sequence**) უწოდებენ.

თავისუფალი მონოიდის ცნება ერთნაირად მოხერხებულია როგორც ბუნებრივი, ისე ხელოვნური ენებისათვის. თუმცა, ბუნებრივ ენებს შედარებით დიდი ლექსიკონები აქვთ: ცხადია, რომ თუკი ენის ყველა განსხვავებული 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 არაღირებულია 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.

<sup>1</sup> ფორმალურ ენათა თეორიაში ხელოვნური ენის ლექსიკონს ზოგჯერ ალფაბეტს უწოდებენ, სიტყვას – ასოს, წინადადებას – სიტყვას. ლინგვისტური თვალსაზრისით ეს პრაქტიკა შეცდომაში შემყვანია და იგი არც არანაირი აუცილებლობით არ არის განპირობებული. ამგვარად, ჩვენი მიდგომებით, როგორც ბუნებრივი, ისე ხელოვნური ენის ძირითადი გამოსახულება სიტყვად იწოდება (მაშინაც კი, როცა ეს სიტყვა შედგება მხოლოდ ერთი 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  $aaabbb$ )

<sup>2</sup> სხვა სიტყვებით:  $LX$  სიმრავლეზე განსაზღვრული თავისუფალი მონოიდი იგივეა რაც  $LX^+ \cup \{\varepsilon\}$  (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 ვსაზღვრავთ  $a^k b^k$  ( $k \geq 1$ ) 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.

$a^k b^k$  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  $a^k b^k$  ენის ზემოაღწერილ სტრუქტურასთან. მაგალითად, ისეთი გამოსახულებები, როგორებიცაა ვთქვათ  $a, b, ba, bbaa, abab$ , და ა.შ.. არ არიან  $a^k b^k$  ენის მართებულად აგებული გამოსახულებები. ანალოგიურად  $a^k b^k$  ხელოვნური ენის ამ აღწერისა შეგვიძლია აღვწეროთ ისეთი ხელოვნური ენებიც როგორებიცაა  $a^k b^k c^k, a^k b^m c^k b^m$ , და ა.შ..

არამხოლოდ სასრულ ლექსიკონზე განსაზღვრული თავისუფალი მონოიდი შეიცავს უსასრულოდ ბევრ გამოსახულებას: ასევე შესაძლებელია არსებობდეს ენები – როგორც თავისუფალი მონოიდის ქვესიმრავლებები – რომლებიც აგრეთვე შეიცავენ უსასრულოდ<sup>3</sup> ბევრ მართებულად აგებულ გამოსახულებასა და წინადადებას. ამით არის განპირობებული ის, რომ თავისუფალი მონოიდებისაგან ენის ფილტრაციის კრიტერიუმებს უფრო რთული აღნაგობა აქვთ, ვიდრე უბრალო სიებს.<sup>4</sup>

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

$a^k b^k$  ხელოვნური ენის წარმოქმნელი გრამატიკის შემდეგი მაგალითი იყენებს ფორმალიზმს, რომელიც დღეს **PS**-გრამატიკის<sup>5</sup> სახელითაა ცნობილია

### 7.1.3 PS-გრამატიკა $a^k b^k$ ენისათვის

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

<sup>3</sup> ის, რომ უსასრულო სიმრავლეების ქვესიმრავლე თავადვე შეიძლება იყოს უსასრულო მტკიცდება ლუწი 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.

<sup>4</sup> 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.

<sup>5</sup> **PS**-გრამატიკის დეტალური აღწერა მოცემულია მე-8 თავში.

**PS**-გრამატიკის წარმოქმნითი უნარი ემყარება წარმოქმნილი 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** → **aSb** წესი. ისრის მარცხენა მხარეს მდებარე **S** გამოსახულების ისრის მარჯვენა მხარეს არსებული გამოსახულებით ჩანაცვლება იძლევა

**aSb** სტრიქონს.

შემდეგ ისევ პირველი წესი გამოიყენოთ: 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 ვღებულობთ

**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 გამოსავალზე (**output**) (ანუ, ამავე წესით გამოყვანილ (ე.ი. უკვე წარმოქმნილ) გამოსახულებაზე). მეორე მხრივ, 7.1.3 გრამატიკის მეორე წესი ნიმუშია **PS**-გრამატიკის არარეკურსული (**non-recursion**) წესისა.

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

### 7.1.4 წარმოქმნილი გრამატიკის ელემენტარული ფორმალიზმები

1. კატეგორიათა, ანუ **C**-გრამატიკა
2. ფრაზათა სტრუქტურული, ანუ **PS**-გრამატიკა
3. 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**).

<sup>6</sup> სამაგალითოდ, შეადარეთ ერთმანეთს  $a^k b^k$  ენის **PS**- და **C**-გრამატიკული ანალიზები, რომლებიც 7.1.3 და 7.4.4 განაწერებშია გაკეთებული.



## 7.1.5 ალგებრული განსაზღვრება

წარმომქმნელი გრამატიკის ალგებრული განსაზღვრება ცხადად წარმოგვიდგენს სისტემის ძირითად შემადგენლებს იმით, რომ იძლევა ამ შემადგენლებისა და მათი სტრუქტურული ურთიერთმიმართებების 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, წარმომქმნელი 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

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

ხაზგასასმელია ისიც, რომ უკვე იყო გამოთქმული მოსაზრებები **PS-** და **C-** გრამატიკების ერთ წარმოებულ ფორმალიზმში 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 შეხვედრით, ივარაუდება, რომ წარმოებული ფორმალიზმები იმყოფებიან უპირატეს მდგომარეობაში და რომ მათი შესაბამისი ელემენტარული ფორმალიზმების უკვე შესწავლილმა თვისებებმა უნდა გაგვიადვილოს მათი ანალიზი. მიუხედავად ამისა, აღმოჩნდა, რომ წარმოებული ფორმალიზმების მათემატიკური ანალიზი მათი შესაბამისი ელემენტარული ფორმალიზმების მათემატიკურ ანალიზზე

არანაკლებ რთულია. ამ სახის კვლევითი პროცესები ხშირად ათწლეულები გრძელდებოდა და, შესაბამისად, არც თუ იშვიათად იყო შეცდომებიც.<sup>7</sup>

ძირეული და წარმოებული ფორმალიზმების გარდა არსებობენ აგრეთვე წარმომქმნელი გრამატიკის ნახევრად ფორმალური სისტემები (**semi-formal systems**). რომელთა აღწერაც არ არის დაფუძნებული ალგებრულ განსაზღვრებებზე. მაგალითად, ასეთებია: **დამოკიდებულებათა გრამატიკა (dependency grammar)** (ტენიერი, 1959) და **სისტემური გრამატიკა (systemic grammar)** (ჰალიდეი, 1985). ასეთი სისტემების მკვლევარები ეყრდნობიან იმ ჰიპოტეტურ მოსაზრებას, რომ შესასწავლი სისტემის თვისებები გარკვეული ხარისხით სიმძლავრულია 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 ძირეული, ანუ ელემენტარული ფორმალიზმებისთვის. საკითხი დაისმის იმის თაობაზე, თუ ამ ძირეული ფორმალიზმებიდან რომელია უფრო ვარგისი ბუნებრივი ენის (i) ლინგვისტური ანალიზისა და (ii) ავტომატური პარსერის კონსტრუირების მიზანთა კრილში. ამავე მიზეზით იხსნება კონცეპტუალური და ფორმალური მიზეზები იმისა, თუ რატომ იგება ასე მრავლად ძირეული ფორმალიზმების ახალ-ახალი წარმოებული ფორმალიზმები.

## 7.2 წარმომქმნელი გრამატიკის უპირატესობათა განმსაზღვრელი მეთოდოლოგიური მიზეზები

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

### 7.2.1 გრამატიკულად მართებულად აგებული 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.

<sup>7</sup> მაგალითად: ნ. 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 შემოტანას **GPSG** სისტემაში არ უნდა გაეზარდა ამ სისტემის თავდაპირველი კონტექსტისაგან თავისუფალი სირთულე, რაც უარყოფილი იქნა 1986 წელს ჰ. უშკორეიტისა და ს. პეტერსის კვლევებით.

## 7.2.2 ბრამატიკულად არამართებულად აგებული გამოსახულება

**\*earlier slept have dogs little the<sup>8</sup>**

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, გასაკები ხდება, რომ წარმომქმნელი გრამატიკების გამოყენება თანამედროვე ლინგვისტურ მეთოდოლოგიებში სრულიად აუცილებელია.

## 7.2.3 წარმომქმნელი გრამატიკის მეთოდოლოგიური მიდგომები

**ემპირიკული:** იძლევა ცხადი ჰიპოთეზების ჩამოყალიბების საშუალებას:

ნებისმიერი წარმომქმნელი გრამატიკული ანალიზი ჯამდება ისეთი ფორმალური წესების სისტემით, რომელიც შეიცავს ცხად ჰიპოთეზებს, იმის შესახებ თუ რომელი შესავალი **(input)** გამოსახულებაა მართებულად აგებული და რომელი არა. ეს ცხადი სახის ჰიპოთეზები 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 თვისებების<sup>9</sup> ანალიზის საშუალებას, როგორიცაა ამონხნადობა, სირთულე და წარმომქმნითი უნარი (ე. ი. შესაძლებლობა). გრამატიკული ფორმალიზმის მათემატიკური თვისებები კი, თავის მხრივ, საშუალებას იძლევიან განვსაზღვროთ რამდენად გამოყენებადია იგი კონკრეტული პრაქტიკული აღწერებისა და გამოთვლითი რეალიზაციებისათვის.

**გამოთვლითი:** პარსერის აღწერილობითი განსაზღვრება:

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

<sup>8</sup> ლინგვისტიკაში არაგრამატიკული სტრუქტურის მქონე მაგალითები მონიშნულია ვარსკვლავით \*. ეს შეთანხმება დაამკვიდრა ლ. ბლუმფილდმა 1933 წელს.

<sup>9</sup> ცხადია, რომ არაფორმალური აღწერის მათემატიკური თვისებები შეუძლებელია გამოკვლეული იქნას მათი სტრუქტურების არასაკმარისი სიღრმის გამო

<sup>10</sup> ჯერ კიდევ არსებობენ პროგრამები, რომლებიც არ არიან დაფუძნებულნი 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 თანდათანობით გაუმჯობესების 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 კონკრეტული გრამატიკის შემოწმებისათვის.

რამდენადაც კარგი მეთოდოლოგია აუცილებელი წინაპირობაა იმისა, რომ მივიღოთ სოლიდური აღწერილობითი შედეგი, ბუნებრივი ენების ანალიზისას წარმომქმნელი გრამატიკების სისტემური გამოყენება როგორც წესი იძლევა დადებით 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 ფორმალიზების 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 გამოსადეგი იქნება განხილვის ქვეშ მყოფი მიზნებისათვის, არ გვარწმუნებს.

ესა თუ ის ფორმალიზმი ამ თუ იმ კონკრეტულ ენასთან მიმართებაში ფასდება ემპირიკულად ვარგისად, თუ მას აქვს საკმარისი გამომხატველობითი ძალა იმისათვის, რომ აღწეროს ამ ენის ყველა შესაძლო სტრუქტურა. მათემატიკური თვალსაზრისით ფორმალიზმი რაც შეიძლება ნაკლებ რთული 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 შემეცნებითი 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 წარმომქმნელი გრამატიკის ადეკვატურობა

წარმომქმნელ გრამატიკას უწოდებენ მოცემული ენის სრულად აღმწერს, ანუ დესკრიფციულად ადეკვატურს (**descriptively adequate**), თუ ის წარმოქმნის ამ ენის გრამატიკულად მართებულად აგებულ ყველა გამოსახულებას და მხოლოდ მათ. არაადეკვატური 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 (ზემწარმოებლობა (**overgeneration**)). წარმოქმნილი გრამატიკა არასრულია, თუ არსებობს ერთი მაინც ისეთი გრამატიკულად მართებულად აგებული გამოსახულება, რომელსაც იგი ვერ წარმოქმნის (ნაკლებწარმოებლობა (**undergeneration**)).

დაუშვათ უკვე გვაქვს ისეთი ბუნებრივი ენის ფორმალური გრამატიკა, რომელზეც ჩვენ არ ვლაპარაკობთ (მაგ. კერუას ენის<sup>11</sup>) და გვინდა გავარკვიოთ ამ ენის რომელიმე მოცემულ გამოსახულებაზე არის თუ არა იგი მართებულად აგებული. ამისთვის გასაანალიზებელი გამოსახულების სიტყვა ფორმებს ვუსაბამებთ მათ შესაბამის კატეგორიებს კერუას ენის ამ ფორმალური გრამატიკის ლექსიკონის მიხედვით. შემდეგ ვცდილობთ გამოვიყვანოთ (ე.ი. წარმოვქმნათ) გასაანალიზებელი გამოსახულება ამავე ფორმალური გრამატიკის წესების გამოყენებით. გამოსახულება, რომელსაც ვაანალიზებთ, მხოლოდ იმ შემთხვევაში იქნება მართებულად აგებული, თუ ასეთი მიდგომებით მისი გამოყვანის არსებობა დადასტურდება.

ამგვარად, ბუნებრივი ენის გამოსახულების ფორმალური გამოყვანის არსებობა გამოსახულების გრამატიკულად მართებულად აგებულობის დამადასტურებელია, თუ ცნობილია, რომ ეს წარმოქმნილი გრამატიკა არის ამ ენის სრულად აღმწერი, ანუ დესკრიფციულად ადეკვატური. ამდენად, როცა არსებობს რაიმენაირი ეჭვი იმ გრამატიკის დესკრიფციულ ადეკვატურობაზე, 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 წარმოქმნება 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 მიიჩნევა, რომ გამოყენებადი ფორმალური გრამატიკა მაღალი მათემატიკური სირთულისაა. წარმოქმნილი 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 წარმოქმნილი 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 ნაკლებსაგან, ბუნებრივი ენის სინტაქსური ანალიზი უნდა იყოს:

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

<sup>11</sup> 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.

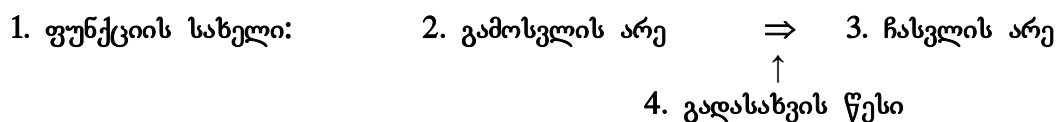
ცხადია, რომ ეს მოთხოვნები უნდა შესრულდეს ერთდროულად: რა სარგებლობა აქვს კარგად ჩამოყალიბებულ სინტაქსს, თუ მისი სირთულე ან ამოუხსნადი, ან ექსპონენციალურია? რა მიზანი აქვს მათემატიკურად სწორად ფორმირებულსა და ეფექტურ ფორმალიზმს, თუ ის აღმოჩნდება სტრუქტურულად არათავსებადი ბუნებრივი კომუნიკაციის მექანიზმებისათვის? 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 (**category grammar**). ის შეიმუშავეს პოლონელმა ლოგიკოსებმა ლეშნისკიმ (1929) და ევლუკევიჩმა (1935) იმ მიზნით, რომ ფორმალური ენების ანალიზისას თავი დაედწით რასელის პარადოქსისაგან. პირველად C-გრამატიკა ბუნებრივი ენის დამუშავების მიზნით ი. ბარ-ჰილელმა<sup>12</sup> 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 უზრუნველყოფს იმას, რომ ნებისმიერი შესაძლო არგუმენტი, მაგალითად: 1, 2, 3, 4, 5, ... შესაბამისად აისახება ზუსტად ერთ და კერძოდ 1, 4, 9, 16, 25, ... მნიშვნელობებზე<sup>13</sup>.

<sup>12</sup> ამ საკითხების კარგი არაფორმალური შეჯამება 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.

<sup>13</sup> კვადრატული ფესვი არაა ფუნქცია. რადგან იგი განსაზღვრის არის ზოგიერთ არგუმენტს მნიშვნელობათა არეში ორ განსხვავებულ მნიშვნელობას უთანადებს. მაგალითისთვის 4-დან კვადრატულ ფესვს აქვს ორი მნიშვნელობა, ესენია: 2 და -2. ამის გამო, 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.<sup>14</sup>

## 7.4.2 C-გრამატიკის ალგებრული განსაზღვრება

**C**-გრამატიკა არის ხუთეული  $\langle W, C, LX, R, CE \rangle$ . სადაც

- W** არის სიტყვითი გამოსახულებების სასრული სიმრავლე.
- C** არის ისეთი კატეგორიების სიმრავლე, როგორებიცაა
  - ბაზისი:  $u$  და  $v \in C$ <sup>15</sup>;
  - ინდექსია: თუ  $X$  და  $Y \in C$ , მაშინ  $(X/Y)$  და  $(X|Y) \in C$ ;
  - ჩაკეტვა: არაფერი სხვა გარდა (ა) და (ბ) წესებით განსაზღვრულებისა არ არის **C** სიმრავლის წევრი.
- LX** სასრული სიმრავლეა. ამასთან,  $LX \subset (W \times C)$ .
- R** არის სიმრავლე, რომელიც შეიცავს შემდეგ ორ სქემატურ წესს:
 
$$\alpha(x/y) \circ \beta(y) \Rightarrow \alpha\beta(x)$$

$$\beta(y) \circ \alpha(yx) \Rightarrow \beta\alpha(x)$$
- 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  $\langle W, C, LX, R, CE \rangle$  ხუთეულით და, ამასთან, 1-5 წინადადებებით მოცემულია ამ კომპონენტების სიმრავლური განსაზღვრებები.

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

**C** სიმრავლე განსაზღვრულია რეკურსულად. შესაბამისად, საბაზისო **u** და **v** წევრებთან ერთად ინდექციური პირობის ძალით **C** სიმრავლეში არიან აგრეთვე  $(u|v)$ ,  $(v|u)$ ,  $(u/v)$  და  $(v/u)$  კატეგორიებიც. თავის მხრივ, ისევ და ისევ ინდექციური 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 რეკურსიის 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, ინდექსად მიეწერება წყვილის პირველ კომპონენტს, როგორც ეს არის მაგალითად  $a((u/v)|v)$  გამოსახულებაში. ამ გამოსახულების ძირითადი ნაწილი (ე.ი. **W** სიმრავლის

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

<sup>15</sup> ელემენტარული, ანუ ძირეული კატეგორიების სახელები და რაოდენობა (აქ **u** და **v**) აუცილებლობით და ცალსახობით არ ისაზღვრება. მაგალითისათვის ევდუკვიჩი იყენებდა მხოლოდ ერთ ელემენტარულ კატეგორიას, გიჩი და მონტევიუ ორს, სხვები სამს.

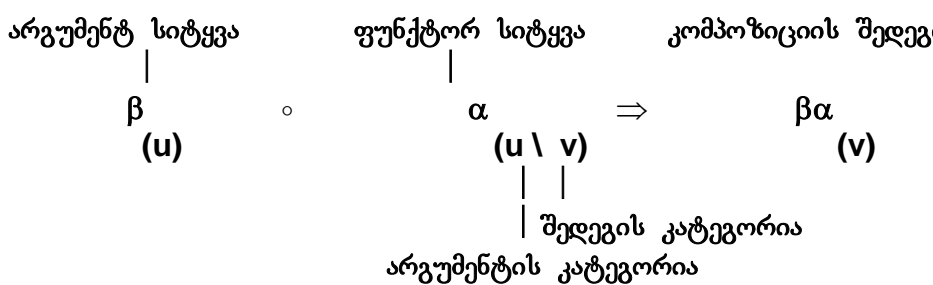
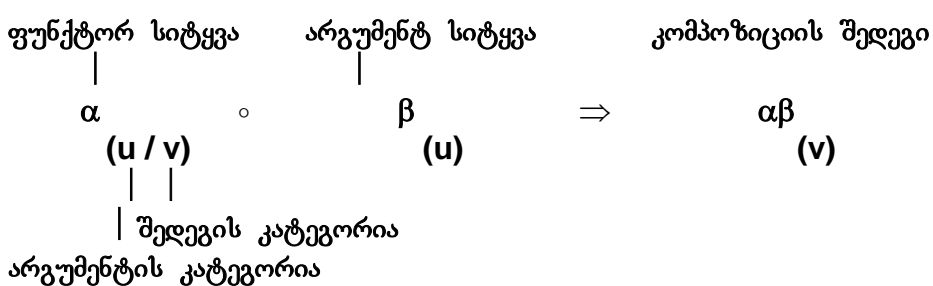
წვერი) ინდექსად იღებს **C** სიმრავლის რომელიმე წვერს, ანუ კატეგორიას, რაც იმას ნიშნავს, რომ გამოსახულების ძირითადი წვერი ამ კატეგორიის გამოსახულებაა, რაც მთლიანობაში, ცხადად მოიცემა **LX** სიმრავლის შემადგენელი წევრებისა და აქ გაკეთებული შეთანხმებების საფუძველზე.

**R** სიმრავლე შეიცავს ორ სქემატურ წესს. ამ წესებში  $\alpha$  და  $\beta$  ცვლადები შესაბამისად გამოიყენება ფუნქტორისა და არგუმენტის ჩანაწერობრივი სახით წარმოსადგენად. შესაბამისად **X** და **Y** ცვლადები წარმოგვიდგენენ მათ კატეგორიალურ მახასიათებლებს. პირველი სქემა აკომბინირებს ფუნქტორსა და არგუმენტს  $\alpha\beta$  რიგით. ამასთან, **Y**, როგორც  $\beta$  არგუმენტის კატეგორიის მანიშნებელი, აუქმებს  $\alpha$  ფუნქტორის კატეგორიალურ მარჯვენებელში **Y** კატეგორიას. მეორე სქემა აკომბინირებს ფუნქტორსა და არგუმენტს პირიქითი, ანუ  $\beta\alpha$  რიგით. ამ დალაგების ფორმალური გასაღებია  $\alpha$  ფუნქტორის კატეგორიალურ მარჯვენებელში უკუხაზად (**backslash**) წოდებული  $\backslash$  სიმბოლო (ნაცვლად წინა წესში გამოყენებადი ხაზად (**slash**) წოდებული  $/$  სიმბოლოსი). ამ ტიპის **C**-გრამატიკებს ბარ-ჭილელმა 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 იძლევიან კატეგორიათა დაწყვილების ნიმუშებს (იხ. თავი 17 და 18).

### 7.4.3 დაწყვილების ცხადი ნიმუშები C-გრამატიკის კომბინაციაში





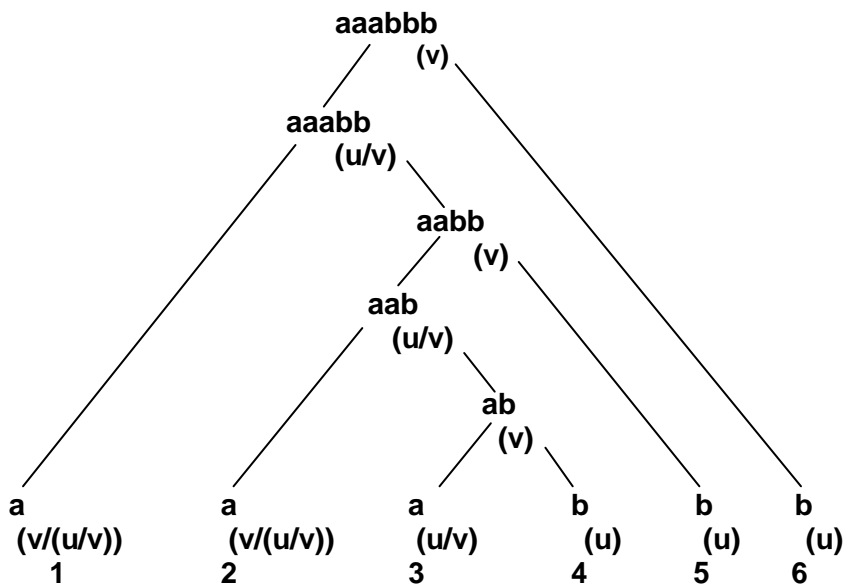
ცხადია, რომ ეს სქემატური წესები და ამ წესებით განსაზღვრული ეს კომბინაციები ნაგულისხმევი უნდა იყოს **C**-გრამატიკის კატეგორიალურ სტრუქტურებში. აქედან გამომდინარე, ენა ითვლება საკმარისად განსაზღვრულად, თუ მოცემულია ამ ენის (i) **LX** ლექსიკონი და (ii) დასრულებული გამოსახულებების **CE** სიმრავლე. სამაგალითოდ 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

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

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

#### 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 შუალედური 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, ითხოვს ლექსიკური ორაზროვნების მაღალ ხარისხს. ამის შედეგია ის, რომ ხშირად ექსპერტებსაც კი უჭირთ **C**-გრამატიკის მოცემული სტრიქონისთვის 7.4.5 გამოყვანის მსგავსი გამოყვანის აგება, ანდა იმის დამტკიცება, რომ ასეთი გამოყვანა არ არსებობს. – განსაკუთრებით ეს ჭირს ისეთ არამარტივ ენებში, როგორც  $a^k b^k$  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  $a^k b^k$  ენისათვის ისეთი **C**-გრამატიკის შექმნა, რომელსაც შეეძლება ნებისმიერი წინადადების წარმოქმნა დროში წრფივი მეთოდით.

## 7.5 C-გრამატიკა ბუნებრივი ენისათვის

**C**-გრამატიკა არის პროტოტიპი ლექსიკური მიდგომისა: ენის ყველა კომბინაციური თვისება კოდირებულია მისი ძირითადი გამოსახულებების კატეგორიებში. ეს დასტურდება როგორც 7.4.4 პუნქტში გაკეთებული  $a^k b^k$  ხელოვნური ენის **C**-გრამატიკული განსაზღვრებით, ასევე ინგლისური ენის მცირე ფრაგმენტის<sup>16</sup> ქვემოთ მოყვანილი განსაზღვრებითაც.

