ΚΕΦΑΛΑΙΟ 10

ΣΥΝΤΑΚΤΙΚΗ ΑΝΑΛΥΣΗ

Σκοπός

Στο κεφάλαιο αυτό θα αναφερθούμε στο πρόβλημα της συντακτικής ανάλυσης, δηλαδή της ανακατασκευής της ακολουθίας των κανόνων σύμφωνα με τους οποίους παράγεται μία συμβολοσειρά. Θα δούμε πως κατασκευάζεται ένας συντακτικός αναλυτής ξεκινώντας από ένα αυτόματο στοίβας για την αντίστοιχη γλώσσα και θα μελετήσουμε δύο βασικές μεθόδους συντακτικής ανάλυσης.

Προσδοκώμενα Αποτελέσματα

Όταν θα έχετε μελετήσει αυτό το κεφάλαιο θα μπορείτε να:

- αναφέρετε τι είναι συντακτική ανάλυση και πως σχετίζεται με την κατασκευή συντακτικών δέντρων.
- αναφέρετε ποια είναι τα κύρια είδη της συντακτικής ανάλυσης.
- εξηγήσετε τι είναι αριστερή παραγοντοποίηση μίας γραμματικής και πότε χρειάζεται να εφαρμοστεί.
- περιγράψετε πως λύνετε το πρόβλημα της αριστερής αναδρομής σε μία γραμματική κατά τη διάρκεια της συντακτικής ανάλυσης από πάνω προς τα κάτω.
- περιγράψετε τις βασικές λειτουργίες ενός αναλυτή από κάτω προς τα πάνω.
- αναφέρετε πως γίνεται η επιλογή μεταξύ μετακίνησης ενός συμβόλου στο σωρό και αναγωγής μίας συμβολοσειράς από το σωρό στην ανάλυση από κάτω προς τα πάνω.

Έννοιες Κλειδιά

- Συντακτική ανάλυση.
- Είδη συντακτικής ανάλυσης.
- Μετατροπή ενός αυτόματου στοίβας σε συντακτικό αναλυτή.
- Προβλήματα κατά τη φάση της ανάλυσης από πάνω προς τα κάτω.
- Αριστερή αναδρομή και παραγοντοποίηση μίας γραμματικής.
- Βασικές λειτουργίες ενός αναλυτή από κάτω προς τα πάνω.
- Σχέσεις προτεραιότητας στην συντακτική ανάλυση.

Εισαγωγικές παρατηρήσεις

Το κεφάλαιο αυτό αποτελείται από δύο βασικές ενότητες. Στην πρώτη ενότητα περιγράφεται ο τρόπος σύμφωνα με τον οποίο ένα μη ντετερμινιστικό αυτόματο στοίβας μπορεί να μετατραπεί σε έναν αναλυτή που χτίζει αριστερές παραγωγές ή αλλιώς συντακτικά δέντρα από πάνω προς τα κάτω. Αναφέρονται τα κύρια προβλήματα που συναντούνται σε αυτό το είδος της συντακτικής ανάλυσης και η επίλυσή τους με μικρές τροποποιήσεις της αντίστοιχης γραμματικής. Στη δεύτερη ενότητα, γίνεται αναφορά στο άλλο είδος της συντακτικής ανάλυσης η οποία χτίζει ένα συντακτικό δέντρο χρησιμοποιώντας μία προσέγγιση από κάτω προς τα πάνω. Περιγράφονται οι δύο βασικές λειτουργίες ενός τέτοιου αναλυτή, της αναγωγής και μετακίνησης συμβόλων από και προς το σωρό, και εξηγείται πως ο αναλυτής επιλέγει μεταξύ αυτών των δύο κινήσεων.

10.1 Ανάλυση από πάνω προς τα κάτω

Είδαμε πόσο σημαντικό είναι να γνωρίζουμε την ακολουθία των κανόνων σύμφωνα με την οποία παράγεται μία συμβολοσειρά από μία γραμματική. Για παράδειγμα, η έκφραση $x+x*y$ μπορεί να ερμηνευθεί σαν $x+(x*y)$ ή σαν $(x+x)*y$ ανάλογα με την παραγωγή από την οποία έχει προκύψει. Φυσικά η γραμματική θα πρέπει να μην είναι διφορούμενη, αλλά όπως αναφέραμε στο Κεφάλαιο 7 αυτό δεν είναι πάντα δυνατό.

