Οι κανονικές γλώσσες χαρακτηρίζονται από κάποιες ιδιότητες. Οι ιδιότητες αυτές μπορούν να χρησιμοποιηθούν για να αποδείξουμε ότι μία γλώσσα δεν είναι κανονική. Οι ιδιότητες αποδεικνύονται με την βοήθεια του Λήμματος Αντλησης είτε με απλούστερες μεθόδους.

'Ενα σύνολο ονομάζεται κλειστό ως προς κάποια πράξη αν το αποτέλεσμα της πράξης ανάμεσα σε στοιχεία του συνόλου είναι στοιχείο του συνόλου. Παρ. Οι φυσικοί αριθμοί είναι σύνολο κλειστό ως προς την πρόσθεση και τον πολλαπλασιασμό, δεν είναι κλειστό ως προς την αφαίρεση. Οι τετραγωνικοί πίνακες με ορίζουσα ίση με 0 δεν είναι σύνολο κλειστό ως προς την πρόσθεση πινάκων

Οι κανονικές γλώσσες είναι κλειστές ως προς τις πράξεις παράθεση, κλειστότητα Kleene, τομή, ένωση, συμπλήρωμα, συνολοθεωρητική διαφορά. Χρησιμοποιούμε τις ιδιότητες των κανονικών γλωσσών με τον εξής τρόπο: Αν κατασκευάσουμε μία μη κανονική γλώσσα με τις πράξεις για τις οποίες οι κανονικές γλώσσες είναι κλειστές, τότε μία από τις αρχικές γλώσσες δεν είναι κανονική.

Η απόδειξη γίνεται κατασκευάζοντας ΠΑ ώστε να αναγνωρίζουν τις σχετικές γλώσσες.

Για μία συμβολοσειρά $x = a_1 a_2 \dots a_{n-1} a_n$ ορίζουμε την αντίστροφη (ανάστροφη) της x και την συμβολίζουμε με x^r να είναι η συμβολοσειρά $a_n a_{n-1} \dots a_2 a_1$. Για μία γλώσσα L ορίζουμε την αντίστροφή της $L^r = \{\Sigma^* : x^r \in L\}$. Ισχύει ότι για κανονικές γλώσσες η αντίστροφη γλώσσα είναι επίσης κανονική γλώσσα.

Κατασκευάζουμε ΠΑ ώστε να αναγνωρίζει τις αντίστροφες συμβολοσειρές της γλώσσας L. Αν το ΠΑ M αναγνωρίζει την L τότε αντιστρέφουμε τα βέλη στο διάγραμμα, αλλάζουμε την αρχική κατάσταση του M σε τελική, προσθέτουμε μία νέα αρχική κατάσταση και e μεταβάσεις στις τελικές καταστάσεις του M και έχουμε αυτόματο M ώστε να αναγνωρίζει την L^r. Δηλ. αν $\delta(q, a) = r$ ορίζουμε $\delta(r, a) = q$.

Ομομορφισμός. Οι ομομορφισμοί είναι συναρτήσεις που διατηρούν κάποια δομή. Παρ. οι γραμμικές συναρτήσεις στην γραμμική άλγεβρα διατηρούν τις πράξεις πολλαπλασιασμός με βαθμωτό και πρόσθεση. Η συνάρτηση λογάριθμος από τους πραγματικούς αριθμούς με πράξη την πρόσθεση, στους θετικούς πραγματικούς με πράξη τον πολλαπλασιασμό, $\log(x + y) = \log(x) \log(y)$.

Για δύο γλώσσες L_1 στο αλφάβητο Σ_1, L_2 στο αλφάβητο Σ_2 αν έχουμε την συνάρτηση $h : \Sigma_1 \to \Sigma_2^*$ την επεκτείνουμε με επαγωγή σε συμβολοσειρές της L_1 σαν $h(e) = e$, $h(ax) = h(a)h(x)$, με $a \in \Sigma_1$, $x \in \Sigma_1^*$. Οι ομομορφισμοί στις γλώσσες διατηρούν την παράθεση συμβολοσειρών.

Παρ. Αν $h(0) = ab$, $h(1) = e$, τότε $h(0110) = abab$.

Ανάλογα ορίζουμε την συνάρτηση h σε γλώσσες με τον τύπο $h(L) = \{h(x) : x \in L\}$.

Θεώρημα: Αν $L \subseteq \Sigma_1^*$ είναι κανονική και η συνάρτηση h είναι ομομορφισμός τότε το σύνολο $h(L)$ είναι κανονική γλώσσα.

