



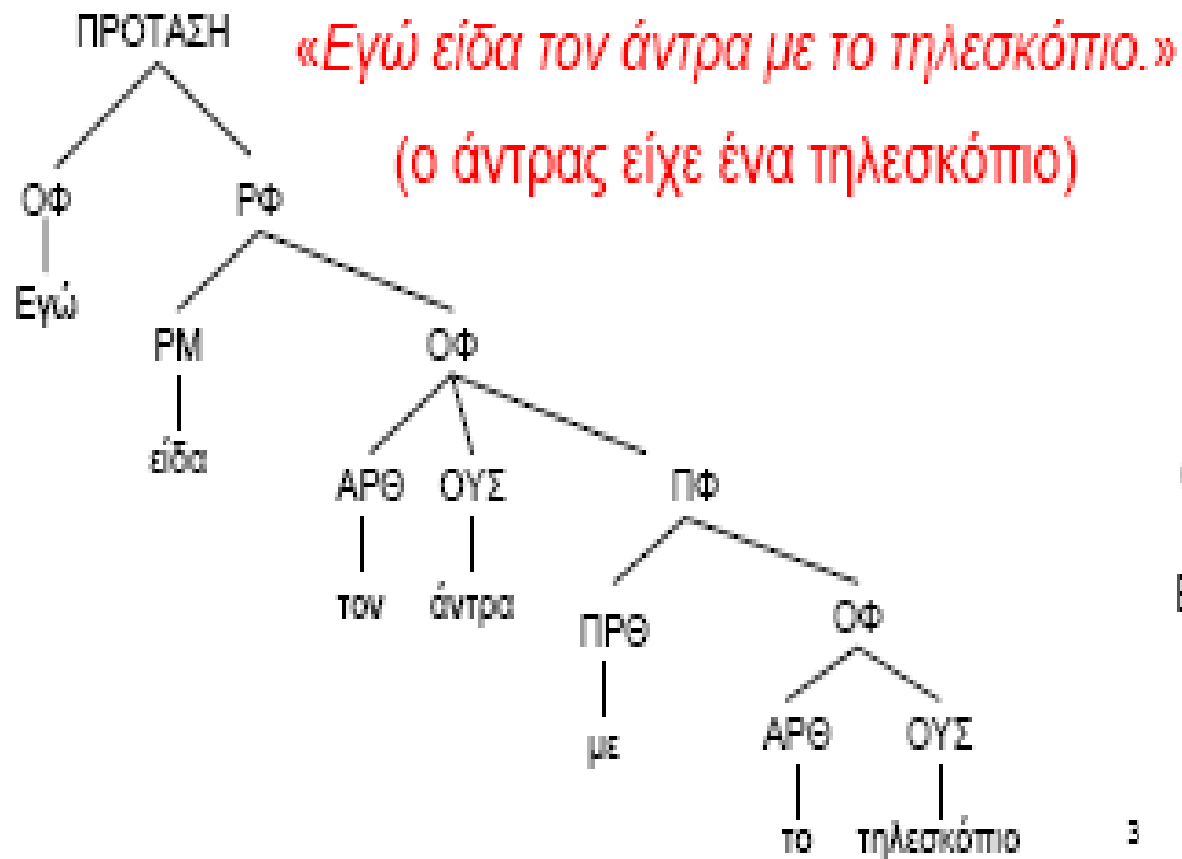
ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

