Θεωρία Υπολογισμού

Ασυμφραστικές Γλώσσες (1)

Στην ενότητα αυτή θα μελετηθούν τα εξής επιμέρους θέματα:

Ασυμφραστικές Γραμματικές (2.1)

- *Τυπικός Ορισμός*
- *Σχεδιασμός Ασυμφραστικών Γραμματικών*
- *Πολυτροπία*
- *Κανονική Μορφή Chomsky*

Ασυμφραστικές Γλώσσες

- Στην προηγούμενη ενότητα παρατηρήσαμε ότι υπάρχουν γλώσσες οι οποίες *δεν* είναι κανονικές.
  - Δεν μπορούν να τύχουν περιγραφής από κανονικές εκφράσεις
  - Δεν υπάρχει πεπερασμένο αυτόματο που να τις αναγνωρίζει
- Παράδειγμα μη κανονική γλώσσας:
  - $B = \{0^k 1^k \mid k \geq 0\}$
- Υπάρχει δυνατότερο υπολογιστικό μοντέλο που να μπορεί να αναγνωρίζει μη κανονικές γλώσσες όπως την B;
- Ναι! Οι *ασυμφραστικές γλώσσες* – ιδιότητες με αναδρομικό χαρακτήρα
  - Περιγραφή: *Ασυμφραστικές Γραμματικές*
  - Μηχανή: *Αυτόματα με στοίβα*
- Εφαρμογές: σχεδίαση και μετάφραση γλωσσών προγραμματισμού
  - συντακτική ανάλυση προγραμμάτων

Ασυμφραστικές Γραμματικές

- *Ασυμφραστικές γραμματικές* (context-free grammars, CFG)
  - Συλλογή κανόνων για την παραγωγή λέξεων
- *Κανόνας αντικατάστασης* / *κανόνας παραγωγής*: $\boldsymbol{S \rightarrow w}$
  - S: μεταβλητή
  - w: λέξη, αποτελεί συναρμογή μεταβλητών και τερματικών σύμβολων
- Παράδειγμα:

$$\left.\begin{array}{llll} S & \rightarrow & aSa & (1) \\ S & \rightarrow & T & (2) \\ T & \rightarrow & \# & (3) \end{array}\right\} \text{Κανόνες παραγωγής}$$

- S, T: μεταβλητές (μη τερματικά σύμβολα)
- a, #: τερματικά σύμβολα
- Εναρκτήρια Μεταβλητή: μεταβλητή πρώτου κανόνα

Παραγωγή Λέξεων

- Μια γραμματική μας επιτρέπει να παραγάγουμε λέξεις.

- Βήματα παραγωγής λέξης:
    1. Γράφουμε την εναρκτήρια μεταβλητή
    2. Επιλέγουμε ένα μη-τερματικό σύμβολο που υπάρχει στη λέξη που έχουμε γραμμένη και το αντικαθιστούμε με το δεξί μέλος ενός κανόνα που αναφέρεται στο σύμβολο αυτό.
    3. Επαναλαμβάνουμε το 2 μέχρις ότου η λέξη να περιέχει μόνο τερματικά σύμβολα.

Παραγωγή Λέξεων

- Έστω η ασυμφραστική γραμματική

$$\begin{array}{lll} S \rightarrow aSa & & (1) \\ S \rightarrow T & & (2) \\ T \rightarrow \# & & (3) \end{array}$$

- Για την εναρκτήρια μεταβλητή έχουμε (ανάμεσα σε άλλα):
  - $S \Rightarrow aSa$ από τον κανόνα παραγωγής (1)

- Μετασχηματισμοί μπορούν να εφαρμοστούν και στο εσωτερικό μιας λέξης χρησιμοποιώντας τους κανόνες:
  - $aSa \Rightarrow aTa$ από τον κανόνα παραγωγής (2)
  - $aTa \Rightarrow a\#a$ από τον κανόνα παραγωγής (3)

- Συμβολισμός: $S \overset{*}{\Longrightarrow} a\#a$
  - Το S μπορεί να μετασχηματιστεί στη λέξη $a\#a$ σε ένα ή περισσότερα βήματα.