Απόδ. Αν $h(a) = s_1 s_2 \dots s_k$, στο αυτόματο το οποίο αναγνωρίζει την L αντικαθιστούμε για κάθε μετάβαση με σύμβολο το a μετάβαση με τα σύμβολα $s_1 s_2 \dots s_k$. Προσθέτουμε νέες καταστάσεις ώστε να έχουμε ανάγνωση ενός συμβόλου και όχι της ακολουθίας συμβόλων $s_1 s_2 \dots s_k$. Διατηρούμε τις ίδιες τελικές και αρχική κατάσταση.

Αντίστροφος ομομορφισμός.

Για μία γλώσσα L και ομομορφισμό h ορίζουμε $h^{-1}(L) = \{x \mid h(x) \in L\}$. Παρ. απαλείφουμε το σύμβολο c από μία γλώσσα, δηλ. $h(a) = a$, $h(b) = b$, $h(c) = e$. Για την γλώσσα $L = (abc)^*$ έχουμε $h(L) = (ab)^*$. Ποια είναι η μεγαλύτερη γλώσσα L_1 ώστε $h(L_1) = L$; Μία περιγραφή είναι η $h^{-1}(L)$ η οποία δεν είναι η L_1 πάντοτε αλλά ίσως μεγαλύτερο σύνολο . $h^{-1}\left[(ab)^*\right] = c^*\left(c^*ac^*bc^*\right)^* c^*$.

Θεώρημα: Αν $L \subseteq \Sigma_1$ είναι κανονική γλώσσα και $h : \Sigma_2 \to \Sigma_1^*$ είναι ομομορφισμός, τότε η γλώσσα $h^{-1}(L)$ είναι κανονική.

Απόδειξη: Αν $(Q, \Sigma_1, \delta, q_0, F)$ είναι ένα ΠΑ το οποίο αναγνωρίζει την L, το ΠΑ $(Q, \Sigma_2, \gamma, q_0, F)$ με $\gamma(q, a) = \widehat{\delta}(q, h(q))$ αναγνωρίζει την $h^{-1}(L)$.

Αν δ είναι η συνάρτηση μετάβασης ενός ΠΑ, ορίζουμε την συνάρτηση $\widehat{\delta}$ σε συμβολοσειρές με επαγωγή $\widehat{\delta}(q, a) = \delta(q, a)$ για $a \in \Sigma$, $\widehat{\delta}(q, ax) = \widehat{\delta}(q_1, x)$ όπου $\delta(q, a) = q_1$. Η συνάρτηση $\widehat{\delta}$ ονομάζεται επέκταση της δ σε συμβολοσειρές και σχετίζεται με την μεταβατική κλειστότητα.

Συνέχεια της απόδειξης. Με επαγωγή στο μήκος μιάς συμβολοσειράς x μπορούμε να αποδείξουμε ότι $\widehat{\gamma}(q,x) = \widehat{\delta}(q,h(x))$.

Αλγόριθμοι απόφασης για κανονικές γλώσσες.

Τα ΠΑ είναι απλοί μηχανισμοί σχεδιασμένοι ώστε να απαντούν σε ερωτήματα της μορφής $x \in L$ με απάντηση ΝΑΙ – ΟΧΙ. Θα δούμε προβλήματα τα οποία εκφράζονται με μηχανές Turing όπου η απάντηση είναι μόνο ΝΑΙ με κάποια έννοια. Συγκεκριμένα αν η απάντηση είναι ΝΑΙ έχουμε την σωστή απάντηση, αν η απάντηση είναι ΟΧΙ είναι δυνατόν να μην το ανακαλύψουμε. Παρ. σε εξισώσεις μαντεύουμε την λύση και επαληθεύουμε, τι κάνουμε αν δεν υπάρχει αλγόριθμος για να βρούμε την λύση και αναζητούμε την λύση με εξαντλητική αναζήτηση και δεν υπάρχει λύση;

Μία άλλη κατηγορία ερωτημάτων είναι σχετικά με ιδιότητες της γλώσσας που αναγνωρίζει ένα ΠΑ.

Θεώρημα: Αν ένα ΠΑ M έχει n καταστάσεις τότε η γλώσσα που αναγνωρίζεται από το αυτόματο M είναι μη κενή αν υπάρχει συμβολοσειρά x με μήκος μικρότερο του n

ώστε η x να γίνεται δεκτή από το M. Η γλώσσα που αναγνωρίζεται από το αυτόματο M είναι άπειρη αν υπάρχει συμβολοσειρά y με μήκος μεταξύ του n και του $2n$ ώστε η y να γίνεται δεκτή από το M.