Κάτια Κερμανίδου kerman@ionio.gr

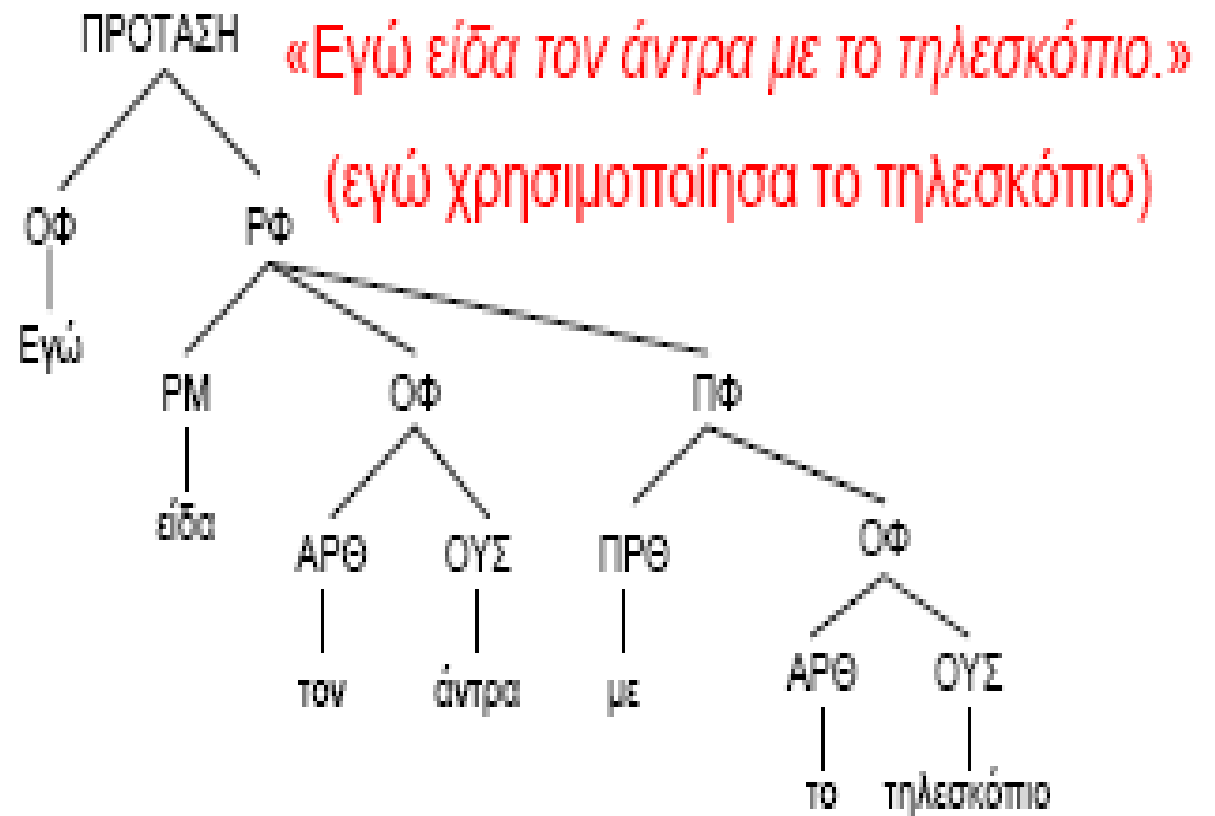
ΣΥΝΤΑΞΗ

Ορισμένα στοιχεία είναι δανεισμένα από τον W. Sun, Πανεπιστήμιο του Cambridge και από την διάλεξη της Natalie Parde

Τι είναι η Σύνταξη



Οι λέξεις μιας πρότασης είναι οργανωμένες σε εμφωλευμένες (nested) δομές (blocks).



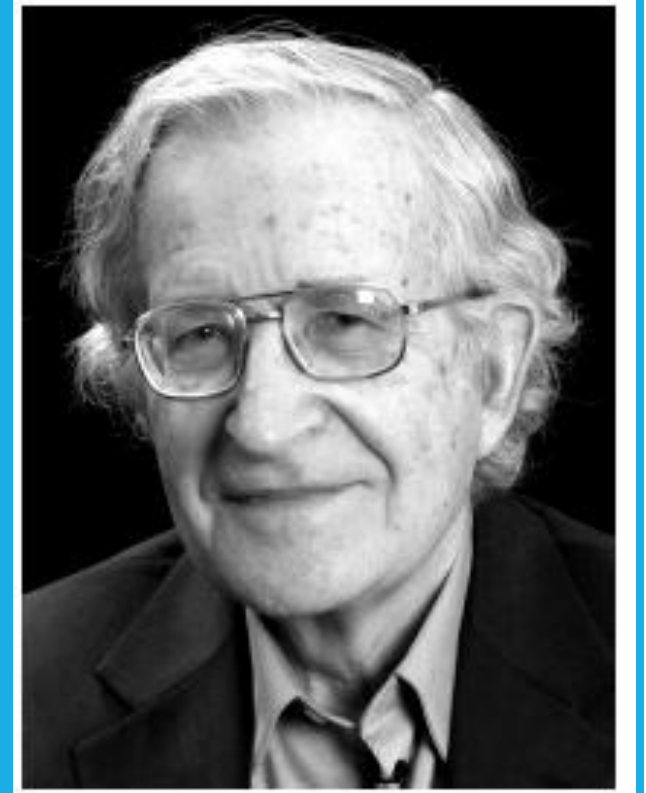
Διαφορετική δομή = Διαφορετικό νόημα

Noam Chomsky

“I think the deepest property of language and puzzling property that’s been discovered is what is sometimes called **structure dependence**.”

