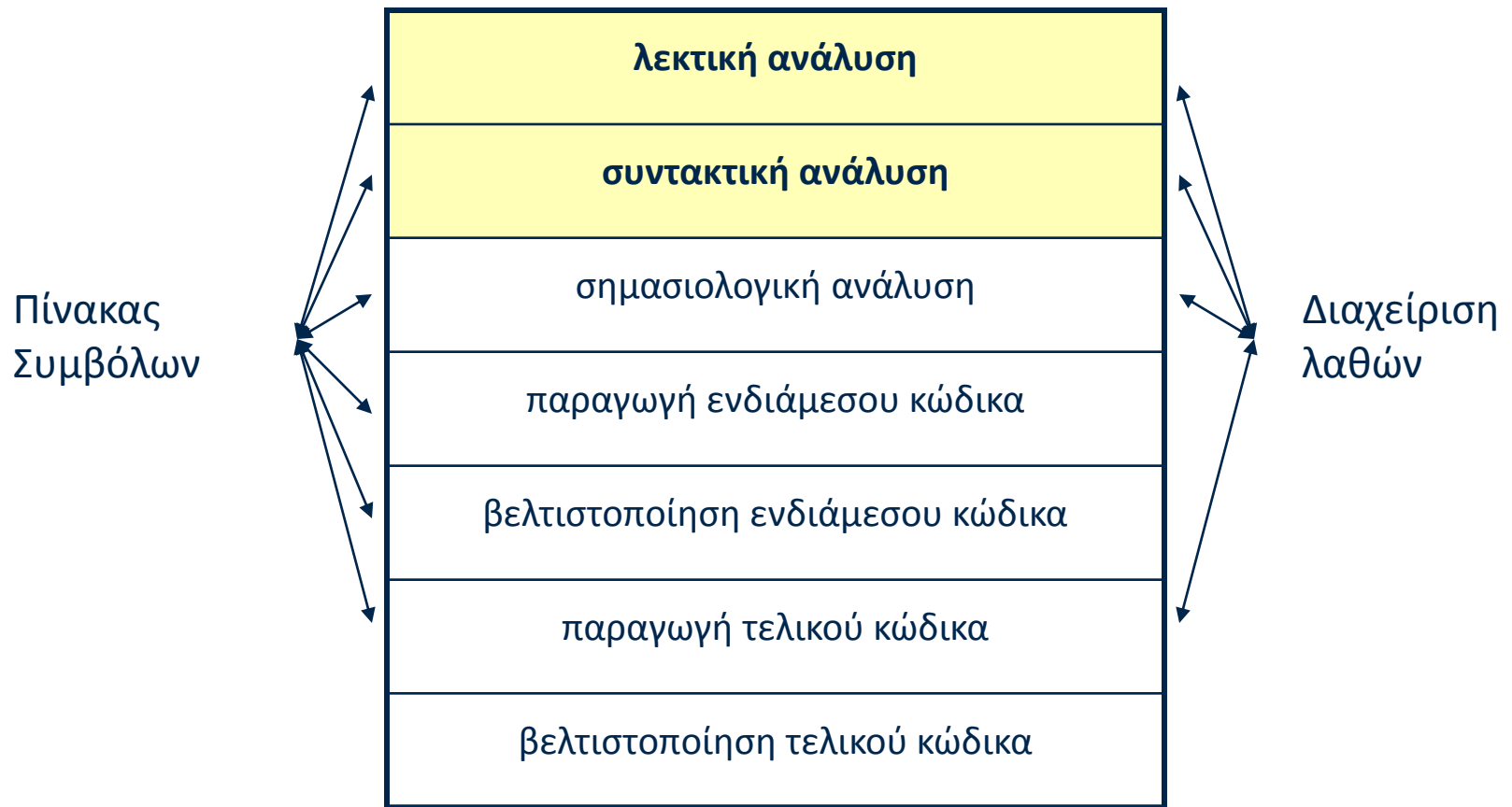

Γραμματικές



Διαλέξεις στο μάθημα: Μεταφραστές
Γιώργος Μανής

Στις Φάσεις της Μεταγλώττισης



Γραμματικές

- # T: αλφάβητο τερματικών συμβόλων
- # N: αλφάβητο μη τερματικών συμβόλων
 - $T \cap N = \emptyset$
- # P: πεπερασμένο σύνολο κανόνων παραγωγής
 - $P \subseteq (T \cup N)^* \times (T \cup N)^*$
 - συμβολισμός: $\alpha \rightarrow \beta$
- # S: αρχικό σύμβολο