### 7.5.1 C-გრამატიკა ინგლისური ენის მცირე ფრაგმენტისათვის

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

სადაც

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

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

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

7.4.4 გრამატიკასთან ამ გრამატიკის შედარება გვიჩვენებს ორ ნოტაციურ განსხვავებას: პირველი, ლექსიკური მონაცემები თავმოყრილია სიტყვათა კატეგორიზებულ (ანუ,  $W_{\text{cat}}$ ) კლასებში. ეს მიდგომა, რომელიც მონტეგიუდან მოდის, აადვილებს ლექსიკონის აგებას, რადგან კატეგორიები მიეწერება მთლიან კლასებს ნაცვლად ცალკეული სიტყვა ფორმებისა. მეორე, 7.4.2 პუნქტში  $u, v$  სიმბოლოებით მონიშნული ძირეული კატეგორიები  $e, t$  სიმბოლოებად გადასახელდნენ. მონტეგიუს თანახმად  $e$  ანზოგადებს საგნობრივ,  $t$  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

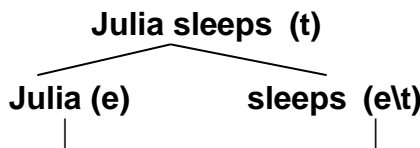
<sup>16</sup> რ. მონტეგიუ იყენებდა ფრაგმენტის ცნებას იმ მიზნით, რომ გამოყო ბუნებრივი ენის ის ქვესიმრავლე, რომელიც უკვე შემუშავებული ფორმალური გრამატიკით უნდა გაანალიზებულიყო.

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 არეც. 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 სიის ჩამოწერით) არის ერთ-ერთი ძირითადი პრობლემა მოდელურ-თეორიული სემანტიკებისა.<sup>17</sup>

**Julia sleeps** (ჯულიას სძინავს) წინადადების **C**-გრამატიკული ანალიზის შინაარსული ინტერპრეტაცია ნებისმიერად ალებულ **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 პრტიდროული სინტაქსური და სემანტიკური ანალიზი



მნიშვნელობები (**M** მოდელში): არსი (საგანი) {საგანთა სიმრავლე}

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, მონტევიუმ<sup>18</sup> გადაიხზრა საგანთა სიმრავლის ჭეშმარიტულ მნიშვნელობათა სიმრავლეში ამსახველ მახასიათებელ ფუნქციად, ანუ (**e/t**) კატეგორიად. აქ მან გადახრილი ხაზი გამოიყენა არსებითი სახელებისა და ინტრანზიტული ზმნების სინტაქსური განსხვავებულობის ხაზგასასმელად.

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,

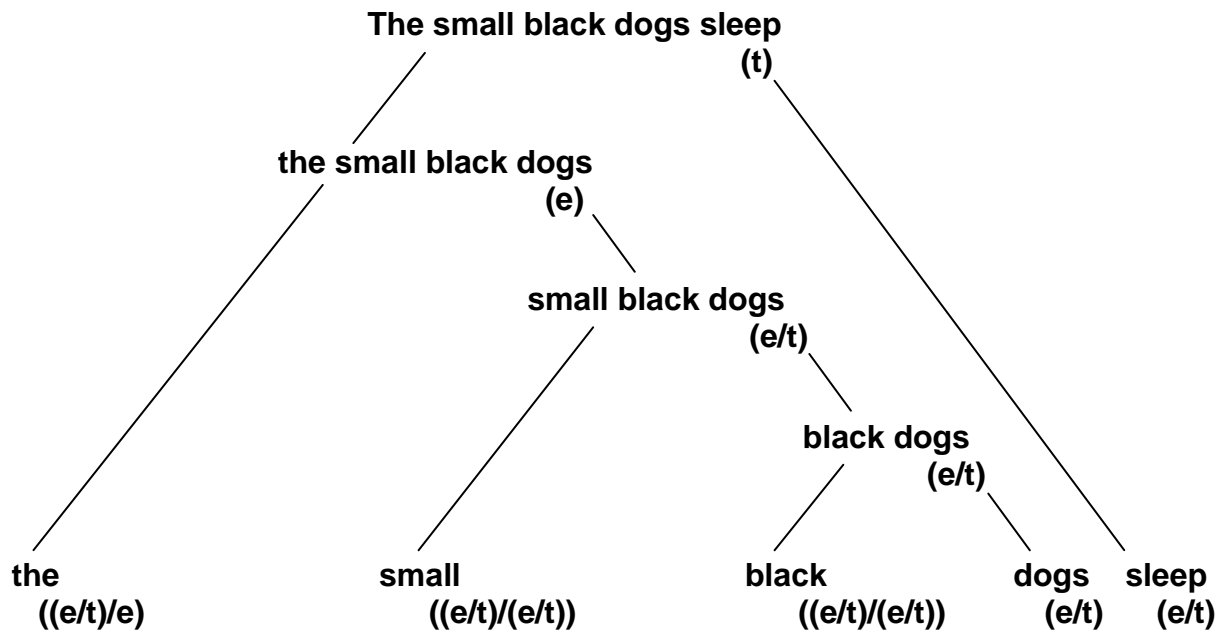
<sup>17</sup> იხილეთ 19.4 ქვეთავი და **Col.**, გვ 292-295.

<sup>18</sup> სიმარტივისა და შეთანხმებულობისათვის: მონტევიუმს აღნიშვნებისგან განსხვავებით, ჩვენ 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 რეკურსიულ 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 წინადადებების 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

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

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

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

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

$$W_{((e/t)/t)} = \{a, the, every \dots\}$$

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

ამ **C**-გრამატიკის ლინგვისტური მოტივაციაა ფუნქტორ-არგუმენტის სტრუქტურების სემანტიკური მიმართებებისა და სინტაქსური მკომბინირებელი 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.<sup>19</sup>

პრაქტიკული გადაწყვეტებისას **C**-გრამატიკის სტრუქტურები მიიღება საკმაოდ რთული გამოყვანებით, რომლებსაც იგივე მახასიათებლები აქვთ, რაც თეორემათა მამტიკებელ 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 ჭარბწარმოქმნელობით: 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 ძაღლები)** და **\*all dog (ყველა ძაღლი)**.

თეორიაში ყოველი ასეთი სირთულე დაძლეულ უნდა იქნას ფორმალიზმის შესაბამისი გაფართოებით. თუმცა, პრაქტიკაში, ნებისმიერი გაფართოება ითხოვს საკმარისი მოცულობის აღწერითი სამუშაოების<sup>20</sup> 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 წარმოქმნელ გრამატიკასთან.
3. რა განსხვავებაა დადებით და კლინის ჩაკეტვებს შორის?
4. რა გაგებით შეიძლება იქნეს განხილული წარმოქმნელი გრამატიკა, როგორც ფილტრი?

<sup>19</sup> იხ. დისკუსია მეტაენაზე დაფუძნებული სემანტიკის თაობაზე 19-ე და 20-ე თავებში.

<sup>20</sup> საცდელად იხილეთ **SCG**.

5. ახსენით რეკურსიის როლი 7.13 მიდგომებით **aaaabbbb** გამოსახულების გამოყვანაში.
6. რა არის ალგებრული განსაზღვრება და რა არის მისი მიზანი?
7. რა განსხვავებაა ძირეულ (ე.ი. 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 ადეკვატური?
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. რატომაა **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 გამოსახულებას?
8. რატომ იწოდება 7.4.4 პუნქტში განსაზღვრული გრამატიკული ფორმალიზმი ორმომართულებიან **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 რთული 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. გააფართოვეთ 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 იმარაქი

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

ბუნებრივი ენისათვის პირველი გადამწერი სისტემა, სახელწოდებით ფრაზის სტრუქტურული გრამატიკა<sup>1</sup>, მოკლედ **PS**-გრამატიკა, 1957 წელს ნ. ჩომსკის მიერ იქნა აღწერილი. **PS**-გრამატიკაზე დაყრდნობით ჩომსკიმ და სხვებმა ააგეს მთელი რიგი წარმოებული ფორმალიზმებისა, რომლებსაც თავდაპირველად ტრანსფორმაციულ გრამატიკებს უწოდებდნენ.

8.1 თავში მოყვანილია **PS**-გრამატიკის ალგებრული განსაზღვრება და აღწერილია ის შეზღუდვები, რომელთა **PS**-წესებზე დამატებით მიიღება წესიერი, კონტექსტისგან თავისუფალი, კონტექსტზე დამოკიდებული და შეუზღუდავი **PS**-გრამატიკები. 8.2 თავში ახსნილია სირთულის ხარისხის ოთხი ძირითადი ღონე 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.

<sup>1</sup> 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 აღრიცხვების ზოგად სტრუქტურებში მისი კარგამხსნელი 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. ჰარესსა და ი. ბარ-ჰილელს 1947 წლიდან ახლო მეცნიერული კონტაქტები ჰქონდათ. გარდა ამისა, გასათვალისწინებელია 1951 წლის დასაწყისში ბარ-ჰილელისა და ჩომსკის დისკუსიაც „დაუსრულებელი საუბრები ლინგვისტიკაზე, ლოგიკასა და მეთოდოლოგიაზე“. (ბარ-ჰილელი, 1964, გვ.16).



## 8.1.1 PS-გრამატიკის ალგებრული 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 გადაწერის წესების (rewrite rules) სიმრავლე, სადაც  $\alpha$  არის  $V^+$  სიმრავლის,  $\beta$  კი –  $V^*$  სიმრავლის<sup>2</sup> 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 გარეწედაპირები (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 შემდგენიერად 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  $\gamma$  და  $\delta$  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 მიიღება  $\gamma\alpha\delta$  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-სემანტურ ფონემაზე

0. შეუზღუდავი PS-წესები (unrestricted PS-rules): 0 ტიპის წესის მარჯვენა და მარცხენა მხარე შეიძლება იყოს ტერმინალური და არატერმინალური სიმბოლოების ნებისმიერი მიმდევრობა.

1. კონტექსტზე 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 სიგრძე არ აჭარბებდეს მარჯვენა მხარეს განთავსებული გამოსახულების სიგრძეს. მაგალითი:  $ABC \rightarrow ADEC$

<sup>2</sup>  $V^+$  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.შ.<sup>3</sup>

3. წესიერი PS-წესები (**regular PS-rules**): 3 ტიპის წესის მარცხენა მხარე მხოლოდ ერთი ცვლადისგან უნდა შედგებოდეს, მარჯვენა მხარე კი უნდა შეიცავდეს ერთ, და მხოლოდ ერთ ტერმინალურ სიმბოლოს და არ უნდა შეიცავდეს ერთზე მეტ ცვლადს. მაგალითი<sup>4</sup>:  $A \rightarrow b$ ,  $A \rightarrow bC$

რადგან წესის ტიპის რიგითი ნომრის ზრდასთან ერთად (დაწყებული 0 ტიპიდან და დამთავრებული 3 ტიპით) შემზღვეველ 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 ამავედროულად 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 შემზღვეველ 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, წესიერი (**regular**), შესაბამისად კონტექსტისგან თავისუფალ (**context-free**), შესაბამისად კონტექსტზე დამოკიდებულ (**context sensitive**) PS-გრამატიკად (**PS-grammar**) იწოდება; PS-გრამატიკა, რომლის წესებიც არანაირად არ იზღუდება, შეუზღვეველად 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**); შეუზღვეველი PS-გრამატიკით წარმოიქმნება რეკურსულად გადათვლადი ენები (**recursively enumerable languages**).

<sup>3</sup> 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.90, თეორემა 4.3). ჩვენი მიდგომებით 2 ტიპის წესებისთვის ვითხოვთ მარჯვენა მხარეს განთავსებული მიმდევრობის არაკარიელობას, რათა ფორმალურად შესაძლებელი იყოს კონტექსტისგან თავისუფალი წესების განხილვა კონტექსტზე დამოკიდებული წესების კერძო შემთხვევად.

<sup>4</sup> ეს არის მარჯვნივ წრფივი PS-გრამატიკის (**right linear PS-grammar**) მაგალითი.  $A \rightarrow b$ ,  $A \rightarrow bC$  წესებით განსაზღვრული PS-გრამატიკა არის ზემოაღწერილი მარჯვნივ წრფივი გრამატიკის ინვერსული მარცხნივ წრფივი PS-გრამატიკა (**left linear PS-grammar**). მარცხნივ და მარჯვნივ წრფივი გრამატიკები ურთიერთეკვივალენტურებია (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 იძლევიან მხოლოდ შეზღუდული კონტროლის დამყარების საშუალებას წარმოსაქმნელ სტრუქტურებზე<sup>5</sup>.

## 8.2 ენობრივი კლასები და გამოთვლითი სირთულე

გრამატიკული ტიპით წარმოქმნილი ენების კლასზე და ამ გრამატიკული ტიპის წარმოქმნის უნარზე 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 სირთულებისა (**degrees of computational complexity**).

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

<sup>5</sup> 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 ეკუთვნის  $a^k b^k$  ენას.

წარმოქმნითი გრამატიკული ფორმალიზმის სირთულე იზომება იმ ალგორითმის სირთულით, რომლის საფუძველზეც ეს გრამატიკა, როგორც ოპერაციული პროცედურა, რეალიზდება აბსტრაქტულ ავტომატზე (მაგალითად ისეთზე, როგორებიცაა ტიურინგის მანქანა, წრფივად ბმული ავტომატი, ძირს მბიძგი ავტომატი, სასრულ მდგომარეობიანი ავტომატი და.ა.შ.). ამ ალგორითმის სირთულე კი იმ **პრიმიტიული ოპერაციების<sup>6</sup> (primitive operations)** რიცხვით ისაზღვრება, რომელსაც ითხოვს დასამუშავებლად შემოსული გამოსახულების ანალიზი ყველაზე უარეს შემთხვევაში (ზედა ზღვარი<sup>7</sup> (**upper bound**)). ამასთან, ამ პრიმიტიული ოპერაციების ეს რიცხვი ისაზღვრება დასამუშავებელი გამოსახულების სიგრძესთან ურთიერთკავშირში.

წარმოქმნილი გრამატიკების ელემენტარულ ფორმალიზმებში, ვთქვათ ისეთებში, როგორებიცაა **PS**-გრამატიკა ან **LA**-გრამატიკა, სირთულე ისაზღვრება (ე.ი. ითვლება) მხოლოდ მათი ისეთი მკაცრად განსაზღვრული ქვეტიპებისათვის როგორებიცაა, ვთქვათ, **PS**-გრამატიკის წესიერი, კონტექსტისგან თავისუფალი, კონტექსტზე დამოკიდებული და შეუზღუდავი 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**):  
 $n, 2n, 3n$ , და.შ.
2. პოლინომიალური სირთულე (**Polynomial complexity**):  
 $n^2, n^3, n^4$ , და.შ.
3. ექსპონენციალური სირთულე (**Exponential complexity**):  
 $2^n, 3^n, 4^n$ , და.შ.
4. არაამოხსნადი (ამოუხსნადი) (**Undecidable**):  
 $n \times \infty$

<sup>6</sup> 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 მხრივ, **LA**-გრამატიკის **C-LAG** კლასის ქვეკლასი პრიმიტიულ ოპერაციად იყენებს წესის გამოყენებას (იხ. 11.4).

<sup>7</sup> აქ საუბარია დროით სირთულეზე.

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

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

### 8.2.2 პოლინომიალური და ექსპონენციალური ალგორითმების შედარება

დროითი სირთულე	პრობლემის ზომა n		
	10	50	100
<b>n<sup>3</sup></b>	.001 წამი	.125 წამი	1.0 წამი
<b>2<sup>n</sup></b>	.001 წამი	35.7 წელი	10 <sup>15</sup> 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 შედგება.<sup>9</sup> ამგვარად, გარეი და ჯონსონის გაზომვებით გამოდის, რომ **2<sup>n</sup>** ხარისხის ექსპონენციალური სირთულის გრამატიკული ალგორითმით კორპუსის ავტომატური ანალიზი ყველაზე ცუდ შემთხვევაში **10<sup>15</sup>** საუკუნეზე მეტ ხანს უნდა გაგრძელდეს. ეს დრო ბევრად მეტია

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

<sup>9</sup> ეს მონაცემები გამოაქვეყნა მარკუს შულცმა **CLUE**-ში.

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

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

### 8.2.3 ფორმალური ენების PS-გრამატიკული იერარქია

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	შეუზღუდავი PSG	რეკურ. გადათვლადი ენები	არაამონუნადი

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 ფორმალური 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 გამომდინარე 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.<sup>10</sup> წესიერი გრამატიკის წარმოქმნითი უნარი ერთი წარმოქმნითი ოპერაციის ფარგლებში ემყარება ყოველგვარი თანამდევნი რეკურსული მოვლენების გარეშე მხოლოდ ერთი, ცალკე აღებული სიტყვის რეკურსული გამეორების შესაძლებლობას.

<sup>10</sup> 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 ნაჩვენებია 11.5 ქვეთავში, **C1**-ენების კლასი შეიცავს როგორც წესიერი ენებს, ისე ისეთ დეტერმინისტულ კონტექსტისაგან თავისუფალ ენებს, რომლებიც გამოცნობადია ეფსილონ-თავისუფალი **DPDA**-ს მიერ, და, ასევე, ბევრ კონტექსტზე დამოკიდებულ ენებს.

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

### 8.3.1 $ab^k$ ენის წესიერი **PS**-გრამატიკა ( $k \geq 1$ )

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

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

### 8.3.2 $\{a,b\}^+$ ენის წესიერი **PS**-გრამატიკა

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

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

### 8.3.3 $a^m b^k$ ენის წესიერი **PS**-გრამატიკა ( $k, m \geq 1$ )

$$\begin{aligned} V &=_{\text{def}} \{S, S_1, S_2, a, b\} \\ V_T &=_{\text{def}} \{a, b\} \\ P &=_{\text{def}} \{S \rightarrow aS_1, S_1 \rightarrow aS_1, S_1 \rightarrow bS_2, S_2 \rightarrow b\} \end{aligned}$$

$a^m b^k$  ენა წესიერია, რადგანაც  $a$  და  $b$  სიმბოლოების რიცხვი  $a^m b^k$  გამოსახულებაში ერთმანეთისაგან დამოუკიდებელია. – ამას ერთმანეთისაგან განსხვავებული ზედა  $m$  და  $k$  ინდექსები უზრუნველყოფენ. მეორე მხრივ,  $a^k b^k$  ენა წესიერი **PS**-გრამატიკის წარმოქმნითი უნარის, ანუ შესაძლებლობის მიღმა რჩება, რადგანაც ის მოითხოვს  $a$  და  $b$  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  $a^k b^k$  ენა.

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

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

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

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

### 8.3.4 $a^k b^{3k}$ ენის კონტექსტისგან 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 მარცხენა მხარე შესდგება მხოლოდ ერთი გამხოლოებული ცვლადისაგან. – ე.ი. სხვა „კონტექსტურად“ მოსაზღვრე ნიშნებისა და სიმბოლოების გარეშე<sup>12</sup>.

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**).<sup>13</sup>

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. **dcba**).<sup>14</sup>

### 8.3.5 $WW^R$ ენის კონტექსტისგან თავისუფალი **PS**-გრამატიკა

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

<sup>11</sup> იხ. Hopcroft and Ullman, 1979, გვ.55.

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

<sup>13</sup> ყოველი კონტექსტისგან თავისუფალი ენა არის წესიერი სიმრავლის და ნახევრად დიკ სიმრავლის გადაკვეთის ჰომომორფული (ჩომსკი-შუტცენბერგერის თეორემა). იხ. M. Harrison 1978, გვ. 317.

<sup>14</sup> ზედა  $W^R$  ინდექსი  $WW^R$  გამოსახულებაში სიმბოლურად გვანიშნებს 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 შეზღუდულია<sup>15</sup>. 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 შედგება. 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 დაწვეილებული ინვერსულობის სტრუქტურულ ბუნებაში, რაც ჩვენთვის ცნობილია  $a^k b^k$  და  $WW^R$  ენების უკვე განხილული მაგალითებით.

კიდევ ერთი ენა, რომელიც სცდება კონტექსტისგან თავისუფალი **PS**-გრამატიკების წარმოქმნითი უნარით განსაზღვრულ შესაძლებლობებს არის **WW** ენა. აქ **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 დაწვეილებული ინვერსულობის ამკარად 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, მიუხედავად  $WW^R$  და **WW** ენების ამკარა მსგავსებისა, **WW** ენისთვის 8.3.5 სახის **PS**-გრამატიკა არ არსებობს.

$a^k b^k c^k$  და **WW** ენების მეშვეობით ინტუიციურად დასაბუთდა, რომ არსებობს ენები, რომელთა წარმოქმნაც შეუძლებელია კონტექსტისგან თავისუფალი **PS**-გრამატიკით. – ამის ფორმალური

<sup>15</sup> 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 პოლინომურ კლასს. როგორც ნაჩვენებია 12.4 ქვეთავში **C2**-ენების კლასი შეიცავს უმეტეს, მაგრამ არა ყველა კონტექსტისგან თავისუფალ ენას, ასევე ბევრ კონტექსტზე-დამოკიდებულ ენას.

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

შემდეგი უფრო ფართო ენობრივი კლასი **PS**-გრამატიკულ იერარქიაში არის კონტექსტზე დამოკიდებული ენები, რომლებსაც **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 „ვინმემ შეიძლება ივარაუდოს“, რომ კონტექსტზე დამოკიდებული ენებისათვის მტუმბავი ლემა არ მუშაობს. 1 ტიპის, ანუ კონტექსტზე დამოკიდებული წესების სტრუქტურა აღწერილია 8.3.6 განაწერში:

### 8.3.6 კონტექსტზე დამოკიდებული ფისების სტანდარტული სქემა

$$\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2, \text{ სადა } \beta \text{ არ არის ცარიელი მიმდევრობა.}$$

**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 გამხოლოებული ცვლადის განთავსება, კონტექსტზე დამოკიდებული 1 ტიპის წესის მარცხენა მხარეში ნებადართულია ცვლადის განთავსება გარემუზობელ (ე.ი. მის ორივე მხარეს მდებარე) ტერმინალურ სიმბოლოებთან ერთად. ეს, როგორც ამას 8.3.6 პუნქტში აღწერილი სქემაც გვიჩვენებს, 1 ტიპის წესს კონტექსტზე დამოკიდებულს ხდის, რადგან ამ სქემის მიხედვით, **A** ცვლადი შეიძლება ჩანაცვლდეს (ე.ი. გადაიწეროს)  $\beta$  სიმბოლოთი მხოლოდ  $\alpha_1 \text{---} \alpha_2$  ‘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 კონტექსტზე დამოკიდებული $a^k b^k c^k$ ენის PS-გრამატიკა

$V =_{\text{def}} \{S, B, C, D_1, D_2, a, b, c\}$		
$V_T =_{\text{def}} \{a, b, c\}$		
$P =_{\text{def}} \{$	$S \rightarrow a S B C,$	<i>წესი 1</i>
	$S \rightarrow a b C,$	<i>წესი 2</i>
	$C B \rightarrow D_1 B,$	<i>წესი 3a</i>
	$D_1 B \rightarrow D_1 D_2,$	<i>წესი 3b</i>

<sup>16</sup> 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<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** ენის გამოსახულებებში არსებული სამჯერადი ურთიერთშესაბამისობის გაკონტროლებასა და რეალიზაციაში კონტექსტის გადაწყვეტ როლს თვალნათლივ წარმოაჩენს **aaabbbccc** 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. <b>S</b>	
2. <b>a S B C</b>	(1)
3. <b>a a S B C B C</b>	(1)
4. <b>a a a b C B C B C</b>	(2)
5. <b>a a a b B C C B C</b>	(3)
6. <b>a a a b B C B C C</b>	(3)
7. <b>a a a b B B C C C</b>	(3)
8. <b>a a a b b B C C C</b>	(4)
9. <b>a a a b b b C C C</b>	(4)
10. <b>a a a b b b c C C</b>	(5)
11. <b>a a a b b b c c C</b>	(6)
12. <b>a a a b b b c c c</b>	(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 ამავედროულად განაპირობებს გამოთვლითი სირთულის მაღალ ხარისხს. იმისათვის, რომ ავტომატური საშუალებებით ზუსტად განვსაზღვროთ კონტექსტზე დამოკიდებული წესების ის მიმდევრობა, 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 ღირებულების 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.<sup>17</sup> 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 ენები<sup>18</sup>.

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, ამავედროულად, არ არის კონტექსტზე დამოკიდებული, წარმოადგენს აკერმანის ფუნქცია.<sup>19</sup>

**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 არამოხსნადია. იმის გარკვევამ, არის თუ არა რეკურსულად გადათვლადი ენის გამოსახულება სწორად აგებული, შეიძლება წაიღოს არა ძალიან დიდი დრო, არამედ სამუდამო.<sup>20</sup>

## 8.4 **PS-გრამატიკა ბუნებრივი ენისათვის**

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

<sup>17</sup> Hopcroft and Ullman ,1979, გვ.228, თეორემა 9.8.

<sup>18</sup> **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)**.

<sup>19</sup> Hopcroft and Ullman , გვ.175, 7.4.

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

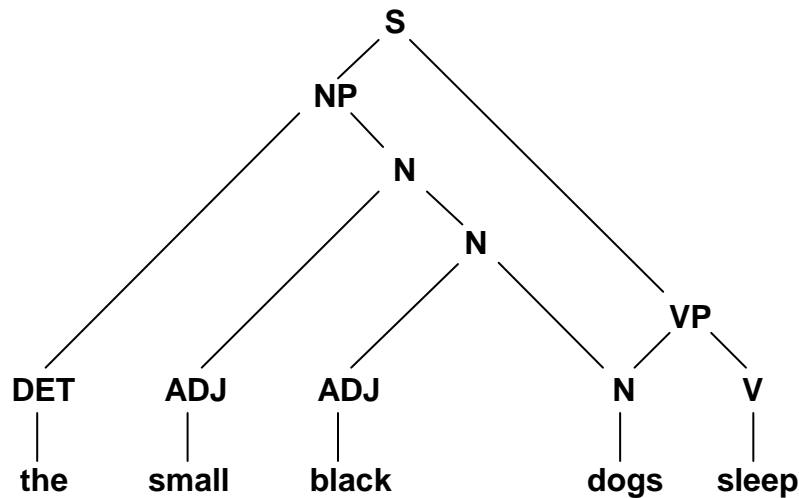
### 8.4.1 PS-გრამატიკა 7.5.3 მავალითისათვის

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

ამ PS-გრამატიკის ფორმა კონტექსტისგან თავისუფალია: ის ჯერ კიდევ არ არის კონტექსტზე დამოკიდებული. ეს იმიტომ, რომ წესების მარცხენა 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 წინადადების PS-გრამატიკული ანალიზის მავალითი



ასეთ ხეებს PS-გრამატიკაში ფრაზათა სტრუქტურებს უწოდებენ. ფრაზის სტრუქტურის ხეში კატეგორიათა აღმნიშვნელი 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 წინადადების 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 **შემაღგენელი სტრუქტურების (constituent structure)** ამომწურავ გამოკვლევას.

