ΟΜΑΛΕΣ (ή ΚΑΝΟΝΙΚΕΣ) ΓΡΑΜΜΑΤΙΚΕΣ ΚΑΙ ΓΛΩΣΣΕΣ.

1. Ομαλές γραμματικές: Το εάν μια γραμματική είναι ομαλή ή όχι, φαίνεται στην μορφή της, στον τρόπο με τον οποίο είναι γραμμένη: πρέπει (i) τα παραγωγικά σύμβολα («κεφαλαία») να μετατρέπονται μόνον σε μια λέξη λ από τερματικά σύμβολα («πεζά»), ακολουθούμενη από ακριβώς ένα παραγωγικό ή κανένα.

   Ερώτημα: Ποιές από τις εξής γραμματικές είναι ομαλές;

$I \to \alpha\beta K \mid \varnothing$ $K \to \Lambda \mid M$ $\Lambda \to \beta\beta K \mid \varnothing$ $M \to \gamma M \mid \Lambda$ (α)	$I \to \alpha\beta K$ $K \to \Lambda \mid K$ $\Lambda \to \beta\gamma K$ $M \to \gamma M \mid \Lambda$ (β)	$I \to \alpha\beta K\gamma \mid \varnothing$ $K \to \Lambda \mid M$ $M \to \gamma M \mid \Lambda \mid \varnothing$ (γ)
$I \to \alpha\beta K \mid \varnothing$ $K \to \alpha\alpha\Lambda M$ $\Lambda \to \gamma\beta\Lambda \mid \varnothing$ (δ)	$I \to \alpha\beta K \mid \varnothing$ $\alpha K \to \Lambda \mid M$ $\Lambda \to \beta\beta K \mid \varnothing$ $M \to \alpha\Lambda \mid I$ (ε)	$I \to \alpha\beta\Lambda \mid \varnothing$ $K \to \alpha\beta\gamma \mid \varnothing$ $\Lambda \to \gamma\gamma K \mid \Lambda$ $M \to M\gamma\beta \mid \Lambda$ (στ)

   ΑΠΑΝΤΗΣΗ:
   (α) Είναι ομαλή: τηρούνται οι κανόνες μορφής: αριστερά του «→» έχουμε =1 παραγωγικό σύμβολο, και δεξιά μια λέξη (λ.χ. 'αβ') και ακριβώς 1 ή 0 παραγωγικά σύμβολα. Προσέξτε ότι ένας κανόνας «$X \to \varnothing$» περιέχει (τρόπος του λέγειν) την κενή λέξη και κανένα (επόμενο) παραγωγικό σύμβολο.
   (β) Είναι ομαλή. Προσέξτε ότι δεν μας πειράζει που δεν περιέχει απαλειφές της μορφής $X \to \varnothing$ - απλά η γραμματική δεν παράγει καμμία λέξη – (δηλαδή δεν παράγει *ούτε κάν την κενή λέξη*, δηλαδή *παράγει την κενή γλώσσα*).
   (γ) Δεν είναι ομαλή. Ο 1ος κανόνας περιέχει τερματικά σύμβολα και από τις δύο πλευρές του Κ.
   (δ) Δεν είναι ομαλή. Ο 2ος κανόνας περιέχει δύο παραγωγικά σύμβολα.
   (ε) Δεν είναι ομαλή. Ο 2ος κανόνας δεν έχει ένα (και ακριβώς) παραγωγικό σύμβολο αριστερά του «→».
   (στ) Δεν είναι ομαλή. Ο 4ος κανόνας έχει τα τερματικά σύμβολα μετά το παραγωγικό. Θα πρέπει να είναι παντού πριν από αυτό (ή παντού μετά, αλλά όχι τα δύο ανάμικτα).
   ■

2. Ομαλές γλώσσες: Το εάν μια *γλώσσα* είναι ομαλή ή όχι, εξαρτάται από το εάν μπορούμε ή όχι να βρούμε μια ομαλή γραμματική που να την παράγει. Στις απλές περιπτώσεις αυτό το κάνουμε «κατ' ευθείαν». Στις πιο σύνθετες χρησιμοποιούμε τους κανόνες κλειστότητας.

   Ερώτημα: Δείξτε ότι οι εξής γλώσσες Λ (επί του Σ = {α, β, γ } αν δεν δίδεται κάτι άλλο) είναι ομαλές:
   α) Λ = «όλες οι λέξεις που περιέχουν ως υπόλεκτο το ***ββαβα*** »
   β) Λ = «όλες οι λέξεις που _δεν_ περιέχουν το υπόλεκτο ***βββ*** »
   γ) Λ = «όλες οι λέξεις που περιέχουν 3 ***α*** στη σειρά αλλά και 4 ***β*** στη σειρά ».
   δ) Λ = «όλες οι λέξεις που περιέχουν 3 ***α*** στη σειρά αλλά δεν περιέχουν 4 ***β*** στη σειρά ».
   ε) Λ = «όλες οι λέξεις που _αν_ περιέχουν το ***αβγ*** τότε περιέχουν και το ***βγα*** »

   ΑΠΑΝΤΗΣΗ:

(α) Η πιο απλή και θεμελιακή περίπτωση: Λ = Σ* || {ββαβα} || Σ*, δηλαδή η παράθεση τριών προφανώς ομαλών γλωσσών.
(β) Η χρησιμότητα της «άρνησης» ή «συμπληρώματος»: παίρνουμε την γλώσσα που έχει το υπόλεκτο «βββ» και εξ αυτής το συμπλήρωμα: Λ = Σ* – (Σ* {βββ} Σ*).
(γ) Η χρησιμότητα της «σύζευξη» ή «τομής»: παίρνουμε την γλώσσα $Λ_{3α}$ = (Σ* {ααα} Σ*) και την $Λ_{4β}$ = (Σ* {ββββ} Σ*), και θέλουμε να είμαστε στην $1^η$ _και_ $2^η$, άρα στην τομή: Λ = $Λ_{3α} \cap Λ_{4β}$. Οι $Λ_{3α}$, $Λ_{4β}$ είναι εμφανώς ομαλές και τα άλλα προκύπτουν από την κλειστότητα ως προς την τομή.
(δ) Η χρησιμότητα της «διαφοράς»: παίρνουμε την γλώσσα $Λ_{3α}$ = (Σ* {ααα} Σ*) και την $Λ_{4β}$ = (Σ* {ββββ} Σ*), και θέλουμε να είμαστε στην $1^η$, *αλλά όχι στη* $2^η$, άρα θέλουμε την διαφορά Λ = $Λ_{3α} – Λ_{4β}$. Οι $Λ_{3α}$, $Λ_{4β}$ είναι εμφανώς ομαλές και τα άλλα προκύπτουν από την κλειστότητα ως προς την διαφορά.
(ε) Αυτό είναι πιο πονηρό: απαιτεί απλά στοιχεία προτασιακού λογισμού: έχουμε την συνθήκη (P ⇒ Q), όπου P = «περιέχει αβγ», Q = «περιέχει βγα». Επειδή η έκφραση (P ⇒ Q) ισοδυναμεί με την έκφραση (¬P ∨ Q) έχουμε ισοδυνάμως την ένωση [(δεν «έχει αβγ») ***ή*** «έχει βγα»].
Έστω οι γλώσσες $Λ_{αβγ}$ = (Σ* {αβγ} Σ*), $Λ_{βγα}$ = (Σ* {βγα} Σ*), και Λ = (Σ* – $Λ_{αβγ}$) ∪ $Λ_{βγα}$. Οι $Λ_{αβγ}$, $Λ_{βγα}$ είναι εμφανώς ομαλές και το ζητούμενο για την προκύπτει από την κλειστότητα ως προς συμπλήρωμα «Σ* –», και την ένωση.

■

Στα παραπάνω εφαρμόζουμε κυρίως τους κανόνες κλειστότητας ως προς τομή, ένωση, συμπλήρωμα, για να αναλύσουμε τις ιδιότητες που έχουν οι λέξεις της γλώσσας. Έχουμε όμως στη διάθεσή μας και τους «γραμματικούς» κανόνες κλειστότητας ως προς παράθεση, επανάληψη και κατοπτρισμό.

<u>Ερώτημα:</u> Δείξτε ότι οι εξής γλώσσες Λ (επί του Σ = {α, β, γ } αν δεν δίδεται κάτι άλλο) είναι ομαλές:
α) Λ = «οι λέξεις που έχουν την μορφή μιας σειράς φυσικών δεκαδικών αριθμών (≥ 1), χωρισμένων με παύλες και εγκλεισμένων σε αγκύλες, π.χ. **[23 – 45 – 6 – 12]** »
β) Λ = «οι δεκαδικοί γραμμένοι σε ‘επιστημονική’ γραφή, λ.χ. –0.314159E+2 »

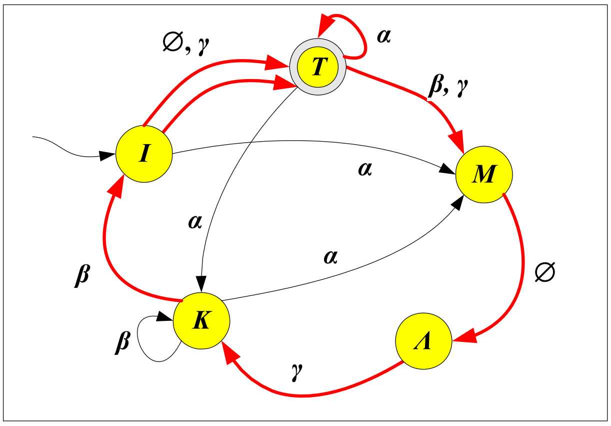
ΑΠΑΝΤΗΣΗ:
(α) Η γλώσσα Φ των φυσικών δεκαδικών είναι (εύκολα) ομαλή, και μια σειρά σαν την παραπάνω αρχίζει με [Φ (*παράθεση*) συνεχίζει με μια *αόριστη επανάληψη* (– Φ)* και τελειώνει με] . Η όλη γλώσσα είναι ομαλή από την κλειστότητα ως προς παράθεση και αόριστη επανάληψη.
(β) Οι αριθμοί αυτοί αποτελούνται από 7 μέρη: ένα προαιρετικό πρόσημο, ένα ακέραιο μέρος, μια τελεία, ένα δεκαδικό μέρος, τον εκθέτη «Ε», ένα προαιρετικό πρόσημο, και έναν φυσικό αριθμό. Όλα αυτά εκφράζονται εύκολα με ομαλές γλώσσες και ο όλος αριθμός δίδεται από την παράθεσή τους. Τα «προαιρετικά» μέρη μπορούν να εκφραστούν από την κενή λέξη (και σε επίπεδο αυτομάτου) από κενές μεταβάσεις.

