Κεφάλαιο Ι

Γλώσσες

1. ΑΛΦΑΒΗΤΑ

Ας είναι X ένα πεπερασμένο σύνολο, το ***αλφάβητό*** μας. Τα στοιχεία του X θα τα ονομάζουμε ***γράμματα***. Αν γράψουμε κάποια γράμματα του X το ένα δίπλα στο άλλο σχηματίζουμε μια ***λέξη*** από το X.

Ας πάρουμε για παράδειγμα το ελληνικό αλφάβητο

$$X = \{\alpha, \ldots, \omega\}.$$

Τότε, οι

ψωμι, πεπονι, υπονομος

είναι μερικές λέξεις που μπορούμε να σχηματίσουμε.

Επίσης, αν πάρουμε το αλφάβητο του Morse

$$X = \{-, \quad \cdot, \quad |\}$$

τότε τα σήματα Morse δεν είναι παρά λέξεις από το παραπάνω αλφάβητο.

Γενικά λοιπόν, αν X είναι ένα οποιοδήποτε αλφάβητο, μια λέξη από το X είναι μια πεπερασμένη ακολουθία γραμμάτων του X:

$$x_1 \cdots x_n$$

Δύο λέξεις

$$x_1 \cdots x_m \quad \text{και} \quad y_1 \cdots y_n$$

είναι ***ίσες***, αν

$$m = n \quad \text{και} \quad x_k = y_k, \quad k = 1, \ldots, m.$$

Στη θεωρία που θ' αναπτύξουμε παρακάτω είναι απαραίτητο να εισάγουμε την ***κενή*** λέξη, δηλαδή τη λέξη χωρίς γράμματα. Τη λέξη αυτή τη συμβολίζουμε με το ελληνικό γράμμα ε.

Τώρα, το σύνολο όλων των λέξεων που μπορούμε να σχηματίσουμε από ένα αλφάβητο X, το συμβολίζουμε X^*, δηλαδή

$$X^* = \{\varepsilon\} \cup \{\, x_1 \cdots x_n \mid x_k \in X, \quad k = 1, 2, \ldots \}.$$

Παράδειγμα 1.1. Ας πάρουμε ένα αλφάβητο που ν' αποτελείται από δύο μόνο γράμματα:

$$X = \{a,\ b\}.$$

Τότε οι

$$aaba,\quad bba,\quad aa,\quad ababa,\quad b$$

είναι μερικές λέξεις που μπορούμε να σχηματίσουμε απ' αυτό.

$\square$

Παράδειγμα 1.2. Ας πάρουμε τώρα το φτωχότερο αλφάβητο που μπορούμε να φανταστούμε, δηλαδή το αλφάβητο με ένα μόνο γράμμα!

$$X = \{a\}.$$

Τότε οι

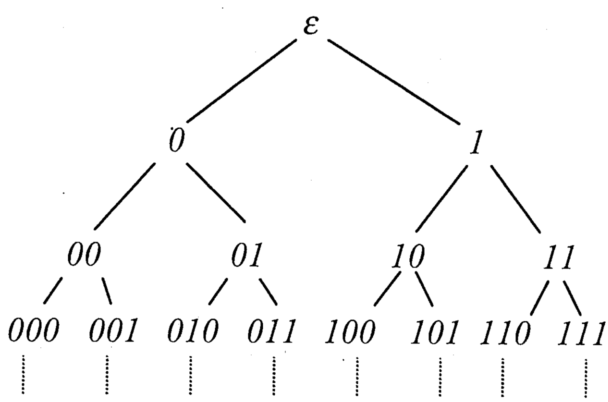
$$\varepsilon, a, aa, aaa, \ldots$$

είναι όλες οι λέξεις που μπορούμε να σχηματίσουμε από το X, δηλαδή

$$X^* = \{a\}^* = \{\varepsilon, a, aa, aaa, \ldots\}.$$

$\square$

Παράδειγμα 1.3. Μερικές φορές για να παριστάνουμε τις λέξεις του X^* χρησιμοποιούμε ένα "δέντρο", όπως αυτό που δίνουμε παρακάτω για $X = \{0, 1\}$:



$\square$

Παράδειγμα 1.4. Όπως γνωρίζουμε από τη στοιχειώδη Θεωρία Αριθμών, κάθε μη μηδενικός φυσικός αριθμός n γράφεται κατά μοναδικό τρόπο με τη μορφή

$$n = a_0 \cdot 10^k + a_1 \cdot 10^{k-1} + \ldots + a_{k-1} \cdot 10 + a_k,$$

όπου

$$0 \le a_j < 10, \quad j = 0, \ldots, k \quad \text{και} \quad a_0 > 0.$$

Η λέξη

$$a_0 a_1 \cdots a_{k-1} a_k$$

που σχηματίζεται από τους παραπάνω συντελεστές ονομάζεται ***παράσταση του*** n ***στο δεκαδικό σύστημα*** και συμβολίζεται $[n]_{10}$, δηλαδή

$$[n]_{10} = a_0 a_1 \cdots a_{k-1} a_k.$$

Οι φυσικοί λοιπόν αριθμοί μπορούν να παρασταθούν από λέξεις του αλφαβήτου

$$X = \{0, 1, 2, \ldots, 9\}.$$

❐

☞ **Παράδειγμα 1.5.** Κάθε μη μηδενικός φυσικός αριθμός n γράφεται κατά μοναδικό τρόπο με τη μορφή

$$n = a_0 \cdot 2^k + a_1 \cdot 2^{k-1} + \ldots + a_{k-1} \cdot 2 + a_k,$$

όπου

$$a_0 = 1, \quad 0 \le a_j < 2, \quad j = 1, \ldots, k.$$

Η λέξη

$$a_0 a_1 \cdots a_{k-1} a_k$$

ονομάζεται ***παράσταση του*** n ***στο δυαδικό σύστημα*** και συμβολίζεται $[n]_2$, δηλαδή