Τα αντίστοιχα προβλήματα για την γενική περίπτωση δηλ. μηχανές Turing δεν έχουν καταφατική απάντηση, δηλ. δεν υπάρχει αλγόριθμος ώστε να απαντάει με ΝΑΙ ΟΧΙ απάντηση στα παραπάνω ερωτήματα για γλώσσες γενικά.

Μη κενή γλώσσα.

Αν το αυτόματο M έχει n καταστάσεις και υπάρχει συμβολοσειρά x με μήκος μικρότερο του n που γίνεται δεκτή από το M τότε η γλώσσα είναι μη κενή. Για το αντίστροφο αν x είναι η μικρότερη συμβολοσειρά που αναγνωρίζεται από το M και η x έχει μήκος μεγαλύτερο του n τότε από το Λήμμα Αντλησης $x = uvw$ με v μη κενή. Τότε η συμβολοσειρά $xv^0w = xw$ ανήκει στην γλώσσα αντίφαση άρα το μήκος είναι μικρότερο από το n.

Απειρη γλώσσα.

Αν η x αναγνωρίζεται από το αυτόματο M και $|n| < |x| < 2n$ τότε η L είναι άπειρη, αφού οι συμβολοσειρές της μορφής $uv^m w$ ανήκουν στην L. Αν η L είναι άπειρη τότε υπάρχει συμβολοσειρά $x \in L$ με $|x| \geq n$. Αν x είναι η μικρότερη συμβολοσειρά σε μήκος και ισχύει $|x| \geq 2n$ εφαρμόζοντας το Λήμμα Αντλησης με $x = uvw$, με v μη κενή συμβολοσειρά, έχουμε ότι η συμβολοσειρά $uw \in L$ και έχει μήκος μικρότερο από το ελάχιστο μήκος που υποθέσαμε για την x.

Γλώσσες ανεξάρτητες συμφραζομένων ΓΑΣ – Context Free Languages CFL.

Οι κανονικές γλώσσες εκφράζουν το πρώτο βήμα για την εκτέλεση ενός προγράμματος την λεκτική ανάλυση, είναι απλές γλώσσες. Δεν απαιτείται μνήμη περισσότερη από τις καταστάσεις του αυτόματου το οποίο αναγνωρίζει την γλώσσα.

Παρ. η γλώσσα $0^n 1^n$ δεν είναι κανονική. Η γλώσσα αυτή είναι ένα απλό παράδειγμα όπου έχουμε ζευγάρια συμβόλων που θέλουμε να ταιριάζουν, (αριστερή δεξιά παρένθεση, `begin`, `end` σε γλώσσες προγραμματισμού

κλπ). Οι συμβολοσειρές που είναι παλινδρομικές δεν αποτελούν κανονική γλώσσα. Οι δύο παραπάνω γλώσσες αναγνωρίζονται με την προσθήκη στοίβας (stack) σε ένα αυτόματο. Για την γλώσσα 0^n1^n όσο διαβάζουμε 0 το τοποθετούμε στην στοίβα, όταν αρχίσουν τα σύμβολα 1 εξάγουμε για καθε 1 ένα 0 από την στοίβα. Αν καταλήξομε με άδεια στοίβα στο τέλος της συμβολοσειράς έχουμε την σωστή μορφή.

Για τα παλίνδρομα μία περιγραφή είναι $x = x^r$. Η περιγραφή αυτή δεν μας δίνει αλγόριθμο ώστε να ελέγχουμε αν έχουμε την σωστή μορφή. Μία άλλη προσέγγιση είναι ότι οι παλινδρομικές συμβολοσειρές είναι συμμετρικές ως προς το μέσον της συμβολοσειράς. Αν έχουμε στην διάθεσή μας στοίβα και μαντέψουμε το μέσον της συμβολοσειράς (μη αιτιοκρατία) τότε με την βοήθεια της στοίβας μπορούμε να αναγνωρίσουμε τις παλινδρομικές συμβολοσειρές. Τοποθετού στην στοίβα τα σύμβολα και μαντεύουμε αν φτάσαμε στο μέσον. Φτάνοντας στο μέσον της συμβολοσειράς εξάγουμε τα σύμβολα της στοίβας ένα ένα και συγκρίνουμε με το σύμβολο που διαβάζουμε. Αν δεν ταυτίζονται είτε περισσέψουν σύμβολα δεν έχουμε παλίνδρομο. Αν αδειάσει η στοίβα και ταυτόχρονα τελειώσει η συμβολοσειρά έχουμε παλίνδρομο.

Οι παλινδρομικές συμβολοσειρές δεν αποτελούν κανονική γλώσσα.

Ο μηχανισμός της στοίβας είναι μία μορφή μνήμης άπειρη σε χωρητικότητα αλλά με περιορισμό στην πρόσβαση, μπορούμε να δούμε μόνο το τελευταίο στοιχείο.