■

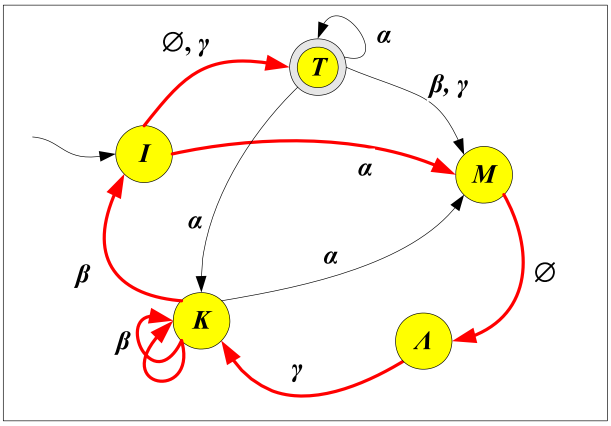
3. <u>Συμπερίληψη λέξης σε ομαλή γλώσσα</u>: Το εάν μια λέξη ανήκει σε μια ομαλή γλώσσα πιστοποιείται μόνον εάν δώσουμε την σειρά των κανόνων με τους οποίους παράγεται. Εναλλακτικά έχουμε μια δυνατότητα απεικονιστικής βεβαίωσης δίνοντας στο αντίστοιχο διάγραμμα μεταβάσεων μια *διαδρομή* από την αφετηριακή κατάσταση σε μια αποδεκτική κατάσταση η οποία σχηματίζει αυτή την λέξη.

<u>Ερώτημα:</u> Ποιές διαδρομές παράγουν τις αντίστοιχες λέξεις λ ; (*i*) λ = ***α γ γ β***. (*ii*) λ = ***α γ β β β***.

ΑΠΑΝΤΗΣΗ:
(*i*) Η εξής διαδρομή αρκεί: Ι – Τ – Μ – Λ – Κ – Ι – Τ .



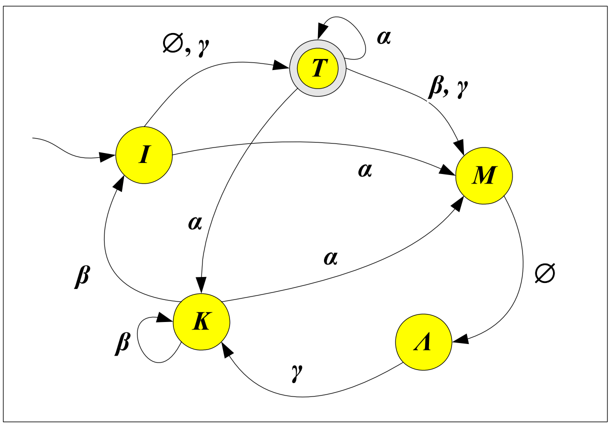
(*ii*) Η εξής διαδρομή αρκεί: Ι – Μ – Λ – Κ – Κ – Κ – Ι – Τ.



■

4. <u>Αποκλεισμός μιας λέξης από ομαλή γλώσσα</u>: Για να δείξουμε ότι μια λέξη _δεν_ ανήκει σε μια ομαλή γλώσσα πρέπει να πιστοποιήσουμε ότι δεν υπάρχει *καμμία διαδρομή* από την αφετηριακή κατάσταση σε μια αποδεκτική κατάσταση η οποία σχηματίζει αυτή την λέξη. Αυτό σημαίνει ότι πρέπει να εξετάσουμε – μέ κάποιο βολικό και σωστό τρόπο – *όλες τις δυνατές διαδρομές για αυτή την λέξη*. Αυτό είναι πολύ εύκολο στα αιτιοκρατικά αυτόματα, αλλά όχι στα μη-αιτιοκρατικά. Για τα δεύτερα ένας τρόπος είναι το να σημειώνουμε σε κάθε βήμα όλες τις καταστάσεις στις οποίες θα μπορούσαμε να είχαμε βρεθεί διαβάζοντας άλλο ένα γράμμα από την λέξη. Αν – στο τέλος – θα μπορούσαμε να είχαμε φθάσει σε *έστω μία* αποδεκτική κατάσταση τότε – και μόνον τότε – η λέξη ανήκει στη γλώσσα.

<u>Ερώτημα:</u> Δείξατε ότι το εξής (μη-αιτιοκρατικό) αυτόματο _δεν_ αποδέχεται τη λέξη: ***α α β α β*** .



ΑΠΑΝΤΗΣΗ:

Για να γίνει αποδεκτή η λέξη *ααβαβ* αρκεί να υπάρχει *έστω ένας* περίπατος από την αφετηριακή κατάσταση Ι σε μια αποδεκτική κατάσταση (εδώ την Τ) που να σχηματίζει αυτή τη λέξη. Πρέπει λοιπόν να δείξουμε ότι δεν υπάρχει κανένας τέτοιος περίπατος – άρα θα πρέπει να ελέγξουμε μεθοδικά όλες τις δυνατές διαδρομές που θα μπορούσε να ακολουθήσει το αυτόματο. Προς τούτο υπάρχει απλός και ταχύς συστηματικός τρόπος:

- Σχηματίζουμε έναν πίνακα με τόσες στήλες όσες οι καταστάσεις του αυτομάτου, και τόσες γραμμές όσες μία για την αρχή και από μία για κάθε σύμβολο που θα διαβάσουμε.
- Συμπληρώνουμε την πρώτη γραμμή σημειώνοντας με '✓' τις καστάσεις στις οποίες θα μπορούσε να βρεθεί το αυτόματο αφετηριακά δηλαδή την αφετηριακή και συνεχίζοντας με μία ή περισσότερες κενές μεταβάσεις.
- Συμπληρώνουμε κάθε γραμμή που αντιστοιχεί σε σύμβολο σ, εξετάζοντας το σε ποιές καταστάσεις θα μπορούσε να ήταν το αυτόματο (από την προηγούμενη γραμμή) και σημειώνοντας όσες θα μπορούσε να βρεθεί σε αυτές διαβάζοντας σ – και συνεχίζοντας με μία ή περισσότερες κενές μεταβάσεις.
- Ελέγχουμε εάν στην τελική γραμμή υπάρχει *έστω μία* αποδεκτική κατάσταση.