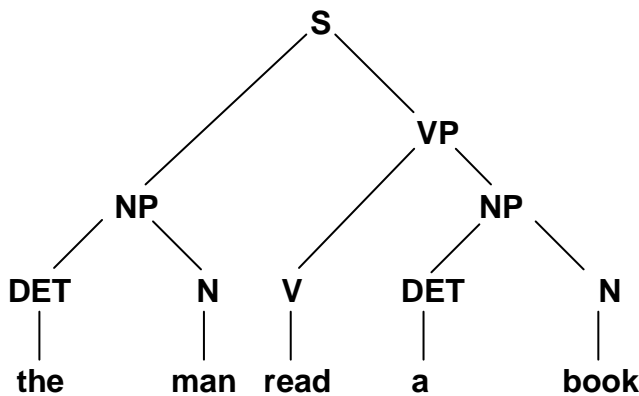
შემაღგენელი სტრუქტურები ასახავენ ლინგვისტურ ინტუიციებს, იმის შესახებ, თუ წინადადების რომელ ნაწილებს შორის არის ყველაზე უფრო მჭიდრო შინაარსობრივი კავშირები. შემაღგენელი სტრუქტურების ეს ინტუიციური პრინციპი ფორმალურად გადააზრებულია ფრაზის სტრუქტურული ხეების თვისებებად.

### 8.4.3 შემაღგენელი სტრუქტურის განსაზღვრება

1. სიტყვები და/ან შემაღგენლები, რომლებიც შინაარსობრივად ერთმანეთს მიეკუთვნებიან, უნდა იყვნენ უშუალოდ და სრულად დომინირებულნი რომელიმე კვანძის მიერ.
2. შემაღგენელი სტრუქტურის ხაზები არ უნდა იკვეთებოდნენ (**არაჩახლართულობის პირობა (nontangling condition)**).

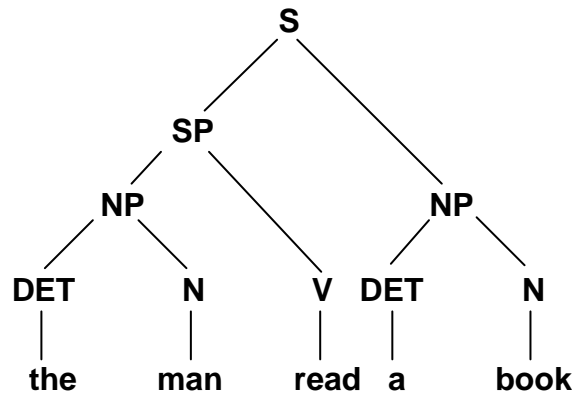
ამ განსაზღვრების თანახმად, **the man read a book** წინადადების შემდეგი ანალიზი ლინგვისტურად კორექტულია.

### 8.4.4 შემაღგენელი სტრუქტურის კორექტული ანალიზი



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

#### 8.4.5 შემადგენელი სტრუქტურის არაკორექტული 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).

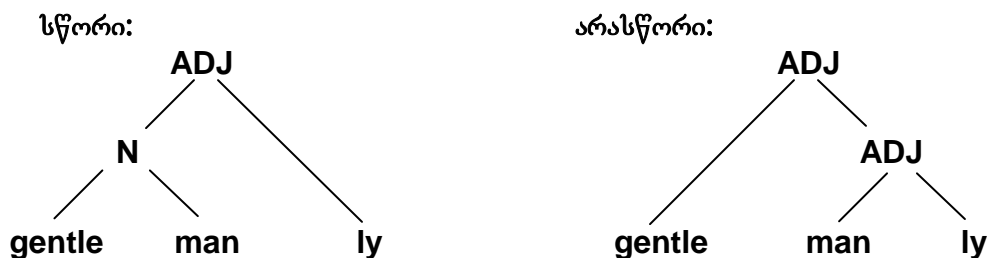
ისტორიულად, შემადგენელი სტრუქტურის ცნება არც ისე ძველია. ის წარმოიშვა ამერიკელი სტრუქტურალისტის ლ. ბლუმფილდის (**L. BLOOMFIELD (1887-1949)**) უშუალო შემადგენელთა ანალიზისა და მისი სტუდენტის – ზ. ჰარისის (**Z. Harris**) დისტრიბუციული ტესტირების საფუძველზე შემუშავებული მიდგომების შეჯერებით. ბლუმფილდის 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 ხისებრი სტრუქტურით გამოიხატება:

#### 8.4.6 უშუალო შემადგენლები PS-გრამატიკაში:



**gentlemanly** სიტყვის ეს მაგალითი არის აგრეთვე ჰარისის 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:*

*არაკორექტული ჩანაცვლება:*

Suzan has [eaten] an apple

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 სწორადაა 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. წინადადებების სეგმენტაცია და მისი თანამდევი ჰიპოთეზები წინადადებების ერთმანეთთან უფრო მეტად, თუ უფრო ნაკლებად დაკავშირებულ ნაწილებზე საჭირო გახდა რათა ერთმანეთისაგან განესხვაებინათ ლინგვისტურად კორექტული და არაკორექტული ფრაზის სტრუქტურის ხეები.

ასეთი განსხვავება აუცილებელია, რადგანაც ნებისმიერი სასრული სტრიქონით განსაზღვრული შესაძლო ფრაზის სტრუქტურების რაოდენობა უსასრულოა<sup>21</sup>. ეს ქმნის ხელისშემშლელი სიჭარბის

<sup>21</sup> თუ **A-B-C...A** სახის ფრაზის სტრუქტურებს გამოვრიცხავთ, განსხვავებული ფრაზის სტრუქტურული ხეების რაოდენობა შეყვანილი მიმდევრობის სიგრძის ზრდასთან ერთად მანც ექსპონენციალურად გაიზრდება. 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 ვიდრე 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 გადაჩანაცვლების 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. ორმოცდაათი წლის მნიშვნელოვნად ფინანსირებადმა კვლევებმა ვერ მოგვცა ვერცერთი ბუნებრივი ენის სრული ანალიზი, რაც შეიძლება მივიღოთ მინიშნებად, რომ ეს მიდგომა არასაკმარისად ოპტიმალურია.

## 8.5. შიდადგენილი სტრუქტურის პარადოქსი

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.

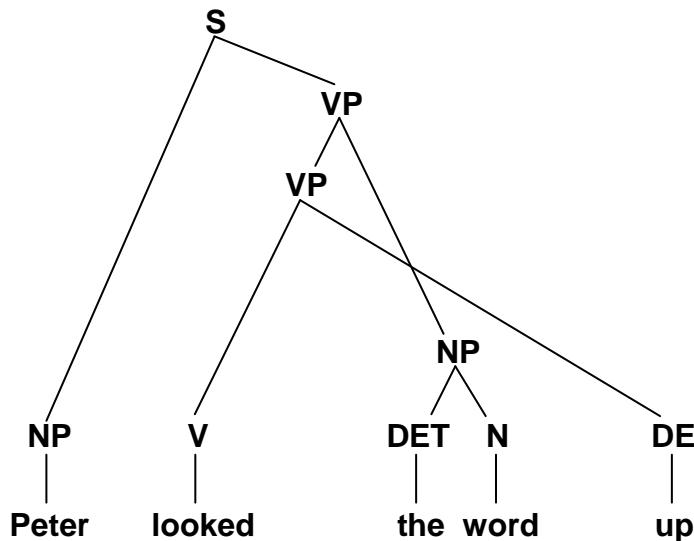
მაგალითად, საერთო შეთანხმებით წინადადებაში

**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)** ეწოდა<sup>22</sup>.

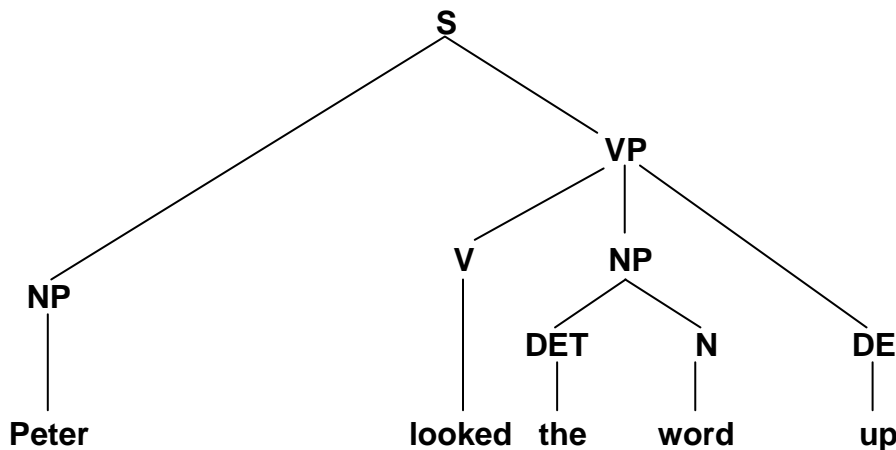
**8.5.1 8.4.3 ბანაწმირის მემორა პირობის უარყოფელი მაგალითი**



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.

<sup>22</sup> CoL, გვ. 24.

## 8.5.2 8.4.3 განაწერის პირველი პირობის უარყოფელი მაგალითი



ამ ხის ტოტები არ იკვეთებიან და ამით იგი აკმაყოფილებს 8.2.1 განაწერის მეორე პირობას. თუმცა, ამ ანალიზისას ირღვევა პუნქტის პირველი პირობა, რადგანაც სემანტიკურად მიკავშირებული **looked** და **up** გამოსახულებები, ანუ **V** და **DET** კვანძები სრულად და უშუალოდ არ დომინირებიან 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 ცნობილი<sup>23</sup>.

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. ჰარისის მეთოდოლოგიით მოტივირებული ჩანაცვლებისა და გადანაცვლების ტესტები გადაამუშავა წარმომქმნელ წესებად. ამ წესებს ტრანსფორმაციები ეწოდა და ისინი აუცილებელ, 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.

<sup>23</sup> ი. ბარ-ჰილელი 1960 [1964, გვ.102] წელს წერდა, რომ მან მიატოვა **C**-გრამატიკის 1953 წლის ნამუშევრები დისკონტინიალურ შემადგენლების ანალოგიური სიძნელის გამო **He gave it up** წინადადებაში.

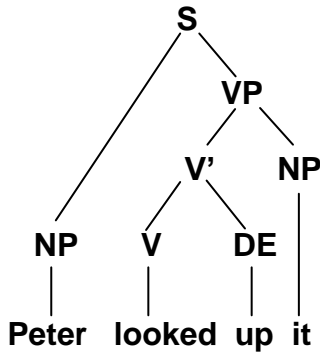
### 8.5.3 ფორმალური ტრანსფორმაციის მაგალითი

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

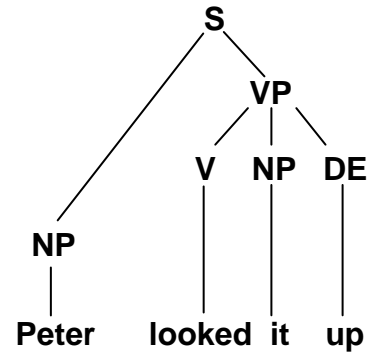
ამ ტრანსფორმაციის გამოყენება ნაჩვენებია 8.5.4 პუნქტში.

### 8.5.4 8.5.3 ტრანსფორმაციის გამოყენება

სიღრმისეული სტრუქტურა:  
(deep structure)



ზედაპირული სტრუქტურა:  
(surface structure):



ტრანსფორმაცია

⇒

სტანდარტულ თეორიის (ST, ჩომსკი, 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 შეესაბამებოდნენ გრამატიკულ მიმდევრობებს (როგორც ეს არის 8.5.4 პუნქტში აღწერილი მარცხენა ფრაზის სტრუქტურულ ხეში).

თუ ტრანსფორმაციულ წესში შემავალი გამოსახულება (ინდექსირებული ფრჩხილებიანი 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 მიმდევრობასთან, 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 აღიწერება მოსაუბრე-მსმენელის თანდაყოლილი ცოდნა კომუნიკაციური ფუნქციონირების გარეშე.

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

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

ნ. ჩომსკი იმედოვნებდა, რომ გვერდს აუვლიდა ამ არასასურველ ვითარებას **აღდგენის უნარიან წაშლებად (recoverability of delation)** 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-ხის აღდგენის საშუალებასაც იძლევა.

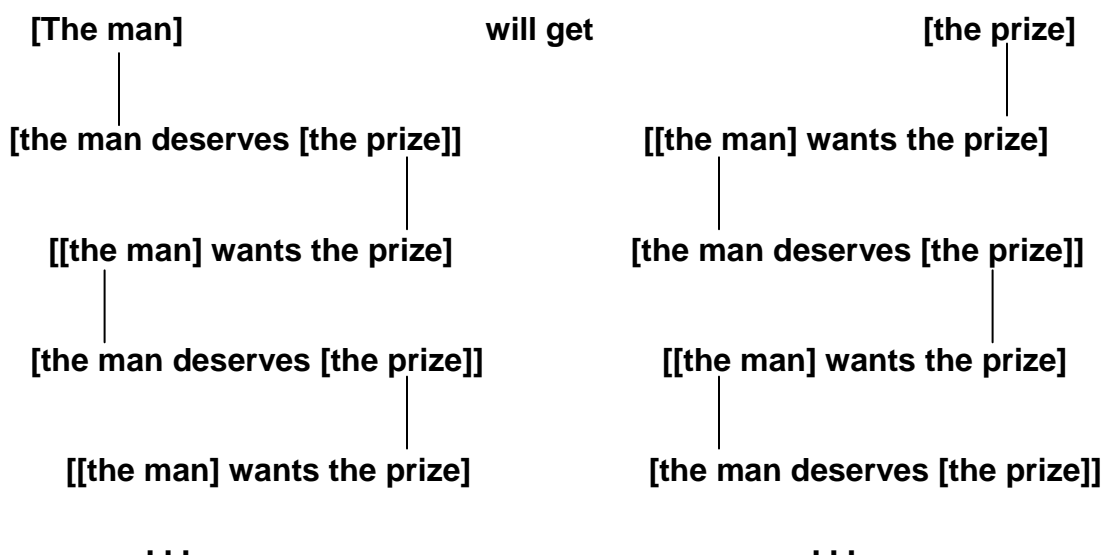
ბახ-პეტერსის წინადადებები გვარწმუნებენ, რომ აღდგენის უნარიანი წაშლები ყოველთვის არ იძლევა სასურველ ეფექტს.

### 8.5.5 ბახ-პეტერსის წინადადების ერთი 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 წინადადების 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 ადღენის უნარიანი წაშლების ბიჯის მოლოდინში.

ასე მაგალითად, დავეშვათ 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, და.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 რიჭის<sup>24</sup> მიერ (გამოქვეყნდა 1973 წელს).

ნეტივიზმის (**nativism**) გვიანი ვარიანტები, როგორებიცაა **LFG**, **GPSG** და **HPSG**, ტრანსფორმაციული მიდგომების გარეშე ცდილობენ გადაწყვიტონ შემაღგენელი სტრუქტურის პარადოქსი. **PS**-გრამატიკის ამ წარმოებულ ფორმალიზმებში პირობები შემაღგენელი სტრუქტურებისგან ითხოვენ მხოლოდ იმას, რაც დაშვებულია ენობრივად შესატან მოცემულობაზე. სხვაგვარად რომ ვთქვათ (როგორც დისკონტინიალური შემაღგენლების 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, შემაღგენელ სტრუქტურათა მეთოდმა ვერ შეძლო გაემართლებინა მისი ნავარაუდები 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 კიდევ, თუ რატომ დგანან ისეთი სისტემები, როგორებიცაა **GPSG**, **LFG** და **HPSG**, შემაღგენელი სტრუქტურების გრამატიკებთან ერთად. ამის გარდა, ეს გვიანი სისტემები სირთულის თვალსაზრისით ტრანსფორმაციულ გრამატიკებზე ნაკლებ სრულყოფილებია: ისინი წარმოქმნიან რეკურსულად გადათვლად ენათა კლასს და არიან არაამოხსნადები<sup>25</sup>.

<sup>24</sup> ამ პერიოდში ბობ რიჭის ჰქონდა პირადი კავშირები და აქტიური კონსულტაციები ნ. ჩომსკისთან.

<sup>25</sup> იხ. **B. Barton, R.C. Berwick da E.S. Ristad** 1987, თუ კონტექსტისგან თავისუფალი წესის ამდაგვარი  $A \rightarrow B \rightarrow .$   $. \rightarrow A$  გაკვანძვა იკრძალება, მაშინ **LFGs** სირთულე ექსპონენციალურია, რადგანაც ასეთი კვანძები ფორმალურად ლეგალურია კონტექსტისგან თავისუფალ **PS**-გრამატიკაში. სირთულის თეორიის თვალსაზრისით ეს შეზღუდვა არ არის დაკანონებული. უფრო მეტიც, მაჩვენებლიანი სირთულე გაცილებით მეტია გამოთვლითი გამოყენებისათვის.

# სავარჯიშოები

## 8.1 ქვეთავი

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

## 8.2 ქვეთავი

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

## 8.3 ქვეთავი

1. 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  $a^k b^m c^k$  ენების გრამატიკების წარმოქმნის უნარები. რომელი მათგანია მაღალი და რატომ?
2. რა მიზეზებით იხსნება **PS**-გრამატიკაში არსებობა ტერმინისა კონტექსტისგან თავისუფალი?
3. რა სახის სტრუქტურები წარმოიქმნება კონტექსტისაგან თავისუფალი **PS**-გრამატიკებისგან?
4. დაასახელეთ ორი ხელოვნური ენა, რომლებიც არ არიან კონტექსტისგან თავისუფალი. რატომ აჭარბებენ ეს ენები კონტექსტისგან თავისუფალი გრამატიკების წარმოქმნით შესაძლებლობებს?
5. განსაზღვრეთ **PS**-გრამატიკა  $a^k b^{2k}$  ენისათვის. ახსენით თუ რატომ ემორჩილება ეს ენა წყვილების ინვერსირების კონტექსტისგან თავისუფალ სქემას.
6. განსაზღვრეთ **PS**-გრამატიკა  $ca^m dyb^n$  ენისთვის. მოიყვანეთ ამ ხელოვნური ენის დასრულებული გამოსახულებების მაგალითები. ახსენით თუ რატომ არის ეს ენა წესიერი ენა.
7. რატომ არღვევს 8.3.5 კონტექსტზე-დამოკიდებული **PS**-გრამატიკის წესის განსაზღვრებას, მაშინ როცა  $\beta$  ნოლია?
8. რა არის მტუმბავი ლემა?
9. რატომ არ არსებობს კონტექსტზე დამოკიდებული ენებისათვის მტუმბავი ლემა?
10. არიან თუ არა რეკურსულად გადათვლადი ენები რეკურსული?

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

#### 8.4 ქვეთავი

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

#### 8.5 ქვეთავი

1. აღწერეთ დისკონტინიალური ელემენტის ცნება ბუნებრივ ენაში და ახსენით, თუ რატომ იწვევს დისკონტინიალური ელემენტი შემადგენელი სტრუქტურის პარადოქსს.
2. რა ხერხებით ცდილობს ტრანსფორმაციული გრამატიკა დისკონტინიალური ელემენტების მიერ წარმოქმნილი პრობლემის გადაჭრას?
3. შეადარეთ ტრანსფორმაციული გრამატიკის და გამოთვლითი ლინგვისტიკის მიზნები.
4. რა არის ტრანსფორმაციული გრამატიკის წარმოქმნის უნარი?
5. 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  $a^k b^k$  ენის ერლეს 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 ფსიქოლოგიკური თვისებები, რომლებიც არ შეიძლება არ ჰქონდეს ბუნებრივი ენის ემპირიკულად აღეკვატურ წარმომქმნელ გრამატიკას.

### 9.1 ანალიზის დეკლარაციული და პროცედურული ასპექტები

ხელოვნური ენების ანალიზატორები, ანუ პარსერები<sup>1</sup> (**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

<sup>1</sup> როგორც განმარტებულია 1.3 ქვეთავში, პარსერი არის კომპიუტერული პროგრამა, რომელიც აბუშავებს ენობრივ გამოსახულებებს, როგორც შემავალ (**input**) მოცემულობებს, და გამოსავალში (**output**) იძლევა ახალ, გადამუშავებულ გამოსახულებებს, რაც დასამუშავებლად შესული გამოსახულების გარკვეული სახის სტრუქტურული ანალიზის შედეგია.

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 კომბინატორიკის 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 ამოცნობას 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 მათთვის შემუშავებული გამოთვლითი ალგორითმით.

სხვანაირად რომ ვთქვათ, თანამედროვე პარსერს შეუძლია გამომთვლელში (**computer**) ჩატვირთოს სხვადასხვა ფორმალის მქონე **G<sub>i</sub>, G<sub>j</sub>, G<sub>k</sub>** 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<sub>j</sub>** ენა მასთან მიკავშირებულ **G<sub>j</sub>** გრამატიკულ ინტერპრეტაციაში. **L<sub>j</sub>** ენის გამოსახულებების ავტომატური ანალიზისას ცხადად განირჩევა ერთმანეთისაგან დატვირთვების

გადაწილება (i)  $G_j$  გრამატიკაზე და (ii) პარსერზე  $G_i, G_j, G_k$  ფორმალური გრამატიკების მთლიანი კლასისათვის.<sup>2</sup>

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

### 9.1.1 დეკლარაციული & პროცედურული ასპექტები ენათმეცნიერებაში

– ენის გამოთვლითი ანალიზატორის, ანუ პარსერის **დეკლარაციული ასპექტი (declarative aspect)** მოიცემა იმ ენისათვის დაწერილი წარმომქმნელი გრამატიკით, რომელიც უნდა გაანალიზდეს ზოგად, მათემატიკურად მკაცრად განსაზღვრულ ფორმალიზმში.

– ენის გამოთვლითი ანალიზის, ანუ პარსერის **პროცედურული ასპექტი (procedurai 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:

---

<sup>2</sup> 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. პროგრამისტი ინდივიდუალური განმასხვავებელი თავისებურების ან პროგრამული გარემოს შედეგი) და რომელია თეორიულად აუცილებელი (ე.ი. აღწერილი ენის ფორმალური ანალიზის შედეგი). ამგვარი მიდგომის კიდევ ერთი უარყოფითი მხარე იმაში მდგომარეობს, რომ ასეთი პროგრამა შეიძლება 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 ქვემოთგამლად 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 მომავლესრიგებელი გადაწყვეტები, რომლებიც გადის ფორმალური გრამატიკის დეკლარაციული შემადგენლების მიღმა, ან რომლებიც – მანალიზებული 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 მანალიზებული 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  $a^k b^k$  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 ბრამატიკის მორგება 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 კონტექსტისგან 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\_ფრაზა<sub>1</sub> სახელური\_ფრაზა<sub>2</sub> ... ზმნური\_ფრაზა<sub>2</sub> ზმნური\_ფრაზა<sub>1</sub>“,

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 **n<sup>3</sup>** 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 საკმარისი 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 ანალოგიით – ის ცდილობდა დაესაბუთებინა, რომ ბუნებრივი ენები, სულ ცოტა, კონტექსტზე დამოკიდებულნი არიან.

## 9.2.2 კონტექსტზე 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 სირთულისაა და რომ გრძელი წინადადების ანალიზი უარეს შემთხვევაში მილიარდობით საუკუნეს მოგვთხოვდა (იხ. 8.2.2).

ამ წარმოუდგენელი შედეგის თავიდან ასაცილებლად ჰერმანმა და გაზდარმა, შესაბამისად 1963 და 1982 წლებში წარმოადგინეს საკმარისი ზომის **PS**-გრამატიკული სისტემები, რომლებიც გამიზნული იყო იმის საჩვენებლად, რომ ბუნებრივ ენაში არ არსებობს ისეთი სტრუქტურა, რომელიც არ შეიძლება არ იქნას დაყვანილი კონტექსტისგან თავისუფალ ფორმაზე. ჰერმანის განაცხადი გაკეთდა იმ დროს, როდესაც სირთულეთა თეორია არ იყო ფართოდ გამოკვლეული და ტრანსფორმაციულ გრამატიკებში მოღვაწეობდა მხოლოდ ერთუზიასტთა მცირე ჯგუფი. ამავდროულად, ჰერმანის 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**-წესების<sup>3</sup> კომბინირება, იმისათვის, რომ შესაძლებელი ყოფილიყო იმ ლინგვისტურ განზოგადებათა ფორმალიზება, რომელთაც იმ დროს გაზდარიც და სხვებმაც მნიშვნელოვნად მიიჩნევდნენ.

გარდა ამისა, იმელოვნებდნენ, რომ **GPSG** ბუნებრივ ენებს გააანალიზებდა სირთულის იმ ხარისხით, რომელიც გამოთვლითი თვალსაზრისებით პრაქტიკულად ღირებულად მიიჩნევა, ე.ი. 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 გამოკვლევამ<sup>4</sup> 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:

<sup>3</sup> "დაახლოებით ტრილიონობით წესი", ს. შებერი, ს. სტაკი, ჰ. უშკორეტი & ჯ. რობინსონი 1983

<sup>4</sup> ჰ. უშკორეტი & ს. პეტერსი, 1986.



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

2. **PS**-გრამატიკა არაა წარმოქმნილი 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, პირველი 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 შემდეგი სამიდან მხოლოდ ერთი შესაძლო მიმართება გამოიკვეთება.