Η περιγραφή γλωσσών με αυτόματα μας δίνει αλγόριθμο που απαντάει στο ερώτημα αν ένα στοιχείο ανήκει σε κάποιο σύνολο (ικανοποιεί κάποιες προδιαγραφές). Μία εναλλακτική προσέγγιση είναι να παράγουμε τα στοιχεία ενός συνόλου. Η προσέγγιση με παραγωγή των στοιχείων ενός συνόλου είναι με γραμματικές – κανόνες παραγωγής.

Οι Γραμματικές Ανεξάρτητες Συμφραζομένων ΓΑΣ (Context Free Languages) περιγράφουν και απλές συντακτικές μορφές σε φυσικές γλώσσες.

Παρ. Οι παλινδρομικές συμβολοσειρές στο αλφάβητο $\{0, 1\}$ εκφράζονται με του κανόνες

$$P \to 0P0,$$
$$P \to 1P1,$$
$$P \to e$$

Οι συμβολοσειρές της μορφής 0^n1^n εκφράζονται με του κανόνες

$$P \longrightarrow 0P1,$$
$$P \longrightarrow e$$

Οι κανόνες αυτοί πχ. $P \longrightarrow 1P1$, ονομάζονται παραγωγές. Ένα σύνολο από κανόνες αυτής της μορφής ονομάζεται Γραμματική. Έχουμε μεταβλητές (variables), το σύμβολο P και ίσως κάποια άλλα σύμβολα (συνήθως κεφαλαία γράμματα του Λατινικού αλφάβητου) και τερματικά σύμβολα τα a, b, 0, 1, έχουμε δηλ. έμμεση απόδοση τυπων.

Με τις ΓΑΣ έχουμε κάποια μορφή αναδρομής αφού το ίδιο σύμβολο εμφανίζεται και στα δύο μέλη ενός κανόνα. Δεν έχουμε σαφές κριτήριο τερματισμού για την αναδρομή αλλά έχουμε τερματικά. Επειδή είναι δυνατόν να εφαρμόσουμε περισσότερους από ένα κανόνα έχουμε μη αιτιοκρατία (non determinism) Η στοίβα μας εκφράζει την έννοια της αναδρομής σε επίπεδο μηχανισμού.

Μία γραμματική ανεξάρτητη συμφραζομένων – ΓΑΣ είναι μία πεντάδα (V, T, P, S) όπου V είναι ένα σύνολο από μη

τερματικά σύμβολα, T είναι ένα σύνολο από τερματικά σύμβολα, τα σύνολα V, T είναι ξένα μεταξύ τους. Το σύνολο P είναι ένα σύνολο από παραγωγές $P \subseteq V \times (V \cup T)^*$, το σύμβολο S είναι το σύμβολο πρότασης (sentence) είτε τα αρχικό σύμβολο (start). Αν το σύμβολο S παραλείπεται τότε θεωρούμε το πρώτο μη τερματικό σύμβολο του πρώτου κανόνα σαν το αρχικό σύμβολο. Οι παραγωγές είναι της μορφής $A \longrightarrow aBc$. Το σύμβολο A στο αριστερό μέλος ονομάζεται κεφαλή (head) της παραγωγής, το σύνολο των συμβόλων στο δεξιό μέλος ονομάζεται σώμα (body) της παραγωγής. Οι παραγωγές στο αριστερό έχουν μόνο ένα σύμβολο. Ανεξάρτητα από την θέση του συμβόλου αυτού εφαρμόζουμε την παραγωγή, δεν εξαρτάται από την θέση του συμβόλου μέσα στην έκφραση, συμφραζόμενα.

Συμβολισμοί - συμβάσεις Τα σύμβολα a, b, c, 0, 1, ... είναι τερματικά. Τα σύμβολα A, B, C, ... είναι μη τερματικά.

Παρ. Η ακόλουθη ΓΑΣ παράγει τις κανονικές εκφράσεις

$$
\begin{aligned}
&S \to R,\\
&R \to \emptyset|a|b,\\
&R \to (R),\\
&R \to (R+R),\\
&R \to (RR),\\
&R \to R^*
\end{aligned}
$$

Αν έχουμε περισσότερες από μία παραγωγές με ίδιο αριστερό μέλος για συντομία τις συμπτύσσουμε στην μορφή $R \to \emptyset|a|b$ (Backus Naur μορφή).

Παράδειγμα παραγωγής της ΚΕ $(a+ba)^*$. $R \to (R) \to (R)^* \to (R+R)^* \to (R+(RR))^* \to (a+(RR))^* \to (a+(bR))^* \to (a+(ba))^*$.