Παραδείγματα

- Έστω η ασυμφραστική γραμματική

$$\begin{array}{llll} S & \rightarrow & aSa & (1) \\ S & \rightarrow & T & (2) \\ T & \rightarrow & \# & (3) \end{array}$$

- $$\begin{array}{llll} S & \Rightarrow & aSa & \text{από τον κανόνα παραγωγής (1)} \\ & \Rightarrow & aaSaa & \text{από τον κανόνα παραγωγής (1)} \\ & \Rightarrow & aaTaa & \text{από τον κανόνα παραγωγής (2)} \\ & \Rightarrow & aa\#aa & \text{από τον κανόνα παραγωγής (3)} \end{array}$$

  Άρα η S *παράγει* τη λέξη *aa#aa*

- $$\begin{array}{llll} S & \Rightarrow & T & \text{από τον κανόνα παραγωγής (2)} \\ & \Rightarrow & \# & \text{από τον κανόνα παραγωγής (3)} \end{array}$$

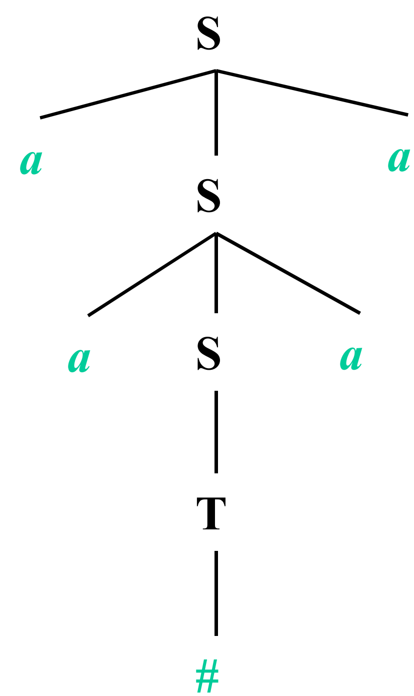
  Άρα η S παράγει τη λέξη #

- Οι λέξεις που παράγει η S είναι οι $\{a^n\#a^n \mid n \geq 0\}$

Συντακτικά Δέντρα (parse tree)

- Η παραγωγή οποιασδήποτε λέξης μπορεί να τύχει αναπαράστασης ως ένα *συντακτικό δένδρο*:
  - *Ρίζα* του δέντρου είναι η εναρκτήρια μεταβλητή
  - *Παιδιά κάθε εσωτερικού κόμβου* είναι τα σύμβολα (τερματικά ή μη) με τα οποία το αντικαθιστούμε:

$$
\begin{aligned}
\mathrm{S} &\Rightarrow a\mathrm{S}a \\
&\Rightarrow aa\mathrm{S}aa \\
&\Rightarrow aa\mathrm{T}aa \\
&\Rightarrow aa\#aa
\end{aligned}
$$



Ασυμφραστικές Γραμματικές – Ορισμοί

<u>ΟΡΙΣΜΟΣ</u>

Ασυμφραστική Γραμματική είναι μια τετράδα (V, Σ, R, S) όπου

1. V είναι ένα σύνολο μεταβλητών,
2. Σ είναι ένα αλφάβητο τερματικών συμβόλων
3. $R \subseteq V \times (\Sigma \cup V)^*$ είναι το σύνολο κανόνων
4. $S \in V$ είναι η αρχική μεταβλητή

<u>ΟΡΙΣΜΟΣ</u>

Οι γλώσσες που παράγονται από ασυμφραστικές γραμματικές ονομάζονται *ασυμφραστικές γλώσσες* (context free languages).

Τυπικός ορισμός παραγωγής λέξεων

- Έστω
  - u, v, w λέξεις από μεταβλητές και τερματικά σύμβολα
  - $A \rightarrow w$ κανόνας της γραμματικής

Τότε

- Η λέξη uAv *αποδίδει* τη λέξη uwv.

  Αυτή την εφαρμογή του κανόνα τη συμβολίζουμε ως $uAv \Rightarrow uwv$

- Η λέξη u *παράγει* τη λέξη v αν η u μπορεί να μετασχηματιστεί στη v σε μηδέν ή περισσότερα βήματα:

$$u \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \ldots \Rightarrow u_k \Rightarrow v$$

  Αυτή τη διαδοχική εφαρμογή κανόνων τη συμβολίζουμε ως $u \overset{*}{\Longrightarrow} v$

- *Γλώσσα της γραμματικής* G είναι το σύνολο

$$L(G) = \{w \in \Sigma^* \mid S \overset{*}{\Longrightarrow} w\}$$

Συγχώνευση Κανόνων

- Μπορούμε να συγχωνεύσουμε όλους τους κανόνες που έχουν στο αριστερό τους μέλος την ίδια μεταβλητή μέσω του συμβόλου ‘|’.