### 9.2.3 შესაძლო მიმართებები ორ გრამატიკულ ფორმალიზმს შორის

#### - 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 თავისუფალი 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 მივხვდი, რომ **BCG**ები [ბიდირექციული კატეგორიული გრამატიკები (იხ. 7.4.1)] არიან დაახლოებით იგივე სიძლიერის როგორც **CFPSG**ები [კონტექსტისგან თავისუფალი ფრაზების სტრუქტურული გრამატიკები]. მათი ეკვივალენცია დაამტკიცა კაუფმანმა 1959 წლის ივნისში. . . . გრამატიკის ამ განსხვავებული ტიპების ეკვივალენცია 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 ჰარისის [1951] და ჰოკეტის [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 ბიდირექციული **C** გრამატიკა და კონტექსტისგან თავისუფალი **PS** გრამატიკა ეკვივალენტურები არიან, გააფერკმათაა იმ ფაქტმა, რომ კონტექსტისგან თავისუფალი გრამატიკები წარმოქმნიან იმ ყველაზე უფრო ფართო ენობრივ კლასს, რომელიც ჯერ კიდევ ექვემდებარება გამოთვლას. ამის გათვალისწინებით, ბიდირექციული **C** გრამატიკისა და კონტექსტისგან თავისუფალი **PS** გრამატიკის სუსტი ეკვივალენცია, 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.<sup>5</sup>

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<sup>k</sup>b<sup>m</sup>**, **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>**, **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>d<sup>k</sup>**, **{a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>}\***, **WW<sup>R</sup>**, **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

<sup>5</sup> მართლაც, კონტექსტისგან თავისუფალი **a<sup>k</sup>b<sup>k</sup>** (იხ. 7.1.3, 10.2.2) და კონტექსტზე დამოკიდებული **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** (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<sup>R</sup>** (იხ. 8.3.4, 11.1.5) და კონტექსტზე დამოკიდებული **WW** ენები **LA**-გრამატიკაში კლასიფიცირებული არიან როგორც **C2-LAG**-ების ერთი და იგივე პოლონომური (**n**) 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 საერთო...ყალიბის...ფარგლებში...სპეციფიკური...ენის...აღსაწერად განსაზღვრული...წესთა სპეციფიკური სისტემა. მაგალითად, **PS** გრამატიკა, როგორც ზოგადი ფორმალური ყალიბი, განსაზღვრულია  $\langle V, V_T, S, P \rangle$  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 ფორმალურ ფორმალიზმზე, ასოცირებული ენის კლასები არ არსებობენ თავისთავად, ისინი გამოსახავენ გამოყენებული გრამატიკული ტიპის ფორმალურ თვისებებს. მაგალითად, კონტექსტისგან თავისუფალი ენების წყვილ-ინვერსირებადი სპეციფიკაცია პირდაპირ გამოდინარეობს 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 ფორმალური ფორმალიზმების განსაზღვრული შეზღუდვების შედეგია.

### - პარსერები (Parsers)

ანალიზატორები ენის ავტომატურად მანალიზებული 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 მანალიზებული 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 კონტექსტზე 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 ყურდნობიან ენის სპეციფიკურ სტრუქტურულ თვისებებს (რაიმე 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 რთული ენებია. შესაბამისად, ეს ენები ნებისმიერ შესაძლო გრამატიკულ ფორმალიზმში აუცილებლად მაღალი სირთულის კლასში იქნებიან (აქ 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, შეესაბამება 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 პარსერებს შორის

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

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

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 თანამიმდევრობითაც 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 გამოშვალ 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 მაწარმოებელი, ანუ გადამწერი სისტემა ააგო რეკურსიის თეორიაში ეფექტური გამოთვლადობის (**effective computability**) მათემატიკური ცნების განსასაზღვრავად.<sup>6</sup> პოსტის სისტემაში დერივაციული რიგი ემყარებოდა 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 მაწარმოებელი 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 ეკვივალენტიაში (**input-output equivalent**). – ეს იმას ნიშნავს, რომ ვერ იარსებებს ისეთი პარსერი, რომელსაც შეეძლება **PS**-გრამატიკის წესების პირდაპირი გამოყენება.

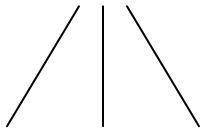
პარსერში კონტექსტისგან თავისუფალი **PS**-გრამატიკის გამოყენების სტრუქტურული პრობლემების სადემონსტრაციოდ ქვემოთ წარმოდგენილია **aaabbb** გამოსახულების ნაბიჯ-ნაბიჯითი დერივაცია (**derivation**) 7.1.3 განაწერში  $a^k b^k$  ენისათვის განსაზღვრულ **PS** გრამატიკაში.

### 9.3.2 aaabbb გამოსახულების ქვემოთ მიმართული გამოყვანა

<sup>6</sup> მაგალითისთვის იხილეთ **A. Church**, გვ. 52, სქოლიო № 119

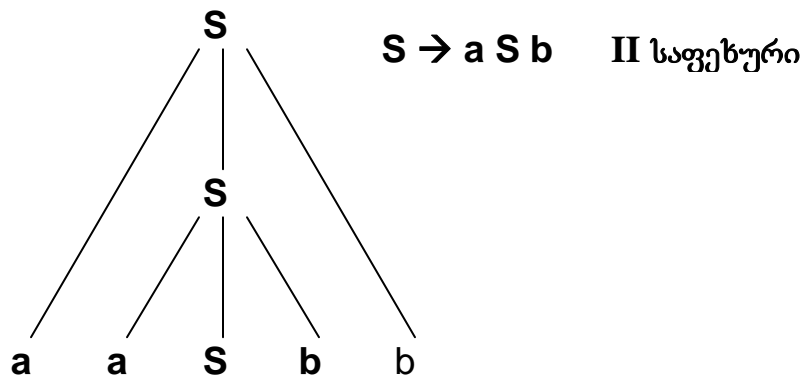
**S**

**S → a S b** I საფეხური



**a S b**



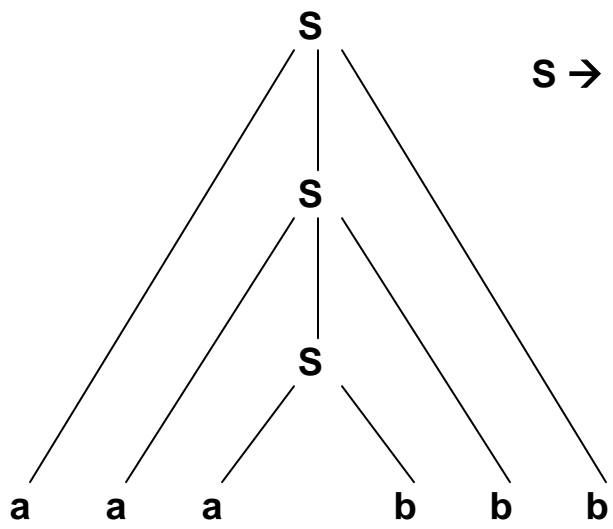












$S \rightarrow a b$  III საფეხური









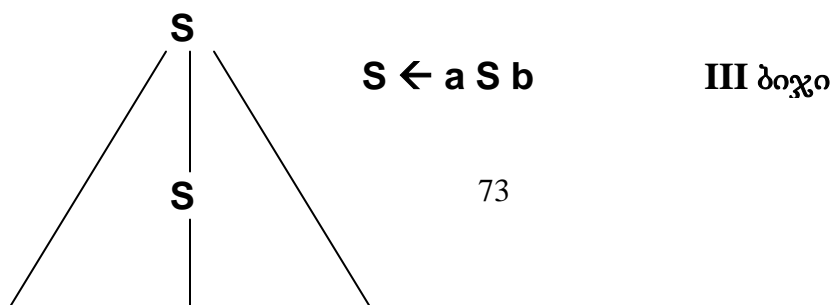
ამ ჩანაცვლებაზე დაფუძნებული გამოყვანის (**substitution-based derivation**) წესების მარცხენა მხარის **S** ცვლადი მრავალჯერაა ჩანაცვლებული წესების მარჯვენა მხარეს განთავსებული ნიშნების სტრიქონით. გამოყვანა იწყება **S** ცვლადით და მთავრდება ყველა ცვლადის ტერმინალური ნიშნებით ჩანაცვლების შემდეგ. ცვლადების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

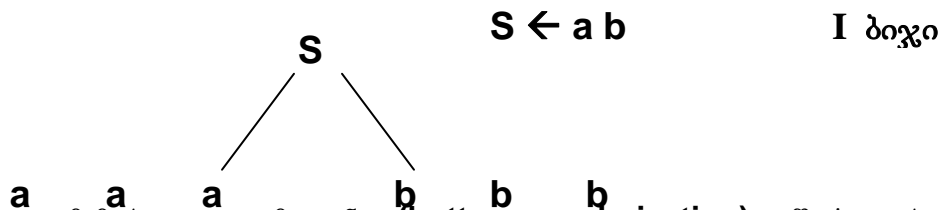
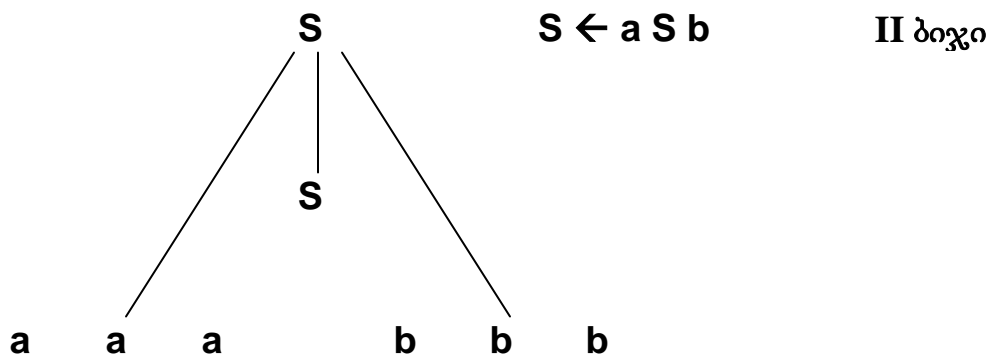
გამოთვლითი პროცედურა **aaabbb** შემაგალი გამოსახულების ანალიზისას პირდაპირ ვერ იყენებს **PS** გრამატიკის დერივაციულ მექანიზმს, რადგანაც **PS**-გრამატიკა გამოძვალ სტრიქონში ტერმინალურ სიმბოლოებს ყოველთვის ორ სხვადასხვა ადგილას სვამს. **PS** გრამატიკისგან განსხვავებით კომპიუტერულ პროგრამას დასამუშავებელი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

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

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

### 9.3.3 aaabbb გამოსახულების ზემოთ მიმართული გამოყვანა





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

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

$a^k b^k$  ენის შემთხვევაში პროგრამას ცენტრის განსასაზღვრად შეუძლია გამოიყენოს საზღვარი **a** და **b** სიტყვებს შორის, მაგრამ ასეთი მიდგომა 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  $a^k b^{3k}$  (იხ. 8.3.4) სახის კონტექსტისგან

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

იმის გამო, რომ ვაჩვენოთ, თუ როგორ შეიძლება გადალახოს პარსერმა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ლოგიკური წარმოების წესრიგთან შექმნილი სირთულეები ზოგადად მოქმედი მეთოდით, სვობს განვიხილოთ ანალიზი **ერლეს** 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 რესტრუქტურისაციის მიმოხილვა.

### 9.3.4 $a^k b^k$ ენის მრღმის მანანალიზებელი ალგორითმი

```
.aaabbb
.S
      a.aabbb
.ab  -> a.b
.aSb -> a.Sb
      aa.abbb
      a.abb  -> aa.bb
      a.aSbb -> aa.Sbb
      aa.aabb  aaab.bb
      aa.abb  -> aaa.bbb -> aaab.bb -> ...
      aa.aSbbb -> aaa.Sbbb
```

**ერლეს**<sup>7</sup> ალგორითმი იყენებს სამ ოპერაციას, რომლებიც არ არიან **PS**-გრამატიკის ალგებრული განსაზღვრების შემდგენლები (იხ. 8.1.1). ესენია **პრედიკატორ-** (**predicator-**), **სკანერ-** (**scan-**) და **კომპლეტორ-** (**completor-**) 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.ი.

<sup>7</sup> **ჯეი ერლეიმ** აავო თავისი მანალიზებელი ალგორითმი **PS**-გრამატიკებისთვის თავის დისერტაციაში კარნეგ-მელონის უნივერსიტეტის (პიტსბურგი, 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.

შემდეგ ციკლში წერტილი კვლავინდებურად წაიძრა კიდევ ერთი პოზიციით მარჯვნივ. შედეგად 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 შედეგად იძლევა მდგომარეობებს

**a a. b b** და **a a. S b b**

რომელთაგანაც ორივეს **სკან-ოპერაცია** უსაბამებს დასამუშავებელ სტრიქონს. შემდეგ **კომპლექტორ-ოპერაცია** წერტილს კვლავ წაძრავს ერთი პოზიციით მარჯვნივ. შედეგად ვლტულობთ

**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 შედეგად 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.<sup>8</sup> წერტილი კიდევ ერთხელ წაიძრა მარჯვნივ და შედეგად მოგვცა

**a a a b b. b**

<sup>8</sup> ამ სტადიაში მეორე გამოსახულებაც ასევე ჯერ კიდევ აქტიურია, რადგან ის ნებას რთავს მომდევნო **პრედიკატორ-ოპერაციას**, რომელიც შედეგად გვაძლევს ახალ მდგომარეობებს

**a a a a. b b b b**

**a a a a. S b b b b**

თუმცა ეს მდგომარეობები მიმდინარეობს **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 სტრიქონსა 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, რითაც 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 ტრანსფარენტული 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 რუტინაზე. მიუხედავად ამისა, თუ ამ მიდგომებით ხერხდება არსებული განსხვავებების სრული კორექცია, ანუ თუ **PS**-გრამატიკა ადეკვატურად არის აღწერილი, ასეთი გადაწყვეტები გამართლებულად მიიჩნევა.

ამის საპირისპიროდ, თუ პარსერი გამოყენებულია აღწერილად ადეკვატურ 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) ვერისტიკული 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 რეინტერპრეტაციების კონსტრუირება დაკავშირებულია დიდ ძალისხმევებთან და დანახარჯებთან. ამასთან, ირკვევა, რომ სასურველი შედეგის მისაღწევად არასაკმარისია პროგრამისტული დამუშავება და რომ რიგ შემთხვევებში საჭირო ხდება მექანიზმების გაზრდა.

## 9.4 შემავალ-გამომავალი ეკვივალენცია მსმენელ-მოლაპარაკე სისტემასთან

ნეიტრისტების მხარდამჭერებს ჰქონდათ მცდელობა მიეჩქმალათ **PS**-გრამატიკის სტრუქტურული სისუსტეები, როგორც პროგრამირების პრობლემა, რომელთა გადაწყვეტა კომპიუტერულ მეცნიერებათა ამოცანაა. ისინი ამბობდნენ, რომ მათი ნამდვილი მიზანი 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 არამართებულია. 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 ფორმალისთვის 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.

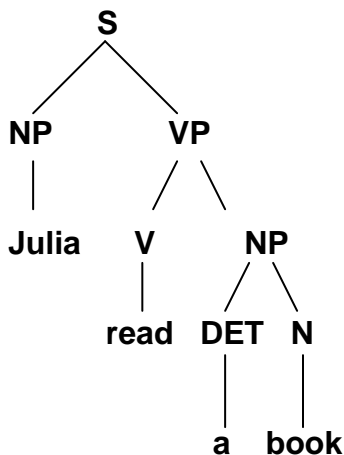
### 9.4.1 კონტექსტისგან თავისუფალი PS-გრამატიკა ინგლისური ენის მარტივი წინადადებისათვის

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

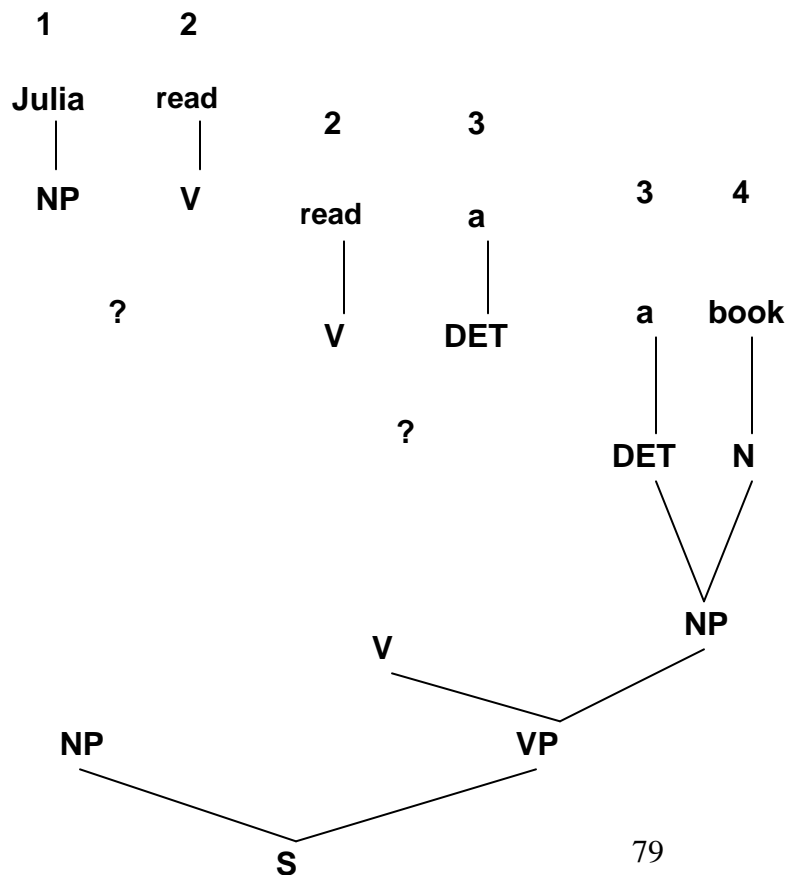
ცვლის **DET (determiner)** და **N (noun)** კვანძებით. საბოლოოდ, წესი 6 ცვლის **DET** კვანძს სიტყვით **a**, და წესი 7 ცვლის **N** კვანძს სიტყვით **book**.

### 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. შემაჯავალ-გამომავალი პირობების თვალსაზრისით **PS**-გრამატიკებსა და მოლაპარაკე-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.ი. სიტყვები იკითხება მარცხნიდან მარჯვნივ ერთი მეორეს მიყოლებით.<sup>9</sup> 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 ჩასანაცვლებლად ეძებს **PS**-წესს. თუმცა, რადგან ასეთი სახის წესის არსებობას ვერ უზრუნველფოს 9.4.1 გრამატიკა, ქვეზე, 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** კატეგორიის კომბინირებას ამ მესამე სიტყვის **DET** კატეგორიასთან უფრო მაღალი კვანძში. კვლავაც, 9.4.1 გრამატიკაში შესაბამისი წესის არ არსებობის გამო, მეორე სიტყვის შესაბამისი ქვეზე დროებით გვერდზე უნდა გადაიდოს 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 უზრუნველფოს 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 გვერდზე 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 ეცდება **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 არატრანსფარენტულობა 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 ნეიტრისტური გრამატიკა **მიზნად არ ისახავდა**<sup>10</sup> მოლაპარაკე-მსმენელის საკომუნიკაციო ურთიერთობების აღწერას და რომ ის

<sup>9</sup> ტრადიციული ბერძნულ-რომაული წერის შეთანხმების მიხედვით.

<sup>10</sup> ნ. ჩომსკი დაუღალავად აკეთებდა აქცენტირებას, რომ მისი ნეიტრისტური პროგრამის მიზანი არ იყო მოლაპარაკე-მსმენელს შორის კომუნიკაციის პროცედურის მოდელირება. მაგალითისთვის იხილეთ ჩომსკი 1965, გვ.9.

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

## 9.5 გრამატიკის უკმარისობა კონვერგენციის მისაღწევად

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

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“<sup>12</sup>. 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.<sup>13</sup> მეოთხე: ბუნებრივი ენის დამამუშავებელი

<sup>11</sup> მეცნიერების ისტორიაში კვლევათა ველი განხილულია როგორც დადებითად განვითარებადი თუ მისი სხვა და სხვა არეები იკრიბება. ე.ი. თუ გაუმჯობესება ერთ არეში იწვევს გაუმჯობესებას სხვა არეებში. ამის საპირისპიროდ, კვლევათა ველი განხილულია როგორც უარყოფითად განვითარებადი თუ გაუმჯობესება ერთ არეში იწვევს გაუარესებას სხვა არეებში.

<sup>12</sup> ეს რაოდენობა გამოთვლილია ენის ნეიტრალური თეორიის ფარგლებში ღია ალტერნატივების საფუძველზე. მაკოლის გამოთვლით დღეში საშუალოდ **2055** განსხვავებული გრამატიკული თეორია იქმნებოდა, ანუ თითო ახალი თეორია ყოველ 42 წამში, რაც მთლიანობაში 40 წელი გრძელდებოდა (1957 წლიდან - ანუ, **სინტაქტიკური სტრუქტურების** გამოჩენიდან - 1997 წლამდე).

<sup>13</sup> **PS-გრამატიკა** შემადგენელი სტრუქტურის ანალიზთან კომბინაციაში აჩვენებს აღწერით **ეიფორიას**, მაგ., ამას გვიჩვენებს გერმანულის ენის ისეთი ძირითადი კლაუზა, როგორცაა **Peter hat die Tür geschlossen**, რომელიც მისი დისკონტინიუალური 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 **სიმდიდრის 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 სტრუქტურის 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, კერძოდ 18.5 ქვეთავში.

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

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

ქვემოთ დამწეები ლინგვისტიკისათვის მოცემულია სხვადასხვა ფორმალიზმებზე ხანგრძლივი მუშაობის შედეგად მიღებული გამოცდილების საფუძველზე გაკეთებული რჩევები<sup>14</sup>: **(i) თქვენ უნდა გახსოვდეთ, რომ თქვენი მიზანია** ბუნებრივი კომუნიკაციის მექანიზმის მოდელირება, ამასთან, საჭიროა **(ii)** თქვენს მიერ შემუშავებული მოდელების გამოთვლითი ვერიფიცირება, და **(iii)** მათი უტილიზება ისეთ პრაქტიკულ გამოყენებებში, როგორცაა ადამიანი-მანქანის კომუნიკაციები. – თქვენი მოდელების 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 ( $n^3$ ), 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.

<sup>14</sup> 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 იქონიოს დიდი სოციალური გამოძახილი 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 ხდება, ერთი ფორმალიზმების ცვლა შესაძლოა გადაიქცეს გადარჩენის აუცილებელ საშუალებად. ამ შესაძლებლობების გაშუქებისას ყველაზე ხანგრძლივი სტრატეგია სამეცნიერო არჩევანთა შეფასება და გრამატიკული ფორმალიზმის რაციონალური არჩევაა. როგორც ყველა კარგ მეცნიერებაში, ლინგვისტური კვლევა, უნდა იმართებოდეს კვლევის ველის ფართო და ღრმა ცოდნაზე დამყარებული რწმენით.

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

### 9.5.2 მოთხოვნები წარმომქმნელ გრამატიკულ ფორმალიზმებზე

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 რეკურსიული (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 ეკვივალენტიაში მოლაპარაკე-მსმენელის 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 მოთხოვნებს. საპირისპიროა ვითარება **C**- და **PS**- გრამატიკულ ფორმალიზმებში, რომლებშიც გამოყვანები დაფუძნებულია შესაძლო ჩასმების (**possible substitutions**) პრინციპზე. **LA**-გრამატიკაში გამოყვანები დაფუძნებულია შესაძლო გაგრძელებათა (**possible continuations**) პრინციპზე.

## საკვარჯიშოები

### 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 ალტერნატივები.
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 კონტექსტისაგან 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 დამოკიდებულნი?
7. არიან თუ არა ენობრივი კლასები გრამატიკის ფორმალიზმზე დამოკიდებულნი?
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 ტრანსფარენტულობის ცნება.
2. რა მიზნით ააგო პოსტმა 1936 წელს თავისი მაწარმოებელი სისტემა?
3. როდის არიან გრამატიკული ფორმალიზმი და პარსერი შემავალ-გამომავალ ეკვივალენტიაში?
4. რა განსხვავებაა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ქვემოთ მიმართულ და ზემოთ მიმართულ გამოყვანებს შორის?
5. რატომ არ არის კონტექსტისგან თავისუფალი **PS**-გრამატიკა შემავალ-გამომავალ ეკვივალენტიაში მისსავე პარსერებთან? თქვენი ახსნები დააფუძნეთ ქვემოთ მიმართულ და ზემოთ მიმართულ გამოყვანებს.
6. ახსენით **ერლეის** ალგორითმის ფუნქციონირება  $a^k b^k$  ენის გამოსახულებების გამოყენებით. როგორ აკომპენსირებს **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  $a^k b^k c^k$  ენის გაანალიზება **ერლეის** ალგორითმით?
9. არსებობს თუ არა კონტექსტისგან თავისუფალი **PS**-გრამატიკის ტრანსპარენტული პარსერები?
10. დაასახელეთ ორი პრაქტიკული მოუხერხებლობა იმ ავტომატური ანალიზისა, რომელიც არაა ტრანსფარენტული ტიპის.

## 9.4 ქვეთავი

1. შემაჯავლ-გამომავალი ეკვივალენტურობის ცნების გამოყენებით ახსენით, რატომაა **PS**-გრამატიკაზე დამყარებული ენის ნეიტრალისტური თეორია შეუთავსებელი პრინციპთან – **ფორმები განაპირობებენ ფუნქციებს (form follows function)**.
2. წარმოაჩინეთ მაგალითით, რომ **PS**-გრამატიკის წარმოების რიგი შეუთავსებელია ბუნებრივი ენის დროით-საზოგადოებრივ სტრუქტურასთან.
3. 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

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

## 10 თავი

### მარცხნივ ასოცირებადი 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)<sup>1</sup> 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, რომელიც შემაჯავლ-გამომავალ ეკვივალენტიაში იქნებოდა მოლაპარაკე-მსმენელის სისტემასთან.

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

<sup>1</sup> ოფიციალური პუბლიკაცია არის Hausser, ჟურნალში თეორიული კომპიუტერული მეცნიერება (TCS).

კონცეპტუალურ 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 ანალიზისა და ავტომატური გენერაციის ურთირთაკავშირები **LA**-გრამატიკაში და ნაჩვენებია **LA**-გრამატიკის ტიპობრივი ტრანსფარენტულობა. 10.5 ქვეთავში წარმოდგენილია ბუნებრივი ენის **LA**-გრამატიკული ანალიზის ენათმეცნიერული მოტივაციები.

## 10.1 ზუსტა 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 გამოიყენებინან, შეიძლება არ იყოს ცხადი. მაგალითად:  $a+b+c$  შეიძლება ინტერპრეტირდეს, როგორც  $((a+b)+c)$ , ან როგორც  $(a+(b+c))$ . ჩვენ ვამბობთ, რომ  $+$  არის **მარცხნივ ჯგუფდებადი (left-associative)**, თუ ის ოპერანდებს აჯგუფებს მარცხნიდან მარჯვნივ, ისე როგორც  $((a+b)+c)$ . ჩვენ ვამბობთ, რომ  $+$  არის **მარჯვნივ ჯგუფდებადი (right-associative)**, თუ ის ოპერანდებს საპირისპირო მიმართულებით აჯგუფებს, როგორც  $(a+(b+c))$ .