Στο παράδειγμα θα είχαμε τα εξής:

	Ι	Κ	Λ	Μ	Τ
αρχή	✓				✓
α		✓	✓	✓	✓
α		✓	✓	✓	✓
β	✓	✓	✓	✓	✓
α		✓	✓	✓	✓
γ		✓	✓	✓	

– : Σημειώνουμε την αφετηριακή Ι και την Τ, διότι από Ι πάμε Τ με κενή μετάβαση.

α: Από Ι μέσω 'α' πάμε στην Μ και από εκεί με κενή στην Λ.
Από Τ μέσω 'α' στην Τ, και στη Κ.
Σημειώνουμε τις Κ, Λ, Μ, Τ.

α: Από Κ μέσω 'α' στην Μ και από εκεί με κενή/κενές στην Λ.
Από Λ μέσω 'α' δεν πάμε πουθενά.
Από Μ μέσω 'α' δεν πάμε πουθενά.
Από Τ μέσω 'α' στην Τ και στην Κ.
Σημειώνουμε τις Κ, Λ, Μ, Τ.

β: Από Κ μέσω 'β' στην Κ και στην Ι, και από εκεί με κενή στην Τ.
Από Λ μέσω 'β' δεν πάμε πουθενά.
Από Μ μέσω 'β' δεν πάμε πουθενά.
Από Τ μέσω 'β' στην Μ και από εκεί με κενή στην Λ.
Σημειώνουμε Ι, Κ, Λ, Μ, Τ (όλες...).

α: Από Ι μέσω 'α' στην Μ και από εκεί με κενή στην Λ.
Από Κ μέσω 'α' στην Μ και από εκεί με κενή την Λ.
Από Λ μέσω 'α' δεν πάμε πουθενά.
Από Μ μέσω 'α' δεν πάμε πουθενά.
Από Τ μέσω 'α' στην Τ και στην Κ.
Σημειώνουμε τις Κ, Λ, Μ, Τ.

γ: Από Κ μέσω 'γ' δεν πάμε πουθενά.
Από Λ μέσω 'γ' στην Κ.
Από Μ μέσω 'γ' δεν πάμε πουθενά.
Από Τ μέσω 'γ' στην Μ και από εκεί με κενή στην Λ.
Σημειώνουμε τις Κ, Λ, Μ.

Η αποδεκτική κατάσταση Τ δεν περιλαμβάνεται στην σημειωμένες καταστάσεις της τελευταίας γραμμής, άρα όποια διαδρομή και εάν ακολουθούσε το αυτόματο δεν θα αποδεχόταν την λέξη $\boldsymbol{\alpha\ \alpha\ \beta\ \alpha\ \gamma}$.

■

5. Ανώμαλες γλώσσες – λήμμα άντλησης: Για να δείξουμε ότι μια γλώσσα είναι ομαλή εξαρτάται από το να βρούμε μια ομαλή γραμματική για αυτήν. Για να δείξουμε όμως ότι _δεν_ είναι ομαλή πρέπει να δείξουμε ότι _καμμία_ ομαλή γραμματική Γ *δεν την παράγει ακριβώς*, ότι δηλαδή η Γ
είτε (i) προσθέτει λέξεις που δεν έχει η Λ,
είτε (ii) δεν παράγει λέξεις που έχει η Λ.
Το εργαλείο εδώ για το (i) είναι το λήμμα άντλησης: δείχνουμε ότι (λόγω απείρου πλήθους λέξεων) *αν η Λ ήταν ομαλή* τότε θα έπρεπε να *περιέχει (μεταξύ άλλων)* και όλες λέξεις της μορφής $\boldsymbol{x\,\mu^{(\kappa)}\,y}$, για κάποιες συγκεκριμμένες σταθερές λέξεις $\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{y}$, $\boldsymbol{\mu} \neq \varnothing$, κ = 0, 1 , 2, 3, ... και δείχνουμε ότι (*οποιαδήποτε* μορφή και εάν είχαν τα $\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{y}$), για κάποιο κ θα συνέβαινε $\boldsymbol{x\,\mu^{(\kappa)}\,y} \notin \Lambda$ – πράγμα που οδηγεί σε άτοπο.

Προσέξτε ότι μερικές φορές βολεύει να δείξουμε όχι ότι η Λ είναι ανώμαλη, αλλά ότι ένα κομμάτι της Λ′ είναι ανώμαλο – τέτοιο όμως που *θα έπρεπε να ήταν ομαλό* εάν ήταν και η Λ.

Ερώτημα: Δείξατε ότι οι εξής γλώσσες δεν είναι ομαλές.
α) η γλώσσα Λ επί του Σ = { α, β } όπου πλήθος α = πλήθος β.
β) η γλώσσα Λ επί του Σ = { α, β } όλων των λ για τα οποία $\lambda = \lambda^{(R)}$ (κατοπτρικό).