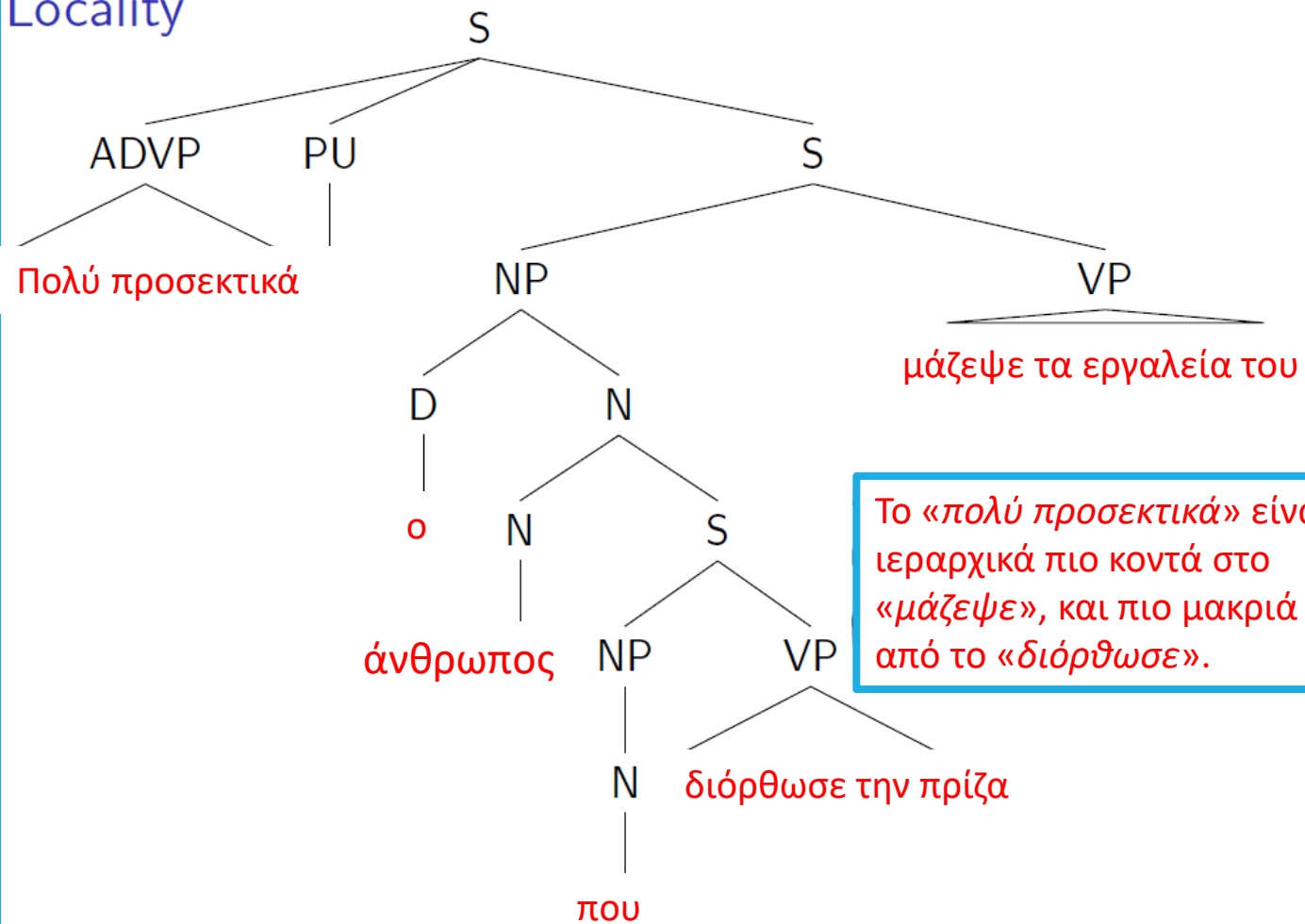
Ο άνθρωπος που διόρθωσε την πρίζα **πολύ προσεκτικά** μάζεψε τα εργαλεία του.

Πολύ προσεκτικά, ο άνθρωπος που διόρθωσε την πρίζα μάζεψε τα εργαλεία του.



Τοπικότητα (Locality)

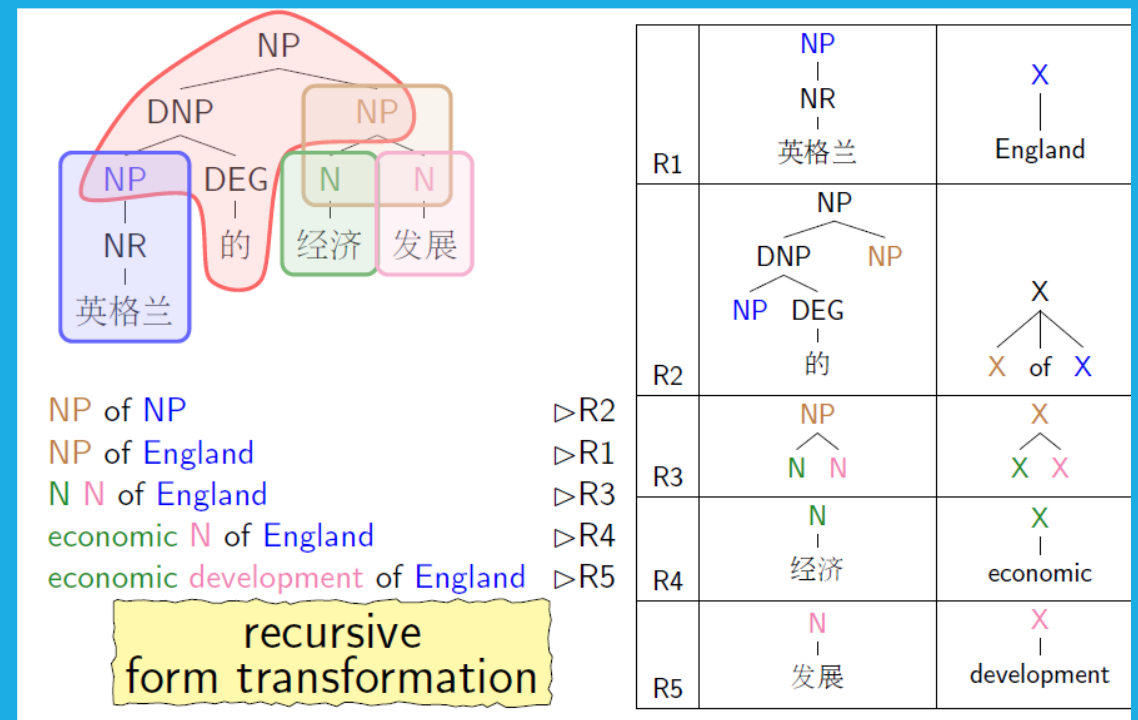
Locality



Χρησιμότητα Συντακτικής Ανάλυσης

Τα αυτόματα εργαλεία συντακτικής ανάλυσης σήμερα δίνουν για αρκετές γνώσεις πληροφορία χρήσιμη για διάφορες εφαρμογές, όπως

- ❖ Μηχανική Μετάφραση
- ❖ Εξαγωγή Πληροφορίας
- ❖ Γραμματικός έλεγχος
- ❖ ...