Η διαδικασία σύμφωνα με την οποία κάποιος προσπαθεί να ανακατασκευάσει την παραγωγή μίας συμβολοσειράς, ή καλύτερα ένα συντακτικό δέντρο για τη συμβολοσειρά, ονομάζεται *συντακτική ανάλυση* και τα εργαλεία που χρησιμοποιούνται για το σκοπό αυτό ονομάζονται *συντακτικοί αναλυτές*. Σε αυτό το κεφάλαιο θα δούμε δύο από τις βασικές μεθόδους συντακτικής ανάλυσης και τον τρόπο μετατροπής ενός αυτόματου στοίβας σε συντακτικό αναλυτή. Οι μέθοδοι αυτές δεν εξαντλούν φυσικά τη θεωρία αλλά αποτελούν την βάση για την κατασκευή αποτελεσματικών συντακτικών αναλυτών.

Θα ξεκινήσουμε τη συζήτηση με ένα παράδειγμα. Θεωρείστε τη γλώσσα $L = \{0^n1^n \mid n \geq 0\}$, η οποία παράγεται από τη γραμματική με κανόνες $S \rightarrow 0S1$ και $S \rightarrow \epsilon$. Δεν είναι δύσκολο να κατασκευάσουμε ένα ντετερμινιστικό αυτόματο στοίβας για την L, επειδή όμως θέλουμε μία γενική μέθοδο κατασκευής αναλυτών, θα αρχίσουμε κατασκευάζοντας πρώτα ένα ισοδύναμο με τη γραμματική μη ντετερμινιστικό αυτόματο χρησιμοποιώντας τις ιδέες του Θεωρήματος 8.2. Στη συνέχεια θα δώσουμε τους γενικούς κανόνες μετατροπής ενός τέτοιου αυτόματου σε ντετερμινιστικό. Θα είναι όμως πιο εύκολο να θεωρήσουμε ότι κάθε συμβολοσειρά της L τελειώνει σε \$, ή με άλλα λόγια ότι η γλώσσα προς αναγνώριση είναι η $L' = L\$$. Η αντίστοιχη γραμματική τώρα δίνεται από τους κανόνες $S \rightarrow T\$$ και $T \rightarrow 0T1 \mid \epsilon$.

Εφαρμόζοντας την κατασκευή του Θεωρήματος 8.2 παίρνουμε το αυτόματο $M = (\{p, q, f\}, \Sigma, \Gamma, p, Z_0, \delta, \{f\})$, όπου $\Sigma = \{0, 1, \$\}$, $\Gamma = \{S, T, Z_0\}$ και η συνάρτηση μετάβασης δ ορίζεται όπως όπως παρακάτω:

Αριθμός Κίνησης	**Κατάσταση**	**Σύμβολο Εισόδου**	**Σύμβολο Σωρού**	**Κινήσεις**	**Επεξηγήσεις**
0	p	ϵ	Z_0	(q, SZ_0)	
1	q	ϵ	S	$(q, T\$)$	$S \rightarrow T\$$
2	q	0	0	(q, ϵ)	Ταίριασμα 0
3	q	1	1	(q, ϵ)	Ταίριασμα 1
4	q	ϵ	T	$(q, 0T1)$	$T \rightarrow 0T1$
5	q	ϵ	T	(q, ϵ)	$T \rightarrow \epsilon$
6	q	\$	\$	(q, ϵ)	Ταίριασμα \$
7	q	ϵ	Z_0	(f, Z_0)	
(Οι υπόλοιποι συνδυασμοί)					(Τίποτα)

Το παραπάνω αυτόματο μιμείται τις αριστερές παραγωγές της γραμματικής αφού κάθε μεταβλητή στην κορυφή του σωρού αντικαθίσταται από το δεξιό μέρος του αντίστοιχου κανόνα. Το μόνο που πρέπει να κάνουμε τώρα είναι να εξοπλίσουμε το αυτόματο με δυνατότητα παραγωγής μηνυμάτων. Έτσι όταν αυτό χρησιμοποιεί μία από τις κινήσεις 1, 4 και 5, θα μπορεί να μας ειδοποιεί για τον αντίστοιχο κανόνα της γραμματικής (δείτε ξανά το Παράδειγμα 8.2 καθώς το αυτόματο αναγνωρίζει τη συμβολοσειρά 0011). Τέτοια αυτόματα με δυνατότητες εξόδου είναι ουσιαστικά οι συντακτικοί αναλυτές. Προσέξτε τώρα ότι ο αναλυτής αυτός δουλεύει *από πάνω προς τα κάτω*: κάθε μεταβλητή που αφαιρείται από το σωρό αντιστοιχεί στην κορυφή ενός υποδέντρου του οποίου τα παιδιά σχηματίζουν το δεξιό μέρος του αντίστοιχου κανόνα. Αυτή η διαδικασία συνεχίζεται αναδρομικά με τα παιδιά του κόμβου, με αποτέλεσμα το δέντρο να κτίζεται από πάνω προς τα κάτω.