$$[n]_2 = a_0 a_1 \cdots a_{k-1} a_k.$$

Κάθε φυσικός αριθμός λοιπόν, μπορεί να παρασταθεί από μια λέξη που σχηματίζεται από μηδενικά και μονάδες, δηλαδή από μια λέξη του αλφαβήτου $X = \{0, 1\}$. Για παράδειγμα αν $n = 2^k$, τότε

$$[2^k]_2 = 1\underbrace{0\ldots0}_{k}$$

διότι

$$2^k = 1 \cdot 2^k + 0 \cdot 2^{k-1} + \ldots + 0 \cdot 2 + 0.$$

❐

Αν

$$u = x_1 \cdots x_m, \quad v = y_1 \cdots y_n$$

είναι δύο λέξεις με γράμματα από ένα αλφάβητο X, τότε μπορούμε να δημιουργήσουμε μια νέα λέξη αν γράψουμε την v δεξιά της u:

$$x_1 \cdots x_m y_1 \cdots y_n$$

Η πράξη αυτή λέγεται ***παράθεση*** λέξεων. Για παράδειγμα αν $X = \{a, b\}$ και

$$u = aa, \quad v = bbb,$$

τότε

$$uv = aabbb \text{ ενώ } vu = bbbaa.$$

Παρατηρούμε ότι $uv \neq vu$.

Πρόταση 1.1. Η παράθεση είναι πράξη προσεταιριστική, δηλαδή για κάθε $\rho, \vartheta, \varphi \in X^*$ έχουμε

$$(\rho\vartheta)\varphi = \rho(\vartheta\varphi).$$

Απόδειξη. Ας είναι

$$\rho = x_1 \cdots x_p, \quad \vartheta = y_1 \cdots y_q, \quad \varphi = z_1 \cdots z_s$$

Τότε,

$$(\rho\vartheta)\varphi = (x_1 \cdots x_p y_1 \cdots y_q) z_1 \cdots z_s = x_1 \cdots x_p y_1 \cdots y_q z_1 \cdots z_s =$$
$$= x_1 \cdots x_p (y_1 \cdots y_q z_1 \cdots z_s) = \rho(\vartheta\varphi)$$

□

Πρόταση 1.2. Για κάθε λέξη $\eta \in X^*$ έχουμε $\eta\varepsilon = \eta = \varepsilon\eta$, όπου ε είναι η κενή λέξη.

Απόδειξη. Φανερή.

□

Λίγο αργότερα θα δούμε την αλγεβρική σημασία των παραπάνω. Τώρα μας χρειάζεται ο ακόλουθος

Ορισμός 1.1. Ονομάζουμε ***μήκος*** μιας λέξης w και το συμβολίζουμε με $|w|$, το πλήθος των γραμμάτων που την απαρτίζουν.

Έτσι, τα μήκη των λέξεων $aaba,\ bba,\ b$ που είδαμε στο παράδειγμα 1.1, είναι αντίστοιχα 4, 3 και 1, δηλαδή

$$|aaba|=4, \quad |bba|=3, \quad |b|=1.$$

Πρόταση 1.3. Ισχύουν οι ισότητες

$$|\varepsilon|=0,$$

$$|uv|=|u|+|v|,$$

για κάθε $u, v \in X^*$, δηλαδή το μήκος της κενής λέξης είναι μηδέν, αφού στερείται γραμμάτων, και το μήκος της λέξης uv που προκύπτει από την παράθεση των u και v ισούται με το άθροισμα των μηκών των u και v.

Απόδειξη. Φανερή. $\square$

Για παράδειγμα, αν $X=\{a, b, c\}$ και

$$u=aabb, \quad v=ccc,$$

οπότε

$$uv=aabbccc,$$

θα έχουμε

$$|u|=4, \quad |v|=3, \quad |uv|=7,$$

δηλαδή

$$|uv|=|u|+|v|.$$

Σε κάθε λέξη από το αλφάβητο X

$$w=x_1\cdots x_n$$

αντιστοιχεί η λέξη

$$w^\rho=x_n\cdots x_1$$

που προκύπτει από τη w αντιστρέφοντας τη σειρά των γραμμάτων της. Η w^ρ ονομάζεται ***είδωλο*** της λέξης w. Για παράδειγμα το είδωλο της λέξης abb είναι η λέξη bba. Αν μια λέξη συμπίπτει με το είδωλό της, την ονομάζουμε ***παλινδρομική***. Π.χ. οι λέξεις aba, $baab$ είναι παλινδρομικές, όπως αμέσως διαπιστώνεται.

Ισχύει η επόμενη

Πρόταση 1.4. Για κάθε $u, v \in X^*$ έχουμε $(uv)^\rho = v^\rho u^\rho$.

Απόδειξη. Αν

$$u = x_1 \cdots x_m, \quad v = y_1 \cdots y_n$$

τότε

$$(uv)^\rho = (x_1 \cdots x_m y_1 \cdots y_n)^\rho = y_n \cdots y_1 x_m \cdots x_1 = v^\rho u^\rho.$$

❐

Για κάθε λέξη $w \in X^*$ μπορούμε να ορίσουμε τη n-οστή ***δύναμή της*** επαγωγικά ως εξής:

$$\begin{aligned} w^0 &= \varepsilon \\ w^1 &= w \\ w^2 &= ww \\ &\vdots \\ w^{n+1} &= w^n w. \end{aligned}$$

Πρόταση 1.5. Ισχύουν οι επόμενοι κανόνες:

$$u^m u^n = u^{m+n}, \quad (u^m)^n = u^{m\,n}$$

για κάθε λέξη $u \in X^*$.