Τα δυο συστατικά της Συντακτικής Ανάλυσης

- ❑ Σχεδόν όλα τα υπάρχοντα συστήματα έχουν δύο συστατικά
 - Γραμματική (Grammar): ρητή αναπαράσταση των συντακτικών κανόνων της γλώσσας
 - Αναλυτής (Parser): συγκρίνει τη γραμματική με τις προτάσεις εισόδου και παράγει συντακτικές δομές

Γραμματική

Μια Γραμματική παράγει (αναλύει σωστά) όλες τις σωστές (αποδεκτές) προτάσεις μιας γλώσσας και μόνο αυτές.

Αποτελείται από

- Ένα κλειστό (πεπερασμένο) σύνολο Σ τερματικών συμβόλων (λέξεων)
- Ένα κλειστό σύνολο N μη τερματικών συμβόλων (συντακτικών κατηγοριών), που είναι διαφορετικά από το Σ
- Ένα μη τερματικό σύμβολο S , που είναι το start symbol
- Ένα κλειστό σύνολο R κανόνων παραγωγής, που αντιστοιχούν (ταιριάζουν, μετατρέπουν) μια ακολουθία συμβόλων σε/με μια άλλη.

$N = \{S, NP, VP, AdjP, AdvP\} \cup \{N, Adj, Adv\}$

$\Sigma = \{colorless, green, ideas, sleep, furiously\}$

R

$S \rightarrow NP VP$	$NP \rightarrow AdjP NP$
$VP \rightarrow VP AdvP$	
$VP \rightarrow V$	$NP \rightarrow N$
$AdvP \rightarrow Adv$	$AdjP \rightarrow Adj$
$Adj \rightarrow colorless$	$Adj \rightarrow green$
$N \rightarrow ideas$	$V \rightarrow sleep$
$Adv \rightarrow furiously$	

$S = S$

Παράδειγμα Ανάλυσης

$N = \{S, NP, VP, AdjP, AdvP\} \cup \{N, Adj, Adv\}$

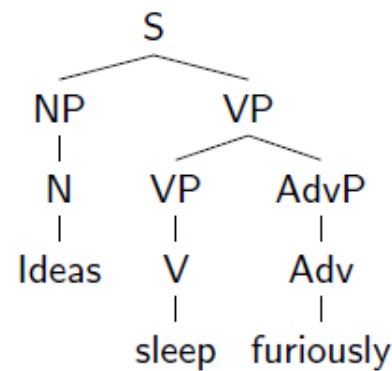
$\Sigma = \{colorless, green, ideas, sleep, furiously\}$

R

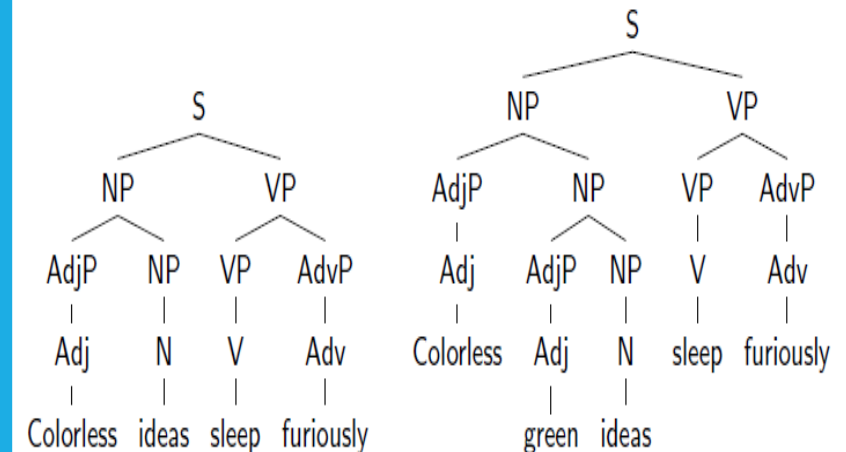
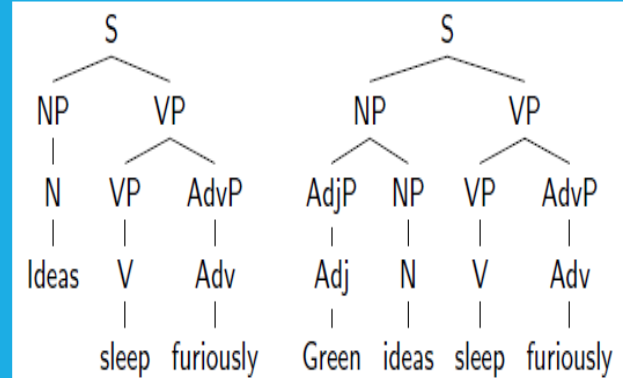
$S \rightarrow NP VP$	$NP \rightarrow AdjP NP$
$VP \rightarrow VP AdvP$	$NP \rightarrow N$
$VP \rightarrow V$	$AdjP \rightarrow Adj$
$AdvP \rightarrow Adv$	
$Adj \rightarrow colorless$	$Adj \rightarrow green$
$N \rightarrow ideas$	$V \rightarrow sleep$
$Adv \rightarrow furiously$	