Όμως το παραπάνω αυτόματο δεν πληρεί όλες τις προϋποθέσεις που πρέπει να έχει ένας συντακτικός αναλυτής. Αν προσέξετε καλύτερα θα δείτε ότι το M δεν είναι ντετερμινιστικό αφού όταν στην κορυφή του σωρού είναι το T το αυτόματο θα πρέπει να μαντέψει ποιες από τις δύο κινήσεις 4 και 5 να χρησιμοποιήσει. Η κίνηση 4 αντικαθιστά το T με $0T1$, το οποίο είναι σωστό αν το *επόμενο* σύμβολο εισόδου είναι το 0, ενώ σε οποιαδήποτε άλλη περίπτωση θα πρέπει να αφαιρεθεί το T από το σωρό. Φαίνεται λοιπόν ότι μπορούμε να εξαλείψουμε το μη ντετερμινισμό κοιτώντας "μπροστά" ένα σύμβολο, χρησιμοποιώντας δηλαδή το επόμενο σύμβολο εισόδου για να εντοπίσουμε τη σωστή κίνηση. Έτσι αν το σύμβολο εισόδου είναι 0, το αυτόματο θα πρέπει να αντικαταστήσει το T με $0T1$, ενώ αν το σύμβολο είναι 1, θα πρέπει να αφαιρέσει το T από το σωρό. Ο τρόπος με τον οποίο το M καταφέρνει να *προβλέπει* τη σωστή κίνηση είναι ενσωματώνοντας το επόμενο σύμβολο σε μία από τις καταστάσεις του.

Αριθμός Κίνησης	**Κατάσταση**	**Σύμβολο Εισόδου**	**Σύμβολο Σωρού**	**Κινήσεις**	**Επεξηγήσεις**
0	p	ϵ	Z_0	(q, SZ_0)	
1	q	ϵ	S	$(q, T\$)$	$S \to T\$$
2	q	0	ϵ	(q_0, ϵ)	Πρόβλεψη 0
2.1	q_0	ϵ	0	(q, ϵ)	Ταίριασμα 0
3	q	1	ϵ	(q_1, ϵ)	Πρόβλεψη 1
3.1	q_1	1	ϵ	(q, ϵ)	Ταίριασμα 1
4	q_0	ϵ	T	$(q_0, 0T1)$	$T \to 0T1$
5	q_1	ϵ	T	(q_1, ϵ)	$T \to \epsilon$
6	q	\$	\$	(q, ϵ)	Ταίριασμα \$
7	q_1	ϵ	Z_0	(f, Z_0)	
(Οι υπόλοιποι συνδυασμοί)					(Τίποτα)

Από την κατάσταση q το αυτόματο διαβάζει ένα σύμβολο (0 ή 1) και μπαίνει σε μία από τις νέες καταστάσεις q_0, q_1 (κινήσεις 2 και 3). Από αυτές τις καταστάσεις ξέρει ότι θα πρέπει να ταιριάξει αντίστοιχα το 0 ή 1 στην κορυφή του σωρού (κινήσεις 2.1 και 3.1). Αν όμως στην κορυφή βρίσκεται το σύμβολο T, τότε από την πρώτη κατάσταση θα σπρώξει το $0T1$ στο σωρό (κίνηση 4), ενώ από τη δεύτερη απλά θα αφαιρέσει το T (κίνηση 5). Άρα το αυτόματο είναι ντετερμινιστικό και αντιστοιχεί σε ένα αναλυτή από πάνω προς τα κάτω με *πρόβλεψη ενός συμβόλου*.

Δυστυχώς η παραπάνω ιδέα δεν μπορεί να εφαρμοστεί σε όλα τα αυτόματα που προκύπτουν από τη εφαρμογή του Θεωρήματος 8.2. Ο λόγος είναι ότι υπάρχουν γλώσσες που γίνονται δεκτές από μη ντετερμινιστικά αυτόματα αλλά από κανένα ντετερμινιστικό. Έτσι δεν μπορούν να προκύψουν σωστοί αναλυτές ακόμα και αν χρησιμοποιήσουμε πρόβλεψη περισσοτέρων συμβόλων. Υπάρχουν όμως περιπτώσεις στις οποίες το αυτόματο μπορεί να γίνει ντετερμινιστικό αν η γραμματική τροποποιηθεί λιγάκι.