Απόδειξη. Θ' αποδείξουμε μόνο την πρώτη ισότητα με χρήση επαγωγής επάνω στο n. Για $n = 0$ η ισότητα ισχύει φανερά, διότι

$$u^m u^0 = u^m \varepsilon = u^m = u^{m+0}.$$

Ας υποθέσουμε ότι η πρότασή μας ισχύει και για το φυσικό αριθμό n, δηλαδή ότι

$$u^m u^n = u^{m+n}.$$

Θα δείξουμε ότι ισχύει και για τον $n+1$:

$$u^m u^{n+1} = u^m \left(u^n u\right) \overset{(1)}{=} \left(u^m u^n\right) u \overset{(2)}{=} u^{m+n} u = u^{m+n+1},$$

όπου η ισότητα (1) προέκυψε από την προσεταιριστική ιδιότητα της παράθεσης, ενώ η ισότητα (2) από την υπόθεση της επαγωγής

❐

Η χρήση των δυνάμεων βοηθά στην απλοποίηση της γραφής. Έτσι, αν για παράδειγμα διαθέτουμε το αλφάβητο $X = \{a, b, c\}$, τότε τη λέξη

$$aaabbccccaa$$

μπορούμε να τη γράψουμε απλά

$$a^3 b^2 c^4 a^2$$

και τη λέξη

$$bbbbbbbccccc$$

να τη γράψουμε απλά

$$b^7 c^5.$$

Πρόταση 1.6. Για κάθε λέξη $w \in X^*$ και για κάθε φυσικό αριθμό n ισχύει

$$\left|w^n\right| = n \cdot |w|.$$

Απόδειξη. Είναι

$$\left|w^n\right| = \left|\underbrace{w \cdots w}_{n}\right| = \underbrace{|w| + \ldots + |w|}_{n} = n \cdot |w|,$$

όπως το θέλαμε.

❐

2. ΓΛΩΣΣΕΣ

Ας είναι X ένα αλφάβητο. Ένα οποιοδήποτε υποσύνολο του X^* ονομάζεται ***γλώσσα***. Αφού λοιπόν οι γλώσσες από το X είναι υποσύνολα του X^*, μπορούμε να μιλάμε για την ***ένωση***, την ***τομή*** τους, κ.λ.π.

Αν $L_1,\ L_2 \subseteq X^*$ είναι δύο γλώσσες, τότε μπορούμε να δημιουργήσουμε μια νέα γλώσσα αν παραθέσουμε κάθε λέξη της L_1 δίπλα σε κάθε λέξη της L_2. Τη γλώσσα που έτσι σχηματίζεται ονομάζουμε ***παράθεση*** των L_1, L_2 και τη συμβολίζουμε $L_1 L_2$, δηλαδή

$$L_1L_2 = \left\{ uw \,\middle|\, u \in L_1,\ \ w \in L_2 \right\}.$$

☞ **Παράδειγμα 2.1.** Ας είναι $X = \{a, b, c\}$ και ας θεωρήσουμε τις γλώσσες

$$L_1 = \{ab,\ ba\},\quad L_2 = \{ac,\ ca\}.$$

Τότε

$$L_1L_2 = \{abac,\ abca,\ baac,\ baca\}.$$

❐

☞ **Παράδειγμα 2.2.** Ας είναι $X = \{a, b\}$ και ας πάρουμε τις γλώσσες

$$L_1 = \{a, a^3, a^5, \ldots\},\quad L_2 = \{b\}.$$

Τότε

$$L_1L_2 = \{ab,\ a^3b,\ a^5b,\ \ldots\}$$

ενώ

$$L_2L_1 = \{ba,\ ba^3,\ ba^5,\ \ldots\}.$$

❐

Δύο βασικές ιδιότητες που έχει η παράθεση γλωσσών είναι ο προσεταιρισμός και ο επιμερισμός ως προς την ένωση.

Πρόταση 2.1. Αν $L_k \subseteq X^*$, $k = 1, 2, 3$, είναι τρεις οποιεσδήποτε γλώσσες, τότε

$$L_1(L_2L_3) = (L_1L_2)L_3,$$

$$L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3,$$

$$(L_1 \cup L_2)L_3 = L_1L_3 \cup L_2L_3.$$

Απόδειξη. Για την πρώτη έχουμε:

$$L_1(L_2L_3) = \{\, u(vw) \mid u \in L_1,\ v \in L_2,\ w \in L_3 \,\} =$$
$$= \{\, (uv)w \mid u \in L_1,\ v \in L_2,\ w \in L_3 \,\} = (L_1L_2)L_3$$

Ας πάρουμε τώρα τη δεύτερη:

$$\begin{aligned} uw \in L_1(L_2 \cup L_3) &\Leftrightarrow \{\, u \in L_1 \ \wedge\ w \in (L_2 \cup L_3) \,\} \Leftrightarrow \\ &\Leftrightarrow \{\, u \in L_1 \ \wedge\ (w \in L_2 \ \vee\ w \in L_3) \,\} \Leftrightarrow \\ &\Leftrightarrow \{\, (u \in L_1 \ \wedge\ w \in L_2) \ \vee\ (u \in L_1 \ \wedge\ w \in L_3) \,\} \Leftrightarrow \\ &\Leftrightarrow \{\, uw \in L_1L_2 \ \vee\ uw \in L_1L_3 \,\} \Leftrightarrow uw \in L_1L_2 \cup L_1L_3 \end{aligned}$$

δηλαδή

$$L_1(L_2 \cup L_3) = L_1L_2 \cup L_1L_3 .$$

Όμοια εργαζόμαστε και για την τρίτη. □

Μπορούμε τώρα να ορίσουμε επαγωγικά την $n-$οστή δύναμη μιας γλώσσας $L \subseteq X^*$, ως εξής:

$$L^0 = \{\varepsilon\},\ L^1 = L,\ L^2 = LL,\ \ldots,\ L^{n+1} = L^nL .$$

000

Οι τύποι που αποδείξαμε για τις δυνάμεις μιας λέξης εξακολουθούν να ισχύουν και στην προκειμένη περίπτωση.