Παραγωγές μίας γραμματικής

Μία αναλυτική (top down) προσέγγιση ώστε να περιγράψουμε την γλώσσα που παράγεται από μία ΑΣΓ είναι οι παραγωγές της γραμματικής. Ορίζουμε $\alpha A\beta \to \alpha\gamma\beta$ αν υπάρχει κανόνας - παραγωγή της μορφής $A \to \gamma$. Ορίζουμε την σχέση $\Rightarrow^*$ ανάμεσα σε δύο συμβολοσειρές x, y όταν μπορούμε να παράγουμε σε κάποιο αριθμό βημάτων σύμφωνα

με τους κανόνες την συμβολοσειρά y από την συμβολοσειρά x με επαγωγή.

1. $x \Rightarrow^* x$ για κάθε συμβολοσειρά x,

2. Αν $x \Rightarrow^* y$ και $y \Rightarrow^* z$, τότε $x \rightarrow^* z$.

3. Τίποτα άλλο δεν ανήκει στην σχέση $\Rightarrow^*$.

Για μία γραμματική G ορίζουμε την γλώσσα που παράγει η γραμματική $L(G) = \{w \mid w \in T^* \wedge S \rightarrow^* w\}$. Η γλώσσα η οποία αντιστοιχέι σε μία ΓΑΣ ονομάζεται Γλώσσα Ανεξάρτητη Συμφραζομένων (Context Free Language – CFL). Για την γλώσσα για τα παλίνδρομα έχουμε $P \rightarrow^* abba$ αφού $P \rightarrow aPa \rightarrow abba$.

Μία γραμματική μπορεί να παράγει μία συμβολοσειρά με περισσότερες από μία παραγωγές. Εξετάζουμε την γραμματική $G1$

$$\text{EXPR} \rightarrow \text{EXPR} + \text{TERM} \mid \text{TERM},$$
$$\text{TERM} \rightarrow \text{TERM} \times \text{FACTOR} \mid \text{FACTOR},$$
$$\text{FACTOR} \rightarrow (\ \text{EXPR}\ ) \mid a.$$

Δύο συμβολοσειρές που μπορούν να παραχθούν από την παραπάνω γραμματική είναι οι $(a + a) \times a$ και η $a + a \times a$.

Κατά την διαδικασία μετάφρασης - μεταγλώττισης ο μεταφραστής παράγει συμβολοσειρές κατάλληλες για την εκτέλεσης από τον Η/Υ. Για τον σκοπό αυτό ο μεταφραστής προσπαθεί να εξάγει την σημασιολογία του κώδικα μέσω της διαδικασίας της συντακτική ανάλυσης (parsing). Μία παράσταση της ερμηνείας - σημασίας μίας έκφρασης είναι μέσω της κατασκευής μίας δομής δεδομένων η οποία δομή παριστάνει την έκφραση (δενδρική δομή συνήθως).

Η γραμματική που περιγράφτηκε παράγει αριθμητικές εκφράσεις. Στην δενδρική δομή για την έκφραση $a + a \times a$ ο τελεστής $\times$ έχει σαν τελεστέους - ορίσματα τις εκφράσεις a, a, δηλ. έχουμε την συνηθισμένη προτεραιότητα των πράξεων. Στην έκφραση $(a + a) \times a$ έχουμε αλλαγή της προτεραιότητας των πράξεων μέσω των παρενθέσεων. Η γραμματική έχει δυνατότητα να εκφράσει την αλλαγή της προτεραιότητας αυτή.

Δέντρο συντακτικής ανάλυσης (parse trees)

Η γραμματική

$$\begin{aligned} \mathsf{A} &\to \mathsf{0\ A\ 1} \\ \mathsf{A} &\to \mathsf{B} \\ \mathsf{B} &\to \# \end{aligned}$$

παράγει την γλώσσα $L = \{0^n\#1^n \mid n \in \mathbb{N}\}$. Παράδειγμα παραγωγής $A \to 0A1 \to 00A11 \to 000A111 \to 000B111 \to 000\#111$. Μία εναλλακτική περιγραφή της παραγωγής μίας συμβολοσειράς είναι + & δέντρο συντακτικής ανάλυσης της συμβολοσειράς (parse tree). Σχηματίζουμε ένα δένδρο όπου οι εσωτερικοί κόμβοι αντιστοιχού σε παραγωγές οι οποίες περιέχουν μη τερματικά σύμβολα, τα φύλλα αντιστοιχούν σε παραγωγές οι οποίες περιέχουν μόνο τερματικά σύμβολα. Τα φύλλα του δένδρου μας δίνουν από αριστερά προς τα δεξιά την παραγόμενη συμβολοσειρά