$S = S$

$S \Rightarrow NP VP$
 $\Rightarrow N VP$
 $\Rightarrow ideas VP$
 $\Rightarrow ideas VP AdvP$
 $\Rightarrow ideas V AdvP$
 $\Rightarrow ideas sleep AdvP$
 $\Rightarrow ideas sleep Adv$
 $\Rightarrow ideas sleep furiously$



Η γλώσσα ορίζεται εφαρμόζοντας όλες τις δυνατές παραγωγές:



Η Ιεραρχία του Chomsky (Chomsky Hierarchy)

- ❑ Κανονική γραμματική (Regular Grammar) ή τύπου 3: στα αριστερά ένα μη-τερματικό σύμβολο, στα δεξιά ένα τερματικό σύμβολο και πιθανώς ένα μη-τερματικό
 - $A \rightarrow aB$ ή
 - $A \rightarrow a$
- ❑ Γραμματική ελεύθερης σύνταξης (Context-free Grammar) ή τύπου 2: στα αριστερά ένα και μόνο ένα μη-τερματικό σύμβολο, στα δεξιά οποιοσδήποτε συνδυασμός τερματικών και μη-τερματικών συμβόλων
- ❑ Γραμματική ευαίσθητης σύνταξης (Context-sensitive Grammar) ή τύπου 1: κανόνες της μορφής $X A Y \rightarrow X B Y$
- ❑ Ελεύθερη Γραμματική (Free ή Recursively Enumerable Grammar) ή τύπου 0: Οτιδήποτε κανόνες

Κάθε γραμματική τύπου 3 είναι και τύπου 2, 1 και 0. Κάθε γραμματική τύπου 2 είναι και τύπου 1 και 0, κ.ο.κ.

Συντακτική Ανάλυση - Parsing

□ Top-down

- Ξεκινώντας από το αρχικό σύμβολο (S), εφαρμόζουμε τους γραμματικούς κανόνες προς τα εμπρός (αντικαθιστώντας το αριστερό σύμβολο ενός κανόνα με το δεξί) μέχρι οι τερματικοί κόμβοι του δέντρου να αντιστοιχούν σε λέξεις της πρότασης.
- Καθοδηγούμενο από τον στόχο (goal-driven)

□ Bottom-up

- Ξεκινώντας από τις λέξεις της πρότασης εφαρμόζουμε τους γραμματικούς κανόνες αντίστροφα (αντικαθιστώντας το δεξί τους μέρος με το αριστερό) μέχρι να φτάσουμε στο αρχικό σύμβολο (S).
- Καθοδηγούμενη από τα δεδομένα (την πρόταση) – data driven

Top-Down Parsing: Example

Input Sentence:

Book that flight.

Grammar:

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP

Lexicon:

Det → that | this | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Pronoun → I | she | me
Proper-Noun → Houston | NWA
Aux → does
Preposition → from | to | on | near | through

Top-Down Parsing: Example

Book that flight.

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP



15 συνδυασμοί



15 συνδυασμοί



5
συνδυ
ασμοί

Top-Down and Bottom-Up Parsing



Natalie Parde

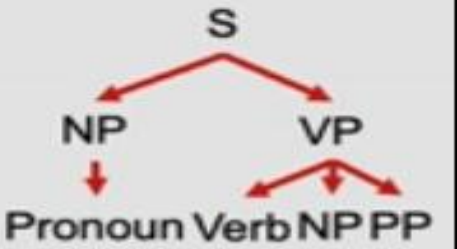
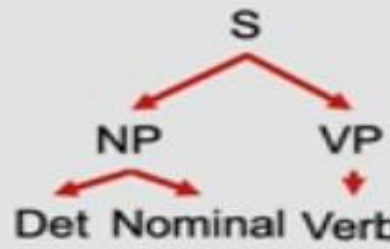
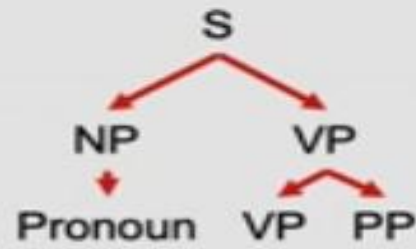
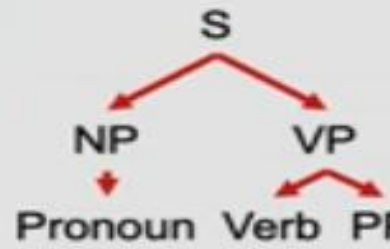
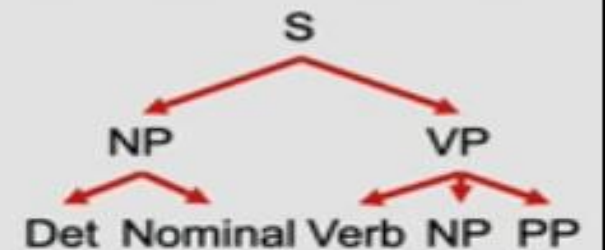
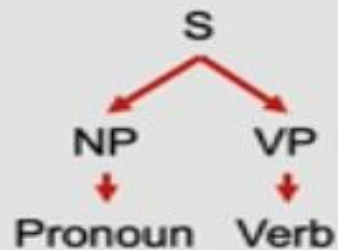
1.21K subscribers

Subscribe

Top-Down Parsing: Example

Book that flight.