Θέτουμε

$$L^* = \{\varepsilon\} \cup L \cup L^2 \cup \ldots = \bigcup_{k=0}^{\infty} L^k .$$

Η γλώσσα L^* που έτσι ορίσαμε ονομάζεται ***θήκη*** της γλώσσας L. Μ' άλλα λογια η θήκη μιας γλώσσας L περιέχει κάθε λέξη που σχηματίζεται αν παραθέσουμε οποιονδήποτε αριθμό λέξεων της L. Μερικές φορές αντί της L^* χρησιμοποιούμε τη γλώσσα

$$L^+ = L^* - \{\varepsilon\} = L \cup L^2 \cup \ldots = \bigcup_{k=1}^{\infty} L^k .$$

☞ **Παράδειγμα 2.3.** Ας είναι $X = \{a, b\}$ και ας είναι L η γλώσσα που έχει μια μόνο λέξη, την ba, δηλαδή $L = \{ba\}$. Θα έχουμε διαδοχικά

$$L^0 = \{\varepsilon\},$$

$$L^1 = L = \{ba\},$$
$$L^2 = LL = \{baba\},$$
$$L^3 = L^2 L = \{bababa\},$$
$$\vdots$$

Έτσι λοιπόν

$$L^* = \{ba\}^* = \{\varepsilon,\ ba,\ baba,\ bababa,\ \ldots\} = \left\{ \varepsilon,\ ba,\ (ba)^2,\ (ba)^3,\ \ldots \right\}.$$

Τέλος, το ***είδωλο*** μιας γλώσσας $L \subseteq X^*$, είναι η γλώσσα που απαρτίζεται από τα είδωλα όλων των λέξεων της L, δηλαδή

$$L^\rho = \left\{ w^\rho \mid w \in L \right\}.$$

3. ΡΗΤΕΣ ΓΛΩΣΣΕΣ

Ας είναι

$$X = \{\ a, b, c, \ldots, \}$$

ένα αλφάβητο. Στο εξής, χάριν απλότητας θα ταυτίζουμε τη γλώσσα $\{a\}$ με το ίδιο το γράμμα a, τη γλώσσα $\{b\}$ με το ίδιο το γράμμα b, κ.λ.π. Έτσι, αντί να γράφουμε

$$\{a\} \cup \{b\}$$

θα γράφουμε $a \cup b$, αντί να γράφουμε

$$\{a\}\{b\}$$

θα γράφουμε ab, αντί να γράφουμε

$$\{a\}^*$$

θα γράφουμε απλά a^*, κ.λ.π. Με την παραπάνω συμφωνία η γλώσσα

$$\{a\} \cup \left[\{b\}\{c\}^*\right]$$

γράφεται απλά

$$a \cup bc^*$$

ενώ η γλώσσα

$$\{a\}^* \cup \left[\{b\}^*\{c\} \cup \{c\}^*\right]^*$$

γράφεται απλά

$$a^* \cup \left(b^* c \cup c^*\right)^*,$$

κ.λ.π.

Ορισμός 3.1. Μια γλώσσα $L \subseteq X^*$ θα την λέμε ***ρητή***, εάν προκύπτει από κάποια γράμματα του αλφαβήτου X, κάνοντας μια ή περισσότερες φορές τις ακόλουθες πράξεις:

(α) ***ένωση***,
(β) ***παράθεση***,
(γ) ***θήκη***.

Έτσι, αν $X = \{a, b, c\}$, η γλώσσα $a^* \cup b^*$ είναι ρητή. Πράγματι, ξεκινώντας από τα γράμματα a και b σχηματίζουμε αρχικά τις γλώσσες a^* και b^* (πράξη (γ)) και στη συνέχεια ενώνουμε τις δύο αυτές γλώσσες για να πάρουμε την $a^* \cup b^*$ (πράξη (α)). Όμοια, η γλώσσα

$$\left(b^{*}c \cup c^{*}b\right)^{*}$$

είναι ρητή. Πράγματι, ξεκινώντας από τα γράμματα b και c σχηματίζουμε αρχικά τις γλώσσες b^{*} και c^{*} (πράξη (γ)) και στη συνέχεια τις γλώσσες $b^{*}c$ και $c^{*}b$ (πράξη (β)). Κατόπιν τις ενώνουμε παίρνοντας την $b^{*}c \cup c^{*}b$ (πράξη (α)) και τελικά παίρνουμε τη θήκη της λευταίας (πράξη (γ)).

Συμβολίζουμε με $Rat(X)$ το σύνολο όλων των ρητών γλωσσών από το αλφάβητο X. Αν $L_k \subseteq X^{*}$, $k = 1, 2,$ είναι ρητές γλώσσες, δηλαδή και οι δυο τους προκύπτουν από κάποια γράμματα του αλφαβήτου X κάνοντας μία ή περισσότερες φορές τις πράξεις της ένωσης, παράθεσης και θήκης, τότε και οι γλώσσες $L_1 \cup L_2$, L_1L_2 και L_1^{*} προκύπτουν φανερά με τέτοιες διαδικασίες, δηλαδή είναι και αυτές ρητές γλώσσες. Έχουμε λοιπόν

$$L_1, L_2 \in Rat(X) \Rightarrow L_1 \cup L_2 \in Rat(X),$$

$$L_1, L_2 \in Rat(X) \Rightarrow L_1L_2 \in Rat(X),$$

$$L \in Rat(X) \Rightarrow L^{*} \in Rat(X).$$

Στην πραγματικότητα ισχύει κάτι πολύ ισχυρότερο.

Πρόταση 3.1. Η $Rat(X)$ είναι η μικρότερη κλάση γλωσσών που περιέχει τις πεπερασμένες γλώσσες και είναι κλειστή ως προς την ένωση, την παράθεση και τη θήκη.