Ας εξετάσουμε την πρώτη απο αυτές τις περιπτώσεις με τη χρήση ενός παραδείγματος. Θεωρείστε τη γραμματική με κανόνες

$$S \rightarrow aSb \mid abS \mid aSbS \mid ab$$

η οποία προκύπτει από τη γραμματική

$$S \rightarrow aSbS \mid \epsilon$$

αφαιρώντας τις ϵ-παραγωγές. Ένας αναλυτής για αυτή τη γλώσσα θα είχε πρόβλημα να αποφασίσει για τη σωστή κίνηση μιας και όλοι οι κανόνες αρχίζουν με το ίδιο σύμβολο a. Έτσι, ακόμα και αν κοιτάξουμε μπροστά ένα σύμβολο, αν το σύμβολο στο σωρό είναι το S δεν θα μπορέσουμε να διαλέξουμε το σωστό κανόνα αφού και οι τέσσερις αρχίζουν με a. Η λύση εδώ είναι να *παραγοντοποιήσουμε* τα δεξιά μέρη των κανόνων και να αντικαταστήσουμε τα μη κοινά μέρη με νέες μεταβλητές. Η παραπάνω γραμματική γίνεται λοιπόν:

$$\begin{array}{l} S \rightarrow aT \\ T \rightarrow Sb \mid bS \mid SbS \mid b \end{array}$$

Τώρα όμως χρειάζεται να παραγοντοποιήσουμε το δεξί μέρος των κανόνων που προκύπτουν από το T γιατί κάποιοι από αυτούς αρχίζουν με μία κοινή συμβολοσειρά, την "Sb". Εισάγοντας μία νέα μεταβλητή X και παραγοποιώντας το "Sb" παίρνουμε τους κανόνες

$$\begin{array}{l} S \rightarrow aT \\ T \rightarrow SbX \mid bX \\ X \rightarrow S \mid \epsilon \end{array}$$

για τους οποίους μπορεί εύκολα να προκύψει ένας ντετερμινιστικός αναλυτής χρησιμοποιώντας πρόβλεψη ενός συμβόλου. Η παραπάνω μέθοδος εξάλειψης του μη ντετερμινισμού ονομάζεται *αριστερή παραγοντοποίηση* και μπορεί να περιγραφεί ως παρακάτω:

Αριστερή Παραγοντοποίηση:

Όταν σε μία γραμματική υπάρχουν οι κανόνες $A \rightarrow \alpha\beta$ και $A \rightarrow \alpha\gamma$, όπου $\alpha \neq \epsilon$, τότε εισάγουμε μία νέα μεταβλητή A' και τους αντικαθιστούμε από τους κανόνες $A \rightarrow \alpha A'$ και $A' \rightarrow \beta \mid \gamma$.

Θα προχωρήσουμε τώρα στη δεύτερη των περιπτώσεων στις οποίες μπορεί εύκολα να εξαλειφθεί ο μη ντετερμινισμός με μία μικρή τροποποίηση της γραμματικής. Θεωρείστε τη γραμματική που παράγει ένα υποσύνολο των σωστών αλγεβρικών εκφράσεων:

$$\begin{array}{l} S \to S+T \\ S \to T \\ T \to T*x \\ T \to x \end{array}$$

Έστω ότι το αυτόματο βρίσκεται σε μία κατάσταση στην οποία το επόμενο σύμβολο εισόδου είναι το x και στην κορυφή του σωρού βρίσκεται το S. Το αυτόματο τώρα έχει τις παρακάτω επιλογές: η επόμενη κίνηση μπορεί απλά να αντικαταστήσει το S με T αν η είσοδος είναι το x. Ή μπορεί να αντικαταστήσει το S με $S+T$ και ξανά το νέο S με T αν η είσοδος είναι το $x+x$. Ή τέλος να αντικαταστήσει το νέο S με $S+T$ αν η είσοδος είναι το $x+x+x$, κοκ. Φαίνεται λοιπόν ότι όσο μπροστά και να κοιτάξει το αυτόματο πάντα μπορεί να υπάρξει μία είσοδος η οποία να το οδηγήσει σε λάθος πρόβλεψη. Το πρόβλημα εδώ είναι οι κανόνες της μορφής $S \to S+T$ στους οποίους η μεταβλητή στο αριστερό μέρος εμφανίζεται στην *αρχή* του δεξιού μέρους. Το φαινόμενο αυτό λέγεται *αριστερή αναδρομή* και μπορεί να εξαλειφθεί με τη χρήση των παρακάτω κανόνων.

Αριστερή Αναδρομή:

Όταν σε μία γραμματική υπάρχουν οι κανόνες $A \to A\alpha_i$, $1 \leq i \leq m$ και $A \to \beta_i$, $1 \leq i \leq n$ όπου τα β_i δεν αρχίζουν με τη μεταβλητή A, τότε εισάγουμε μία νέα μεταβλητή A' και τους αντικαθιστούμε από τους κανόνες $A \to \beta_i A'$, $A' \to \alpha_i A'$ και $A' \to \epsilon$, για κάθε i όπως παραπάνω.