ΑΠΑΝΤΗΣΗ:
(α) Η Λ έχει οσοδήποτε μεγάλες λέξεις επομένως, από οποιαδήποτε ομαλή γραμματική και εάν παραγόταν, θα έπρεπε για κάποια συγκεκριμμένα $\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{y}$ όλες οι λέξεις $\boldsymbol{x\,\mu^{(\kappa)}\,y}$, $\kappa \geq 0$, να ανήκουν στην Λ, δηλαδή να έχουν ίσο πλήθος από ‘α’ και ‘β’. Αλλά αυτό είναι δυνατόν! Αρκεί λ.χ. τα x και y να έχουν ίσο πλήθος από ‘α’ και ‘β’, και το ίδιο να ισχύει για το ‘μ’. Δεν είναι η επανάληψη του μ που μπορεί να κάνει (εδώ) ζημιά στο πλήθος των ‘α’ και ‘β’. Γι’ αυτό χρησιμοποιούμε την γλώσσα $\{ \alpha^{(\kappa)}\beta^{(\lambda)} : \kappa, \lambda \geq 0 \}$. Αυτή είναι ομαλή (εύκολα), και η τομή της Λ ‘ με την Λ είναι η γλώσσα $\{ \alpha^{(\kappa)}\beta^{(\lambda)} : \kappa, \lambda \geq 0, \kappa = \lambda\}$. Αυτή γνωρίζουμε ότι δεν είναι ομαλή – αλλά ας επαναλάβουμε το επιχείρημα:
Η Λ′ έχει οσοδήποτε μεγάλες λέξεις, επομένως, από οποιαδήποτε ομαλή γραμματική και εάν παραγόταν, θα έπρεπε για κάποια συγκεκριμμένα $\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{y}$ όλες οι λέξεις $\boldsymbol{x\,\mu^{(\kappa)}\,y}$, $\kappa \geq 0$, να ανήκουν στην Λ, δηλαδή να έχουν ίσο πλήθος από ‘α’ και ‘β’, με όλα τα ‘α’ στην αρχή και όλα τα ‘β’ στο τέλος. Αλλά αυτό είναι (τώρα) αδύνατον:
(i) αν το μ έχει ίσο πλήθος από ‘α’ και ‘β’ τότε η επανάληψη του θα τα ανεμίγνυε και δεν θα είχαμε όλα τα ‘α’ στην αρχή και όλα τα ‘β’ στο τέλος.
(ii) αν το μ έχει διαφορετικό πλήθος από ‘α’ και ‘β’ τότε η επανάληψη του – έστω και μία φορά – θα κατέστρεφε το ισοζύγιο και δεν θα είχαμε ίσο πλήθος από ‘α’ και ‘β’.
Επομένως η Λ′ δεν παράγεται από καμμία ομαλή γραμματική, άρα ούτε και η Λ (διότι τότε και η Λ′ θα ήταν ομαλή (από θ. κλειστότητας τομής)).

(β) Έστω Χ η γλώσσα των λέξεων με ένα και μόνον ‘β’. Η Χ είναι προφανώς ομαλή. Εάν η Λ ήταν ομαλή θα ήταν και η $\Lambda' = \Lambda \cap X$. Αλλά εάν $\lambda = \lambda^{(R)}$ και η λ έχει ένα ακριβώς ‘β’, τότε $\lambda = \alpha^{(\kappa)}\beta\alpha^{(\kappa)}$, $\kappa \geq 0$. Γνωρίζουμε ότι αυτή τη γλώσσα δεν είναι ομαλή – αλλά ας επαναλάβουμε το επιχείρημα: Η Λ′ έχει οσοδήποτε μεγάλες λέξεις, επομένως, από οποιαδήποτε ομαλή γραμματική και εάν παραγόταν, θα έπρεπε για κάποια συγκεκριμμένα $\boldsymbol{x}$, $\boldsymbol{\mu}$, $\boldsymbol{y}$ όλες οι λέξεις $\boldsymbol{x\,\mu^{(\kappa)}\,y}$, $\kappa \geq 0$, να ανήκουν (*μεταξύ άλλων*) στην Λ, δηλαδή να έχουν ένα ‘β’ και ίσο πλήθος από ‘α’ πριν και μετά από το ‘β’. Αλλά αυτό είναι αδύνατον.
Έστω ότι η $\lambda = \boldsymbol{x\,\mu\,y}$ ανήκει στην Λ′, ότι έχει δηλαδή την μορφή $\lambda = \alpha^{(\kappa)}\beta\alpha^{(\kappa)}$.

(i) αν το μ περιέχει το 'β' τότε η $\boldsymbol{x\mu\mu y}$ θα είχε 2 'β' και η λέξη λ δεν θα ήταν στην Λ'.
(ii) αν το μ δεν περιέχει 'β' τότε ανήκει όλο στο 1ο μέρος πριν το 'β', είτε όλο στο 2ο μέρος, και η επανάληψη του στην λέξη $\boldsymbol{x\mu\mu y}$ θα κατέστρεφε το ισοζύγιο των 'α' ανάμεσα στο 1ο και 2ο μέρος της λέξης λ.
Επομένως η Λ' δεν παράγεται από καμμία ομαλή γραμματική, άρα ούτε και η Λ (διότι τότε και η Λ' θα ήταν ομαλή (από θ. κλειστότητας τομής)).

■

ΑΣΥΜΦΡΑΣΤΙΚΕΣ ΓΡΑΜΜΑΤΙΚΕΣ ΚΑΙ ΓΛΩΣΣΕΣ.

1. Ασυμφραστικές γραμματικές: Το εάν μια γραμματική είναι ασυμφραστική ή όχι φαίνεται στην μορφή της, στον τρόπο με τον οποίο είναι γραμμένη: πρέπει κάθε παραγωγικό σύμβολο («κεφαλαίο») να μετατρέπεται σε μια λέξη λ, από τερματικά ή παραγωγικά σύμβολα ανάμεικτα, (ή να απαλείφεται μετατρεπόμενο σε «κενό» ∅).