Απόδειξη. Αποδεικνύουμε πρώτα ότι κάθε πεπερασμένη γλώσσα είναι ρητή. Προς τούτο, ας είναι $w = a_1 \cdots a_n$ μια λέξη από το αλφάβητο X. Τότε, η γλώσσα $\{w\}$ είναι ρητή αφού προκύπτει από τα γράμματα $a_1, \ldots, a_n$ του X με την πράξη της παράθεσης. Άρα, κάθε πεπερασμένη γλώσσα

$$L = \{w_1, \ldots, w_n\}$$

είναι ρητή ως ένωση ρητών γλωσσών:

$$L = \bigcup_{k=1}^{n} \{w_k\}.$$

Ας είναι τώρα $\Im$ ένα σύνολο από γλώσσες του αλφαβήτου X, τέτοιο ώστε κάθε πεπερασμένη γλώσσα ν' ανήκει σ' αυτό και επιπρόσθετα

$$L_1, L_2 \in \Im \Rightarrow L_1 \cup L_2 \in \Im,$$
$$L_1, L_2 \in \Im \Rightarrow L_1L_2 \in \Im,$$

$$L \in \Im \Rightarrow L^* \in \Im .$$

Θα δείξουμε ότι

$$Rat(X) \subseteq \Im ,$$

δηλαδή ότι η κλάση $Rat(X)$ είναι η ελάχιστη κλάση γλωσσών που περιέχει τις πεπερασμένες γλώσσες και είναι κλειστή ως προς την ένωση, την παράθεση και τη θήκη.

Πράγματι, αφού το $\Im$ περιέχει όλες τις πεπερασμένες γλώσσες του X, θα περιέχει και τις γλώσσες $\{a\}$ $(a \in X)$, δηλαδή θα περιέχει όλα τα γράμματα του αλφαβήτου X. Αφού όμως μια ρητή γλώσσα προκύπτει από κάποια γράμματα του αλφαβήτου X με εφαρμογή των πράξεων της παράθεσης, της ένωσης και θήκης και αφού το $\Im$ από την υπόθεση είναι κλειστό ως προς αυτές τις πράξεις, προκύπτει ότι κάθε ρητή γλώσσα ανήκει στο $\Im$, πράγμα που θέλαμε ν' αποδείξουμε. ❐

4. ΚΩΔΙΚΟΠΟΙΗΣΕΙΣ - ΑΝΤΙΚΑΤΑΣΤΑΣΕΙΣ

Ας είναι X και Y δύο αλφάβητα. Ονομάζουμε ***κωδικοποίηση*** του X στο Y ένα κανόνα που σε κάθε γράμμα του αλφαβήτου X αντιστοιχεί μια λέξη από το αλφάβητο Y. Μ' άλλα λόγια, μια κωδικοποίηση του X στο Y δεν είναι παρά μια απεικόνιση

$$f : X \to Y^* .$$

Αν

$$x_1 \cdots x_m$$

είναι μια τυχούσα λέξη του X^*, τότε αυτή μετασχηματίζεται από την κωδικοποίηση $f : X \to Y^*$ στη λέξη

$$f(x_1) \cdots f(x_m)$$

του Y^*. Έτσι λοιπόν κάθε κωδικοποίηση $f : X \to Y^*$ επεκτείνεται (επαγωγικά) σε μια απεικόνιση

$$f^* : X^* \to Y^*$$

με τον ακόλουθο τρόπο

$$\begin{cases} f^*(\varepsilon) = \varepsilon \\ f^*(x_1 \cdots x_m) = f(x_1) \cdots f(x_m) \end{cases}$$

☞ **Παράδειγμα 4.1.** Ας είναι

$$X = \{ a, b \} \text{ και } Y = \{ 0, 1 \}$$

και ας πάρουμε την ακόλουθη κωδικοποίηση του X στο Y:

μετασχηματίζεται στη λέξη

$$10100110$$

ενώ η λέξη

$$bbba$$

μετασχηματίζεται στη λέξη

$$01010110.$$

Ας πάρουμε τώρα την κωδικοποίηση

$$a \mapsto 1, \quad b \mapsto 1.$$

Τότε, η λέξη $baba$ μετασχηματίζεται στη λέξη 1111 κ.τ.λ.

□

Ας υποθέσουμε τώρα ότι διαθέτουμε μια απεικόνιση

$$g : X^* \to Y^*.$$

Πώς μπορούμε να ελέγξουμε αν αυτή προέρχεται από κάποια κωδικοποίηση; Απάντηση σ' αυτό το ενδιαφέρον ερώτημα δίνει το επόμενο

Θεώρημα 4.1. Αν για μία απεικόνιση $g : X^* \to Y^*$ ισχύει

(α) $g(\varepsilon) = \varepsilon$,

δηλαδή η g στέλνει την κενή λέξη του X^* στην κενή λέξη του Y^* και

(β) $g(uw) = g(u)g(w)$

για όλες τις λέξεις $u, w \in X^*$, τότε υπάρχει μία μοναδική κωδικοποίηση

$$f : X \to Y^*$$

της οποίας η επέκταση είναι η g:

$$f^* = g.$$

Απόδειξη. Ορίζουμε μια κωδικοποίηση $f : X \to Y^*$ με τον ακόλουθο τρόπο

$$f(x) = g(x), \quad \text{για κάθε } x \in X,$$

$$f(x) = g(x), \text{ για κάθε } x \in X,$$

δηλαδή η f στέλνει κάθε γράμμα $x \in X$ στη λέξη $g(x) \in Y^*$. Θα δείξουμε ότι $f^* = g$, δηλαδή ότι

$$f^*(w) = g(w), \text{ για κάθε } w \in X^*.$$

Για $w = \varepsilon$, το πράγμα είναι φανερό

$$f^*(\varepsilon) = \varepsilon = g(\varepsilon),$$

ενώ για $w = x_1 \cdots x_m$, $m \geq 1$, έχουμε

$$\begin{aligned} f^*(w) &= f^*(x_1 \cdots x_m) = f(x_1) \cdots f(x_m) = \\ &= g(x_1) \cdots g(x_m) = g(x_1 \cdots x_m) = g(w) \end{aligned}$$

Βρήκαμε, λοιπόν, μια κωδικοποίηση f του X στο Y η οποία όταν επεκταθεί μας δίνει την g.