Ας προσπαθήσουμε να εφαρμόσουμε τις παραπάνω ιδέες για να εξαλείψουμε την αριστερή αναδρομή από τους κανόνες $S \to S+T \mid T$. Εδώ το ρόλο του α παίζει η συμβολοσειρά $+T$ και του β η T. Εισάγοντας λοιπόν μία νέα μεταβλητή S' παίρνουμε τους κανόνες

$$\begin{array}{l} S \to TS' \\ S' \to +TS' \\ S' \to \epsilon \end{array}$$

Αντίστοιχα για τους κανόνες $T \to T*x \mid x$ παίρνουμε τις ισοδύναμες εκφράσεις:

$$\begin{array}{l} T \to xT' \\ T' \to *xT' \\ T' \to \epsilon \end{array}$$

Έτσι η γραμματική μετατρέπεται σε μία ισοδύναμη, η οποία δεν περιέχει αριστερή αναδρομή και επομένως μπορεί να αναγνωριστεί από ένα ντετερμινιστικό αυτόματο με πρόβλεψη ενός συμβόλου. Γραμματικές σαν την παραπάνω που προκύπτουν με εξάλειψη της αναδρομής και πιθανώς αριστερή παραγοντοποίηση ονομάζονται *LL(1)* γιατί μπορούν να παράγουν αποτελεσματικούς αναλυτές με πρόβλεψη ενός συμβόλου. Αν απαιτούνται $k \geq 1$ σύμβολα για τη σωστή αναγνώρισή τους, τότε ονομάζονται *LL(k)*.

Άσκηση Αυτοαξιολόγησης 10.1 Για την ακόλουθη γραμματική, βρείτε μία ισοδύναμη $LL(1)$ γραμματική.

$$
\begin{aligned}
S &\to aAb \mid aAA \mid aB \mid bbA \\
A &\to aAb \mid ab \\
B &\to bBa \mid ba
\end{aligned}
$$

10.2 Ανάλυση από κάτω προς τα πάνω

Στην προηγούμενη ενότητα είδαμε πως μπορεί να προκύψει ένας αναλυτής ο οποίος χτίζει μία αριστερή παραγωγή από πάνω προς τα κάτω, δηλαδή αντικαθιστώντας τη μεταβλητή στην κορυφή του σωρού (τη ρίζα ενός υποδέντρου) με το δεξιό μέρος του αντίστοιχου κανόνα (τα παιδιά αυτού του κόμβου).

Στην αντίθετη προσέγγιση η κατασκευή του δέντρου γίνεται από κάτω προς τα πάνω. Όταν στο σωρό υπάρχουν τα παιδιά ενός κόμβου (το δεξιό μέρος ενός κανόνα) αυτά αντικαθίστανται από τη ρίζα του υποδέντρου (τη μεταβλητή στο αριστερό μέρος του κανόνα). Η λειτουργία αυτή επιτυγχάνεται με δύο τύπους κινήσεων: τη "μετακίνηση" των συμβόλων εισόδου στο σωρό μέχρι να σχηματιστεί το δεξιό μέρος κάποιου κανόνα και την "αναγωγή" αυτής της ακολουθίας συμβόλων, δηλαδή την αντικατάστασή της από τη μεταβλητή στο αριστερό μέρος του κανόνα. Επειδή όμως στο τέλος κάθε αναγωγής η μεταβλητή στην κορυφή του σωρού είναι η *δεξιότερη* της παραγόμενης συμβολοσειράς, η παραγωγή που εξομοιώνεται είναι μία δεξιά παραγωγή. Ένα τέτοιο αυτόματο με τους δύο αυτούς τύπους κινήσεων ονομάζεται *αναλυτής μετακίνησης-αναγωγής* και μπορεί να σχηματιστεί για κάθε γραμματική με τρόπο ανάλογο του Θεωρήματος 8.2.