Ερώτημα: Ποιές από τις εξής γραμματικές είναι ασυμφραστικές;

$I \to \alpha\beta K\gamma \mid \varnothing$ $K \to \Lambda M \mid M\Lambda$ $\Lambda \to \beta I \beta K \mid \varnothing$ $M \to \gamma MM \mid \Lambda$ (α)	$I \to \gamma\gamma\gamma\Lambda$ $K \to \Lambda \mid K \mid \varnothing$ $\Lambda \to K\beta\beta\beta \mid \alpha\beta$ $M \to M\alpha\alpha \mid \Lambda$ (β)	$I \to K$ $K\alpha \to \Lambda \mid M$ $M \to \gamma M \mid \Lambda$ (γ)

ΑΠΑΝΤΗΣΗ:
(α) Η πρίπτωση αυτή τηρεί όλους τους κανόνες και είναι ασυμφραστική.
(β) Η περίπτωση αυτή μοιάζει να είναι ακόμα και ομαλή γραμματική γιατί δεξιά εμφανίζεται πάντοτε ένα μόνον παραγωγικό σύμβολο, αλλά αυτό είναι παραπλανητικό διότι τα τερματικά σύμβολα εμφανίζονται και δεξιά και αριστερά των παραγωγικών.
(γ) Δεν είναι ασυμφραστική διότι ο 2ος κανόνας δεν περιέχει αριστερά μόνον ένα παραγωγικό σύμβολο. Κανένα παραγωγικό σύμβολο δεν απαλείφεται («$X \to \varnothing$») αλλά αυτο δεν επηρρεάζει την μορφή και δεν αποτελεί αιτία «ασυμφραστικότητας» ή μη.

■

2. Ασυμφραστικές γλώσσες: Το εάν μια γλώσσα είναι ομαλή ή όχι, εξαρτάται από το εάν μπορούμε να βρούμε μια ομαλή γραμματική γι' αυτήν ή όχι. Στις απλές περιπτώσεις αυτό το κάνουμε «κατ' ευθείαν». Στις πιο σύνθετες χρησιμοποιούμε τους κανόνες κλειστότητας. Προσοχή όμως: οι ασυμφραστικές γλώσσες _δεν_ είναι κλειστές ως προς την τομή, συμπλήρωμα ή διαφορά. Είναι κλειστές ως προς ένωση, παράθεση, επανάληψη, και κατοπτρισμό. Είναι επίσης κλειστές ως προς την τομή ***αλλά με ομαλή γλώσσα***.

Ερώτημα: Δείξτε ότι οι εξής γλώσσες Λ (επί του Σ = {α, β, γ } αν δεν δίδεται κάτι άλλο) είναι ασυμφραστικές:
α) Λ = «όλες οι λέξεις που αρχίζουνε με μια σειρά από 'α' και τελειώνουν με μια σειρά από 'β' ή 'γ' (ανακατεμένα) και πλήθος(α) = πλήθος(β) + πλήθος(γ)»
β) Λ = «όλες οι λέξεις από ισορροπημένες παρενθέσεις που _δεν_ περιέχουν τρείς απλές παρενθέσεις στη σειρά (δηλαδή το '**( ) ( ) ( )**')»
γ) Λ = «όλες οι λέξεις που περιέχουν 3 ***α*** στη σειρά αλλά δεν περιέχουν 3 ***β*** στη σειρά ».

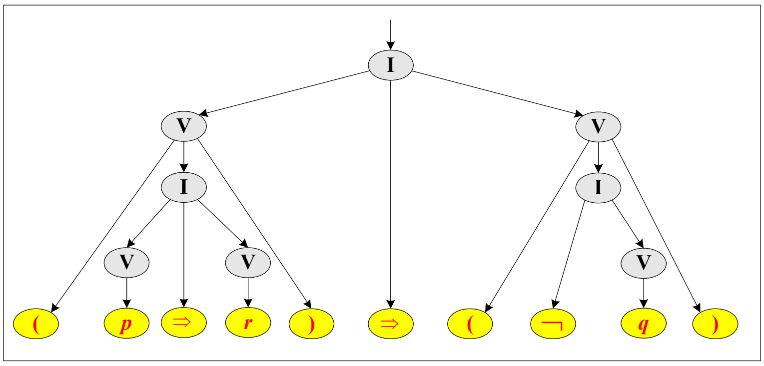
ΑΠΑΝΤΗΣΗ:
(α) Οι εξής κανόνες αρκούν: $I \to \alpha I T \mid \varnothing$, $T \to \beta \mid \gamma$.
(β) Γνωρίζουμε ότι η γλώσσα Λ_I των ισορροπημένων παρενθέσεων είναι ασυμφραστική, και γνωρίζουμε ότι η γλώσσα $\Lambda' = \Sigma^* - [\ \Sigma^* ()()() \Sigma^*\ ]$ είναι ομαλή ως συμπλήρωμα ομαλής. Η τομή $\Lambda = \Lambda_I \cap \Lambda'$ είναι ασυμφραστική ως *τομή ασυμφραστικής* *με ομαλή γλώσσα*, και είναι η ζητουμένη.
(γ) Είδαμε προηγουμένως ότι η Λ είναι ομαλή άρα είναι και ασυμφραστική, αφού ΟΜΑΛΕΣ ΓΛΩΣΣΕΣ ⊆ ΑΣΥΜΦΡΑΣΤΙΚΕΣ ΓΛΩΣΣΕΣ.

■

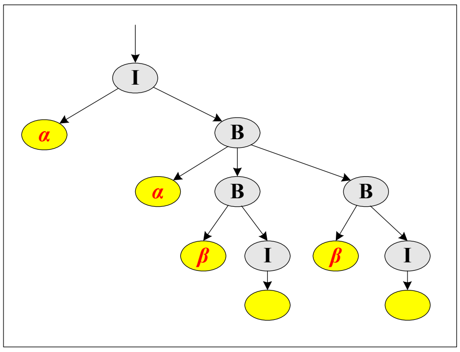
3. <u>Συμπερίληψη λέξης σε ασυμφραστική γλώσσα</u>: Το εάν μια λέξη ανήκει σε μια ασυμφραστική γλώσσα πιστοποιείται μόνον εάν δώσουμε την σειρά των κανόνων με τους οποίους παράγεται. Εναλλακτικά έχουμε μια δυνατότητα απεικονιστικής βεβαίωσης δίνοντας ένα κάποιο (ή «το») συντακτικό δένδρο (ΣΔ) που παράγει την λέξη.