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:*

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

*მარჯვნივ ჯგუფდებადი:*

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

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 ამავედროულად 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 ჯგუფებადი ფრჩხილებიანი სტრუქტურების გარდა, არსებობს ამავე ტიპის უამრავი არარეგულარული სტრუქტურა. მაგალითად:

$((a + b) + (c + d)) + e$   
 $((a + b) + ((c + d) + e))$   
 $(a + ((b + c) + (d + e)))$   
 $((a + (b + c)) + (d + e))$   
 $((a + b) + c) + (d + e)$

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 წარმოშობენ **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 ჯგუფებად ფრჩხილებიან სტრუქტურებსა (იხ. 10.1.1) და მათ შესაბამის ხეებს (იხ. 10.1.6).

### 10.1.2 შესაძლო გაგრძელებათა პრინციპი

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 სპეციფიური სქემატური წესით.



### 10.1.3 LA-გრამატიკის მარცხნივ ჯგუფდებიანი სქემატური წესი

$$r_1: cat_1 cat_2 \Rightarrow cat_2 rp_i$$

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 (დაწვეილებულია  $cat_1$ -თან) და **nw** მომდევნო სიტყვას (დაწვეილებულია  $cat_2$ -თან) ასახავს წინადადების ახალ **ss'** საწყისში (აღნიშნულია  $cat_3$ -ით).  $r_1$  წესის წარმატებული გამოძვავალი არის **მდგომარეობა (state)**, რომელიც (**ss' rp<sub>i</sub>**) დალაგებული წყვილით მოიცემა.

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

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

### 10.1.4 C-გრამატიკის გაბათილების წესის სქემა

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

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

### 10.1.5 PS-გრამატიკის გადაჭერის წესის სქემა

$$A \rightarrow BC$$

**A** სიმბოლოს ჩანაცვლება **B** და **C** სიმბოლოთი შეისაბამებს ხეობრივ სტრუქტურას, რომელშიც **B** და **C** სიმბოლოები უშუალოდ 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

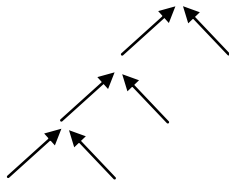
### 10.1.6 სამი აბსტრაქტული წარმომავალი დალაგება

LA-გრამატიკა

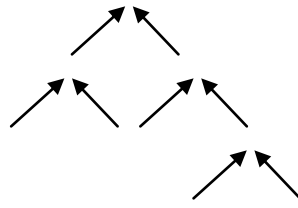
C-გრამატიკა

PS-გრამატიკა

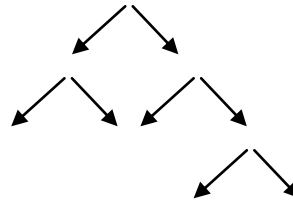




ზემოთ მიმართული მარცხნივ ჯგუფდებადი



ზემოთ მიმართული შევიწროებადი



ქვემოთ მიმართული გაფართოებადი

**C-** და **PS-** გრამატიკები, რომლებიც შესაძლო ჩანაცვლებათა პრინციპს ეფუძნებიან, განსხვავდებიან ერთმანეთისაგან მათი კონცეპტუალური<sup>2</sup> დერივაციული რიგის მიმართულებებით. **C-**გრამატიკაში ორი კატეგორიზებული გამოსახულება შეცვლილია ერთი კატეგორიზებული გამოსახულებით (ზემოთ მიმართულობა). **PS-**გრამატიკაში ერთი ნიშანი შეცვლილია ერთი, ორი ან მეტი ნიშნით (ქვემოთ მიმართულობა).

შესაძლო ჩანაცვლებათა პრინციპი წარმოშობს არარეგულარულ დერივაციულ სტრუქტურებს, რომლებიც იძლევიან არარეგულარულ ფრაზათა სტრუქტურულ ხეებს. ენის **SLIM** თეორიის თვალსაზრისით შესაძლო ჩანაცვლებათა პრინციპზე დაფუძნებული დერივაციული დალაგება წინააღმდეგობაში მოდის ენის საურთიერთობო ფუნქციასთან (ირლვევა 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 შემაღგენელ სტრუქტურებზე დაფუძნებული **C-** და **PS-**გრამატიკების ლინგვისტური თვალსაზრისებით არც კი იყოს საინტერესო. ენის **SLIM** თეორიის თვალსაზრისით ის (i) ამოღელირებს ბუნებრივი ენის ფუნდამენტურ დროში წარფივ სტრუქტურას, (ii) იძლევა შემავალ-გამომავალ ეკვივალენციას გრამატიკულ და მოლაპარაკე-მსმენელ სისტემებს შორის და (iii) უზრუნველყოფს ტიპობრივ ტრანსფარენტულობას გრამატიკასა და პარსერს შორის.

ამგვარად, გამოდის, რომ შემაღგენელი სტრუქტურის ფორმალური ცნება, განსაზღვრული ფრაზათა სტრუქტურული ხეების ტერმინებით (იხ. 8.4.3), არ გამოიყენება **LA-**გრამატიკაში. თუმცა, ბუნებრივი ენის მაგალითების ინტუიციური ანალიზისას ეს ცნება არაფორმალურად შეიძლება გამოვიყენოთ **შეღგენილი 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

<sup>2</sup> **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 შუა ნაწილიდან (კუნძულოვანი სინტაქსური პარსერები (**island parsers**)); ზოგიერთი იყენებს მარცხენა კუთხის, სხვები – მარჯვენა კუთხის დალაგებას და ა.შ. ეს ალტერნატიული დალაგებები, როგორც წესი, სისტემის პროცედურული ნაწილის შემაღგენლებია და ემსახურებიან ენის განსაკუთრებული თვისებების გამოყენებით პარსერის ეფექტურობის გაუმჯობესების მიზნებს.

## The little dog found a bone

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

### 10.2 LA-გრამატიკის ფორმალიზმი

**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  $C^* \cup \{\perp\}$ ,<sup>3</sup> 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 ქვესიმრავლეების მიმდევრობა და მას წესების პაკეტს (**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 დამასრულებელი 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<sub>s</sub>** (იხ. 6),

<sup>3</sup> თეორიული მიზნები განაპირობებენ იმას, რომ ეს კატეგორიალური ოპერაციები არიან სრულად განსაზღვრული (ტოტალური) ფუნქციები. პრაქტიკულად კი, ისინი გაიგებიან როგორც  $(C^* \times C^+)$  სიმრავლის მარტივად გამოცნობადი ქვესიმრავლეები, რომელთა გარეთ დარჩენილი განსაზღვრის არის ნებისმიერი წევრი "**don't care**" პრინციპის საფუძველზე აისახება ( $\perp$ ) სიმბოლოთი აღნიშნულ მნიშვნელობაზე, რითაც ეს კატეგორიალური ოპერაციები გარდაიქმნებიან სრულად განსაზღვრულ ფუნქციებად.

- (iii)  $r_i$  წესების 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<sub>i</sub>** კატეგორიალური ოპერაციის გამოყენებას. თუ მიღებული დასამუშავებელი გამოსახულებების კატეგორიები შესაბამისად უწყვილდება (უთანხმდება) **cat<sub>1</sub>** და **cat<sub>2</sub>** ნიმუშებს, მაშინ  $r_i$  წესის გამოყენება წარმატებით სრულდება გამოსავალი მოცემულობის წარმოქმნით. გამოსავალი შედგება წყვილისგან (**ss rp<sub>i</sub>**), სადაც **ss** არის დასამუშავებელი წინადადების ახლად წარმოქმნილი საწყისი და **rp<sub>i</sub>** არის  $r_i$  წესის მაპაკეტებელი. თუ დასამუშავებელი გამოსახულებების კატეგორიები არ წყვილდება **cat<sub>1</sub>** და **cat<sub>2</sub>** ნიმუშებთან, მაშინ  $r_i$  წესის გამოყენება არ სრულდება წარმატებით და, შესაბამისად, არც გამოსავალი არ წარმოიქმნება.

წესის მაპაკეტებელი **rp<sub>i</sub>** შეიცავს ყველა იმ წესს, რომლებიც შეიძლება წარმატებით იქნენ გამოყენებულნი  $r_i$  წესის შემდეგ. წესის მაპაკეტებელი არის წესების სახელების სიმრავლე, სადაც წესის სახელი არის **CO** მიმდევრობაში მისი **co<sub>g</sub>** 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 წინადადიდან იღებს ახალ მომდევნო სიტყვას (თუკი ის არსებობს) და იყენებს წესის მაპაკეტებლის წესებს წინადადების ახლად წარმოქმნილ საწყისზე და ამ ახლად აღებულ მომდევნო სიტყვაზე. ამგვარი მიდგომით **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<sup>k</sup>b<sup>k</sup>** ენისათვის წარმოდგენილია 10.2.2 პუნქტში.

### 10.2.2 LA-გრამატიკა **a<sup>k</sup>b<sup>k</sup>** ენისათვის

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

**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 სეგმენტებს<sup>4</sup>. ეს განპირობდა სიტყვებისა და მათი ენობრივი გამოსახულებების გაიგივებით.

**ST<sub>S</sub>** საწყისი მდგომარეობა გვანიშნებს იმას, რომ პირველი სიტყვა უნდა იყოს **(a)** კატეგორიის, ანუ იმას, რომ იგი უნდა იყოს **a** და რომ თავდაპირველად **r<sub>1</sub>**, ან **r<sub>2</sub>** წესი უნდა იქნეს გამოყენებული. ამგვარად, ალგორითმის პირველ ბიჯში შეიძლება იქნეს გამოყენებული ნებისმიერი წესი<sup>5</sup>, მაგრამ არა ნებისმიერი სიტყვა.

წესების კატეგორიალური ნიმუშები, ანუ მოდელები იყენებენ **X** ცვლადს (ის წარმოგვიდგენს ნულს, ან მეტ კატეგორიულ სეგმენტს), და **a** და **b** სეგმენტურ კონსტანტებს. **r<sub>1</sub>** წესი დასამუშავებლად იღებს წინადადების ნებისმიერი კატეგორიის საწყისს (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.

**r<sub>2</sub>** 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.

**r<sub>1</sub>** წესის მაპაკეტებელი, სახელად **rp<sub>1</sub>**, შეიცავს **r<sub>1</sub>** და **r<sub>2</sub>** წესებს. ამასთან, სანამ მომდევნო სიტყვა არის **a**, **r<sub>1</sub>** არის წარმატებული, ხოლო **r<sub>2</sub>** წარუმატებელი, რადგან ის მომდევნო სიტყვად მოითხოვს **b** სიტყვას. როგორც კი შესასვლელში პირველი **b** გამოჩნდება, **r<sub>2</sub>** წარმატებული ხდება, ხოლო **r<sub>1</sub>** წარუმატებელი. მაპაკეტებელი წესი **rp<sub>2</sub>** შეიცავს მხოლოდ ერთ წესს, სახელად **r<sub>2</sub>**. 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 **r<sub>2</sub>** წესის გამოყენების შემდეგ. ამას განაპირობებს 10.2.2 პუნქტში მოცემული **ST<sub>F</sub>** დამასრულებელი მდგომარეობა, რომელიც მოითხოვს  $\epsilon$  სიმბოლოს (ე.ი. ცარიელ მიმდევრობას), როგორც დამასრულებელ რეზულტატურ კატეგორიას.

შემდეგი LA-გრამატიკა წარმოშობს კონტექსტზე დამოკიდებულ **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** ენას.

### 10.2.3 LA-გრამატიკა **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** ენისათვის

$$\begin{aligned}
 \mathbf{LX} &=_{\text{def}} \{ [ \mathbf{a} (a) ] , [ \mathbf{b} (b) ] , [ \epsilon (\epsilon) ] \} \\
 \mathbf{ST}_S &=_{\text{def}} \{ [ (a) \{r_1, r_2\} ] \} \\
 r_1 &: (X) (a) \Rightarrow (aX) \{r_1, r_2\} \\
 r_2 &: (aX) (b) \Rightarrow (Xb) \{r_2, r_3\} \\
 r_3 &: (bX) (c) \Rightarrow (X) \{r_3\} \\
 \mathbf{ST}_F &=_{\text{def}} \{ [ \epsilon rp_2 ] \}
 \end{aligned}$$

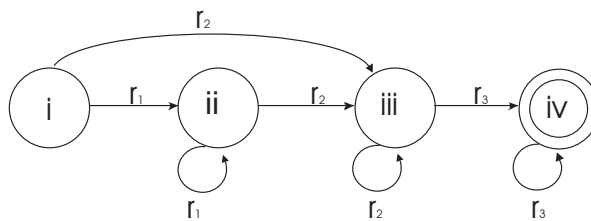
<sup>4</sup> უფრო რთული ენების LA-გრამატიკებში ლექსიკური კატეგორიები შედგება რამდენიმე სეგმენტისაგან.

<sup>5</sup> უფრო რთულ ენებში საწყისი მდგომარეობა დაწვრილებით 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 გარდამქმნელ 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.<sup>6</sup>

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



ეს **FSN** შედგება **i-iv** წრეებით წარმოდგენილი ოთხი მდგომარეობისგან. თითოეული მდგომარეობა განსაზღვრულია, როგორც დალაგებული წყვილი, რომელიც შედგება კატეგორიის ნიმუშისა და წესის პაკეტისგან. **i** მდგომარეობა შეესაბამება **ST<sub>s</sub>** სქემის საწყის მდგომარეობას, მაშინ როდესაც **ii**, **iii** და **iv** მდგომარეობები შესაბამისად შეესაბამებიან **r<sub>1</sub>**, **r<sub>2</sub>** და **r<sub>3</sub>** წესების გამოსავალს. **iv** მდგომარეობას აქვს მეორე წრეც, რომელიც აღნიშნავს იმას, რომ **iv** მდგომარეობა შეიძლება იყოს დამასრულებელი (იხ. **ST<sub>F</sub>** 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 გარდამქმნელი წესების პაკეტებში გაერთიანებული წესების კატეგორიალური ოპერაციების გაერთიანებით.

მაგალითად, **i** მდგომარეობიდან გამომავალი გარდაქმნები განსხვავდებიან ერთმანეთისაგან (კერძოდ ესენია, **r<sub>1</sub>** და **r<sub>2</sub>**) და შეესაბამება **ST<sub>s</sub>** სქემის საწყისი მდგომარეობების წესის პაკეტს. **ii** მდგომარეობაში შემავალი გარდაქმნები არიან ერთი და იგივე და გამოხატავენ **r<sub>1</sub>** წესის გამოყენებებს განსხვავებულ წინა მდგომარეობებზე. შესაბამისად, **ii** მდგომარეობიდან გამომავალი გარდაქმნები განსხვავდებიან ერთმანეთისაგან და შეესაბამებიან **r<sub>1</sub>** წესის პაკეტს. გარდაქმნები, რომლებსაც მივყავართ **iii** მდგომარეობამდე, ერთი და იგივენი არიან და განისაზღვრებიან **r<sub>2</sub>** წესის გამოყენებით სამ განსხვავებულ წინა მდგომარეობაზე, და ასე შემდეგ.

<sup>6</sup> უფრო დეტალური აღწერა ნახეთ **CoL**, ქვეთავი 8.2.

შემავალის 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).

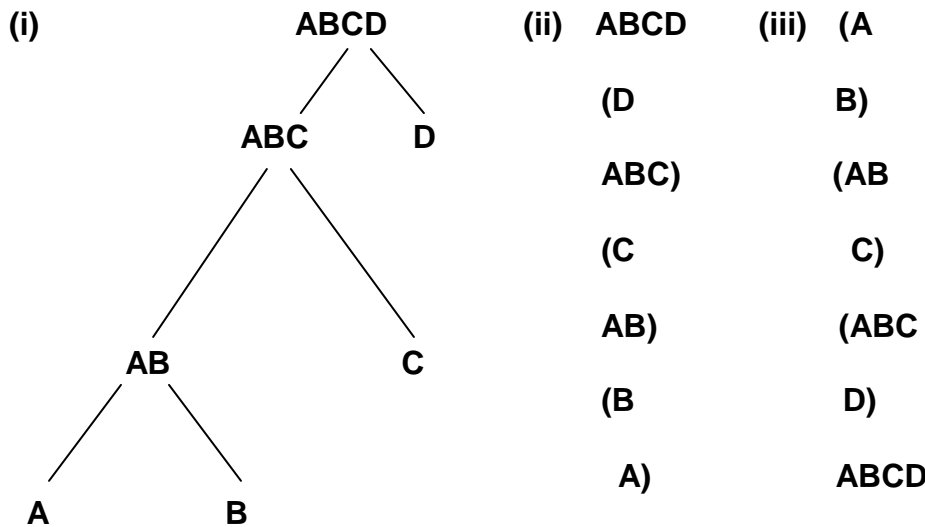
ეს უზარმაზარი წარმოქმნითი უნარი შეიძლება ძუნწად იქნეს გამოყენებული. მაგალითად,  $a^k b^k$  (იხ. 10.2.2) და  $a^k b^k c^k$  (იხ. 10.2.3) ენების **LA**-გრამატიკები არიან უმდაბლესი სირთულის: ისინი ორივე აპარსერებენ წრფივ დროში.<sup>7</sup>

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

### 10.3 დროში წრფივი ანალიზი

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

#### 10.3.1 LA-ხმები როგორც სტრუქტურირებული ცხრილები



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

<sup>7</sup> 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 $a^k b^k$ მინსათვის ( $k=3$ )

```

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
      (NIL) 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 მოჰყვება მაკომპლექტებული რიცხვი 1, წინადადების საწყისი (შეღებნილი პირველი სიტყვისგან), მომდევნო (ანუ მეორე) სიტყვა, და ბოლოს იმ წესის სახელი (**RULE-1**), რომელიც აკომბინირებს წინადადების ამ საწყისსა და ამ მომდევნო სიტყვას. მომდევნო მაკომპლექტებულ რიცხვს მოჰყვება ჯერ კიდევ დამუშავების პროცესში მყოფი შედეგი:

მოქმედი წესის პაკეტი:	<b>*START-0</b>
მაკომპლექტებული რიცხვი:	<b>1</b>
წინადადების საწყისი:	<b>(A) A</b>
მომდევნო სიტყვა:	<b>(A) A</b>
წარმატებული წესი:	<b>*RULE-1</b>
მომდევნო მაკომპლექტებული რიცხვი:	<b>2</b>
შედეგი:	<b>(A A) A A</b>

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:	<b>*RULE-1</b>
კომპლექტაციის რიცხვი:	<b>2</b>
წინადადების საწყისი:	<b>(A A) A A</b>
მომდევნო სიტყვა:	<b>(A) A</b>
წარმატებული წესი:	<b>*RULE-1</b>
მომდევნო კომპლექტაციის რიცხვი:	<b>3</b>
შედეგი:	<b>(A A A) A A A</b>

10.3.2 პროტოკოლში აღწერილი დერივაციის ფორმა შედგენილია წესების კატეგორიული ოპერაციების დასახასიათებლად. მიიჩნევა, რომ ეს დერივაცია საზღვრავს LA-გრამატიკის პირველი სიღრმის 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 ტრანსფარმენტულობა

LA-გრამატიკული ალგორითმი თანაბრად ვარგისია პარსირებისა და გენერირებისთვის. 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 გამოყენებულ წესებს (ქვემოთ გამოყენებადი წესების რიცხვი უფრო ნაკლებია ვიდრე 2n).

#### 10.4.1 aaabbbccc გამოსახულების 1 პარსირება წესთა 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 გამოყენების 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**)<sup>8</sup> 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 წარმომადგენელი ნიმუში $a^k b^k c^k$ –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)

```

<sup>8</sup> CoL, p. 193 ff. სხვა ვერსიით 'gram-gen' მოხსენიებულია მაქსიმალურ ზედაპირის სიგრძესთან 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 C C C**  
 1 1 1 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 შუალედურ გამოსახულებას, რომელთა სიგრძე არ აღემატება 12-ს. ენის დასრულებული გამოსახულებები გამოიცილება 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 სწორად-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 ( $\epsilon$   $r_p$ ), 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<sub>F</sub> სიმრავლის წევრია.

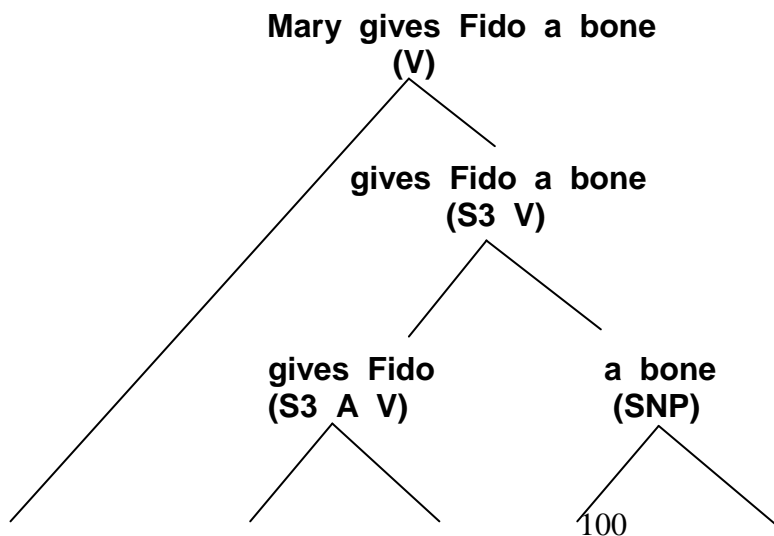
10.2.3-ში განსაზღვრულ LA-გრამატიკას, 10.4.1-ში ილუსტრირებულ LA-პარსერსა და 10.4.2-ში ილუსტრირებულ LA-გენერატორს შორის ურთიერთკავშირები ცხად დემონსტრირებას უკეთებს გრამატიკულ ფორმალიზმს, პარსერსა და გენერატორს შორის არსებული სრული ტიპობრივი ტრანსფარენტულობის მიმართებას. ამგვარად, LA-გრამატიკა არის ერთადერთი 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 (როგორცაა 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 შემაღვენელ 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 თანმხლები გარემოებებით იგი უფრო ახლოსაა **C**-გრამატიკულ ანალიზთან.

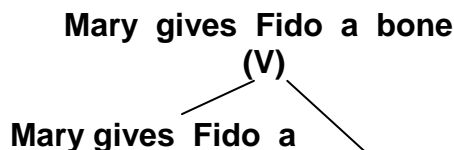
**C**-გრამატიკის ალგებრულ განსაზღვრებასთან შედარებით (იხ. 7.4.1) აქ (იხ. 10.5.1) გამოყენებული კატეგორიები განსაკუთრებულად მარტივი სახისანი არიან: ისინი წარმოგვიდგებიან **კატეგორიული სეგმენტების სიების სახით** – და არ შეიცავენ ხაზზე (ე.ი. / სიმბოლოზე) დამყარებულ კატეგორიულ ფუნქტორ-არგუმენტულ სტრუქტურებს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

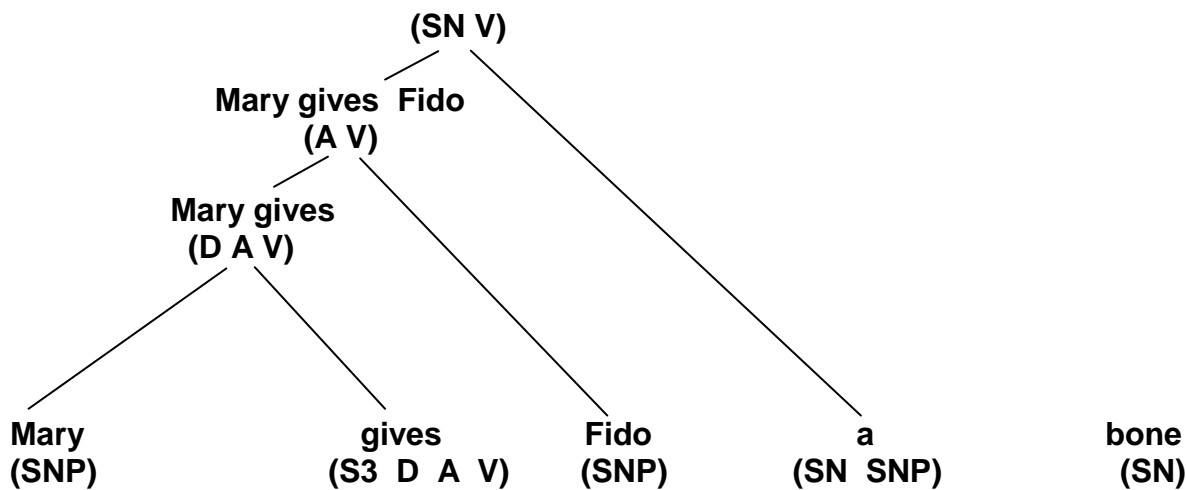
ეს გამარტივებული კატეგორიები სრულიად საკმარისია ვალენტობის ენათმეცნიერულად მეტად ღირებული პრინციპის მნიშვნელობის წარმოსაჩენად. ასე მაგალითად, **gives** სიტყვის კატეგორია (**S3 D A V**) აღნიშნავს ზმნას (**V**), რომელიც იღებს მესამე პირის მხოლოდობითი რიცხვის ნომინატივის (**nominative**) (**S3**), დატივის (**dative**) (**D**) და აკუზატივის (**accusative**) (**A**) ბრუნვებს არგუმენტებად. შესაბამისად, **a** დეტერმინერის კატეგორია (**SN SNP**) აღნიშნავს ფუნქტორს, რომელიც მოქმედებს მხოლოდობით ნაუნზე (**singular noun**) (**SN**) და მოქმედების შედეგად იძლევა მხოლოდობით 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 დროში საზოგადო 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 ჯგუფდება 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:

$$(SNP) (N D A V) \Rightarrow (D A V).$$

**SNP** კატეგორიალური სეგმენტი (შეესაბამება მხოლოდითი რიცხვის სახელურ ფრაზას), რომლითაც ხასიათდება სიტყვა **Mary** აბათილებს **gives** სიტყვის კატეგორიის პირველ **N** სეგმენტს. ენათმეცნიერული ერთ 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 ემატება წინადადების საწყისის კატეგორიას და ხსნის ახალ ვალენტურ ადგილს მხოლოდითი რიცხვის ნაუნისთვის.