Θεώρημα 10.1 *Κάθε γλώσσα ανεξάρτητη συμφραζόμενων μπορεί να αναγνωριστεί από ένα αυτόματο στοίβας που λειτουργεί από "κάτω προς τα πάνω".*

Απόδειξη: Έστω $G = (V, \Sigma, S, R)$ μία γραμματική ΑΣ. Το αυτόματο $M = (Q, \Sigma, \Gamma, q_0, Z_0, \delta, F)$ που εξομοιώνει τη G δίνεται από τα παρακάτω:

$Q = \{p, f\}$

$\Gamma = V \cup \Sigma \cup \{Z_0\}$, όπου $Z_0 \notin V \cup \Sigma$

$F = \{f\}$

Η συνάρτηση μετάβασης δ ορίζεται ως εξής:

Αριθμός Κίνησης	Κατάσταση	Σύμβολο Εισόδου	Σύμβολα Σωρού	Κινήσεις	Επεξηγήσεις
1	p	a	ϵ	(p, a)	Μετακίνηση συμβόλων, για κάθε $a \in \Sigma$.
2	p	ϵ	α^R	(p, A)	Αναγωγή α^R με A, για κάθε κανόνα $A \to \alpha$.
3	p	ϵ	S	(f, ϵ)	Τελική κίνηση

Οι κινήσεις τύπου 1 μετακινούν τα σύμβολα εισόδου στο σωρό. Οι κινήσεις τύπου 2 αντικαθιστούν το *αντίστροφο* δεξιό μέρος κάποιου κανόνα στο σωρό με το αντίστοιχο αριστερό μέρος. Αυτό συμβαίνει γιατί καθώς τα σύμβολα εισόδου μετακινούνται στο σωρό, αυτά θα σχηματίζουν το δεξιό μέρος κάποιου κανόνα, αλλά κατά την αντίστροφη φορά. Τέλος, όταν στο σωρό μείνει μόνο η αρχική μεταβλητή S τότε το αυτόματο μπαίνει στην τελική του κατάσταση με την κίνηση 3.

Προσέξτε ότι εδώ έχουμε τροποποιήσει λίγο τον ορισμό της συνάρτησης μετάβασης ενός αυτόματου στοίβας. Οι κινήσεις τύπου 2 επιτρέπουν την αφαίρεση μίας ολόκληρης συμβολοσειράς α^R από το σωρό αντί τη συνήθη αφαίρεση ενός συμβόλου όπως είχαμε ορίσει στην Ενότητα 8.2. Αυτή όμως η αλλαγή δεν είναι ουσιαστική αφού εισάγοντας νέες καταστάσεις μπορούμε να αφαιρέσουμε το α^R ένα σύμβολο κάθε φορά.

Θα δείξουμε τώρα ότι η μηχανή αναγνωρίζει την ίδια γλώσσα με αυτή που παράγεται από τη γραμματική G, δηλαδή $L(M) = L(G)$. Η απόδειξη είναι ανάλογη του Θεωρήματος 8.2. Η μόνη διαφορά είναι ότι τώρα μιλάμε για δεξιές παραγωγές, δηλαδή παραγωγές στις οποίες μόνο η δεξιότερη μεταβλητή αντικαθίσταται κάθε φορά.

Ο ισχυρισμός που επαρκεί για την απόδειξη και αφήνεται σαν άσκηση για τον αναγνώστη (Άσκηση 10.3) είναι ο ακόλουθος:

> Αν $S \stackrel{*}{\Rightarrow} \alpha x$ με μία δεξιά παραγωγή, όπου $x \in \Sigma^*$ και $\alpha \in (V \cup \Sigma)^*$,
> τότε $(p, x, \alpha^R Z_0) \models^* (p, \epsilon, SZ_0)$

Παρατηρείστε ότι θέτοντας $\alpha = \epsilon$, η x παράγεται από τη γραμματική αν και μόνο αν $(p, x, Z_0) \models^* (p, \epsilon, SZ_0) \models (f, \epsilon, Z_0)$, δηλαδή αν και μόνο αν η x γίνεται δεκτή από το αυτόματο. Άρα οι δύο γλώσσες είναι ίδιες. $\diamondsuit$

Ας εφαρμόσουμε τώρα τις παραπάνω ιδέες για να κατασκευάσουμε ένα μη ντετερμινιστικό αναλυτή για τη γραμματική που ακολουθεί:

$$
\begin{aligned}
&S \to E\$ \\
&E \to E + T \\
&E \to T \\
&T \to T * x \\
&T \to x
\end{aligned}
$$

Αριθμός Κίνησης	**Κατάσταση**	**Σύμβολο Εισόδου**	**Σύμβολα Σωρού**	**Κινήσεις**	**Επεξηγήσεις**
1	p	a	ϵ	(p, a)	Μετακίνηση $a \in \{x, +, *, \$\}$
2	p	ϵ	$\$E$	(p, S)	Αναγωγή $S \rightarrow E\$$
3	p	ϵ	$T + E$	(p, E)	Αναγωγή $E \rightarrow E + T$
4	p	ϵ	T	(p, E)	Αναγωγή $E \rightarrow T$
5	p	ϵ	$x * T$	(p, T)	Αναγωγή $T \rightarrow T * x$
6	p	ϵ	x	(p, T)	Αναγωγή $T \rightarrow x$
7	p	ϵ	S	(f, ϵ)	
	(Οι υπόλοιποι συνδυασμοί)				(Τίποτα)