<u>Ερώτημα:</u> Ποιά συντακτικά δένδρα παράγουν τις αντίστοιχες λέξεις;
(α) Γραμματική: $I \to V \mid \neg V \mid V \Rightarrow V$, $V \to (I) \mid p \mid q \mid r \mid s$, και λέξη $\lambda = (p \Rightarrow r) \Rightarrow (\neg q)$.
(β) Γραμματική: $I \to \alpha B \mid \beta A \mid \varnothing$, $A \to \alpha I \mid \beta AA$, $B \to \alpha BB \mid \beta I$, και λέξη $\lambda = \alpha\ \alpha\ \beta\ \beta$.

ΑΠΑΝΤΗΣΗ:
(α) Το εξής ΣΔ:



(β) Το εξής:



Προσέξτε την μορφή ενός ΣΔ:
- *Ρίζα*: η ρίζα περιέχει πάντοτε το <u>αρχικό</u> παραγωγικό σύμβολο (εδώ το I).
- *Κόμβοι*: οι μη-τερματικοί κόμβοι περιέχουν από <u>ακριβώς ένα παραγωγικό</u> σύμβολο (π.χ. το B).
- *Φύλλα*: οι τερματικοί κόμβοι (τα «κίτρινα» φύλλα) περιέχουν <u>τερματικές λέξεις</u>, (από «πεζά» σύμβολα, όπως εδώ 'α' ή 'β', ή και ίσως την κενή λέξη ∅).
- *Ακμές*: κάθε κόμβος K μαζί με τους θυγατρικούς του $\theta_1, \theta_2, \ldots, \theta_\nu$ από αριστερά προς τα δεξιά <u>αντιστοιχεί σε κανόνα</u> της γραμματικής $K \to \theta_1\ \theta_2, \ldots\ \theta_\nu$ (π.χ. $B \to \alpha\ B\ B$).

■

4. <u>Αιτιοκρατική μορφή ασυμφραστικής γλώσσας</u>: .

<u>Ερώτημα:</u> Δώστε μια αιτιοκρατική εκδοχή της εξής ασυμφραστικής γραμματικής:

$$I \to \alpha\beta\gamma X \mid \alpha\beta\delta\Upsilon, \quad X \to X\delta \mid \varepsilon, \quad \Upsilon \to XX \mid I.$$

ΑΠΑΝΤΗΣΗ:
Σε μια αιτιοκρατική ασυμφραστική γραμματική υπάρχει το πολύ ένας εφαρμόσιμος κανόνας κάθε στιγμή, ο οποίος μάλιστα εξαρτάται από τον επόμενο προς ανάγνωση γράμμα της υπό ανάλυση λέξης. Υπάρχουν δύο εμπόδια που πρέπει να απαλειφούν:

- απαλειφή διλημμάτων
- απαλειφή αριστερής αναδρομής (βλ. και σημειώσεις).

Στο παραπάνω παράδειγμα έχουμε το δίλημμα Ι → αβγΧ | αβδΥ. Αυτό απαλείφεται από την αλλαγή Ι → αβ Ι', και Ι' → γΧ | δΥ: έτσι το κοινό μέρος 'αβ' των 'αβγ' και 'αβδ' αποσπάται αφήνοντας το 'γ' να επιλέγει τον 1ο κανόνα, και το 'δ' το 2ο κανόνα.

Η αριστερή αναδρομή Χ → Χδ | ε, απαλείφεται με τους κανόνες Χ → εΧ', Χ' → δΧ' | ∅, αφού κατά τους παραπάνω κανόνες το Χ μετατρέπεται απλά και μόνον σε '***ε δ δ … δ***'.

Τέλος έχουμε το (κρυφό) δίλημμα Υ → ΧΧ | Ι – (και τα δύο αρχίζουν με «κενό»). Για να κάνουμε το επόμενο προς ανάγνωση σύμβολο να καθορίσει την ορθή (ή ακριβέστερα την *μόνη*) επιλογή, προσέχουμε πως εδώ ό,τι παράγεται από Χ αρχίζει με 'ε', ενώ ό,τι παράγεται από Ι αρχίζει με 'αβ', άρα αρκεί να χρησιμοποιήσουμε εναλλακτικά τους κανόνες Υ → εΧ' | αβΙ'. Μια ισοδύναμη αιτιοκρατική μορφή είναι λοιπόν:

Ι → αβ Ι'	Ι' → γΧ | δΥ
Χ → εΧ'	Χ' → δΧ' | ∅
Υ → εΧ' | αβΙ'	

■

5. Κανονική μορφή Chomsky: .

Ερώτημα: Δώστε την κανονική μορφή Chomsky για την εξής ασυμφραστική γραμματική:

Ι → ΑΒΓ | αΒ | ∅
Α → ββΙ
Β → γΙ | ∅
Γ → ΓΓ | βα

Απαντηση:

Μια γραμματική είναι σε μορφή Chomsky εάν και μόνον όλοι οι κανόνες έχουν την εξής μορφή:

- Χ → ΥΖ (ένα παραγωγικό γίνεται ακριβώς δύο παραγωγικά σύμβολα).
- Χ → σ (ένα παραγωγικό γίνεται ακριβώς ένα τερματικό σύμβολο).
- Ι → ∅ (μόνον το αρχικό σύμβολο μπορεί να καταστεί «κενό»).

Για να απαλείψουμε τον κανόνα 'Β → ∅' τον εφαρμόζουμε *προσθέτοντας* για κάθε κανόνα της μορφής Κ → ΧΒΥ (περιέχει δεξιά Β), τον ίδιο κανόνα χωρίς το Β: Κ → Χ Υ.