- Παράδειγμα:

Αν $A \to 0A1$ και $A \to B$ τότε γράφουμε

$$A \to 0A1 \mid B$$

Παραδείγματα

- Παράδειγμα 1:
  - $G_1 = (\{S\}, \{\, (\, ,\,)\, \}, R, S)$ όπου R το σύνολο των κανόνων

    $$S \rightarrow (S) \mid SS \mid \varepsilon$$

  - $L(G_1)$ = {όλες οι ακολουθίες ορθά τοποθετημένων παρενθέσεων}

- Παράδειγμα 2:
  - $G_2 = (\ \{\langle \text{ΕΚΦΡΑΣΗ} \rangle, \langle \text{ΟΡΟΣ} \rangle, \langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle\},$

    $\{\, \boldsymbol{a}, +, \times, (,)\, \},$

    $R,$

    $\langle \text{ΕΚΦΡΑΣΗ} \rangle)$

    και R οι κανόνες:

    $$\langle \text{ΕΚΦΡΑΣΗ} \rangle \rightarrow \langle \text{ΕΚΦΡΑΣΗ} \rangle + \langle \text{ΟΡΟΣ} \rangle \mid \langle \text{ΟΡΟΣ} \rangle$$

    $$\langle \text{ΟΡΟΣ} \rangle \rightarrow \langle \text{ΟΡΟΣ} \rangle \times \langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle \mid \langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle$$

    $$\langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle \rightarrow (\langle \text{ΕΚΦΡΑΣΗ} \rangle) \mid \boldsymbol{a}$$

  - $L(G_2) = ?$

Σχεδίαση Ασυμφραστικών Γραμματικών

- Κατασκευή γραμματικής για σύνθετη γλώσσα
  - Διαιρούμε τη γραμματική σε απλούστερα μέρη
  - Κατασκευάζουμε μια γραμματική για κάθε μέρος
  - Προσθέτουμε τον κανόνα

    $$S \rightarrow S_1 \mid S_2 \mid \ldots \mid S_n$$

    όπου S_i το εναρκτήριο σύμβολο για κάθε μια από τις επιμέρους γραμματικές

Σχεδίαση CFG για ειδικές περιπτώσεις

- Γραμματική για *κανονική γλώσσα* για την οποία είναι γνωστό το αυτόματο που την αναγνωρίζει:
  - Για κάθε κατάσταση q_i εισάγουμε ένα μη τερματικό σύμβολο R_i
  - Για κάθε σύμβολο του αλφάβητου, εισάγουμε ένα τερματικό σύμβολο
  - Για κάθε μετάβαση $\delta(q_i, a) = q_j$ εισάγουμε ένα κανόνα της μορφής

    $$R_i \rightarrow aR_j$$
  - Για κάθε κατάσταση αποδοχής q_i εισάγουμε τον κανόνα $R_i \rightarrow \varepsilon$
  - Εναρκτήρια κατάσταση η R_0 όπου q_0 η εναρκτήρια κατάσταση του αυτομάτου

Σχεδίαση CFG για ειδικές περιπτώσεις

- *Αλληλένδετες υπολέξεις*
  - Εμφάνιση μιας υπολέξης με βάση πόσες φορές εμφανίζεται μια δεύτερη υπολέξη, π.χ. $\{0^k 1^k \mid k \geq 0\}$
  - Χρησιμοποιούμε κανόνες της μορφής $R \rightarrow uRv$
    - Όπου u η πρώτη υπολέξη, και
    - v η υπολέξη που εξαρτάται από την u

- *Αναδρομικές Δομές*
  - Εμφανίζονται ως τμήματα άλλων δομών ή ακόμη και του εαυτού τους, π.χ.