Παρατηρείστε ότι η κίνηση 3, η οποία αφαιρεί το $T + E$ από το σωρό και το αντικαθιστά με E μπορεί να εξομοιωθεί από την παρακάτω ακολουθία κινήσεων:

Αριθμός Κίνησης	**Κατάσταση**	**Σύμβολο Εισόδου**	**Σύμβολο Σωρού**	**Κινήσεις**	**Επεξηγήσεις**
3.1	p	ϵ	T	(p_T, ϵ)	Αφαίρεση T
3.2	p_T	ϵ	$+$	(p_+, ϵ)	Αφαίρεση $+$
3.3	p_+	ϵ	E	(p, E)	Αφαίρεση E και αναγωγή

Το παραπάνω αυτόματο όμως δεν είναι ντετερμινιστικό για δύο κυρίως λόγους. Ο πρώτος λόγος είναι ότι δεν είναι ξεκάθαρο πότε πρέπει να γίνει η μετακίνηση ενός συμβόλου στο σωρό και πότε η αναγωγή μίας συμβολοσειράς στην κορυφή του σωρού. Για παράδειγμα, αν T είναι το σύμβολο στο σωρό, αυτό μπορεί να είναι το T στο δεξιό μέρος του κανόνα $T \rightarrow T * x$, οπότε η μετακίνηση είναι η σωστή κίνηση, ή να είναι το T στον κανόνα $S \rightarrow S + T$, οπότε η αναγωγή είναι απαραίτητη. Ο δεύτερος λόγος είναι ότι ακόμη και στην περίπτωση που ξέρουμε ότι πρέπει να ανάγουμε μπορεί να υπάρχουν περισσότερες από μία δυνατές αναγωγές, όπως στην περίπτωση των κανόνων $T \rightarrow T * x$ και $T \rightarrow x$ των οποίων τα δεξιά τους μέρη τελειώνουν σε x. Είναι όμως αρκετό κάθε φορά να επιλέγουμε τη μεγαλύτερη συμβολοσειρά στην κορυφή του σωρού που ταιριάζει στο δεξιό μέρος κάποιου κανόνα της γραμματικής. Άρα στην παραπάνω περίπτωση θα πρέπει να χρησιμοποιήσουμε τον πρώτο κανόνα και να ανάγουμε το $T + x$ στο T.

Θα δούμε τώρα πως να επιλέγουμε μεταξύ μετακίνησης ενός συμβόλου στο σωρό ή αναγωγής κάποιας συμβολοσειράς στο σωρό. Έστω ότι στην κορυφή του σωρού βρίσκεται το σύμβολο a, το οποίο μπορεί να είναι μεταβλητή, ενώ το σύμβολο εισόδου είναι το b. Η απόφαση μεταξύ μετακίνησης ή αναγωγής βασίζεται σε μία σχέση $R \subseteq (V \cup \Sigma) \times \Sigma$, η οποία λέγεται σχέση *προτεραιότητας*. Έτσι αν το ζεύγος (a, b) ανήκει στην R θα λέμε ότι το a έχει *προτεραιότητα* ή *προηγείται* του b και η σωστή κίνηση θα είναι η αναγωγή. Για να δούμε γιατί έστω μία δεξιά παραγωγή της μορφής $S \overset{*}{\Rightarrow} \gamma Abx \Rightarrow \gamma\beta abx$. Εφόσον χτίζουμε τις παραγωγές από δεξιά προς τα αριστερά, είναι λογικό να αντικαταστήσουμε τη συμβολοσειρά $(\beta a)^R$ στο σωρό με το A του κανόνα $A \rightarrow \beta a$ όταν το a προηγείται του b, ή με άλλα λόγια όταν $(a, b) \in R$.

Ο αναλυτής για τη γραμματική του παραδείγματός μας, με τα διάφορα είδη κινήσεων, φαίνεται στον Πίνακα 10.1.