Παραγωγή συμβολοσειρών

- # Η β παράγεται από τη α ($\alpha \Rightarrow \beta$) όταν
 - υπάρχουν $\alpha', \beta', \gamma, \delta \in (T \cup N)^*$
 - τέτοιες ώστε ο κανόνας $\alpha' \rightarrow \beta'$ ανήκει στη γραμματική
 - και $\alpha = \gamma\alpha'\delta$ και $\beta = \gamma\beta'\delta$

- # Η $\alpha \in T^*$ παράγεται από τη γραμματική G με αρχικό σύμβολο S όταν:
 - $\alpha_1 = S,$
 - $\alpha_n = \alpha,$
 - $\alpha_i \Rightarrow \alpha_{i+1}$ για $1 \leq i \leq n-1$

Παραγωγή τυπικής γλώσσας

- # Κάθε γραμματική $G=(T,N,P,S)$ ορίζει μία γλώσσα $L(G) \subseteq T^*$
- # Λέμε ότι η γραμματική παράγει τη γλώσσα και ορίζεται ως
 - $L(G) = \{ \alpha \in T^* \mid S \Rightarrow^+ \alpha \}$
 - το σύμβολο \Rightarrow^+ σημαίνει ότι η συμβολοσειρά παράγεται σε ένα ή περισσότερα βήματα
- # Μία γλώσσα λέγεται τυπική όταν υπάρχει γραμματική που την παράγει

Ιεραρχία Chomsky

<u>Γλώσσες</u>	<u>Γραμματικές</u>	<u>Μηχανές</u>
Κανονικές Γλώσσες	Κανονικές Γραμματικές	Πεπερασμένα αυτόματα
Γλώσσες Χωρίς Συμφραζόμενα	Γραμματικές Χωρίς Συμφραζόμενα	Μη-Ντετερμινιστικά Αυτόματα Με Στοίβα
Γλώσσες Με Συμφραζόμενα	Γραμματικές Με Συμφραζόμενα	Γραμμικά Πεπερασμένα Αυτόματα
Υπολογίσιμες γλώσσες	Γραμματικές χωρίς περιορισμούς	Μηχανή Turing

Γραμματικές χωρίς περιορισμούς

- # Γραμματικές χωρίς περιορισμούς – τύπου 0 – προτασιακής σύνταξης
 - κανόνες της μορφής $\sigma \rightarrow \tau$
 - $\sigma \in V^+$, $V = T \cup N$
 - $\tau \in V^*$
- οι γραμματικές αυτές είναι δύσκολο να συνταχθούν και ο τρόπος αυτός παράστασης δεν είναι φιλικός προς τον άνθρωπο. Οι πρακτικές τους εφαρμογές είναι ελάχιστες έως ανύπαρκτες

Γραμματικές με συμφραζόμενα

- # Γραμματικές με συμφραζόμενα – τύπου 1
 - κανόνες της μορφής $\alpha A \beta \rightarrow \alpha \chi \beta$
 - $A \in N$
 - $\chi \in V^+$
 - $\alpha, \beta \in V^*$
 - κατ' εξαίρεση:
 - επιτρέπεται το $S \rightarrow \epsilon$ ώστε να παράγεται η κενή συμβολοσειρά
 - αρκεί το S να μην υπάρχει σε δεξί μέλος κανόνα