$S \rightarrow NP VP$
 $S \rightarrow Aux NP VP$
 $S \rightarrow VP$
 $NP \rightarrow Pronoun$
 $NP \rightarrow Proper-Noun$
 $NP \rightarrow Det Nominal$
 $Nominal \rightarrow Noun$
 $Nominal \rightarrow Nominal Noun$
 $Nominal \rightarrow Nominal PP$
 $VP \rightarrow Verb$
 $VP \rightarrow Verb NP$
 $VP \rightarrow Verb NP PP$
 $VP \rightarrow Verb PP$
 $VP \rightarrow VP PP$
 $PP \rightarrow Preposition NP$



...and many more!

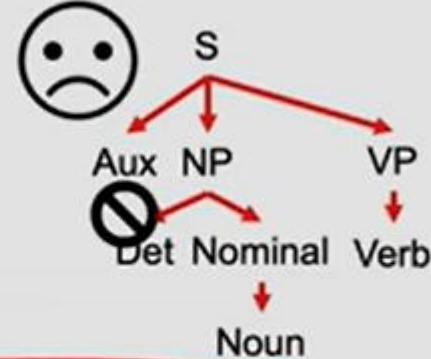
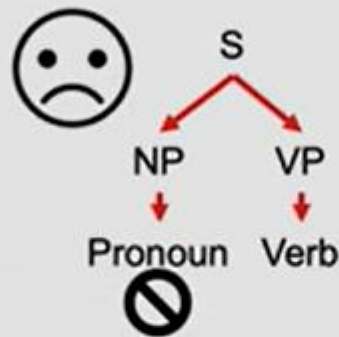
Natalie Parde - UIC CS 421

3 τρόποι να σχηματιστεί το NP x 5 τρόποι να σχηματιστεί το VP = 15 συνδυασμοί συνολικά (στο σχήμα από πάνω φαίνονται μόνο οι 7)

Top-Down Parsing: Example

Book that flight.

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP



Det → that | this | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Pronoun → I | she | me
Proper-Noun → Houston | NWA
Aux → does
Preposition → from | to | on | near | through

...and many, many more not shown!

Natalie Parde - UIUC CS 421

Από όλες αυτές τις εναλλακτικές πολλές απορρίπτονται γιατί βλέπω από νωρίς ότι δεν μπορούν να με οδηγήσουν στην πρότασή μου.

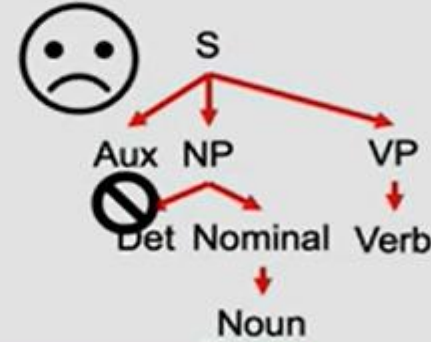
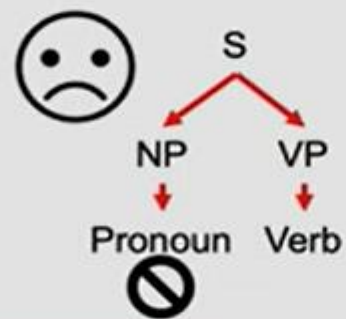
Η αριστερή επιλογή απορρίπτεται γιατί οδηγεί σε πρόταση 2 λέξεων, ενώ η δική μου πρόταση έχει 3 λέξεις.

Η μεσαία απορρίπτεται γιατί ξεκινάει από AUX, το οποίο μπορεί να είναι μόνο η λέξη *does*.

Top-Down Parsing: Example

Book that flight.

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP



Det → **that** | this | a
Noun → book | **flight** | meal | money
Verb → **book** | include | prefer
Pronoun → I | she | me
Proper-Noun → Houston | NWA
Aux → does
Preposition → from | to | on | near | through

...and many, many more not shown!

Natalie Parde - UIC CS 421

- ❑ Η δεξιά επιλογή οδηγεί σε πρόταση 3 λέξεων, με μέρη του λόγου Verb – Det – Noun.
- ❑ Οι λέξεις της πρότασης *Book that flight* είναι 3, και
- ❑ μπορούν, σύμφωνα με την Γραμματική μου, να ανήκουν σε αυτή την ακολουθία μερών του λόγου, οπότε βρήκα λύση, δηλ. σωστή συντακτική ανάλυση της πρότασής μου με αυτή τη Γραμματική.

Bottom-Up Parsing: Example

Book that flight.

Noun	Det	Noun	Verb	Det	Noun
↓	↓	↓	↓	↓	↓
book	that	flight	book	that	flight

Det → that | this | a
Noun → book | flight | meal | money
Verb → book | include | prefer
Pronoun → I | she | me
Proper-Noun → Houston | NWA
Aux → does
Preposition → from | to | on | near | through

- ❑ Ξεκινάω τώρα από τις λέξεις της πρότασής μου.
- ❑ Εφαρμόζω όποιον κανόνα μπορώ από τα δεξιά προς τα αριστερά.
- ❑ Ξεκινάω με τους κανόνες του λεξικού. Το *book* μπορεί να είναι *Noun* ή *Verb*, το *that* μόνο *Det* και το *flight* μόνο *Noun*.

Bottom-Up Parsing: Example

Book that flight.