Αριθμός Κίνησης	**Κατάσταση**	**Σύμβολο Εισόδου**	**Σύμβολο Σωρού**	**Κινήσεις**	**Επεξηγήσεις**
			Μετακίνηση Συμβόλων		
1	p	a	Z	(p, aZ)	$a \in \Sigma$, $Z \in \{Z_0, E, +, *\}$
2	p	a	T	(p, aT)	$a \notin \{+, \$\}$
			Αναγωγή $S \rightarrow E\$$		
3	p	ϵ	$\$$	$(p_\$, \epsilon)$	
4	$p_\$$	ϵ	E	(p, S)	
			Αναγωγή $T \rightarrow T * x$ **ή** $T \rightarrow x$		
5	p	ϵ	x	$(p_{x,1}, \epsilon)$	
6	$p_{x,1}$	ϵ	$*$	$(p_{x,2}, \epsilon)$	
7	$p_{x,2}$	ϵ	T	(p, T)	$T \rightarrow T * x$
8	$p_{x,1}$	ϵ	Z	(p, TZ)	$T \rightarrow x$, $Z \neq *$
			Αναγωγή $E \rightarrow E + T$ **ή** $E \rightarrow T$ **και μετακίνηση συμβόλου** $a \in \{+, \$\}$		
9	p	a	T	$(p_{T,a,1}, \epsilon)$	
10	$p_{T,a,1}$	ϵ	$+$	$(p_{T,a,2}, \epsilon)$	
11	$p_{T,a,2}$	ϵ	E	(p, aE)	$E \rightarrow E + T$
12	$p_{T,a,1}$	ϵ	Z	(p, TZ)	$E \rightarrow T$, $Z \neq +$
			Τελική κίνηση		
13	p	ϵ	S	(f, ϵ)	
		(Οι υπόλοιποι συνδυασμοί)			(Τίποτα)

Πίνακας 10.1: Συντακτικός αναλυτής μετακίνησης-αναγωγής

Όπως φαίνεται από τους παραπάνω πίνακες τα σύμβολα στο σωρό είναι τεσσάρων ειδών: αυτά που απαιτούν μετακίνηση του επόμενου συμβόλου στο σωρό (Z_0, E, $+$ και $*$), αυτά που απαιτούν αναγωγή (x, $\$$), αυτά που οδηγούν στην αναγνώριση της συμβολοσειράς (S) και τέλος αυτά που η σωστή κίνηση εξαρτάται από το επόμενο σύμβολο εισόδου (T).

Γραμματικές όπως η παραπάνω για τις οποίες μπορεί να προκύψουν αναλυτές βασιζόμενοι σε σχέσεις προτεραιότητας, ονομάζονται *γραμματικές προτεραιότητας*. Ειδικά όταν κατά τη διάρκεια μίας αναγωγής αρκεί να ανάγουμε τη μεγαλύτερη δυνατή συμβολοσειρά, τότε η γραμματική λέγεται *ασθενής γραμματική προτεραιότητας*.

10.3 ΑΣΚΗΣΕΙΣ

10.1 Κατασκευάστε ντετερμινιστικά αυτόματα για τις παρακάτω γραμματικές.

a) $S \to aA \quad A \to aA \mid bA \mid \epsilon$

b) $S \to aSb \mid bSa \mid \epsilon$

10.2 Μετατρέψτε τις παρακάτω γραμματικές σε $LL(1)$ και δώστε τους αντίστοιχους συντακτικούς αναλυτές

a) $S \to () \mid x \mid (A) \quad A \to S \mid A, S$

b) $S \to SA \mid ab \quad A \to aAbb \mid ab$

10.3 Ολοκληρώστε την απόδειξη του Θεωρήματος 10.1 αποδεικνύοντας τον ακόλουθο ισχυρισμό:

Αν $S \overset{*}{\Rightarrow} \alpha x$ με μία δεξιά παραγωγή, όπου $x \in \Sigma^*$ και $\alpha \in (V \cup \Sigma)^*$,
τότε $(p, x, \alpha^R Z_0) \models^* (p, \epsilon, SZ_0)$

10.4 Αποδείξτε ότι η μέθοδος εξάλειψης της αριστερής αναδρομής είναι σωστή, έχει σαν αποτέλεσμα δηλαδή τη δημιουργία ισοδύναμων γραμματικών.

10.5 Κατασκευάστε έναν αναλυτή από κάτω προς τα πάνω για τη γλώσσα $L\$$ της Άσκησης 10.2(a). Μπορείτε να βρείτε τη σχέση προτεραιότητας;

10.6 Η παρακάτω γραμματική είναι επίσης ασθενής γραμματική προτεραιότητας. Δώστε έναν αναλυτή από κάτω προς τα πάνω για τη γλώσσα $L\$$.

$$\begin{aligned}
&S \to E\$ \\
&E \to E + T \mid E - T \mid T \\
&T \to T * F \mid T/F \mid F \\
&F \to (E) \mid x
\end{aligned}$$

Σύνοψη

Στο κεφάλαιο αυτό παρουσιάσαμε τις δύο βασικές μεθόδους *συντακτικής ανάλυσης* κατά τις οποίες ανακαλύπτεται η ακολουθία των κανόνων από τους οποίους παράγεται μία