Γραμματικές με συμφραζόμενα

- # Ισοδύναμη μορφή γραμματικών με συμφραζόμενα, καλούνται *μονότονες* γραμματικές
 - κανόνες της μορφής $\alpha \rightarrow \beta$
 - $\alpha, \beta \in V^+$
 - η α περιέχει τουλάχιστον ένα μη τερματικό σύμβολο
 - $|\alpha| \leq |\beta|$
 - κατ' εξαίρεση:
 - επιτρέπεται το $S \rightarrow \epsilon$ ώστε να παράγεται η κενή συμβολοσειρά
 - αρκεί το S να υπάρχει μόνο σε δεξί μέλος κανόνα

Γραμματικές με συμφραζόμενα

- # Παράδειγμα κατασκευής γραμματικής τύπου 1
 - Θα κατασκευάσουμε γραμματική για τη γλώσσα: $a^n b^n c^n$
 - αποδεικνύεται ότι δε μπορεί να κατασκευαστεί γραμματική τύπου 2 για τη γλώσσα αυτή
 - αρχίζοντας από την απλή περίπτωση
 - $S \rightarrow abc$
 - αν προσθέσουμε ένα a στην αρχή θα πρέπει να προσθέσουμε κάτι και στο τέλος
 - $S \rightarrow aSQ$
 - στο τέλος πρέπει να υπάρχει κάτι με b και c αλλά όχι απλά να προσθέσουμε bc αφού όλα τα b πρέπει να βρίσκονται πριν τα c

Γραμματικές με συμφραζόμενα

- κάτι τέτοιο επιτρέπει το b να είναι πάντα πριν το c
 - $bQc \rightarrow bbcc$
- πάλι δεν τελειώσαμε αφού τα Q βρίσκονται δεξιά από τα c , αυτό μπορεί να διορθωθεί με τον κανόνα
 - $cQ \rightarrow Qc$
- η γραμματική ολοκληρώθηκε

Γραμματικές με συμφραζόμενα

Παραγωγή συμβολοσειράς από τη γραμματική

- | | |
|--------------------|------------|
| ▪ S | S → aSQ |
| ▪ aSQ | S → aSQ |
| ▪ aaSQQ | S → abc |
| ▪ aaabcQQ | cQ → Qc |
| ▪ aaabQcQ | bQc → bbcc |
| ▪ aaabbccQ | cQ → Qc |
| ▪ aaabbccQ c | cQ → Qc |
| ▪ aaabbQcc | bQc → bbcc |
| ▪ aaabbbccc | |

Γραμματικές χωρίς συμφραζόμενα

Γραμματικές χωρίς συμφραζόμενα - τύπου 2

- κανόνες της μορφής $A \rightarrow \alpha$
- $A \in N$
- $\alpha \in V^*$

Παράδειγμα

- $\langle \text{name} \rangle ::= \text{tom} \mid \text{john} \mid \text{harry}$
- $\langle \text{sentence} \rangle ::= \langle \text{name} \rangle \mid \langle \text{list} \rangle \text{ and } \langle \text{name} \rangle$
- $\langle \text{list} \rangle ::= \langle \text{name} \rangle, \langle \text{list} \rangle \mid \langle \text{name} \rangle$

Γραμματικές χωρίς συμφραζόμενα

Γραμματική τύπου LL(1)

- L: Left to right - αναφέρεται στον τρόπο που σαρώνεται η συμβολοσειρά εισόδου, από τα αριστερά στα δεξιά
- L: Left most derivation – η διαδικασία κατασκευής του συντακτικού δέντρου αντιστοιχεί στην αριστερότερη παραγωγή
- (1): one look ahead symbol – για να επιλέξουμε ανάμεσα σε πιθανούς κανόνες ποιος είναι ο επόμενος κανόνας που πρέπει να εφαρμοσθεί αρκεί να γνωρίζουμε το επόμενο σύμβολο της συμβολοσειράς εισόδου

Γραμματικές χωρίς συμφραζόμενα

Παράδειγμα γραμματικής τύπου LL(1)