**Mary gives Fido a bone** წინადადების საწყისისა და **bone** მომდევნო სიტყვის კომპოზირებისას ეს ახალი ვალენტური ადგილი ბათილდება შემდეგი სიტყვის (**SN**) კატეგორიით. შედეგი არის (**V**) კატეგორიის მქონე წინადადების საწყისი. ზმნასთან შეუვსებელი 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, როგორებიცაა **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 წინადადებად.

ამავდროულად, ისეთი გამოსახულებები, მაგალითად როგორებიცაა **gives Fido** და **gives Fido a bone**, არ არიან **LA**-გრამატიკის შუალედური გამოსახულებები, რადგან მათი გაგრძელება დასრულებულ, მართებულად აგებულ წინადადებად შეუძლებელია. არადა, **C-** და **PS-**გრამატიკებში ასეთი გამოსახულებები კანონიერ შუალედურ გამოსახულებებად 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 მიყოლებით; 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 სტრუქტურირებული სიის ფორმატს 10.3.2 და 10.4.1 განაწერებში აღწერილი ხელოვნური ენების უკვე ცნობილი გრამატიკული ანალიზისთვის.

### 10.5.3 მარცხნივ ჯგუფდებიანი 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 არამართო სრულ ინფორმაციას 10.5.2 დერივაციის დროში წრფივ ხისებრ ფორმატზე, არამედ დერივაციისას გამოყენებადი თითოეული მარცხნივ ჯგუფდებიანი კომბინაციისათვის დამატებითი შეტყობინების სახით გვაწვდის ამ კომბინირებისას გამოყენებული წესის სახელსაც, რომელიც ამავედროულად 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 შესაბამება 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 (აქ 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 გაბათილება.

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.

არაგრამატიკული შემავალის გრამატიკულად გამართული საწყისის **LA**-პარსერ ანალიზი, რომდენადაც ეს შესაძლებელია, გვაწვდის წინადადების ამ საწყისის ზუსტ გრამატიკულ აღწერას. ამას აქვს პრაქტიკული ღირებულება **LA**-გრამატიკის როგორც გასწორების, ისე გაუმჯობესების თვალსაზრისით. 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 არაგრამატიკული 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 ძვეთაჰი

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

## 10.2 ქვეთავი

1. ახსენით **LA**-გრამატიკის ალგებრული განსაზღვრება.
2. მარცხნივ ჯგუფდებიანი წესის რომელი ნაწილია გამოყენებული ნიმუშების დაწვევლებაში?
3. მარცხნივ ჯგუფდებიანი წესის რომელი ნაწილი იღებს მონაწილეობას სასრულ მდგომარეობიანი 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  $a^{2k} b^k$  ენის **LA**-გრამატიკა.
7. განსაზღვრეთ  $a^k b^{2k}$  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 ჯგუფდებიანი 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 იმარაქია

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 შეუზღუდავი 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 ექპონენციალური, ანუ **C3-** ქვეეიერარქები, რომლებიც ერთმანეთისგან განსხვავდებიან მხოლოდ მათი ორაზროვნების\_ხარისხით.

## 11.1 შეუზღუდავი LAG-ების წარმოქმნითი 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 **ამოხსნად (decidable)** ენას. ამგვარად, რეკურსულ ენათა კლასი წარმოადგენს ენათა იმ ყველაზე ფართო კლასს, რომლის წევრებსაც აქვთ სრულად სპეციფირებადი სტრუქტურა (**completely specifiable structure**).<sup>1</sup>

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

### 11.1.1 შეუზღუდავი LA-გრამატიკის წარმოქმნითი უნარი

შეუზღუდავი **LA**-გრამატიკა წარმოქმნისა და ამუშავებს\_რეკურსულ ენებს და მხოლოდ მათ.

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

ამ მოსახრების შემადგენელი ნაწილებიდან ორივე ჩამოყალიბებულია და დამტკიცებულია თეორემების სახით.<sup>2</sup>

### 11.1.2 თეორემა 1

შეუზღუდავი **LA**-გრამატიკა წარმოქმნის და ამუშავებს მხოლოდ რეკურსულ ენებს.

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

**კომბინაციური ბიჯი 1:** **STs** საწყისი მდგომარეობების სასრული სიმრავლე და პირველი **w<sub>1</sub>** სიტყვის სხვადასხვა შესაძლო წაკითხვები შედეგად იძლევა მართებულად აგებული გამოსახულებების  $WE_1 = \{(ss' \text{ } r p_s) / ss' \in (W^+ \times C^+)\}$  სასრულ სიმრავლეს.

**კომბინაციური ბიჯი n:** კომბინაციური ბიჯი **k-1 (k>1)**, წარმოქმნის მართებულად აგებული გამოსახულებების  $WE_k = \{(ss' \text{ } r p_i) / i \in n, ss' \in (W^+ \times C^+)\}$  და თითოეული **ss'** გამოსახულების სიგრძე არის **k** სასრულ სიმრავლეს. ამასთან, მომდევნო **w<sub>k+1</sub>** სიტყვა ხასიათდება წაკითხვების სასრული ოდენობით. აქედან გამომდინარე, **WE<sub>k</sub>** სასრული სიმრავლის და **w<sub>k+1</sub>** 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 წარმოქმნელი 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 **LA**-გრამატიკასა და არც მის პარსერს არანაირი შემაფერხებელი ვითარება არ ექნება. ეს იმას ნიშნავს, რომ **LA**-გრამატიკა აკმაყოფილებს 3-ე პირობას წარმოქმნითი გრამატიკის 9.5.2 პირობებიდან.

### 11.1.3 თეორემა 2

შეუზღუდავი **LA** გრამატიკა წარმოქმნის და ამუშავებს ყველა რეკურსულ ენას.

**დამტკიცება:**<sup>3</sup> ვთქვათ **W** არის ნებისმიერად აღებული **L** რეკურსული ენის ალფაბეტი. რადგან **L** რეკურსულია, არსებობს **L** ენის მახასიათებელი სრულად განსაზღვრული  $p: W^+ \rightarrow \{0,1\}$  რეკურსული ფუნქცია. ვთქვათ  $LAG^+$  არის შემდგენიარად განსაზღვრული **LA**-გრამატიკა:

$LAG^+$  სიტყვითი გამოსახულებების სიმრავლე არის **W**.  
კატეგორიული სემენტების სიმრავლე არის  $C =_{\text{def}} W U \{0,1\}$ .

<sup>2</sup> Col, თეორემები 1 და 2 გვ. 134.

<sup>3</sup> ეს დამტკიცება ეკუთვნის დიანა სკოტს.

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

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 შესამოწმებლად, უკვე წაკითხული არის თუ არა  $L$  ენის მართებულად აგებული გამოსახულება. ამ უკანასკნელ შემთხვევაში  $p$  ფუნქცია გამოიყენება იმ შემავალი კატეგორიების კონკეტენაციაზე, 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  $p$  მანასიათებელ ფუნქციაზეც. ეს იმას ნიშნავს, რომ ამგვარად განსაზღვრული  $LAG^L$  ამუშავებს და წარმოქმნის თავდაპირველად სრულიად ნებისმიერად ალებულ  $L$  რეკურსულ ენას.

რ.დ.გ.

$LAG^L$ -ის არადამასრულებელი კომბინაციური ბიჯები ემსახურება შემავალი გამოსახულებების შემადგენლების კატეგორიზებულ წაკითხვას. დამასრულებელ კომბინაციურ ბიჯში უკვე სრულად კატეგორიზებული გამოსახულება ანალიზდება ერთი ძალიან რთული კატეგორიალური ოპერაციით, რომელიც დასამუშავებელი ენის  $p$  მანასიათებელ 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 ძირეული, ანუ შეუზღუდავი ფორმალიზმები და მათი შესაბამისი ალგებრული განსაზღვრებები არ არიან ეკვივალენტური.

#### 11.2 $A$ -, $B$ - და $C-LAG$ -ების $LA$ -იერარქია

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

<sup>4</sup> ე. ი. თუ  $p$  შემავალი გამოსახულების  $(X \text{seg}_c)$  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 დამოუკიდებელი არიან და, როგორც წესი, მიყენებდნენ არიან წესებზე დაფუძნებულ ნებისმიერ გრამატიკულ ფორმალიზმზე.

**LA**-გრამატიკაში ალგორითმული ოდენობითი პარამეტრი დამოკიდებულია კატეგორიალურ **CO** 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 შემოვსაზღვროთ ამ წესების გამოყენებისას გამოყენებადი ალგორითმული ბიჯების მაქსიმალური ოდენობა.

**R2:** შეზღუდვები ორაზროვნების ხარისხზე იმ მიზნით, რომ შემოვსაზღვროთ 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 სპეციფიცირება;

**R1.2:** კატეგორიალური ოპერაციების განსაზღვრებებში გამოყენებული მოდელების სპეციფიცირება.

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  $k \cdot n$  –ნამრავლით (**n** არის **E** გამოსახულების სიგრძე, **k** არის კონსტანტა).

**B-LAG**-ების მიერ წარმოქმნილი ენების კლასი ემთხვევა **PS**-გრამატიკულ იერარქიაში კონტექსტზე დამოკიდებული ენების კლასს. დამტკიცება<sup>5</sup> 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

<sup>5</sup> **CoL**, თეორემა 5, გვ. 142.

განპირობებულია იმით, რომ  $CS \subset REC$ .<sup>6</sup>

**C-LAG**-ით აღნიშნულია **B-LAG**-ების ის ქვეკლასი, რომელიც მოიცავს **R1.2** სახის შემზღუდავი პირობით. კერძოდ, **C-LAG**-ებში  $co_i$  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 შემოიზღუდებოდა 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 სქემატური წესები კონსტანტური კატეგორიალური ოპერაციებით

- $r_i: (\text{seg}_1 \dots \text{seg}_k X) \text{ cat}_2 \Rightarrow \text{cat}_3 \text{ rp}_i$
- $r_i: (X \text{ seg}_1 \dots \text{seg}_k) \text{ cat}_2 \Rightarrow \text{cat}_3 \text{ rp}_i$
- $r_i: (\text{seg}_1 \dots \text{seg}_m X \text{ seg}_{m-1} \dots \text{seg}_k) \text{ cat}_2 \Rightarrow \text{cat}_3 \text{ rp}_i$

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 **seg<sub>1</sub>...seg<sub>k</sub>** 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 სქემატური წესი კატეგორიულ ოპერაციათა არაკონსტანტური ოდენობით

$$r_i: (X \text{ seg}_1 \dots \text{seg}_k Y) \text{ cat}_2 \Rightarrow \text{cat}_3 \text{ rp}_i$$

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 ჰქონდეთ ნებისმიერი სიგრძე, ამ წესების გამოყენებისას კატეგორიული ოპერაციის შესასრულებლად საჭირო გამოთვლების ოდენობა დამოკიდებული ხდება **SS**-კატეგორიის მთლიან სიგრძეზე.

<sup>6</sup> თეორემა ჯ. ე. ჰოპკროფტი და ჯ. დ. ულიმანი 1979, თეორემა 9.8 გვ. 228. **CS** აღნიშნავს კონტექსტზე დამოკიდებული ენების კლასს, ხოლო **REC** აღნიშნავს რეკურსულ ენათა კლასს.

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

### 11.2.7 C-LAG-მბის კლასის განსაზღვრება

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

**LA** გრამატიკული კლასები ღღემდე მიიჩნევა, რომ შესდგება შემდეგი იერარქიული კლასებისაგან:

### 11.2.8 A-LAG-მბის, B-LAG-მბისა და C-LAG-მბის 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, დამტკიცებულია<sup>8</sup> იმის საფუძველზე, რომ **C-LAG**-ების კატეგორიული ოპერაციები მოდიფიცირებას უკეთებენ მხოლოდ **cat<sub>1</sub>** 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 დამუშავებადია და წარმოქმნადია იმ **C-LAG**-ების მიერ, რომელთა კატეგორიების სივრძე შემოსაზღვრულია აბსოლიტური **k** მუდმივით.<sup>9</sup>

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

<sup>7</sup> ამ გრამატიკებს მხოლოდ ეს სასრული მუდმივი განსაზღვრავს ერთმანეთისაგან.

<sup>8</sup> **8 CoL**, თეორემა 4, გვ. 138. კონტექსტისგან თავისუფალი **C-LAG**-ები, მოკლეთ **CF-LAG**-ები შედგება მხოლოდ შემდეგი ფორმის წესებისგან:  $r_i(aX)(b) \Rightarrow (\alpha X) r_i$  (აქ,  $a, b \in C$  და  $\alpha \in C^+$ ). ამგვარად, შეზღუდვები **ss** და **ss'** კატეგორიებზე შეესაბამება **PDA**-ს მუშაობის პრინციპს, რომელსაც შეუძლია ერთდროულად ამოიჭლოს 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-LAG $\sigma$**  შეიძლება აიგოს ისე, რომ **L(M)=L( $\sigma$ )**.

აქედან გამომდინარეობს, რომ **CF  $\subseteq$  C<sub>cf</sub>**.

2. ყოველი **CF-LAG  $\sigma$** -ისთვის, **PDA M** შეიძლება აიგოს ისე, რომ **L( $\sigma$ )=L(M)**.

აქედან გამომდინარეობს, რომ **C<sub>cf</sub>  $\subseteq$  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 გვექონდეს კონსტრუქცია იმ პროცედურის რომელიც გარდაქმნის გამონათქვამებს წესებად და წესებს გამონათქვამებად. ეს დიდი, მაგრამ არც ძალიან რთული ამოცანა. – შევნიშნოთ, რომ  $\epsilon$ -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  $\epsilon$ -გადაადგილებების გარეშე მომუშავე **PDA** (ჰარისონი 1978, თეორემა 5.5.1) და ყოველი  $\epsilon$ -თავისუფალი **PDA** -ისთვის შესაძლებელია აიგოს **CF-LAG**.

<sup>9</sup> **CoL**, თეორემა 3, გვ. 138.

<sup>10</sup> განსაზღვრული გზა **LA**-გრამატიკის წარმოქმნითი უნარის განსაზღვრის შედგება – უკიდურეს შემთხვევაში

### 11.3 ორაზროვნება LA-გრამატიკაში

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

#### 11.3.1 წესთა გამოყენების რიცხვის განმსაზღვრელი ფაქტორები

წესთა გამოყენების რიცხვი LA-გამოყვანაში დამოკიდებულია:

1. შემაჯავლი მოცემულობის სიგრძეზე;
2. წესთა პაკეტის იმ წესების ოდენობაზე, რომლებიც გამოიყენება ამ წესის კომბინაციურ ბიჯში შემაჯავლი წყვილის საანალიზოდ (წასაკითხად);
3. თითოეულ კომბინაციურ ბიჯში არსებულ წაკითხვათა<sup>11</sup> ოდენობაზე.

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

ფაქტორი 2 გრამატიკაზე დამოკიდებული კონსტანტაა. მაგალითად: თუ **LAG**-ების ყველაზე დიდი წესთა პაკეტი შეიცავს 5 წესს, მაშინ მაქსიმალური რიცხვი წესთა გამოყენების ცალსახა წარმოებაში იქნება არაუმეტეს 5·n ნამრავლისა, სადაც **n** შემაჯავლი გამოსახულების სიგრძეა.

მხოლოდ 3-ე ფაქტორს შეუძლია გამოიყვანოს წესთა გამოყენების მთლიანი რიცხვი წრფივი ყოფაქცევისგან. მოცემულ მარცხნივ ჯგუფდებად კომპოზიციაში დამატებითი კითხვა იწყება მხოლოდ მაშინ, როდესაც მიმდინარე წესთა პაკეტში არის ერთზე მეტი ისეთი წესი, რომელსაც შეუძლია წარმატებულად იმუშავოს შემაჯავლი მოცემულობაზე. ის, შემაჯავლი მოცემულობაზე ერთზე მეტს წესს შეუძლია თუ არა იყოს წარმატებული, დამოკიდებულია ამ წესის შემაჯავლი პირობებზე.

---

თეორიულად – 10.2.1 ალგებრული განსაზღვრების მე-4 პუნქტის შეცვლით. მაგალითად, თუ კატეგორიულურ ოპერაციებად დაშვებული იქნებოდა **ნაწილობრივ რეკურსული ფუნქციები**, მაშინ LA-გრამატიკა წარმოქმნიდა რეკურსულად გადათვლად ენებს, რაც LA-გრამატიკას გახდიდა **PS** გრამატიკის სუსტად ეკვივალენტურს. მეორეს მხრივ, თუ კატეგორიულურ ოპერაციებად დაშვებული იქნებოდა **პრიმიტიული რეკურსული ფუნქციები**, მაშინ თეორემა 2-ის ანალოგიურად შეგვეძლო გვეჩვენებია, 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 გამართლებულია: კატეგორიულურ ოპერაციების გადაზრება ხან ნაწილობრივ რეკურსულ, ხან სრულად რეკურსულ, ხანაც პრიმიტიულად რეკურსულ ფუნქციებად ძალიან უხეში მეთოდია. მეორე მიზეზი ისაა, რომ პრაქტიკული თვალსაზრისებიდან გამომდინარე ასე მიღებული ენობრივი კლასები ძალიან დიდია: პრიმიტიული რეკურსული ფუნქციები საკუთრივ ქვესიმრავლეა სრულად რეკურსული ფუნქციების. ამასთან, პრიმიტიული რეკურსული ფუნქციები საკუთრივ ქვესიმრავლის სახით შეიცავენ კონტექსტზე დამოკიდებული ენების მთელ კლასს. მესამე მიზეზი: 10.2.1 განაწერში კატეგორიულურ ოპერაციები განსაზღვრულა როგორც სრულად რეკურსული ფუნქციები და ეს მართლდება იმ მნიშვნელოვანი საფუძვლის გათვალისწინებით, რომ LA გრამატიკას აქვს მაქსიმალური წარქმნითი უნარი, იქამდე სანამ ამოხსნაღია.

<sup>11</sup> სიმარტივისთვის აქ ორაზროვნების მხოლოდ სინტაქსური მიზეზებია განხილული. ლექსიკური ორაზროვნება, რომელიც ჩნდება სიტყვის მრავალნაირი გაგების გამო, ფორმალურ ენებში დიდი ნაწილით აღმოფხვრილ იქნა, მაგრამ ამას ვერ ავიცილებთ ბუნებრივი ენის LA-გრამატიკულ ანალიზში. გამოთვლით სირთულებზე ლექსიკური ორაზროვნების შესაძლო ზეგავლენები 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 ბრამატიკების 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.<sup>12</sup>

არაორაზროვანი **C-LAG**-ების მაგალითებია 10.2.2-ში **a<sup>k</sup>b<sup>k</sup>** და 10.3.3-ში **a<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** 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-ბრამატიკების განსაზღვრება

<sup>12</sup> **LA**-გრამატიკა ლექსიკურად ორაზროვანია თუ მისი ლექსიკონი შეიცავს მინიმუმ ორ ისეთ ანალიზებად სიტყვას, რომელთა ზედაპირები, ანუ ენობრივი გამოსახულებები ერთი და იგივეა. არალინგვისტური მაგალითი ლექსიკური ორაზროვნების არის პროპოზიციულ წინადადებათა ალგებრა. 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 ასალი საწყისის წარმოქმნით, ლექსიკურ ორაზროვნებას იწვევს მომდევნო სიტყვის ერთზე მეტი წაკითხვები. – **LA**-გრამატიკა შეიძლება ერთდროულად ხასიათდებოდეს სინტაქსური და ლექსიკური ორაზროვნებით. ამასთან, სინტაქსური ორაზროვნება შესაძლებელია გააზრდეს როგორც ლექსიკური ორაზროვნება, და პირიქით. (იხ. ჰაუზერი 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 წევრებს, 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 საჭაერო 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 მანამდე 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 **ბაღის 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 ორახროვნებების განრჩევა არ ახდენს არსებით გავლენას **LA**-გრამატიკის სირთულზე და ეს ძირითადად კეთდება ლინგვისტური ინტერესების გათვალისწინებით. ±რეკურსული ორახროვნებების განრჩევა, მეორე მხრივ, გადაწყვეტს გავლენას ახდენს ანალიზის სირთულეზე, რადგან მტკიცდება, რომ არარეკურსულ ორახროვნებებთან **LA**-გრამატიკაში ყოველ კომბინაციურ ბიჯზე ერთი რომელიმე წესის გამოყენების მაქსიმალური რიცხვი ლიმიტირდება გრამატიკაზე დამოკიდებული კონსტანტით. (იხ.11.3.7, თეორემა 3).

ორახროვნება არის რეკურსული, თუ ის წარმოიქმნება წესთა გამოყენების რეკურსულ კვანძში. სხვა სიტყვებით: გარკვეულ (**cat, rp<sub>i</sub>**) მდგომარეობას გარკვეულად განსაზღვრული შემდეგი სიტყვისთვის აქვს რამოდენიმე გაგრძელება, თანაც ისეთი, რომ ერთი, ან რამოდენიმე მათგანი ამ გაგრძელებათა შედეგად კვლავ ბრუნდება თავდაპირველ მდგომარეობაში, რითაც წარმოიქმნება ორახროვანი განხილვის კვლავგამეორების იძულება. +რეკურსული ორახროვნების მაგალითებია **SubsetSum** (იხ. 11.5.2) და **WW<sup>R</sup>** და **WW** ენებისთვის 11.5.4 და 11.5.6 განაწერებში აღწერილი **C-LAG**-ები, რომელთაგან პირველი ამავდროულად +გლობალურია, ხოლო ბოლო – -გლობალური.

ორახროვნება –რეკურსულია, თუ არცერთი ორახროვნული გაგრძელება არ ბრუნდება იმ მდგომარეობაში, რომელმაც წარმოშვა იგი. –რეკურსული ორახროვნების მაგალითებია **a<sup>k</sup>b<sup>k</sup>c<sup>m</sup>d<sup>m</sup>** **Ua<sup>k</sup>b<sup>m</sup>c<sup>m</sup>d<sup>k</sup>** ენისთვის 11.5.2 განაწერში აღწერილი **C-LAG**, რომელიც ორახროვანია აგრეთვე +გლობალური თვალსაზრისით, და ბუნებრივი ენის **C-LAG**-ები (იხ. თავი 17, 18), რომლებიც არიან ორახროვნები როგორც +გლობალური, ისე –გლობალური თვალსაზრისით.

### 11.3.7 თეორემა 3

მხოლოდ –რეკურსულ ორახროვნებებთან **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 ორახროვნები, 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 ტრანსფარენტული ტიპის. ისტორიულად, **PS**-გრამატიკული ტიპებისა და მათი ენობრივი კლასების სირთულეები ისაზღვრებოდა არა უშუალოდ გრამატიკული თვისებებით, არამედ ეკვივალენტური აბსტრაქტული

ავტომატების მეშვეობით.<sup>13</sup>

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 აღიწერებიან "პრიმიტიული ოპერაციების" ტერმინებში. მათ გამოსახეს თავიანთი 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 რთული ოპერაციები, რომლის სირთულე დამოუკიდებელია გრამატიკის და შემაგალი სტრიქონის ზომაზე.

ჟ. ერლეი 1970, გვ. 100

ამგვარად, მტკიცებულება "კონტექსტისგან თავისუფალი PS-გრამატიკები აპარსერებს  $n^3$ -ში" პირდაპირ არ ახასიათებს კონტექსტისგან თავისუფალ PS-გრამატიკებს, არამედ გარკვეულ პარსერ ალგორითმს, რომელიც ამუშავებს PS-გრამატიკულ კლასებს, როგორც შემაგალი მოცემულობებს. შესაბამისად, როცა ვალიანტმა (Valiant) 1975 წელს შეძლო კონტექსტისგან თავისუფალი ენების სირთულე  $n^3$ -დან  $n^{2.8}$ -მდე შეემცირებინა, ეს არ იყო PS-გრამატიკის გაუმჯობესების შედეგი, არამედ ეს იყო შედეგი ვალიანტის იმ საკმაოდ გაუმჯობესებული მანალიზებელი ალგორითმის, რომელსაც იგი აბსტრაქტულ ავტომატად განიხილავდა.

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

#### 11.4.1 C-LAG-მბის კლასის პრიმიტიული ოპერაცია

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 კატეგორიული ოპერაციების გამოთვლითი მოთხოვნები კონსტანტით არის შემოსაზღვრული. ამასთან, ის, რომ C-LAG-ების კლასის პრიმიტიული ოპერაცია ამავდროულად

<sup>13</sup> აბსტრაქტული ავტომატები შედგებიან ისეთი კომპონენტებისაგან როგორებიცაა წამკითხავ/დამწერი-თავი, ჩაწერისგან-დაცული შემაგალი ლენტი, მომუშავე ლენტები, ლენტის გასწვრივ უჯრდან უჯრაზე გადაადგილებადი წამკითხავ/დამწერი-თავის, წაკითხვა ან წაშლა უჯრაში მყოფი მოცემულობის და ა.შ.. აბსტრაქტული ავტომატების კლასიკური მაგალითებია ტიურინგის მანქანები (TM), წრფივად შემოსაზღვრული ავტომატები (LBA), და ძირს მბიძგი ავტომატები (PDA), და სასრულ მდგომარეობიანი ავტომატები (FSA). არსებობს კიდევ მრავალი სხვა აბსტრაქტული ავტომატები, თითოეული განსაზღვრულია იმ მიზნით, რომ გადაიჭრას სხვადასხვა განსაკუთრებული ხასიათის 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.<sup>14</sup>

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 უჯრების ის რაოდენობა, რასაც ყველაზე გრძელი დასაშვები გზა საჭიროებს) საკითხი.