$$\langle \text{ΕΚΦΡΑΣΗ} \rangle \rightarrow \langle \text{ΕΚΦΡΑΣΗ} \rangle + \langle \text{ΟΡΟΣ} \rangle \mid \langle \text{ΟΡΟΣ} \rangle$$

$$\langle \text{ΟΡΟΣ} \rangle \rightarrow \langle \text{ΟΡΟΣ} \rangle \times \langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle \mid \langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle$$

$$\langle \text{ΠΑΡΑΓΟΝΤΑΣ} \rangle \rightarrow (\langle \text{ΕΚΦΡΑΣΗ} \rangle) \mid a$$

Παραδείγματα

- Παράδειγμα 1:

  Γλώσσα: $L = \{a^m b^n: \; m \geq \mathrm{n}\}$

  Γραμματική: $G = (V, \Sigma, R, S)$

  $V = \{S\}$

  $\Sigma = \{a, b\}$

  $R = \{ S \rightarrow aSb,$

  $S \rightarrow aS,$

  $S \rightarrow \varepsilon\}$

- Παράδειγμα 2:

  Γλώσσα: $L=\{1^n\#0^n\#\#: n > 0\}$

  Γραμματική: $G = (V, \Sigma, R, S)$

  $V = \{S, A, B\}$

  $\Sigma = \{1, 0, \#\}$

  $R = \{ S \rightarrow A\#\#,$

  $A \rightarrow 1B0,$

  $B \rightarrow 1B0$

  $B \rightarrow \# \}$

Παράδειγμα 3

- Σχεδιάστε μια ασυμφραστική γραμματική για το συμπλήρωμα της γλώσσας $L = \{a^k\, b^k \mid k \geq 0\}$

- Μια λέξη ανήκει στην L αν
  - Όλα τα a πριν από τα b
  - $|a| = |b|$

- Μια λέξη ανήκει στο συμπλήρωμα της L αν ισχύει τουλάχιστον ένα από τα πιο κάτω:
  - Υπάρχει a μετά από b (i)
  - $|a| > |b|$ (ii)
  - $|a| < |b|$ (iii)

- Θα φτιάξουμε τρεις γραμματικές για τις τρεις περιπτώσεις.

Παράδειγμα 3 (συν.)

- Ξεκινούμε με την περίπτωση (i).
- Προσπάθεια 1:

  $S \to TbaT$

  $T \to a$

  $T \to b$

  $T \to \varepsilon$

Προσπάθεια 2:

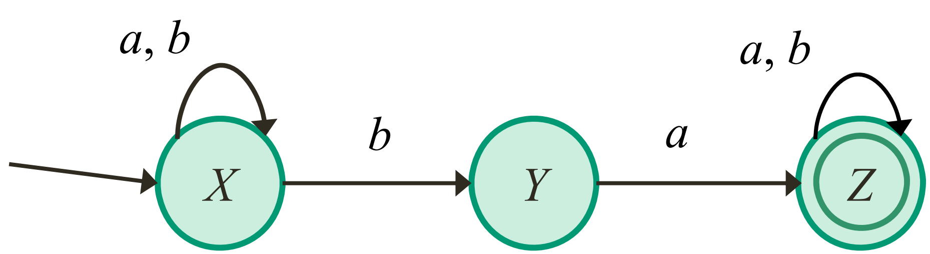
  $S \to TbaT$

  $T \to aT$

  $T \to bT$

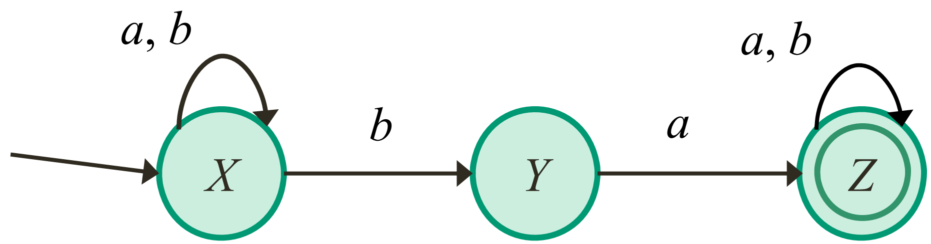
  $T \to \varepsilon$

- Η γραμματική (προσπάθεια 1) παράγει κάθε λέξη στην οποία υπάρχει *a* μετά από κάποιο *b*;
  - Όχι. Για παράδειγμα αδυνατεί να παραγάγει την *aabba*
- Εναλλακτικά, ας κτίσουμε το NFA που αναγνωρίζει τη γλώσσα:



Παράδειγμα 3 (συν.)