- $E \rightarrow TE'$
- $E' \rightarrow +TE' \mid \varepsilon$
- $T \rightarrow FT'$
- $T' \rightarrow x FT' \mid \varepsilon$
- $F \rightarrow (E) \mid id$
- $id \rightarrow a \mid b \mid c \mid d$

Παραγωγή:

- $(a+bx+c+bx(a+c))xd$

Κανονικές γραμματικές

- # Κανονικές γραμματικές – τύπου 3 – γραμματικές πεπερασμένων καταστάσεων
 - κανόνες της μορφής
 - $A \rightarrow \alpha$
 - $A \rightarrow \beta B$
 - $A, B \in N$
 - $\alpha \in T \cup \{\epsilon\}$
 - $\beta \in T$

Κανονικές γραμματικές

Παράδειγμα

- $S \rightarrow a A$
- $A \rightarrow a A$
- $A \rightarrow d A$
- $A \rightarrow \epsilon$

Παραγωγή:

- σειρές από a και d που ξεκινούν υποχρεωτικά από a
-

Παραγωγή συμβολοσειρών από γραμματική

- ένα σύνολο που περιέχει το αρχικό σύμβολο της γραμματικής
- τη διαπερνούμε από αριστερά προς τα δεξιά ψάχνοντας για υποσυμβολοσειρές που ταιριάζουν με το αριστερό μέλος κάποιου γραμματικού κανόνα
- για κάθε τέτοια υποσυμβολοσειρά που βρίσκουμε δημιουργούμε και ένα αντίγραφο της συμβολοσειράς εισόδου και αντικαθιστούμε την υποσυμβολοσειρά που βρίσκουμε στο αριστερό μέλος του κανόνα με αυτό που βρίσκεται στο δεξιό μέλος και την τοποθετούμε στο σύνολο των προς εξέταση συμβολοσειρών
- αν η συμβολοσειρά που παράχθηκε δεν περιέχει μη τερματικά σύμβολα τότε την τοποθετούμε σε ένα σύνολο που περιέχει τις συμβολοσειρές που παράχθηκαν
- στη συνέχεια αντλούμε μία νέα προς εξέταση συμβολοσειρά από το σύνολο των προς εξέταση συμβολοσειρών και επαναλαμβάνουμε τα προηγούμενα βήματα

Παραγωγή συμβολοσειρών από γραμματική

Παράδειγμα της γενικής περίπτωσης για τη γραμματική

- $S \rightarrow abc$
- $S \rightarrow aSQ$
- $bQc \rightarrow bbcc$
- $cQ \rightarrow Qc$

η οποία περιγράφει τη γλώσσα $a^n b^n c^n$

Παραγωγή συμβολοσειρών από γραμματική

Εκτέλεση του αλγορίθμου

- S
- abc aSQ
- aSQ
- aabcQ aaSQQ
- aaSQQ aabQc
- aabQc aaabcQQ aaaSQQQ
- aaabcQQ aaaSQQQ aabbcc
- aaaSQQQ aabbcc aaabQcQ
- aabbcc aaabQcQ aaaabcQQQ aaaaSQQQQ
- aaabQcQ aaaabcQQQ aaaaSQQQQ
- κλπ.

Παραγωγή συμβολοσειρών από γραμματική

Παρατήρηση:

- στην περίπτωση που επιχειρήσουμε να μη φτιάξουμε όλα τα αντίγραφα που προκύπτουν από όλους τους κανόνες αλλά να επιλέξουμε έναν και να κάνουμε εκεί την αντικατάσταση υπάρχει κίνδυνος να χαλάσουμε τα συμφραζόμενα της γραμματικής. Έστω η γραμματική

$S \rightarrow AC$

$A \rightarrow b$

$AC \rightarrow ab$

αντικαθιστώντας με βάση το $A \rightarrow b$ η γραμματική δε θα παράγει τίποτα, ενώ θα μπορούσε να παράγει το ab