Διφορούμενες Γραμματικές. Σε κάποιες περιπτώσεις μία γραμματική παράγει μία συμβολοσειρά με δύο είτε περισσότερες παραγωγές. Μία γραμματική παράγει μία συμβολοσειρά διφορούμενα (διττά) όταν μπορούμε να κατασκευάσουμε δύο διαφορετικά στην δομή τους δένδρα συντακτικής ανάλυσης για την συμβολοσειρά. Είναι δυνατόν να έχουμε διαφορετικές παραγωγές για μία συμβολοσειρά και να έχουμε το ίδιο δένδρο συντακτικής ανάλυσης. Τέτοιες περιπτώσεις παρουσιάζον όταν αλλάζουμε την σειρά εφαρμογής κανόνων για την παραγωγή της συμβολοσειράς αλλά να καταλήξουμε στο ίδιο δένδρο συντακτικής ανάλυσης. Για να αποφύγουμε περιπτώσεις όπου έχουμε διαφορετική σειρά εφαρμογής των κανόνων της γραμματικής ορίζουμε μία ειδική μορφή παραγωγής η οποία αντικαθιστά τις μεταβλητές με ορισμένο τρόπο. Μία παραγωγή μίας συμβολοσειράς ονομάζεται αριστερότερη παραγωγή (leftmost derivation) αν σε κάθε βήμα αντικαθιστούμε την αριστερότερα εμφανιζόμενη μεταβλητή. Μία γραμματική ονομάζεται

Η γραμματική $G2$

$$\text{EXPR} \rightarrow \text{EXPR} + \text{EXPR} \mid \text{EXPR} \times \text{EXPR} \mid (\ \text{EXPR}\ )$$

παράγει την συμβολοσειρά $a + a \times a$ με δύο τρόπους.
Η γραμματική δεν μπορεί να εκφράσει την προτεραιότητα
των πράξεων και μπορεί να εκτελέσει πρώτα είτε τον πολλαπλασιο
είτε την πρόσθεση.

Η γραμματική $G2$ και η γραμματική εκτός από τις παρενθέσεις
παράγουν τις ίδιες συμβολοσειρές. Με την $G1$ έχουμε ένα
μόνο δένδρο για κάθε συμβολοσειρά,

Παράδειγμα διφορούμενης πρότασης από φυσικές γλώσσες
Το αγόρι χάιδεψε το κορίτσι με το λουλούδι. Μπορούμε
να ερμηνεύσουμε με δύο διαφορετικούς τρόπους την πρόταση
και να έχουμε δύο διαφορετικές ερμηνείες – σημασίες.
Κάθε απόδοση ερμηνείας έχει διαφορετικό δέντρο συντακτικής
ανάλυσης.

Μία συμβολοσειρά παράγεται διφορούμενα αν έχει δύο
διαφορετικές αριστερότερες παραγωγές. Μία γραμματική
ονομάζεται διφορούμενη αν περιέχει μία τουλάχιστον διφορούμεν
συμβολοσειρά. Σε αρκετές περιπτώσεις μπορούμε να βρούμε
μη διφορούμενη γραμματική ισοδύναμη με μία διφορούμενη
γραμματική. Υπάρχουν γλώσσες ανεξάρτητες συμφραζομένων
οι οποίες παράγονται από διφορούμενες γραμματικές μόνο.

Τέτοιες γλώσσες ονομάζονται εγγενώς διφορούμενες (inherently ambiguous), τέτοιο παράδειγμα είναι η γλώσσα $\{0^i1^j2^k \mid i = j \vee j = k\}$, όπου έχουμε την ένωση δύο μη διφορούμενων γραμματικών με μη τετριμμένο κοινό τμήμα.

Λήμμα Αντλησης για ΓΑΣ.

'Οπως οι κανονικές γλώσσες δεν εξαντλούν όλες τις δυνατές γλώσσες, όμοια και οι ΓΑΣ δεν εξαντλούν όλες τις δυνατές γλώσσες. Η απόδειξη βασίζεται στην έννοια του πλήθους των στοιχείων ενός συνόλου πεπερασμένου είτε άπειρου αλλά δεν δίνει μία συγκεκριμένη γλώσσα η οποία δεν είναι ΓΑΣ. Για να αποδείξουμε ότι μία γλώσσα δεν είναι ανεξάρτητη συμφραζομένων χρησιμοποιούμε το Λήμμα Αντλησης για ΓΑΣ. Η ιδέα είναι όμοια με το Λήμμα Αντλησης για κανονικές γλώσσες αλλά σε περισσότερο πολύπλοκη μορφή. Συγκεκριμένα η συμβολοσειρά χωρίζεται σε πέντε τμήματα και το δεύτερο και το τέταρτο μπορούν αν επαναληφθούν.