Αν, τώρα, $h : X \to Y^*$ είναι μια άλλη κωδικοποίηση τέτοια ώστε $h^* = g$, τότε για κάθε $x \in X$ θα είναι

$$h(x) = h^*(x) = g(x) = f(x)$$

δηλαδή $h = f$ γεγονός που αποδεικνύει και τη μοναδικότητα.

❐

☞ **Παράδειγμα 4.2.** Η απεικόνιση $g : \{a\}^* \to \{0, 1\}^*$ που ορίζεται από τη σχέση

$$g(a^k) = \begin{cases} 010, & \text{αν } k \text{ αρτιος} \\ 101, & \text{αν } k \text{ περιττος} \end{cases}$$

δεν προέρχεται από μια κωδικοποίηση διότι αν πάρουμε

$$u = a, \quad v = a^2$$

θα έχουμε

$$g(uv) = g(aa^2) = g(a^3) = 101$$

ενώ

$$g(u)g(v) = g(a)g(a^2) = 101010$$

και

$$101 \neq 101010.$$

Αντίθετα, η απεικόνιση $g:\{a\}^* \to \{\,0,1\,\}^*$ που ορίζεται από τη σχέση

$$g\left(a^k\right)=(101)^k\,,\quad k=0,1,\ldots$$

προέρχεται από μια κωδικοποίηση αφού για κάθε $u=a^p\,,\ w=a^q$ έχουμε

$$\begin{aligned} g(uw) &= g\left(a^p a^q\right)=g\left(a^{p+q}\right)=(101)^{p+q} \\ &=(101)^p(101)^q=g\left(a^p\right)g\left(a^q\right)=g(u)g(w) \end{aligned}$$

και

$$g(\varepsilon)=g\left(a^0\right)=(101)^0=\varepsilon\,.$$

Η κωδικοποίηση από την οποία προέρχεται η g είναι η

$$f:\{a\}\to\{\,0,1\,\}^* \quad \text{με} \quad f(a)=101$$

δηλαδή εκείνη που στέλνει το μοναδικό γράμμα του αλφαβήτου $X=\{a\}$ στη λέξη 101 του αλφαβήτου $Y=\{\,0,1\,\}$.

❐

Οι κωδικοποιήσεις λοιπόν από το X στο Y, συμπίπτουν με τις απεικονίσεις

$$g:X^*\to Y^*$$

για τις οποίες

$$g(\varepsilon)=\varepsilon \quad \text{και} \quad g(uw)=g(u)g(w)\,.$$

Ας υποθέσουμε ότι σε κάθε γράμμα x του αλφαβήτου X αντιστοιχεί μια γλώσσα L_x του L^*:

$$x\mapsto L_x$$

Λέμε τότε ότι έχουμε μια ***αντικατάσταση*** του X στο Y. Μια αντικατάσταση ονομάζεται ***ρητή***, αν όλες οι γλώσσες L_x, $x\in X$, είναι ρητές.

☞ **Παράδειγμα 4.3.** Ας είναι $X=\{\,a,b,c\,\}$ και $Y=\{\,0,1\,\}$. Τότε η απεικόνιση που ορίζεται από τις σχέσεις

$$a \mapsto 0^*, \quad b \mapsto 1^* 01^*, \quad c \mapsto 01$$

είναι μια αντικατάσταση του $\{a, b, c\}$ στο $\{0, 1\}$. Η λέξη $abbc$ του X μετασχηματίζεται τότε στη γλώσσα

$$0^*(1^*01^*)(1^*01^*)01$$

ενώ η λέξη cac μετασχηματίζεται στη γλώσσα

$$010^*01.$$

Η παραπάνω αντικατάσταση είναι ρητή, αφού όλες οι γλώσσες

$$0^*, \quad 1^*01^* \quad \text{και} \quad 01$$

είναι ρητές.

❒

Μια αντικατάσταση λοιπόν του X στο Y δεν είναι παρά μια απεικόνιση

$$F : X \to P(Y^*),$$

όπου $P(Y^*)$ συμβολίζει το σύνολο δηλαδή των γλωσσών του Y^*. Αν $x_1 \cdots x_p$ είναι μια λέξη του X^*, τότε αυτή μετασχηματίζεται με τη βοήθεια της απεικόνισης F στη γλώσσα

$$F(x_1) \cdots F(x_p)$$

που προκύπτει παραθέτοντας τις γλώσσες $F(x_k)$, $k = 1, \ldots, p$. Έτσι, κάθε αντικατάσταση $F : X \to P(Y^*)$ επεκτείνεται επαγωγικά σε μια απεικόνιση

$$F^* : X^* \to P(Y^*)$$

με τον ακόλουθο τρόπο:

$$\begin{cases} F^*(\varepsilon) = \{\varepsilon\} \\ F^*(x_1 \cdots x_p) = F(x_1) \cdots F(x_p) \end{cases}$$

Και πάλι, ισχύει το επόμενο

Θεώρημα 4.2. Αν μια απεικόνιση

$$S: X^* \to P(Y^*)$$

έχει τις ακόλουθες δύο ιδιότητες

$$S(\varepsilon) = \{\varepsilon\},$$

$$S(uv) = S(u)S(v), \quad \forall u, v \in X^*,$$

δηλαδή η γλώσσα $S(uv)$ είναι η παράθεση των γλωσσών $S(u)$ και $S(v)$, τότε υπάρχει μοναδική αντικατάσταση

$$F: X \to P(Y^*)$$

της οποίας η επέκταση είναι η S, δηλαδή $F^* = S$.

Απόδειξη. Είναι εντελώς παρόμοια με την απόδειξη του θεωρήματος 4.1. ❐

Οι αντικαταστάσεις λοιπόν από το X στο Y ταυτίζονται με τις απεικονίσεις

$$S: X^* \to P(Y^*)$$

για τις οποίες

$$S(\varepsilon) = \{\varepsilon\} \quad \text{και} \quad S(uv) = S(u)S(v).$$

Σχετικά με τις ρητές αντικαταστάσεις, αρκετά ενδιαφέρουσα είναι η επόμενη

Πρόταση 4.1. Αν $S: X^* \to P(Y^*)$ είναι μια ρητή αντικατάσταση και $L \subseteq X^*$ είναι ρητή γλώσσα, τότε η

$$S(L) = \bigcup_{u \in L} S(u)$$

είναι ρητή γλώσσα του Y^*. Μ' άλλα λόγια, η εικόνα μιας ρητής γλώσσας μέσω μιας ρητής αντικατάστασης είναι πάλι ρητή γλώσσα.