**NTIME** და **NSPACE** სირთულები 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 ეტყვიან 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 არსებობდეს მხოლოდ ერთი გზა, **DTIME** და **DSPACE** სირთულები უნდა ისაზღვრობდეს 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**

<sup>14</sup> **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.<sup>15</sup> ანალოგიური შედეგია ძალაში ტიურინგის მანქანებისთვისაც (**Turing machine (TM)**).<sup>16</sup>

ძირსმბიძგი ავტომატების (**pushdown automata (PDA)**), მეორე მხრივ, დეტერმინირებულ და არადეტერმინირებულ ვერსიებს არ აქვთ ეკვივალენტური წარმოქმნითი უნარი.<sup>17</sup> მაგალითად,  $WW^R$  ენა, სადაც  $\{n=R\}$  მუშავდება არადეტერმინირებული  $PDA$  ავტომატით, მაგრამ არ მუშავდება დეტერმინირებულით. და ბოლოს, 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  $\subset$  NPDA**  
**DLBA ? NLBA**  
**DTM = NTM**

ავტომატები, რომელთა დეტერმინირებულ და არადეტერმინირებულ ვერსიებს აქვთ ტოლი წარმოქმნითი უნარი, წარმოშობენ 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 განსაზღვრებები, განსაკუთრებით კი **NTIME** და **DTIME** სირთულები, 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 – ვეყრდნობით რა მარცხნივ ჯგუფდებად დერივაციულ სტრუქტურებს – ნებისმიერი **n** სიგრძის შემავალი ანალიზდება **n-1** კომბინაციურ ბიჯში. ამასთან, წესთა გამოყენება ყოველ კომბინაციურ ბიჯში საჭიროებს დროს, რომელიც გამოთვლისთვის გამოყოფილი **C** კონსტანტის ტოლია. სხვა სიტყვებით: არაორაზროვანი **C-LAG**-ებს აქვთ დეტერმინირებული წრფივი დროითი სირთულე (**deterministic linear time complexity**),<sup>18</sup> მაშინ როდესაც ორაზროვანი **C-LAG**-ებს აქვთ არადეტერმინირებული წრფივი დროითი სირთულე (**nondeterministic linear time complexity**).

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

<sup>16</sup> ჯ. ე. ჰოპკროფტი და ჯ. დ. ულიმანი 1979, გვ. 164, თეორემა 7.3

<sup>17</sup> ჯ. ე. ჰოპკროფტი და ჯ. დ. ულიმანი 1979, გვ. 113.

<sup>18</sup> მკაცრად თუ ვიტყვით, არაორაზროვანი **C-LAG**-ები (**unambiguous C-LAGs**) უფრო კარგი სირთულით ხასიათდებიან, ვიდრე 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** მიყენებადია მხოლოდ არაორაზროვან **LA**-გრამატიკებზე, ხოლო **NTIME** მხოლოდ ორაზროვან **LA**-გრამატიკებზე. ამასთან, **NTIME** არ იძლევა ნებისმიერად შემოსული შემავალი გამოსახულების გასაანალიზებლად 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 ბუნებრივიც.

### 11.5 C1-,C2-, და C3-LAG-მზის ქვემრავლობა

**A-** და **B-LAG**-ებთან შედარებით **C-LAG**-ები წარმოადგენენ **LAG**-ების ყველაზე მეტად შეზღუდულ კლასს, რომლებიც წარმოქმნიან ენების ყველაზე პატარა **LA**-კლასს. კონტექსტისგან თავისუფალ ენებთან შედარებით (რომლებიც საკუთრივ შედიან **C**-ენებში) **C**-ენების კლასი საკმაოდ ფართოა. აქედან გამომდინარე, თეორიული და პრაქტიკული თვალსაზრისებით მნიშვნელოვანია **C-LAG**-ების ქვეიერარქიის 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 კონტექსტისგან თავისუფალ ენებს, რომლებიც გამოცნობადია **E**-გადანაცვლების ოპერაციის არმქონე **DPDA** ავტომატის მიერ, აგრეთვე – რეკურსული ორაზროვნებით დახასიათებულ კონტექსტისგან თავისუფალ ენებს, მაგ:  $a^k b^k c^k d^k m^k$  ენას, და აგრეთვე კონტექსტზე დამოკიდებულ ენებს, მაგ:  $a^k b^k c^k$ ,  $a^k b^k c^k d^k e^k$ ,  $\{a^k b^k c^k\}^*$ ,  $L^k_{square}$ ,  $L^k_{hast}$ ,  $(a^2)^i$ ,  $a^k b^m c^{km}$  და  $a^i$  ენებს. ამასთან, ხაზგასასმელია, რომ ამ უკანასკნელთაგან ბოლო ინდექსირებული ენაც კი არა არის.<sup>19</sup>

<sup>19</sup>  $a^k b^k c^k d^k m^k$  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 შემდგენიარად ისაზღვრება

### 11.5.1 $a^{2^i}$ ენის კონტექსტზე 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 ამავ ენის აქ აღწერილი **LA** გრამატიკის შედარება ამკარად გვიჩვენებს **LA** გრამატიკის ფორმალურ და კონცეპტუალურ სიმარტივეს.<sup>20</sup>

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.<sup>21</sup>

### 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 პარსირება ხდება წრფივ დროში, რადგან ორაზროვანი გაგრძელებები არ ხდებიან რეკურსული ციკლის ნაწილები:  $rp_2$  და  $rp_5$  არ შეიცავენ  $r_1$  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;  $a^k b^m c^k m$  ენი – ჰუსერის 1992 წელს გამოშვებულ წიგნში გვ. 296;  $a^i$  ენის – 11.5.1-ში;  $a^i$  ენის – ჰუსერის 1992 წელს გამოშვებულ წიგნში (იხ. მინიშნება გვ. 296, შენიშვნა 13).

<sup>20</sup> ჯ. ე. ჰოპკროფტმა და ჯ. დ. ულმანმა 1979-ში წარმოადგინეს კანონიკური კონტექსტზე დამოკიდებული  $a^{2^i}$  ენის **PS**-გრამატიკა (იხ. გვ. 224) და მისი შეუზღუდავი **PS**-გრამატიკული ვერსია (იხ. გვ. 220).

<sup>21</sup> იხ. ჯ. ე. ჰოპკროფტი 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.<sup>22</sup>

**C2-LAG**-ები არიან **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))

+რეკურსიული 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  $L^k_{\text{hast}}$ , და, ამის გარდა, ისეთ კონტექსტზე დამოკიდებულ ენებს, როგორებიცაა:  $WW$ ,  $W^{k \geq 3}$ ,  $\{WWW\}^*$ , და  $W_1W_2W_1^RW_2^R$ .<sup>23</sup>

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

### 11.5.4 კონტექსტისგან თავისუფალი $WW^R$ ენის C2-LAG

$$\begin{aligned}
 LX &=_{\text{def}} \{[a(a)], [b(b)], [c(c)], [d(d)] \dots\} \\
 ST_S &=_{\text{def}} \{[(\text{seg}_c)\{r_1, r_2\}]\}, \text{ სადა } c \in \{a, b, c, d, \dots\} \\
 r_1 : (X) (\text{seg}_c) &\Rightarrow (\text{seg}_c X)\{r_1, r_2\} \\
 r_2 : (\text{seg}_c X) (\text{seg}_c) &\Rightarrow (X)\{r_2\} \\
 ST_F &=_{\text{def}} \{[\varepsilon r p_2]\}
 \end{aligned}$$

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.ი. ის, რომელიც  $r_1$  წესის განმეორებითი გამოყენებებით 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.

<sup>22</sup> ცხადი გამოყვანა მოცემულია **CoL**-ში გვ. 154 ფფ.  
<sup>23</sup> **C2-LAG** განსაზღვრულია  $WW^R$  ენისათვის 11.5.4-ში,  $L^{\infty}_{\text{hast}}$  ენისათვის – შტუბერტი 1993, გვ. 16,  $WW$  ენისათვის – 11.5.6-ში,  $WWW$  ენისათვის – **CoL**-ში, გვ. 216-, და  $W_1W_2W_1^RW_2^R$  ენისათვის – განმარტებულია 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 დაწუნებულია (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 კონტექსტზე დამოკიდებული WW-ენის C2-LAG

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

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<sup>R</sup> ენაში r1 წესი შედეგ კატეგორიაში იძლევა (segc X) ნიმუშს. კონტექსტზე დამოკიდებულ WW და კონტექსტისგან თავისუფალი WW<sup>R</sup> ენებში ყველაზე ცუდ შემთხვევებს ერთი და იგივე შემავალი გამოსახულებები იძლევა: ორივეს შესაბამისი C2-LAG არის SR-ორაზროვანი, და ორივე არის n<sup>2</sup> სირთულის. WW და WW<sup>R</sup> ენებში ანალიზის ყველაზე ცუდ შემთხვევებში წაკითხვების n<sup>2</sup>-ული ზრდა თვალსაჩინოდ არის წარმოდგენილი 11.5.5 პუნქტში.

არადეტერმინირებული კონტექსტზე დამოკიდებული W<sub>1</sub>W<sub>2</sub>W<sup>R</sup><sub>1</sub>W<sup>R</sup><sub>2</sub> ენა n<sup>3</sup> სირთულის C2-LAG ტიპისაა (შედეგი ეკუთვნის ბ. შტუბერტს).

### 11.5.7 კონტექსტზე დამოკიდებული W<sub>1</sub>W<sub>2</sub>W<sup>R</sup><sub>1</sub>W<sup>R</sup><sub>2</sub> ენის C2-LAG

$$\begin{aligned}
 LX &=_{\text{def}} \{[a(a)], [b(b)]\} \\
 ST_S &=_{\text{def}} \{[(\text{seg}_c)\{r_{1a}\}], [(\text{seg}_c)\{r_{1b}\}]\}, \text{ სადა } \text{seg}_c, \text{seg}_d \in \{a, b\} \\
 r_{1a}: (\text{seg}_c) & \quad (\text{seg}_d) \Rightarrow (\# \text{seg}_c \text{ seg}_d) \quad \{r_2, r_3\} \\
 r_{1b}: (\text{seg}_c) & \quad (\text{seg}_d) \Rightarrow (\text{seg}_d \# \text{seg}_c) \quad \{r_3, r_4\} \\
 r_2: (X) & \quad (\text{seg}_c) \Rightarrow (X \text{ seg}_c) \quad \{r_2, r_3\} \\
 r_3: (X) & \quad (\text{seg}_c) \Rightarrow (\text{seg}_c X) \quad \{r_3, r_4\}
 \end{aligned}$$

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

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 ოდენობაზე. ამგვარად, **C2-LAG** ერთი **SR**-რეკურსიული ორაზროვნებით შემძლეა  $n^2$  დროში პარსირების, ხოლო **C2-LAG** ორი **SR** რეკურსიული ორაზროვნებით შემძლეა  $n^3$  დროში 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 დროში 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.<sup>24</sup>

**SubsetSum** ენა მიიჩნევა თვისობრივად რთულ ენად, ანუ ისეთ ენად, რომელიც დასამუშავებლად გამოყენებული ალგორითმის მიუხედავად დასამუშავებლად მოითხოვს ექსპონენციალურ დროს. მისი გამოსახულებები  $y\#a_1\#a_2\#a_3 \dots \#a_n\#$  განსაზღვრულნი 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 ენის C3-LAG

$LX =_{\text{def}} \{[0(0)], [1(1)], [\#(\#)]\}$		
$ST_S =_{\text{def}} \{[\text{seg}_c]\{r_1, r_2\}\}$ , სადაც $\text{seg}_c \in \{0, 1\}$		
$\text{seg}_c \in \{0, 1\}$		

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

<sup>24</sup>  $L_{no}$  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 ჰაუსერთან 1992, მინიშნება 19.

ეს **C3-LAG** (აგებულია დ. ეპლიგეიტის (**D. Applegate**) მიერ) გადაწერს **y**-ს კატეგორიაში. ამის შემდეგ +რეკურსიული ორაზროვნება წარმოიქმნება **y**-ისგან თითოეული **a<sub>i</sub>**-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 განისაზღვრება წესთა სისტემის მზარდი შეზღუდულობის ტერმინებში.

**C1-LAG** ენათა კლასი **C2**-ენების 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**-ენების 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 თვალსაზრისით არ იზღუდება.

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

### 11.5.9 შეზღუდვების ტიპები 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 (მოდულების) 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 იერარქია ფორმალური ენების

ტიპი	LAG-ტიპი	ენები	სირთულე
ტიპი	<b>C1-LAG</b>	<b>C1-ენები</b>	წრფივი
ტიპი	<b>C2-LAG</b>	<b>C2-ენები</b>	პოლინომიალური
ტიპი	<b>C3-LAG</b>	<b>C3-ენები</b>	ექსპონენციალური
ტიპი	<b>B-LAG</b>	<b>B-ენები</b>	ექსპონენციალური
ტიპი	<b>A-LAG</b>	<b>A-ენები</b>	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

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

## ქვეთავი 11.2

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

## ქვეთავი 11.3

1. ახსენით  $\pm$ გლობალური ორაზროვნების ცნებები. რა კავშირშია ისინი ავტომატთა თეორიის  $\pm$ დეტერმინირებული დერივაციის ცნებებთან და ორაზროვნების ცნებებთან **PS**-გრამატიკაში.
2. რას განსაზღვრავს წესთა გამოყენებების რიცხვი **LA**-დერივაციაში?
3. აღწერეთ შემაჯავლი მოცემულობის სამი განსხვავებული პირობა.
4. განსაზღვრეთ ორაზროვანი და არაორაზროვანი **LA**-გრამატიკები.
5. განსაზღვრეთ რეკურსული ორაზროვნების ცნებები **LA**-გრამატიკაში.
6. 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

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

## ქვეთავი 11.5

1. აღწერეთ **C1-**, **C2-**, და **C3-** **LAG**-ების ქვეირარქიები. რა კავშირებია სირთულესა და ორაზროვნებას შორის?
2. 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?
3. ახსენით ერთხელ დაბრუნების პრინციპი.

4. შეადარეთ **WW** და **WW<sup>R</sup>** 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 ნეიტევისტურ თეორიას შორის, ერთი მხრივ, და **LA**-გრამატიკასა და **SLIM** თეორიას შორის, მეორე მხრივ. შესაძლებელია თუ არა, რომ **PS**-გრამატიკა გამოყენებულ იქნას როგორც **SLIM** თეორიის სინტაქსური კომპონენტი? შესაძლებელია თუ არა, რომ **LA**-გრამატიკა გამოყენებულ იქნას როგორც ნეიტევისტური თეორიის სინტაქსური კომპონენტი?

## 12 ტახტი

### LA- და PS- იერარქიების ურთიერთმიმართება

გენერაციული გრამატიკის<sup>1</sup> ორი არაეკვივალენტური ფორმალისმის 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 შექდება **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 შექდება 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

<sup>1</sup> 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 როგორც არის წესიერი (3 ტიპის) **PS**-გრამატიკა – ეს ფენომენი შეუძლებელია აიხსნას მხოლოდ მისივე საშუალებებით. თუმცა, თუ არჩეული ფორმალიზმი ძლიერ დიდია – მაგალითად როგორცაა 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 ‘დიდი’ ფორმალიზმი, მით უფრო ‘მცირე’ ჯდება იგი მათემატიკური სირთულის თვალსაზრისით. კომპიუტერული თვალსაზრისებით კი, ეს სიძვირე ისახება ენობრივი გამოსახულების პარსირებისთვის საჭირო დროზე და მეხსიერებაზე.

ბუნებრივი ენის აღსაწერად ესა თუ ის ფორმალიზმი შეიძლება იყოს არა უბრალოდ ძლიერ დიდი, ან ძლიერ პატარა, არამედ ერთდროულად 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 სიმულარული გამოსახულებების შემოტანის აუცილებლობითაც დასტურდება. თუმცა, ამავე დროს, ის არის ძლიერ დიდი, რასაც აშკარად გვიდასტურებს კონტექსტისგან თავისუფალი **HCFL**<sup>2</sup> 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 ძალისხმევისა **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. მიუხედავად ამისა, ის, რომ ბიდირექციული **C**-გრამატიკა სუსტად ეკვივალენტურია კონტექსტისგან თავისუფალი **PS**-გრამატიკის (იხ. 9.2 ქვეთავი), უკვე დიდი ხანია 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

**PS**- გრამატიკა

არაამოხსნადი

რეკურსულად გადათვლადი ენები

<sup>2</sup> 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, ამავედროულად, არიან განსხვავებულ **LA-** გრამატიკულ კლასებში, და პირიქით. მაგალითად,  $a^k b^k$  და  $WW^R$  ენები **PS-** გრამატიკის ერთი და იგივე კლასში არიან (კონტექსტისგან 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 რთული, არამედ 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 დაიძებნოს ისეთი ახალი ძირეული 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,  $a^k b^k$  ამავედროულად 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 ამავედროულად არის **B-LAG** ტიპის, ანუ კონტექსტზე დამოკიდებული ტიპის ენაც.

## 12.2 ქვესიმრავლეობის მიმართულება ამ ორ იმპარაქტიაში

**X** ენათა კლასი არის **Y** ენათა კლასის ქვესიმრავლე, ანუ ქვეკლასი (ფორმალურად,  $X \subseteq Y$ ), თუ **X** კლასის ნებისმიერი ენა ამავედროულად 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  $\subset$  რეკ. გადათვლადი ენები

ასეთი ქვესიმრავლეობის მიმართება განპირობებულია **PS**-გრამატიკის გადამწერი წესების მზარდი შეზღუდვებით (იხ. 8.2), და იგი საკუთრივი ქვესიმრავლეობის მიმართებად იწოდება.<sup>3</sup>

### 12.2.2 ქვესიმრავლეობის მიმართება LA-იერარქიაში

**C1** ენები  $\subseteq$  **C2** ენები  $\subseteq$  **C3** ენები  $\subseteq$  **B** ენები  $\subset$  **A** ენები

ეს ქვესიმრავლეობის მიმართებაც მზარდი შეზღუდვებითაა განპირობებული. თუმცა, თუ **B**-ენების კლასი **A**-ენების<sup>4</sup> კლასის საკუთრივი ქვესიმრავლეა, სხვა კლასებს შორის საკუთრივი ქვესიმრავლეობის მიმართების არსებობა შეიძლება მხოლოდ ვივარაუდოთ. საკითხი იმის თაობაზე არის თუ არა **C2** ენების კლასი **C3** ენების კლასის საკუთრივი ქვესიმრავლე, დაიყვანება კლასიკური 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  $\cup_{i \geq 1} DTIME(n^i)$ , რომელიც ჩვენთან **P**-თი აღინიშნება. ინტუიციურად მეტად მიმზიდველია ის, რომ **P** არის იმ პრობლემათა კლასი, რომლებიც შეიძლება იქნენ ეფექტურად გადაჭრილნი. თუმცა ვინმე შეიძლება არ დაგვეთანხმოს იმაში, რომ  $n^{57}$  ბიჯიანი ალგორითმი ეფექტურია. 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

<sup>3</sup> იერარქიული ლემა (Hopcroft & Ullman 1979, გვ. 228).

<sup>4</sup> (იხ. Hopcroft & Ullman 1979, გვ. 228, თეორემა 9.8.) **A-LAG** წარმოქმნის რეკურსულ ენებს, მაშინ როდესაც, როგორც ეს ნაჩვენებია 11.1 და 11.2-ში, **A-LAG** წარმოქმნის 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 შესრულებადობის<sup>5</sup> (**Boolean SATisfiability**) პრობლემა. განვიხილოთ შემდეგი ბულისეული გამოსახულება.

### 12.2.3 3SAT ენის სწორად ფორმირებული გამოსახულებები

$$(x \vee \neg y \vee \neg z) \& (y \vee z \vee u) \& (x \vee z \vee \neg u) \& (\neg x \vee y \vee u)$$

$\vee$  ნიშნით აღნიშნულია ლოგიკური ან (ანუ, დიზუნქცია (**disjunction**)),  $\wedge$  ნიშნით აღნიშნულია ლოგიკური და (ანუ, კონუნქცია (**conjunction**)), ასოები გაგებულია პროპოზიციულ, ანუ წინადადებრივ 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 დაენიშნოს **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

$$\mathbf{C2-ენათა\ კლასი} = \mathbf{C3-ენათა\ კლასი}.$$

<sup>5</sup> იხ. Hopcroft & Ullman 1979, გვ. 324.

ეს კი დაადასტურებდა იმას, რომ არსებობს **SubsetSum** ენის **C2-LAG** გრამატიკა. ამასთან, რადგან ყველა **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.<sup>6</sup>

ის, არის თუ არა **C3** ენათა კლასი **B** ენათა კლასის საკუთრივი ქვესიმრავლე, ასევე ღია საკითხია. აქაც ვინმემ შეიძლება სავარაუდო საყრდენად აიღოს **C3-LAG** ენების **NP** სისრულე. ცნობილია, რომ კონტექსტზე დამოკიდებული ენების გამოცნობა (**CS**-გამოცნობა) არის **PSPACE**-სრული (იხ. **Hopcroft & Ullman 1979, გვ. 346,7**). რადგან დაუჯერებელია<sup>7</sup> 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- იმარაქიშის არამკვირვალეობა

დ. ეპლიგეიტის (**D. Applegate**) მიხედვით **L<sub>no</sub>** ენა, რომელიც **PS**-გრამატიკაში კონტექსტისგან თავისუფალია, **LA**-გრამატიკაში კი – **C3**-ტიპის, ხმაურიან-ენად (**noise-language**) იწოდება. **L<sub>no</sub>** ენა წარმოქმნის გამოსახულებებს, რომლებიც შეიცავენ **0** და **1** სიმბოლოებს, და რომლებსაც აქვთ **W' # W<sup>R</sup>** სტრუქტურა. **#** სიმბოლო ერთმანეთისგან გამოყოფს **W'** და **W<sup>R</sup>** გამოსახულებებს, სადაც **W<sup>R</sup>** არის სარკისებული სახე **W**-სი, ხოლო **W'** განსხვავდება **W**-სგან იმით, რომ დამატებით შეიცავს **0** და **1** სიმბოლოების გარკვეულ ოდენობას. **W'**-ის ეს დამატებითი სიტყვები არ განსხვავდება მათგან, რომელთაც აქვთ მეწყვილეები **W<sup>R</sup>**-ში და ისინი ფუნქციონირებენ როგორც ხმაური. ამ ენის კონტექსტისგან თავისუფალი **PS**-გრამატიკა განსაზღვრულია 12.3.1-ში.

#### 12.3.1 L<sub>no</sub> 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, შუა სვეტში განთავსებული წესები აწარმოებენ მხოლოდ **W'**-ის სიტყვებს.

კონტექსტისგან თავისუფალი ენის ერლეისა და **CYK** ალგორითმების მსგავს ტრადიციულ პარსერებს არ უჭირთ გააკეთონ **L<sub>no</sub>** ენის **n<sup>3</sup>** ანალიზი. ეს ასეა რადგან პარსერი იყენებს

<sup>6</sup> ის, რომ **C2**  $\subset$  **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**.

<sup>7</sup> **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 პილინომურ დროში და არც პოლინომურ სიგრძეში





ენები სტრუქტურული თვალსაზრისებით ერთი და იგივენი არიან. ამ ოთხ ენას საერთო აქვთ ის, რომ შემავალი მოცემულობა პირველ ნახევარში შეიძლება შეიცავდეს ისეთი სიტყვების ნებისმიერ ოდენობას, რომლებსთვისაც უცნობია საჭირონი არიან თუ არა ისინი, როგორც მეწყვილები მეორე ნახევარში განთავსებული სიტყვებისთვის.<sup>10</sup>

$L_{no}$  ენის **C-LAG** გაანალიზები 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, თითოეულ 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 ‘ნამდვილია’. 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 $L_{no}$ ენის C3-LAG

$$\begin{aligned}
 LX &=_{\text{def}} \{ [0 (0)], [1 (1)], [\# (\#)] \} \\
 ST_S &=_{\text{def}} \{ [(seg_c), \{r_1, r_2, r_3, r_4, r_5\}] \} \text{ სადაც } seg_c, seg_d \in \{0,1\}. \\
 r_1: (seg_c)(seg_d) &\Rightarrow \varepsilon \quad \{r_1, r_2, r_3, r_4, r_5\} \\
 r_2: (seg_c)(seg_d) &\Rightarrow (seg_d) \quad \{r_1, r_2, r_3, r_4, r_5\} \\
 r_3: (X)(seg_c) &\Rightarrow (X) \quad \{r_1, r_2, r_3, r_4, r_5\} \\
 r_4: (X)(seg_c) &\Rightarrow (seg_c X) \quad \{r_1, r_2, r_3, r_4, r_5\} \\
 r_5: (X)(\#) &\Rightarrow (X) \quad \{r_6\} \\
 r_6: (seg_c X)(seg_c) &\Rightarrow (X) \quad \{r_6\} \\
 ST_F &=_{\text{def}} \{ [\varepsilon rp_6] \}
 \end{aligned}$$

ამ **C3-LAG**-ში არსებული და  $r_3$ -ისა და  $r_4$ -ის მიზეზით წარმოქმნილი +რეკურსული ორაზროვნება, მაგალითად, ხასიათდება **(i)** შემავალ პირობასთან თავსებადობით (ფაქტიური იდენტურობით), **(ii)** თანაშემოსვლებით წესთა პაკეტში, და **(iii)** ერთსა და იმავე საანალიზო გზაზე კვლავ გამოყენებადობით.  $r_3$  წესი იგნორირებას უკეთებს შემდეგ სიტყვას და ამუშავებს მას როგორც ხმაურს.  $r_4$  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 თითოეული  $a_i$  შეიძლება ინტერპრეტირებულ იქნეს, ან როგორც ხმაური, ან როგორც ‘ნამდვილი’ ქვესიმრავლე. **SubsetSum** ენის ეს **C3-LAG** ანალიზი კონტექსტზე დამოკიდებულია, რადგან ზოგიერთი წესი (მაგ.  $r_5$ ) ამოწმებს წინადადების

საზოგადოდ მეტწილ დამოკიდებულია იმაზე, რომ გამოიყენება მრავალენტიანი ტიურინგის მანქანა. სირთულე **C-LAG**-ში კი ისაზღვრება უშუალოდ გრამატიკის საფუძველზე.

2. აბსტრაქტული ავტომატისაგან განსხვავებით არანაირი  $\varepsilon$ -მოძრაობა არ დაიშვება **C-LAG**-ში.

<sup>10</sup> კონტექსტისგან თავისუფალი  $L_{no}$  ენის **C-LAG**-სირთულე იგივეა, რაც კონტექსტისგან თავისუფალი  $L^3_{no}$  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 აგებულნი კონტექსტისგან თავისუფალი ენების ფიქსირებული ინვერსირებად დაწყვილებული სტრუქტურების დასამუშავებლად და  $\varepsilon$ -მოძრაობა არ არის დაშვებული.