$S \rightarrow NP VP$
 $S \rightarrow Aux NP VP$
 $S \rightarrow VP$
 $NP \rightarrow Pronoun$
 $NP \rightarrow Proper-Noun$
 $NP \rightarrow Det Nominal$
 $Nominal \rightarrow Noun$
 $Nominal \rightarrow Nominal Noun$
 $Nominal \rightarrow Nominal PP$
 $VP \rightarrow Verb$
 $VP \rightarrow Verb NP$
 $VP \rightarrow Verb NP PP$
 $VP \rightarrow Verb PP$
 $VP \rightarrow VP PP$
 $PP \rightarrow Preposition NP$

Nominal Nominal
↓ ↓
Noun Det Noun
↓ ↓ ↓
book that flight

 Nominal
 ↓
Verb Det Noun
↓ ↓ ↓
book that flight

□ Το Noun μόνο του μπορεί να δώσει μόνο Nominal.

Bottom-Up Parsing: Example

Book that flight.

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP



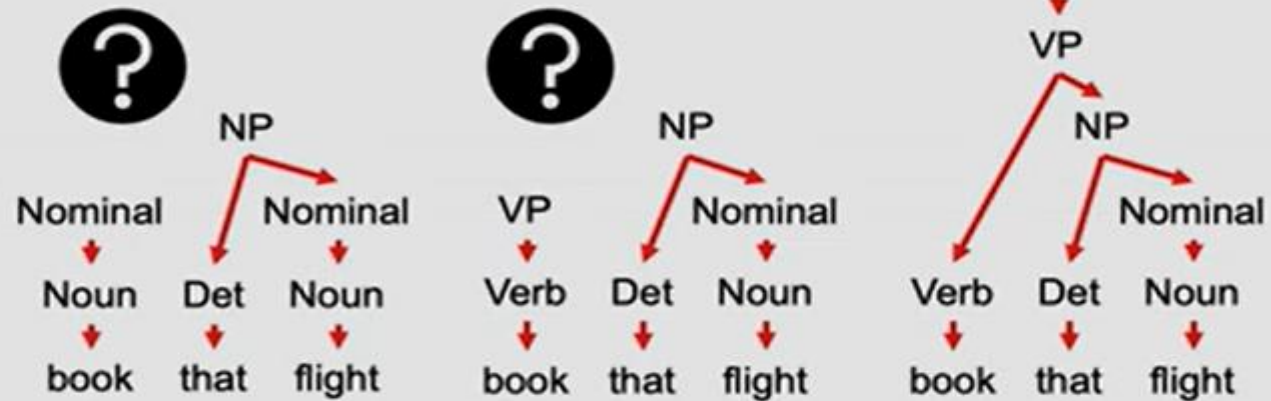
Natalie Parde - UIC CS 421

- ❑ Το DET ακολουθούμενο από Nominal δίνει NP.
- ❑ Οι άλλοι συνδυασμοί
 - ❑ Nominal Det
 - ❑ VP Det
 - ❑ Verb NP
- ❑ δεν δίνουν κάτι

Bottom-Up Parsing: Example

Book that flight.

S → NP VP
S → Aux NP VP
S → VP
NP → Pronoun
NP → Proper-Noun
NP → Det Nominal
Nominal → Noun
Nominal → Nominal Noun
Nominal → Nominal PP
VP → Verb
VP → Verb NP
VP → Verb NP PP
VP → Verb PP
VP → VP PP
PP → Preposition NP



Natalie Parde - UIC CS 421

- ❑ Τα δυο αριστερά δέντρα (Nominal NP και VP NP) δεν δίνουν κάτι
- ❑ Το δεξί (Verb NP) δίνει VP, και αυτό με την σειρά του δίνει S → σωστή συντακτική ανάλυση της πρότασης

Σύγκριση

Top-down vs Bottom-Up Parsing

- ❑ Η ανάλυση top-down δεν χάνει χρόνο εξερευνώντας δέντρα που δεν οδηγούν σε πρόταση.
- ❑ Η ανάλυση top-down δεν εξερευνά υπο-δέντρα που δεν μπορούν να αποτελέσουν τμήμα μιας σωστής πρότασης
- ❑ Η top-down ανάλυση σπαταλάει όμως πολύ χρόνο στην εξερεύνηση δέντρων που οδηγούν μεν σε πρόταση, αλλά είναι άσχετα με την πρόταση προς ανάλυση.
- ❑ Αντίθετα η bottom-up ανάλυση δεν οδηγεί ποτέ σε δέντρα που είναι άσχετα με την είσοδο
- ❑ Η bottom-up ανάλυση σπαταλάει πολύ χρόνο στο να φτιάχνει υπο-δέντρα (sub-trees) τα οποία δεν έχουν καμία ελπίδα να οδηγήσουν σε σωστή πρόταση.