Για την απόδειξη αυτής της πρότασης μας χρειάζεται το επόμενο

Λήμμα 4.1. Αν $S: X^* \to P(Y^*)$ είναι μια αντικατάσταση, τότε για οποιεσδήποτε γλώσσες $L, L_1, L_2 \subseteq X^*$ συμβαίνει (α) $$S(L_1 \cup L_2) = \bigcup_{u \in L_1 \cup L_2} S(u) = \left[\bigcup_{u \in L_1} S(u)\right] \cup \left[\bigcup_{u \in L_2} S(u)\right] = S(L_1) \cup S(L_2).$$

Εύκολα φαίνεται ότι η (α) ισχύει για οσεσδήποτε γλώσσες.

(β) $$S(L_1 L_2) = \bigcup_{\substack{u_1 \in L_1 \\ u_2 \in L_2}} S(u_1 u_2) = \bigcup_{\substack{u_1 \in L_1 \\ u_2 \in L_2}} S(u_1) S(u_2) =$$

$$= \left[\bigcup_{u_1 \in L_1} S(u_1)\right]\left[\bigcup_{u_2 \in L_2} S(u_2)\right] = S(L_1) S(L_2)$$

(γ) Λαμβάνοντας υπόψη τις (α) και (β) έχουμε:

$$S(L^*) = S\left(\bigcup_{k=0}^{\infty} L^k\right) = \bigcup_{k=0}^{\infty} S(L^k) = \bigcup_{k=0}^{\infty} [S(L)]^k = [S(L)]^*.$$

❐

Απόδειξη της πρότασης 4.1. Αφού η γλώσσα L είναι ρητή, θα προκύπτει από κάποια γράμματα x_j, $j=1,\dots k$, του αλφαβήτου X κάνοντας μία ή περισσότερες φορές τις πράξεις της ένωσης, παράθεσης και θήκης. Κατά συνέπεια, σύμφωνα με το προηγούμενο λήμμα, η γλώσσα $S(L)$ θα προκύπτει από τις ρητές γλώσσες $S(x_j)$, $j=1,\dots,k$, του αλφαβήτου Y, κάνοντας τις ίδιες ακριβώς πράξεις. Ώστε, σύμφωνα με την πρόταση 3.1, η $S(L)$ είναι ρητή.

❐

Καθώς μια κωδικοποίηση είναι ειδική περίπτωση ρητής αντικατάστασης, συνάγουμε ότι

Πόρισμα 4.1. Η εικόνα μιας ρητής γλώσσας μέσω μιας κωδικοποίησης είναι πάλι ρητή γλώσσα.

5. ΡΗΤΕΣ ΕΚΦΡΑΣΕΙΣ

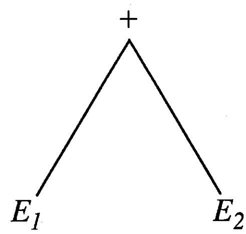
Οι ρητές εκφράσεις είναι "σχήματα" που αντιπροσωπεύουν τις ρητές γλώσσες. Συγκεκριμένα ας θεωρήσουμε ένα αλφάβητο

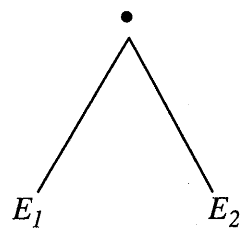
$$X = \{x_1, \dots, x_n\}.$$

Το σύνολο των ***ρητών εκφράσεων*** από το X (που συμβολίζεται $RE(X)$), ορίζεται επαγωγικά ως εξής:

(α) Τα στοιχεία $\varnothing,\ \varepsilon,\ x_1,\ \dots,\ x_n$ είναι ρητές εκφράσεις

(β) Αν E_1, E_2 είναι ρητές εκφράσεις, τότε και τα ***δένδρα***





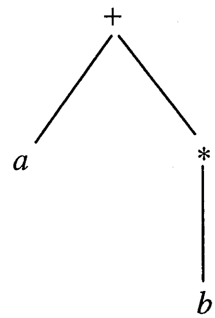
είναι ρητές εκφράσεις.

(γ) Αν E είναι ρητή έκφραση, τότε και το δένδρο

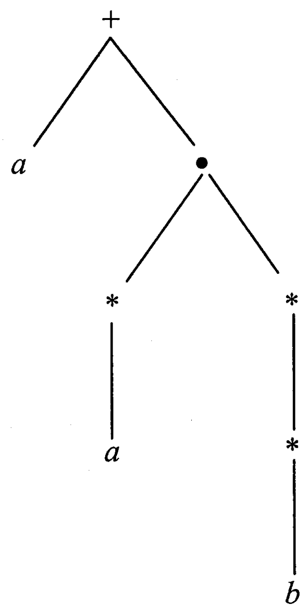
είναι ρητή έκφραση. Έτσι βλέπουμε ότι οι ρητές εκφράσεις από το X δεν είναι παρά δένδρα που σχηματίζονται με τον παραπάνω τρόπο.

Ας δούμε μερικές ρητές εκφράσεις από το αλφάβητο $X = \{a,\ b\}$.