- Μετατρέπουμε το αυτόματο σε ασυμφραστική γραμματική.



- Κανόνες που προκύπτουν:

$$X \to aX$$
$$X \to bX$$
$$X \to bY$$
$$Y \to aZ$$
$$Z \to aZ$$
$$Z \to bZ$$
$$Z \to \varepsilon$$

Παράδειγμα 3 (συν.)

- Προχωρούμε στην περίπτωση (ii): όλες οι λέξεις $a^m b^n$ με $m >$ n

$$P \rightarrow aPb$$
$$P \rightarrow aQ$$
$$Q \rightarrow \varepsilon$$
$$Q \rightarrow aQ$$

- Τέλος θεωρούμε την περίπτωση (iii): όλες οι λέξεις $a^m b^n$ με $m <$ n

$$U \rightarrow aUb$$
$$U \rightarrow bV$$
$$V \rightarrow \varepsilon$$
$$V \rightarrow bV$$

- Συνδυάζουμε τις τρεις γραμματικές μέσω μίας νέας μεταβλητής που θα παίξει τον ρόλο της αρχικής και η οποία θα μετασχηματίζεται σύμφωνα με τον κανόνα:

$$S \quad \rightarrow \quad X \mid P \mid U$$

Πολυτροπία

- *Πολυτροπία* εμφανίζεται σε κάποια γραμματική όταν παράγει την ίδια λέξη με περισσότερους από ένα τρόπους
  - Περισσότερα από ένα συντακτικά δέντρα
- Προσοχή: «περισσότερο από ένα τρόπους» δεν υπονοεί απλά δύο διαφορετικές παραγωγές
  - Οι παραγωγές πρέπει να διαφέρουν ως προς τη δομή τους και όχι απλά ως προς τη σειρά εφαρμογής κανόνων.
  - Για διευκόλυνση της διατύπωσης της έννοιας ορίζουμε:
- *Εξ αριστερών παραγωγή*
  - Σε κάθε βήμα αντικαθιστούμε την αριστερότερη από τις εναπομείναντες μεταβλητές

Πολυτροπία – Ορισμός

<u>ΟΡΙΣΜΟΣ</u>

Για κάθε ασυμφραστική γραμματική G και για κάθε λέξη w, λέμε ότι η w *παράγεται πολύτροπα* στη G εάν υπάρχουν για αυτήν περισσότερες από μια εξ αριστερών παραγωγές στη G.

Η G ονομάζεται *πολύτροπη* αν παράγει κάποια λέξη πολύτροπα.

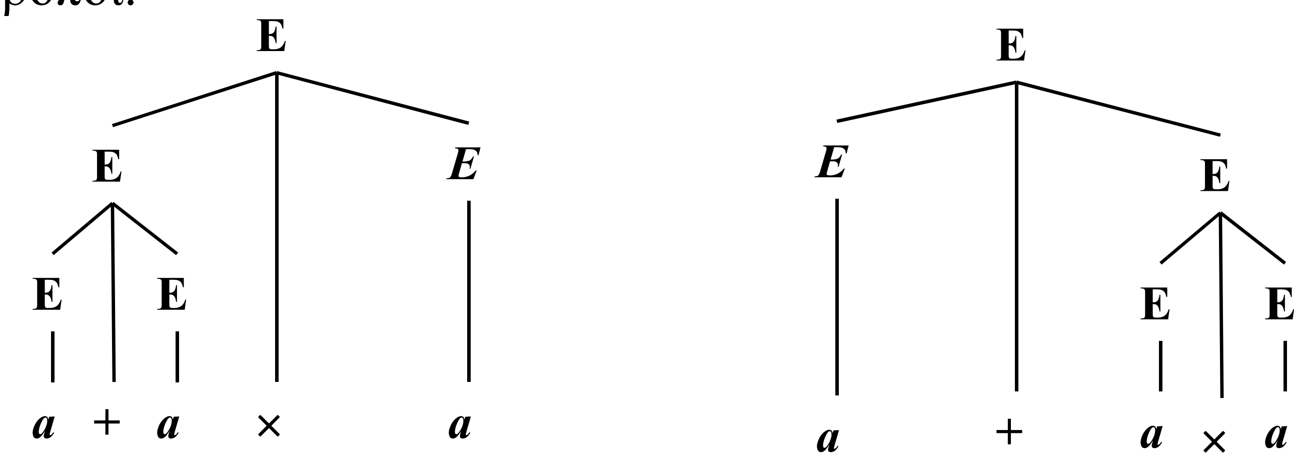
Παράδειγμα Πολυτροπίας

- Έστω η γραμματική:

$$E \;\rightarrow\; E + E \;\mid\; E \times E \;\mid\; (E) \;\mid\; a$$

- Πως παράγεται η $a + a \times a$;