Λήμμα Αντλησης για ΓΑΣ.

Αν A είναι μία γλώσσα ΑΣ τότε υπάρχει αριθμός p (το μήκος άντλησης) ώστε για κάθε συμβολοσειρά x με μήκος

τουλάχιστον p η συμβολοσειρά x μπορεί να γραφτεί $x = uvxyz$ με

1. Για $i \geq 0$, $uv^i xy^i z \in A$,

2. $|vy| > 0$,

3. $|vxy| < p$.

Ιδέα της απόδειξης: Αν A είναι μία γραμματική ΑΣ που παράγει την γλώσσα και η γραμματική αποτελείται από n παραγωγές κανόνες τότε για συμβολοσειρές οι οποίες παράγονται με εφαρμογή $n+1$ είτε περισσότερες παραγωγές κάποια παραγωγή έχει χρησιμοποιηθεί τουλάχιστον δύο φορές γεγονός που μας επιτρέπει να επαναλάβουμε κάποια τμήματα της συμβολοσειράς επαναλαμβάνοντας το δέντρο το οποίο αντιστοιχεί στον κανόνα και μετά την δεύτερη εμφάνιση του κανόνα.

Η συνθήκη $|vy| > 0$, μας εγγυάται ότι τουλάχιστον το ένα από τα δύο τμήματα τα οποία επαναλαμβάνονται είναι μη κενά.

Απόδειξη: Υποθέτουμε ότι G είναι η ΓΑΣ η οποία παράγει την γλώσσα L. Υποθέτουμε ότι b είναι ο μέγιστος αριθμός συμβόλων που εμφανίζονται στο δεξί μέλος των κανόνων της γραμματικής. Ισχύει $b \geq 2$. Σε κάθε δέντρο συντακτικής ανάλυσης ο κάθε κόμβος έχει το πολύ b απογόνους. Δηλ. στο επίπεδο k θα υπάρχουν το πολύ b^k κόμβοι. Αν h συμβολίζει το μέγιστο μήκος μιας διαδρομής στο δένδρο το μέγιστο μήκος μίας συμβολοσειρά που παράγεται από ένα τέτοιο δένδρο είναι b^h. Αν $|V|$ είναι το πλήθος των μεταβλητών ορίζουμε $p = b^{|V|+2}$. Αφού $b \geq 2$ ισχύει $p > b^{|V|+1}$, και κάθε δένδρο συντακτικής ανάλυσης για συμβολοσειρές της γλώσσας με μήκος τουλάχιστον p έχει ύψος τουλάχιστον $|V| + 2$. Για μία συμβολοσειρά s μήκους p είτε μεγαλύτερο με t συμβολίζουμε το δένδρο συντακτικής ανάλυσης της s με τον μικρότερο αριθμό κόμβων. Αφού $|s| > p$ το δένδρο t έχει ύψος τουλάχιστον $|V|+2$ και συνεπώς το μέγιστο μήκος μίας διαδρομής στο δένδρο αυτό θα είναι τουλάχιστον $|V| + 2$. Η διαδρομή αυτή έχει τουλάχιστον $|V| + 1$ μεταβλητές αφού μόνο τερματικά σύμβολα εμφανίζονται στον τελευταίο κόμβο φύλλο και όλοι οι άλλοι κόμβοι έχουν τουλάχιστον μία μεταβλητή. Αφού έχουμε $|V|$ μεταβλητές κάποια μεταβλητή εμφανίζεται δύο φορές. Επιλέγουμε την μεταβλητή που