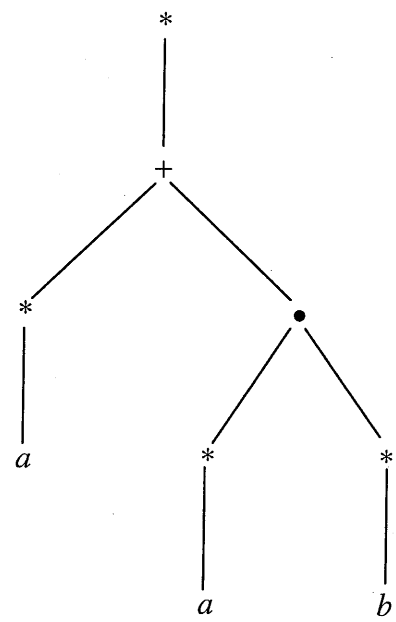
t_1 :



t_2 :



t_3 :



Σε κάθε ρητή έκφραση t από το X, αντιστοιχεί μια ρητή γλώσσα που τη συμβολίζουμε $Rat(t)$. Έτσι, για τις προηγούμενες ρητές εκφράσεις έχουμε

$$Rat(t_1) = a \cup b^*, \quad Rat(t_2) = a \cup a^*\left(b^*\right)^*, \quad Rat(t_3) = \left(a^* \cup a^* b^*\right)^*,$$

κ.λ.π. Βάζοντας αυστηρά τα πράγματα κάτω, η Rat είναι μια απεικόνιση από το σύνολο $RE(X)$ στο σύνολο $P\left(X^*\right)$ όλων των γλωσσών από το X, δηλαδή

$$Rat : RE(X) \to P\left(X^*\right)$$

που ορίζεται με τον ακόλουθο τρόπο:

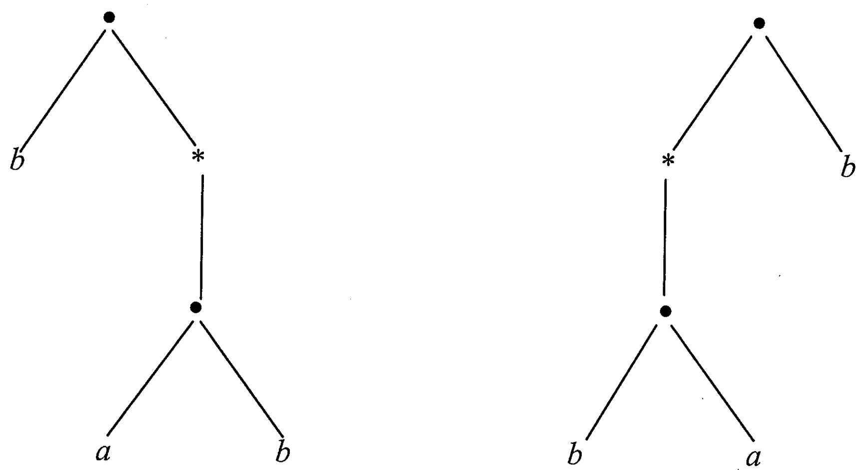
1) $Rat(\varnothing) = \varnothing, \quad Rat(\varepsilon) = \{\varepsilon\}, \quad Rat\left(x_s\right) = \left\{x_s\right\},$

2) $Rat\left(\begin{array}{ccc} & + & \\ & \diagup \diagdown & \\ E_1 & & E_2 \end{array}\right) = Rat\left(E_1\right) \cup Rat\left(E_2\right),$

$$Rat\left(\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ E_1 \qquad E_2 \end{array}\right) = Rat(E_1)Rat(E_2),$$

$$3)\ Rat\left(\begin{array}{c} * \\ | \\ E \end{array}\right) = [Rat(E)]^* .$$

Είναι δυνατό διαφορετικές ρητές εκφράσεις να ορίζουν την ίδια ρητή γλώσσα όπως π.χ. συμβαίνει με τις εκφράσεις

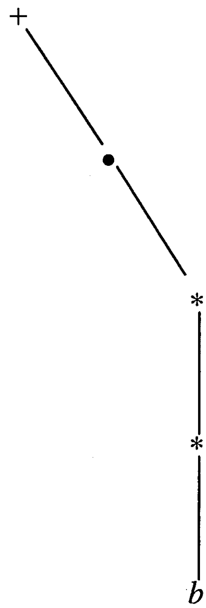


Και οι δύο τους ορίζουν την ίδια ρητή γλώσσα: $b(ab)^* = (ba)^* b$ (άσκηση).

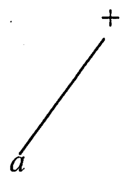
Ας πάρουμε τώρα μια ρητή γλώσσα $L \subseteq X^*$ και όλες τις ρητές εκφράσεις που την ορίζουν, δηλαδή το σύνολο

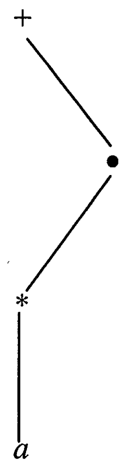
$$\{ t \in RE(X) \mid Rat(t) = L \}.$$

Ένα φυσικό ερώτημα που εγείρεται είναι αν μπορούμε να προσδιορίσουμε αλγοριθμικά μια ρητή έκφραση t της L η οποία να περιέχει τον μικρότερο δυνατό αριθμό αστερίσκων κι ακόμη ισχυρότερα αν μπορούμε να προσδιορίσουμε μια ρητή έκφραση t της L με το μικρότερο δυνατό "αστροβάθος". Το αστροβάθος $h(t)$ είναι ο μέγιστος αριθμός αστεριών που βρίσκονται στον ίδιο κλάδο της ρητής έκφρασης t. Π.χ, για την έκφραση t_2 που είδαμε προηγουμένως είναι $h(t_2) = 2$, αφού ο κλάδος



περιέχει δύο αστερίσκους ενώ οι υπόλοιποι κλάδοι





περιέχουν κανέναν και έναν αστερίσκο αντίστοιχα.

Φορμαλιστικά, η απεικόνιση

$$h : RE(X) \rightarrow N \text{ (= οι φυσικοί αριθμοί)}$$

ορίζεται με τον ακόλουθο τρόπο:

1) $h(\varnothing)=0, \quad h(\varepsilon)=0, \quad h(x_k)=0, \quad k=1,\ldots,n,$