საწყისის კატეგორიის დასაწყისს, სხვა წესები (მაგ. **რ4**) კი ბოლო ნაწილს. ამასთან, **L<sub>no</sub>** 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 კონტექსტისგან 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 კონტექსტისგან თავისუფალი **PS**-გრამატიკები იძლევიან მხოლოდ პირველი რიგის აპროქსიმაციას იმ მრავალი მექანიზმისა, რაც პროგრამული ენების სინტაქსში გამოიყენება.<sup>11</sup>

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**)<sup>12</sup> წოდებულ საკუთრივ ქვესიმრავლეს, რომელიც, თავის მხრივ, ქმნის ინდექსირებულ ენებად<sup>13</sup> წოდებულ ენებს, რაც უკვე კონტექსტზე დამოკიდებული ენების საკუთრივი ქვესიმრავლეა.

ორმოცდაათ წელზე მეტი ხნის **PS**-გრამატიკული ტრადიცია და მრავალხინიანი 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 ფორმალური სხვადასხვაგვარი შეზღუდვები.

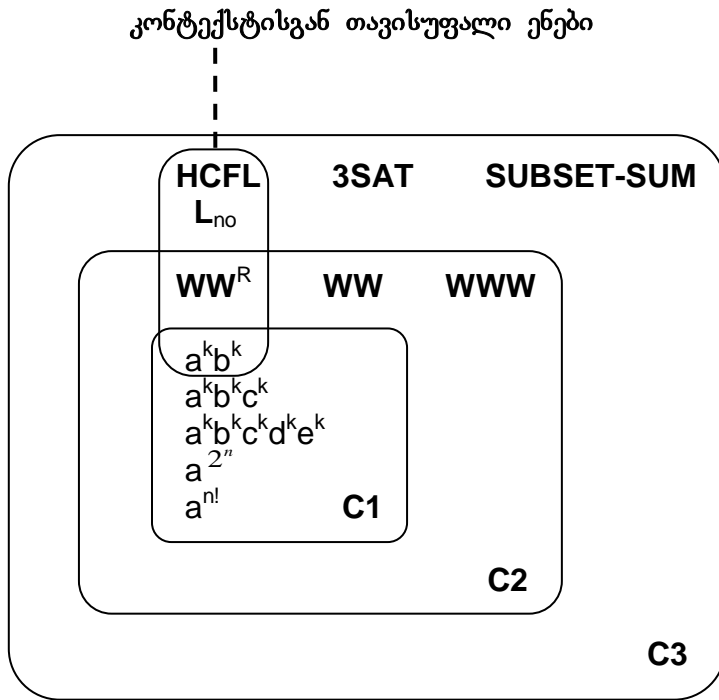
მაგალითად, კონტექსტისგან თავისუფალი ენები განსაზღვრულია იმ შეზღუდვებით, რომლებიც ეხება მხოლოდ გადაწერის წესებს (იხ. 8.1.2). ანალოგიურად **C1**, **C2** და **C3** ენები განსაზღვრულია მხოლოდ იმ შეზღუდვებით, რომლებიც ეხება მხოლოდ **LA**-გრამატიკულ ფორმალიზმს (იხ. 11.5.9) ეს ორი განსხვავებული ტიპის შეზღუდვები შედეგად იძლევა **PS**- და **LA**- გრამატიკულ იერარქიებს, რომლებიც ერთმანეთის ორთოგონალურებია:

<sup>11</sup> იხ. აგრეთვე მ. ჰარისონი (**M. Harrison**) 1978, გვ. 219, სადაც იგივენაირი განწყობაა ასახული.

<sup>12</sup> ა.კ. ჯოში, 1975.

<sup>13</sup> იხ. **Hopcroft & Ullman**, 1979, გვ. 389. მტუმბავი ლემა ინდექსირებული ენებისთვის (დამტკიცება ეკუთვნის ტ. ჰაიშას (**T. Hayashi**) 1973.

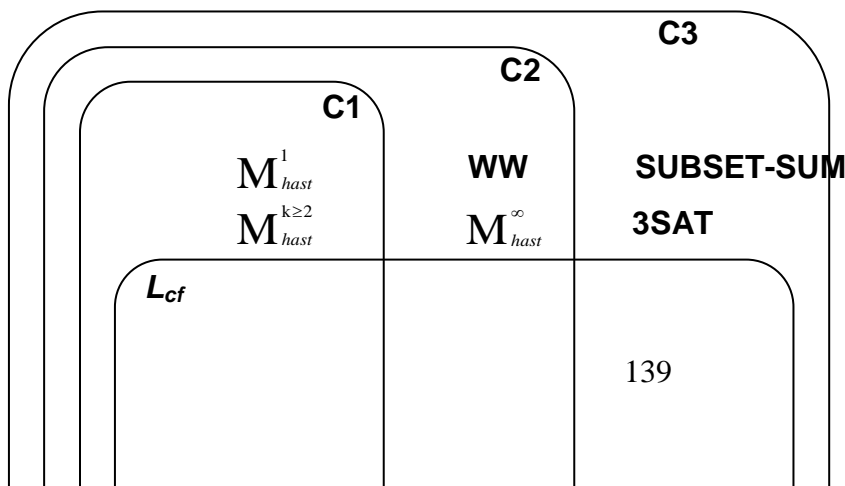
### 12.4.1 ორთოგონალური მიმართება C- და CF- ენებს შორის



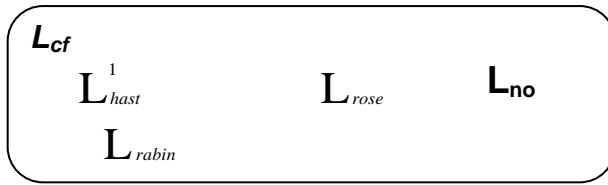
წინათეორიული ხედვებით ვინმე შეიძლება განეწყოს  $a^k b^k$  ენის  $a^k b^k c^k$ ,  $a^k b^k c^k d^k$ , და ა.შ. ენებთან ერთად დაკლასიფიცირებლად, ისევე როგორც  $WW^R$  და  $WW$  ენებისა. – ის, რომ **PS**-იერარქია დებს  $a^k b^k$  და  $WW^R$  ენებს ერთ (კონტექსტისგან თავისუფალ ენათა) კლასში, ხოლო  $a^k b^k c^k$ ,  $a^k b^k c^k d^k$ , და ა.შ. ენებს **WW** ენასთან ერთად მეორე (კონტექსტზე დამოკიდებულ ენათა) კლასში, გაუთვითცნობიერებელი მკითხველისთვის გასაკვირვებელია. ამ თვალსაზრისებით **LA**-იერარქია ინტუიციურად უფრო ბუნებრივად არის მოწესრიგებული, რადგან აქ  $a^k$ ,  $a^k b^k$ ,  $a^k b^k c^k$ ,  $a^k b^k c^k d^k$  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 კონტექსტისგან თავისუფალ ენებსა და  $L_{cf}$  არადეტერმინირებულ კონტექსტისგან 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** კლასიფიკაცია



$L_{hast}^{k \geq 2}$        $L_{hast}^{\infty}$       HCFL  
 $0^k 1^k U 0^k 1^{2k}$        $WW^R$



$L_{dcf}$  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 წრფივი სირთულე

ისეთი კონტექსტზე დამოკიდებული ენის არსებობა, რომელიც არ არის **C-** ენა, დაადასტურებდა **C-** ენების საკუთრივ ქვესიმრავლეობას **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 არადაამაჯერებელია.

თუ ბუნებრივი ენები აღიწერებიან **C-LAG**-ებით, მაშინ შემდეგი ორი ეკვივალენტური შეკითხვებია:

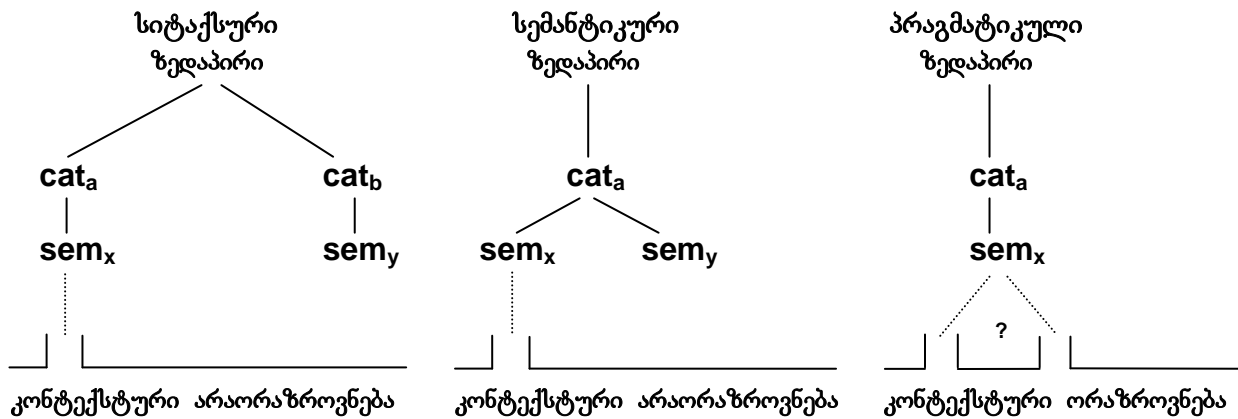
- (i) რამდენად რთულია ბუნებრივი ენა?
- (ii) რამდენად ორაზროვანია ბუნებრივი ენა?

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<sup>2</sup>. 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<sup>2</sup>. პრაგმატულ ორაზროვნებას განაპირობებს ისეთი მნიშვნელობა<sup>1</sup>, რომელსაც მოცემულ კონტექსტში ერთზე მეტი გამოყენება (**use**) აქვს.

## 12.5.1 ორაზროვნების SLIM თეორიული ანალიზი



სინტაქსური და სემანტიკური ორაზროვნებები განპირობებულია და, შესაბამისად, ხასიათდება თავად გამოსახულების ტიპით, პრაგმატული ორაზროვნება კი – იმ მთლიანი გამონათქვამით (**utterance**), რომელშიც ეს ორაზროვნება ვლინდება.

ენობრივი გამოსახულება სინტაქსურად ორაზროვანია, თუ ის ხასიათდება ისეთი ალტერნატიული კატეგორიებით (აქ **cat<sub>x</sub>** და **cat<sub>y</sub>**), რომელთაგან თითოეულის კატეგორიული წაკითხვა იძლევა თავის საკუთარ მნიშვნელობა<sub>1</sub>-ს (აქ **sem<sub>x</sub>** და **sem<sub>y</sub>**). მაგალითად, **won** წარმოქმნის სინტაქსურ ორაზროვნებას კორეული ვალუტის აღმნიშვნელ არსებით სახელსა და **to win** ზმნის წარსულ დროს შორის. ამასთან, თუ კონტექსტი არ ხსნის **sem<sub>x</sub>** და **sem<sub>y</sub>** კატეგორიების ორაზროვნულ გაგებას, მაშინ ეს სინტაქსური ორაზროვნება წარმოქმნის ისეთ გამონათქვამს, რომელსაც აქვს ერთზე მეტი მნიშვნელობა<sub>2</sub>.

ენობრივი გამოსახულება სემანტიკურად ორაზროვანია, თუ ის ხასიათდება მხოლოდ ერთი სინტაქსური ანალიზით, მაგრამ ერთზე მეტი მნიშვნელობა<sub>1</sub>-ით. მაგალითად, ენობრივი გამოსახულება **perch** სემანტიკურად ორაზროვანია: მისი ერთი მნიშვნელობა გადის თევზის სახეობაზე, მეორე კი – ჩამოსაჯდომ ადგილზე. ეს განპირობებს იმას, რომ სინტაქსურად **perch** არ არის ორაზროვანი, რადგან ორივე წაკითხვის შემთხვევაში იგი ნაუნის, ანუ არსებითი სახელის კატეგორიით ხასიათდება.<sup>14</sup> შინაგანი დაწყვილებისთვის, სემანტიკური ორაზროვნება ემსგავსება სინტაქსურს, რადგან ორივე შემთხვევაში გამოსახულებას აქვს ერთზე მეტი მნიშვნელობა<sub>1</sub>.

პრაგმატული ორაზროვნება განპირობებულია მოცემულ კონტექსტში ერთი მნიშვნელობა<sub>1</sub>-ის ალტერნატიული გამოყენებით. მაგალითად, კონტექსტში, სადაც ორი მაგიდა დგას ერთმანეთის გვერდი გვერდ, ატერანსი, ანუ ენობრივი გამონათქვამი **Put the coffee on the table** იქნებოდა პრაგმატულად ორაზროვანი, რადგან არ არის ცხადი ამ ორი მაგიდიდან რომელს გულისხმობს მეტყველი (იხ. 5.2 ქვეთავი). სინტაქსური და სემანტიკური ორაზროვნებისგან განსხვავებით პრაგმატული ორაზროვნების ბუნება არ იხსნება კონტექსტით.

სინტაქსური ორაზროვნება განპირობებს იმას, რომ გამოსახულებას ერთზე მეტი მნიშვნელობა<sub>1</sub> აქვს. ამასთან, გამოსახულება სემანტიკურად ორაზროვანია მხოლოდ მაშინ, თუ ის არ არის

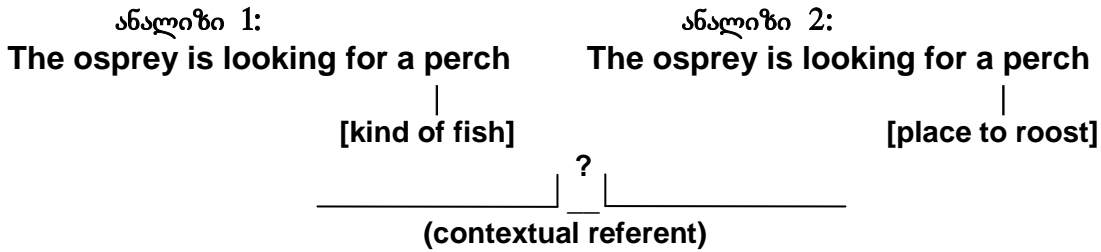
<sup>14</sup> სინტაქსური კატეგორიზაციისას იგნორირება უკეთდება **to perch**-ის შინაარსულ დატვირთვებს.

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 ტიაზე, არამედ იგი წარმოიქმნება გამოსახულების სემანტიკური ინტერპრეტაციისა და კონტექსტის ურთიერთმიმართებისას (იხ. 4.4).

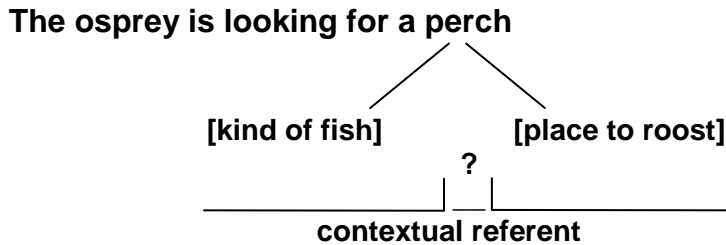
მსგავსად ამისა, არც სემანტიკური ორაზროვნებაა ძალისმიერი ბუნებრივი ენის სირთულის განსაზღვრისას, რადგან სემანტიკური ორაზროვნებები განსაზღვრების ძალით უკავშირდება მხოლოდ სინტაქსურად არაორაზროვან გამოსახულებებს. მაგალითად, იმ ფაქტის გათვალისწინებით, რომ **perch** სიტყვის ორივე წაკითხვას აქვს ერთი და იგივე კატეგორია, გასაგები ხდება, რომ **The osprey is looking for a perch** წინადადებას არ შეიძლება ქონდეს ორი განსხვავებული სინტაქსური ანალიზი:

**12.5.2 სემანტიკური ორაზროვნების არაკორექტული 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 კორექტული ანალიზი**



მაშინაც კი, როცა სინტაქსური დერივაცია არაორაზროვანია, სემანტიკური ინტერპრეტაცია უზრუნველყოფს ორ განსხვავებულ მნიშვნელობა<sub>1</sub>-ს, რომელთაგან თითოეული, თავის მხრივ, კონტექსტურად იძლევა შესაძლებლობას უზრუნველყოს ორი განსხვავებული მნიშვნელობა<sub>2</sub>. 12.5.3 ანალიზში გამოყენებული მეთოდი იწოდება სემანტიკურ გაორმაგებად<sup>15</sup> (**semantic doubling**).

<sup>15</sup> 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 მნიშვნელობა<sup>1</sup>-ის მინიჭება გვარწმუნებს, რომ აუცილებელი არ არის სემანტიკური განსხვავებების სინტაქსურ ღონეზე განრჩევა.

ორაზროვანი გამოსახულება შეიძლება გაანალიზდეს სემანტიკური გაორმაგების ტერმინებში, მაშინ როცა განსხვავება სემანტიკურ ღონეზე 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, წინასწარგანუზღვრულ – 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.

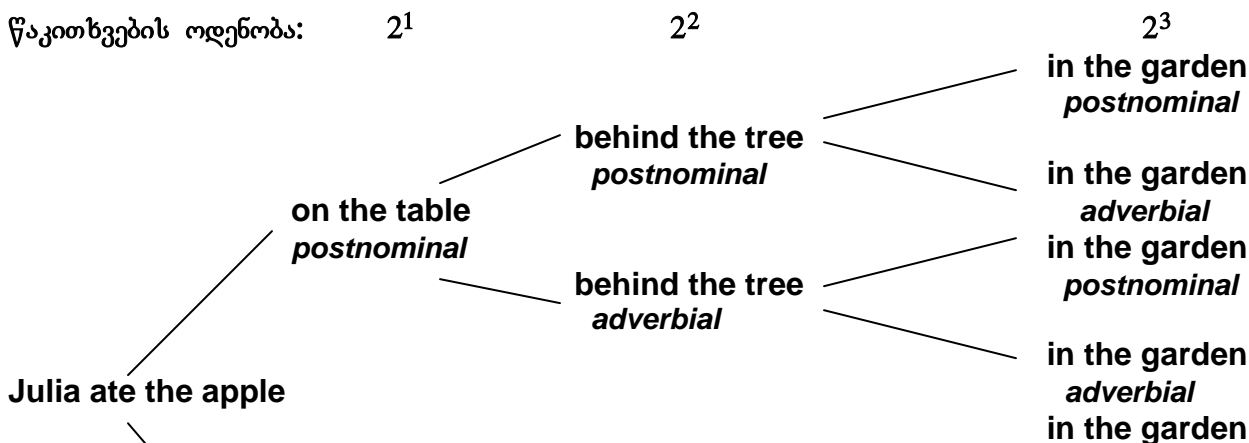
### 12.5.4 პრეპოზიციული ფრაზების მრავალჯერადი ინტერპრეტაციები

**The man saw the girl with the telescop.**  
**Julia ate the apple on the table behind the tree in the garden.**

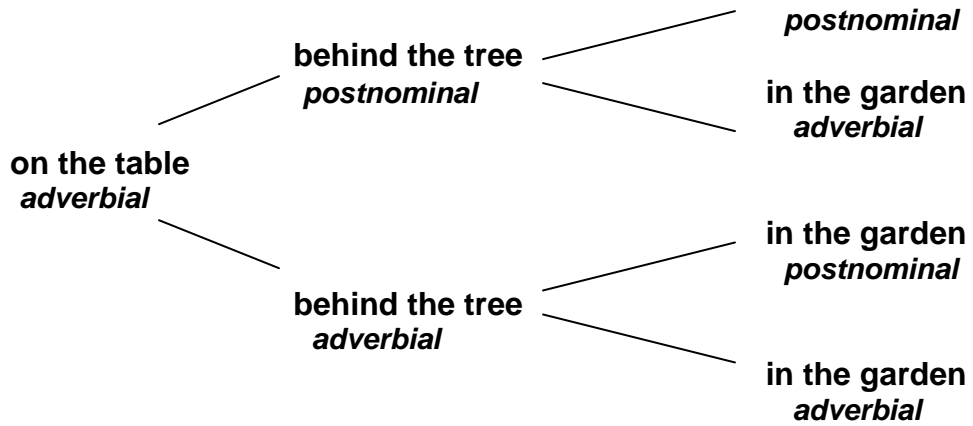
პირველი მაგალითის ინტერპრეტაცია ორ განსხვავებულ მნიშვნელობა<sup>1</sup>-ს იძლევა. ადვერბიალური კითხვისას პრეპოზიციული ფრაზა **with the telescop** მოდიფიცირებას უკეთებს **saw** ზმნას. პოსტნომინალური კითხვისას იგივე პრეპოზიციული ფრაზა მოდიფიცირებას უკეთებს **the girl** კონსტრუქციას.

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 პრეპოზიციული ფრაზების შეცდომით სინტაქსურ დამუშავებას.

### 12.5.5 რეკურსული ფსევდო-ორაზროვნება





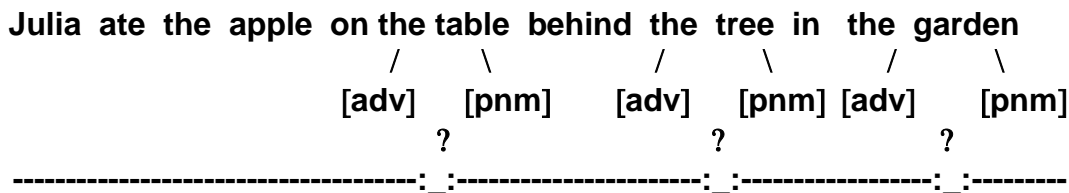


ამგვარი ანალიზების საფუძველზე იყო ის, რომ ბუნებრივი ენები განიხილებოდა სულ ცოტა ექსპონენციალური სირთულის მქონედ.

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.<sup>16</sup> – კარგი სინტაქსური ანალიზი უნდა იმიზნებდეს ლიტერალური მნიშვნელობების ისეთი აბსოლიტური მინიმუმის მოძიებას, რომელიც საკმარისი იქნებოდა ყველა შესაძლო გამოყენების 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 გაიმორმავებზე დამყარებული კორექტული ანალიზი**



წინადადების ენობრივი გამოსახულება გაანალიზებულია როგორც სინტაქსურად არაორაზროვანი.<sup>17</sup> ამასთან, სემანტიკური ინტერპრეტაცია სემანტიკურ დონეზე სისტემურად უნიშნავს თითოეულ პრეპოზიციულ ფრაზას ორ განსხვავებულ ლიტერალურ მნიშვნელობას.

<sup>16</sup> უფრო მეტიც, მაგ., არაორაზროვანი  $a^k b^k$  ენისთვის თავისუფლად შეიძლება დაიწეროს მრავალი ისეთი ორაზროვანი გრამატიკა, რომლებიც სირთულის ხარისხს წრფივიდან ექსპონენციალურამდე, რიგ შემთხვევებში ამოუხსნადლობამდეც კი ზრდიან.

<sup>17</sup> სინტაქსურად, პრეპოზიციონალური ფრაზები კატეგორიზდებიან მრავალკატეგორიზებადი აღნიშვნებით (**multicat notation**), მაგ., როგორც (**adv&pnm**) (იხ. 15.2).

12.5.3 ანალიზის მსგავსად 12.5.6 ეფუძნება ბუნებრივი კომინიკაციების [2+1] დონის სტრუქტურებს. 12.5.6 ანალიზი საკმარისია განსხვავებული მნიშვნელობა-ების 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 დამაწყვილებელი მნიშვნელობა-ების 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 სინტაქსური 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 გაიგონ<sup>18</sup> რთული ტექსტებიც კი. ეს **CoNSyx** ჰიპოტეზა 21 თავში გასრულდება სემანტიკის სირთულის თაობაზე ჩამოყალიბებული ჰიპოტეზით, რომელიც **CoNSem** ჰიპოტეზად იწოდება (იხ. 21.5.2).<sup>19</sup>

## სავარჯიშოები

### ქვეთავი 12.1

<sup>18</sup> სულ ცოტა სინტაქსურ დონეზე მაინც.

<sup>19</sup> **CoNSyx** ჰიპოტეზით ბუნებრივი ენის სინტაქსისთვის განსაზღვრული სტრუქტურული თვისებები შეიძლება და უნდა იქნეს ემპირიულად ტესტირებული მრავალი სხვადასხვა ენის გაფართოებადი ანალიზით. 12.5.7-ის შესაძლო კონტრმაგალითი თუ არის უნდა იყოს +რეკურსული სინტაქსური ორაზროვნების მქონე ბუნებრივ ენაში, რაც, ბუნებრივია, სემანტიკური გაორმაგების პრინციპზე დამყარებული ალტერნატიული ანალიზის გაკეთების საშუალებას არ მოგვცემდა. ბუნებრივი ენების **SLIM** თეორიის ფარგლებში მოცემული ბუნებრივენობრივი კომუნიკაციის მექანიზმები გვაფიქრებინებს ამგვარი ბუნებრივი ენების არ არსებობას.

1. რატომ არიან ენობრივი იერარქიის ქვედა კლასები (ე.ი. შედარებით დაბალი წარმოქმნითი უნარის მქონენი) 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<sup>k</sup>b<sup>k</sup>c<sup>k</sup>** 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 მეთოდით 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 **C2C3** და **C3CB**?

### ქვეთავი 12.3

1. **C**-ენების რომელ ქვეკლასშია **L<sub>no</sub>**-ენა და რატომ?
2. დაწერეთ **PS**- და **LA**- გრამატიკები **L<sup>3</sup><sub>no</sub>**-ენისათვის, რომელიც განსაზღვრულია როგორც **W<sup>#</sup>W<sup>#</sup>W<sup>#</sup>**, სადაც **W'** და **W''** არიან **W**-ს ხმაურიანი ვერსიები.
3. რომელ **PS**-გრამატიკულ კლასში არიან **L<sup>3</sup><sub>no</sub>** და **L<sup>3</sup><sub>no</sub>** ენები?
4. რომელ **LA**-გრამატიკულ კლასში არიან **L<sup>3</sup><sub>no</sub>** და **L<sup>3</sup><sub>no</sub>** ენები?
5. **PS** გრამატიკული კლასიფიკაციის გათვალისწინებით შეადარეთ **L<sup>3</sup><sub>no</sub>** და **L<sup>3</sup><sub>no</sub>** 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<sup>k</sup>b<sup>k</sup>c<sup>k</sup>d<sup>k</sup>e<sup>k</sup>f<sup>k</sup>** არის თუ არა **TAL** ტიპის ენა? რა არის ამ ენის **LA**-კლასი?
3. რა არის სტრუქტურული მიზეზი იმისა, რომ **a<sup>n!</sup>** არ არის ინდექსირებული ენა? ჰოპკროფტისა & 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** ჰიპოტეზა. როგორ შეიძლება მისი უარყოფა ემპირიულად.
11. რა პრაქტიკული ღირებულება აქვს **CoNSyX** ჰიპოტეზას?