- Δύο τρόποι:



- Και στη γραμματική που ακολουθεί;

$$E \;\rightarrow\; E + O \;\mid\; O$$
$$O \;\rightarrow\; O \times P \;\mid\; P$$
$$P \;\rightarrow\; (E) \;\mid\; a$$

Κανονική μορφή Chomsky

<u>ΟΡΙΣΜΟΣ</u>

Μια ασυμφραστική γραμματική G βρίσκεται σε *κανονική μορφή Chomsky* αν κάθε κανόνας της βρίσκεται σε μια από τις μορφές

$$A \rightarrow BC$$

$$A \rightarrow a$$

όπου a οποιοδήποτε τερματικό σύμβολο, A οποιαδήποτε μεταβλητή, και B και C οποιεσδήποτε μεταβλητές διάφορες της εναρκτήριας S. Επιπλέον επιτρέπεται ο κανόνας $S \rightarrow$ ε.

CFG σε Κανονική Μορφή Chomsky

<u>ΘΕΩΡΗΜΑ</u>

Κάθε ασυμφραστική γλώσσα παράγεται από κάποια ασυμφραστική γραμματική σε κανονική μορφή Chomsky

- Πρέπει να δείξουμε ότι:
  - Κάθε ασυμφραστική γραμματική μπορεί να μετατραπεί σε κανονική μορφή Chomsky.

Βήματα Μετατροπής

- Βήμα 1: Προσθέτουμε μια νέα εναρκτήρια μεταβλητή

$$S_0 \rightarrow S$$

- Βήμα 2: Απαλείφουμε όλους τους κανόνες της μορφής:

$$A \rightarrow \varepsilon$$

  – Για κάθε κανόνα που εμφανίζεται η A στο δεξί μέλος κάνουμε τα εξής:

    Για κάθε κανόνα

    $R \rightarrow uAv$ προσθέτουμε τον κανόνα $R \rightarrow uv$

    Αυτό γίνεται επαναληπτικά και για κάθε εμφάνιση της μεταβλητής A.
    Για παράδειγμα:

    Αν $R \rightarrow uAvAw$ τότε προσθέτουμε $R \rightarrow uvAw \mid uAvw \mid uvw$

    Αν $R \rightarrow A$ τότε προσθέτουμε $R \rightarrow \varepsilon$

  – Επαναλαμβάνουμε τα βήματα μέχρι να απαλειφθούν όλοι οι κανόνες ε.

Βήματα Μετατροπής

- Βήμα 3: Απαλείφουμε τους μοναδιαίους κανόνες

$$A \to B$$

  – Για κάθε κανόνα $B \to u$ προσθέτουμε τον κανόνα $A \to u$

- Βήμα 4: Μεταγράφουμε τους υπόλοιπους κανόνες ως εξής:
  – Αντικαθιστούμε κάθε κανόνα της μορφής:

$$A \to u_1 u_2 u_3 \ldots u_k, \quad k \geq 3$$

όπου u_i τερματικό σύμβολο ή μεταβλητή, με τους κανόνες

$$A \to u_1 A_1, A_1 \to u_2 A_2, A_2 \to u_3 A_3 \ldots A_{k-2} \to u_{k-1} u_k$$

  – Για κάθε κανόνα της μορφής

$$A \to u_1 u_2$$

όπου κάποιο u_i είναι τερματικό τότε αντικαθιστούμε στον κανόνα το u_i από μία καινούρια μεταβλητή U_i και προσθέτουμε τον κανόνα

$$U_i \to u_i$$

Παράδειγμα

- Να μεταγάγετε την πιο κάτω ασυμφραστική γραμματική σε κανονική μορφή Chomsky.

$$
\begin{aligned}
S &\rightarrow ASA \mid aB \\
A &\rightarrow B \mid S \\
B &\rightarrow b \mid \varepsilon
\end{aligned}
$$