Πιθανοτικές Γραμματικές (Probabilistic Grammars)

- ❑ Η θεωρία πιθανοτήτων μπορεί να βελτιώσει τις γραμματικές που έχουμε κατασκευάσει χειρονακτικά
- ❑ Μία απλή προσέγγιση είναι να καθορίσουμε μία πιθανότητα για τον κάθε κανόνα γραμματικής με βάση το πόσες φορές χρησιμοποιείται σε ένα σχολιασμένο corpus
- ❑ Πιθανοτικές Γραμματικές Ελεύθερης Σύνταξης (Probabilistic Context-free Grammars - PCFGs)

Μια PCFG

Αποτελείται από

- Ένα κλειστό (πεπερασμένο) σύνολο Σ τερματικών συμβόλων (λέξεων)
- Ένα κλειστό σύνολο N μη τερματικών συμβόλων (συντακτικών κατηγοριών), που είναι διαφορετικά από το Σ
- Ένα μη τερματικό σύμβολο S , που είναι το start symbol
- Ένα κλειστό σύνολο R κανόνων παραγωγής, που αντιστοιχούν (ταιριάζουν, μετατρέπουν) μια ακολουθία συμβόλων σε/με μια άλλη.
- Ένα αντίστοιχο σύνολο πιθανοτήτων για τους κανόνες έτσι ώστε για κάθε i

$$\sum_i P(A_i \rightarrow B_k) = 1$$

PCFGs: Παράδειγμα Υπολογισμού Πιθανοτήτων Κανόνων

- Corpus:

John talks. He laughs. He likes Mary. He puts the book on the shelf.

- Γραμματική:

$VP \rightarrow V$ (1)

$VP \rightarrow V NP$ (2)

$VP \rightarrow V NP PP$ (3)

- Πιθανότητες:

(1)	0.5
(2)	0.25
(3)	0.25

PCFGs: Παράδειγμα Ανάλυσης

□ Έστω η Γραμματική:

$S \rightarrow NP VP$ (1.0)

$PP \rightarrow P NP$ (1.0)

$VP \rightarrow V$ (0.4)

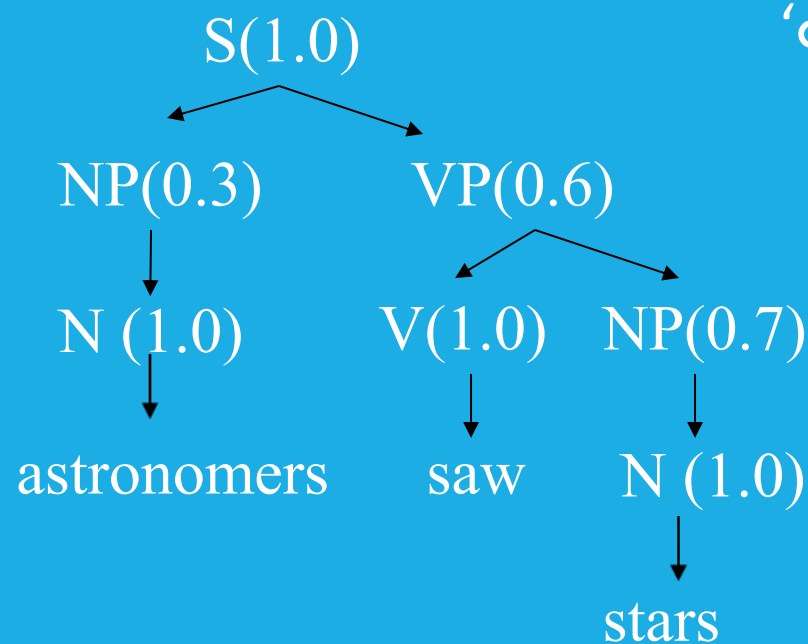
$VP \rightarrow V NP$ (0.6)

$NP \rightarrow N$ (1.0)

$V \rightarrow \text{saw}$ (1.0)

$N \rightarrow \text{astronomers}$ (0.3)

$N \rightarrow \text{stars}$ (0.7)



‘astronomers saw stars’

Η πιθανότητα μιας ανάλυσης ισούται με το γινόμενο των πιθανοτήτων των κανόνων που εμπλέκονται στο συντακτικό δέντρο της ανάλυσης.

$$P(\text{“astronomers saw stars”}) = 1.0 * 0.3 * 0.6 * 1.0 * 0.7 = 0.126$$

Σύγκριση Προσεγγίσεων Βάσει Κανόνων (Λογικής) και Στατιστικών Προσεγγίσεων

- ❑ **Προσεγγίσεις Βάσει Λογικής: Πλεονεκτήματα**
 - ❑ Η γλωσσολογική γνώση και διαίσθηση που έχουμε μπορεί να χρησιμοποιηθεί
 - ❑ Ακρίβεια
- ❑ **Προσεγγίσεις Βάσει Λογικής: Μειονεκτήματα**
 - ❑ Μεγάλο κόστος και χρονική διάρκεια κατασκευής του μοντέλου
 - ❑ Δύσκολα εφαρμόσιμη σε πραγματικά δεδομένα με καλή κάλυψη
 - ❑ Το μοντέλο δεν μπορεί να επεκταθεί εύκολα
- ❑ **Στατιστικές Προσεγγίσεις: Πλεονεκτήματα**
 - ❑ Γρηγορότερη η υλοποίηση του μοντέλου
 - ❑ Καλή κάλυψη
 - ❑ Ευρωστία - Robustness
 - ❑ Μικρές απαιτήσεις σε γνώση (knowledge poor)
 - ❑ Εφαρμόσιμη σε πραγματικά δεδομένα
 - ❑ Το μοντέλο είναι εύκολα επεκτάσιμο
- ❑ **Στατιστικές Προσεγγίσεις: Μειονεκτήματα**
 - ❑ Δεδομένα πολύ χαμηλής συχνότητας (sparse data)
 - ❑ Δυσκολία υπολογισμού στατιστικής συσχέτισης γλωσσολογικών φαινομένων