Ι → ΑΒΓ | ΑΓ | αΒ | α | ∅
Α → ββΙ
Β → γΙ
Γ → ΓΓ | βα

Για να κάνουμε τις απαραίτητες αλλαγές ως προς το πλήθος και είδος των συμβόλων χρησιμοποιούμε βοηθητικά σύμβολα:

$W_{ΒΓ}$	για να παραχθεί το ζεύγος ΒΓ
$W_{βα}$, $W_{ββ}$	για να παραχθούν οι λέξεις βα, ββ
$W_σ$	για να γραφεί το τερματικό σ, όπου σ = α, β, γ.

Οι αντικαταστάσεις στους κανόνες είναι περίπου προφανείς:

Ι → Α $W_{ΒΓ}$ | ΑΓ | $W_α$ Β | α | ∅
Α → $W_{ββ}$ Ι

$B \rightarrow W_\gamma I \mid \varnothing$
$\Gamma \rightarrow \Gamma\,\Gamma \mid W_{\beta\alpha}$

και,

$W_{B\Gamma} \rightarrow B\,\Gamma$
$W_{\beta\beta} \rightarrow W_\beta\, W_\beta$
$W_{\beta\alpha} \rightarrow W_\beta\, W_\alpha$

όπου,

$W_\alpha \rightarrow \alpha$, $W_\beta \rightarrow \beta$, $W_\gamma \rightarrow \gamma$

■

6. Συμφραστικές γλώσσες – λήμμα άντλησης: Για να δείξουμε ότι μια γλώσσα είναι ασυμφραστική εξαρτάται από το να βρούμε μια ασυμφραστική γραμματική που να την παράγει. Για να δείξουμε ότι _δεν_ είναι ασυμφραστική πρέπει να δείξουμε ότι _καμμία_ ασυμφραστική γραμματική Γ δεν την παράγει _ακριβώς_ , ότι δηλαδή η Γ είτε (i) προσθέτει λέξεις που δεν έχει η Λ είτε (ii) δεν παράγει λέξεις που έχει η Λ. Το εργαλείο εδώ για το (i) είναι το λήμμα άντλησης: δείχνουμε ότι (λόγω απείρου πλήθους λέξεων) *αν η Λ ήταν ασυμφραστική* τότε θα έπρεπε να *περιέχει (μεταξύ άλλων)* και όλες λέξεις της μορφής $\boldsymbol{x\,\mu^{(\kappa)}\,y\,\nu^{(\kappa)}\,z}$, για κάποιες λέξεις $\boldsymbol{x, y, z, \mu, \nu}$ όπου $|\boldsymbol{\mu\nu}| > 0$, κ = 0, 1 , 2, ... και δείχνουμε ότι (όποια ειδική μορφή και εάν είχαν οι $\boldsymbol{x, y, z, \mu, \nu}$), για κάποιο κ θα συνέβαινε $\boldsymbol{x\,\mu^{(\kappa)}\,y\,\nu^{(\kappa)}\,z} \notin \Lambda$ – πράγμα που οδηγεί σε άτοπο.

Ερώτημα: Δείξατε ότι η εξής γλώσσα Λ δεν είναι ασυμφραστική: $\Lambda = \{\ \alpha^{n^2} : n \geq 1\ \}$ (δηλαδή: «τόσα ‘α’ όσο ένα οποιαδήποτε τέλειο τετράγωνο») .

ΑΠΑΝΤΗΣΗ:
Έστω ότι η Λ παράγεται από κάποια ασυμφραστική γραμματική Γ. Όσο μεγάλη και εάν είναι η Γ η γλώσσα Λ έχει απείρου πλήθους λέξεις (διότι έχουμε απείρου πλήθους τέτράγωνα n^2) και άρα – κατά το λήμμα της άντλησης – κάποια από αυτές θα έχει την μορφή $\boldsymbol{x\,\mu^{(i)}\,y\,\nu^{(i)}\,z}$, $|\boldsymbol{\mu\nu}| > 0$, τέτοια ώστε όλες οι λέξεις $\lambda = \boldsymbol{x\,\mu^{(\kappa)}\,y\,\nu^{(\kappa)}\,z}$, $\boldsymbol{\kappa \geq 0}$, να ανήκουν στην Λ, δηλαδή το μήκος τους να είναι τέλειο τετράγωνο ή $\lambda = \alpha^{n^2}$.
Αν $p = |\boldsymbol{x}|+|\boldsymbol{y}|+|\boldsymbol{z}|$, και $q = |\boldsymbol{\mu}|+|\boldsymbol{\nu}|$, αυτό σημαίνει ότι για όλα τα κ θα πρέπει: πλήθος(‘α’) = $|\boldsymbol{x\,\mu^{(\kappa)}\,y\,\nu^{(\kappa)}\,z}| = p+\kappa q = n^2$ (για κάποιο κατάλληλο n). Αυτό όμως δεν μπορεί εδώ να συμβαίνει συνεχώς διότι το αριστερό μέλος αυξάνεται «γραμμικά» ενώ το δεξιό «τετραγωνικά»· ***συγκεκριμμένα***: για τις διαδοχικές τιμές του κ = q και κ = q+1, θα είχαμε,
$p + (q)q = n^2$, (δηλαδή $n > q$), αλλά και,
$p + (q+1)q = (n')^2 \geq (n+1)^2$, (που είναι το πλησιέστερο μεγαλύτερο τετράγωνο).
Αφαιρώντας την ισότητα από την ανισότητα θα λαμβάναμε,
$q \geq [(n+1)^2 - n^2]$ ή $q \geq 2n+1$, το οποίο αντιφάσκει με το $n > q$.
Από το άτοπον συμπεραίνουμε ότι δεν υπάρχει ασυμφραστική γραμματική Γ που να παράγει την γλώσσα Λ.

■