εμφανίζεται δύο φορές στο τμήμα της διαδρομής από το φύλλο προς την ρίζα του δένδρου. Χωρίζουμε την συμβολοσειρά s όπως στο σχήμα. Κάθε εμφάνιση της μεταβλητής R ορίζει ένα υποδένδρο το οποίο υποδένδρο παράγει ένα τμήμα της συμβολοσειράς s. Η εμφάνιση της R πλησιέστερα προς το δένδρο ορίζει το μεγαλύτερο τέτοιο δένδρο και παράγει την συμβολοσειρά vxy, και η δεύτερη εμφάνιση της μεταβλητής R καθορίζει ένα δένδρο το οποίο παράγει την x μόνο. Αφού τα δύο δένδρα παράγονται από την ίδια μεταβλητή μπορούμε να αντικαταστήσουμε το ένα με το άλλο και θα έχουμε ένα δένδρο το οποίο παράγει μία συμβολοσειρά της γλώσσας. Αντικαθιστώντας το μικρότερο δένδρο με κάποιο άλλο μας επιτρέπει να παράγουμε συμβολοσειρέ της μορφής uv^ixy^iz με $i > 1$. Αντικαθιστώντας το μεγαλύτερο δένδρο με το μικρότερο έχει σαν αποτέλεσμα την παραγωγή της συμβολοσειράς uxz. Για να εξασφαλίσουμε την συνθήκη $|vy| > 0$ υποθέτουμε ότι και οι δύο υποσυμβολοσειρές είναι κενές. Σε αυτή την περίπτωση το δένδρο συντακτικής ανάλυσης το οποίο θα είχαμε αν αντικαθιστούσαμε το μικρότερο δένδρο με το μεγαλύτερο θα είχε λιγότερους από t κόμβους και θα ήταν ένα δένδρο το οποίο παράγει την s. Το γεγονός αυτό αντιφάσκει με την επιλογή του δένδρου t σαν το δένδρο με τον ελάχιστο αριθμό κόμβων

το οποίο παράγει την s. Για την συνθήκη $|vxy| < p$ στο δένδρο το οποίο παράγει την s η δεύτερη εμφάνιση της R από το φύλλο προς την ρίζα παράγει την vxy. Επιλέγοντας την R ώστε και οι δύο εμφανίζεις της να είναι στις τελευταίες $|V|+1$ εμφανίσεις μεταβλητών στην διαδρομή και επιλέγοντας το μεγαλύτερο δυνατό μήκος διαδρομής στο δένδρο το υποδένδρο το οποίο έχει την R σαν ρίζα και παράγει την vxy έχει ύψος το πολύ $|V| + 2$. Ένα δένδρο με αυτό το ύψος μπορεί να παράγει μία συμβολοσειρά μήκους το πολύ $b^{|V|+2} = p$.

Να δειχθεί ότι η γλώσσα $L = \{ww|w \in \{0,1\}^*\}$ δεν είναι ανεξάρτητη συμφραζομένων.

Υποθέτουμε ότι είναι ΓΑΣ. Η επιλογή της κατάλληλης συμβολοσειρά ώστε να εφαρμόσουμε το λήμμα άντλησης απαιτεί προσοχή. Η πρώτη υποψηφιότητα είναι η 0^p10^p1 αλλά μπορούμε να επαναλάβουμε κάποια τμήματα της συμβολοσε και να παραμείνουμε στην γλώσσα με την διάσπαση $u = 0^{p-1}$, $v = 0$, $x = 1$, $y = 0$, $z = 0^{p-1}1$. Μία επόμενη επιλογή είναι η $s = 0^p1^p0^p1^p$ η οποία εκφράζει κατά κάποιο τρόπο περισσότερο την γλώσσα. Θα δείξουμε με κατάλληλη χρήση του λήμματος άντλησης ότι η συμβολοσειρά

$0^p1^p0^p1^p$ μας οδηγεί σε αντίφαση. Αν p είναι το ελαχιστο μήκος συμβολοσειράς ώστε να ισχύει το λήμμα άντλησης, έχουμε $0^p1^p0^p1^p = uvxyz$ με $|vxy| < p$ και $uv^ixy^iz \in L$. Η συμβολοσειρά vxy περιέχει το μέσον της συμβολοσειράς s. Αν αυτό δεν συμβαίνει τότε θα βρίσκεται είτε στο πρώτο μισό της s είτε στο δεύτερο μισό. Τότε αν βρίσκεται στο πρώτο μισό της s η συμβολοσειρά uv^2xy^2z θα είχε το σύμβολο 1 στην πρώτη θέση του δεύτερου μισού της συμβολοσειράς. Ανάλογα αν βρίσκεται είτε στο δεύτερο μισό της s η συμβολοσειρά uv^2xy^2z θα είχε το σύμβολο 0 στην τελευταία θέση του πρώτου μισού της συμβολοσειράς. Καταλήγουμε ότι η συμβολοσειρά vxy περιέχει το μέσον της s. Στην περίπτωση αυτή η συμβολοσειρά uxz είναι της μορφής $0^p1^i0^j1^p$ όπου οι i, j δεν μπορούν να είναι και οι δύο p. Η συμβολοσειρά αυτή δεν είναι της μορφής ww.

Δέντρα συντακτικής ανάλυσης (parse tree)

Διφορούμενες - διττές γραμματικές (ambiguous grammars)
Μία γραμματική μπορεί να παράγει μία συμβολοσειρά με δύο είτε περισσότερους τρόπους. Παρ.

Αριστερότερες παραγωγές.