Γραμματικές ενεργειών

⌘ Κάθε κανόνας εμφανίζει ως έξοδο κάποια συμβολοσειρά

⌘ Παράδειγμα

Sum -> Digit	{ make it the result }
Sum -> Sum + Digit	{ add it to the previous result }
Digit -> 0	{ take a 0 }
Digit -> 1	{ take a 1 }
...	
Digit -> 9	{ take a 9 }

Με είσοδο 3+5 η γραμματική θα παράξει

take a 3
make it result
take a 5
add it to the previous result

Κατηγορικές γραμματικές

- # κάθε μη τερματικό σύμβολο συνοδεύεται από ένα ή περισσότερα κατηγορήματα
- # κάθε κατηγορημα είναι μία απλή τιμή π.χ. αριθμός, συμβολοσειρά κλπ.
- # σε ένα κανόνα (π.χ μαθηματικό τύπο) συνδυάζουμε τα κατηγορήματα κάποιων από τα μη τερματικά σύμβολα του κανόνα για να παράγουμε κάποια άλλα κατηγορήματα για μη τερματικά σύμβολα
- # αν το κατηγορημα που υπολογίζεται ανήκει σε μη τερματικό σύμβολο που βρίσκεται σε αριστερό μέλος ενός κανόνα τότε το κατηγορημα θεωρείται ότι συντίθεται (derived, synthesized)
- # αν το κατηγορημα που υπολογίζεται ανήκει σε μη τερματικό σύμβολο που βρίσκεται σε δεξιό μέλος ενός κανόνα τότε το κατηγορημα θεωρείται ότι κληρονομείται (inherited)

Κατηγορικές γραμματικές

Παράδειγμα

- στο παράδειγμα για λόγους απλότητας θεωρούμε ένα κατηγορήμα για κάθε μη τερματικό σύμβολο το οποίο συμβολίζεται με A_x , όπου x η σειρά με την οποία συναντάται το μη τερματικό σύμβολο στον κανόνα
 - Sum \rightarrow Digit $\{ A_0 = A_1 \}$
 - Sum \rightarrow Sum + Digit $\{ A_0 = A_1 + A_3 \}$
 - Digit \rightarrow 0 $\{ A_0 = 0 \}$
 - Digit \rightarrow 1 $\{ A_0 = 1 \}$
 - ...
 - Digit \rightarrow 9 $\{ A_0 = 9 \}$

Δυναμικότητα γραμματικών και μεταφραστές

- # Το σύνολο όλων των λεκτικά ορθών προγραμμάτων μπορεί να περιγραφεί από μία κανονική γραμματική, π.χ.:
 - δεν υπάρχουν λευκοί χαρακτήρες ανάμεσα σε δεσμευμένες λέξεις
 - δε συναντάμε end-of-file ενώ έχουν ανοίξει σχόλια και δεν έχουν κλείσει
 - τα αναγνωριστικά ξεκινούν από γράμμα και ακολουθεί μία σειρά γραμμάτων και αριθμών
 - κλπ

Δυναμικότητα γραμματικών και μεταφραστές

- # Το σύνολο των συντακτικά ορθών προγραμμάτων περιγράφεται από μία γραμματική χωρίς συμφραζόμενα:
 - βρέθηκε else χωρίς if
 - εκχώρηση σε αριθμητική σταθερά
 - κλπ.
- # Το σύνολο των σημασιολογικά ορθών προγραμμάτων περιγράφεται από μία γραμματική με συμφραζόμενα
 - η μεταβλητή αυτή έχει δηλωθεί;
 - πρόκειται για απλή μεταβλητή ή πίνακα;
 - γράφουμε μήπως έξω από τα όρια ενός πίνακα;

Δυναμικότητα γραμματικών και μεταφραστές

- ‡ Το σύνολο των προγραμμάτων που τερματίζουν σε πεπερασμένο χρόνο με συγκεκριμένη είσοδο μπορεί να περιγραφεί από μία (πολύ πολύπλοκη) γραμματική χωρίς περιορισμούς
- ‡ Το σύνολο των προγραμμάτων που λύνουν ένα συγκεκριμένο πρόβλημα δε μπορεί να περιγραφεί από καμία γραμματική