

# Αξιοποίηση Ηλεκτρονικών Πόρων στην Ελληνική Φιλολογία-Digital Humanities (9 & 10): Επεξεργασία Φυσικής Γλώσσας

Στέλλα Μαρκαντωνάτου

Υπολογιστική Γλωσσολόγος

Διευθύντρια Ερευνών

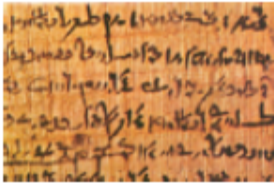
Ινστιτούτο Επεξεργασίας του Λόγου/ΕΚ ΑΘΗΝΑ

stiliani.markantonatou@gmail.com

# Θυμόμαστε;;

## General process

Raw data (analog)



Converting analog to digital  
(e.g., Scanning + OCR,  
Recording + Voice-to-text)



Digital data



Analysis (e.g.,  
machine learning,  
NLP, Computer vision)



**insights**



Visualization

**Graphic interface**

# ΘΥΜΟΣΑΣΤΕ;;

## Όταν φτιάχνω λεξικά, θέλω ένα εργαλείο...

- να μου φέρνει από σώμα κειμένων πολλές και διαφορετικές χρήσεις
- οι διαφορετικές χρήσεις σημαίνουν διαφορετικούς μορφολογικούς τύπους



το εργαλείο μου πρέπει να «ξέρει» τουλάχιστον **μορφολογία**

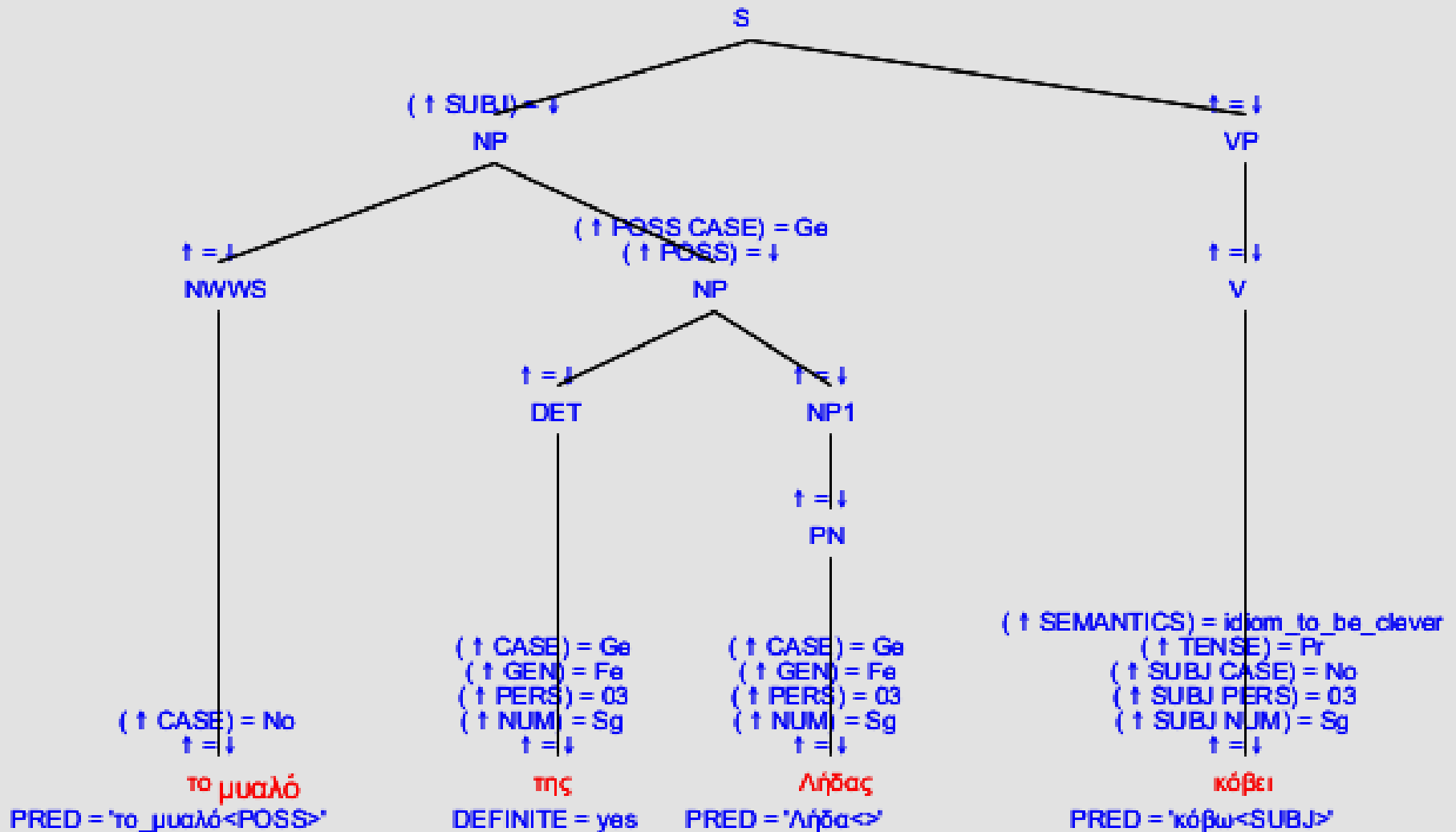
# ΘΥΜΟΣΑΣΤΕ;;

## Και γιατί μας ενδιαφέρουν οι ΠΛΕ;

### Και ολίγη Επεξεργασία Φυσικής Γλώσσας

Σε έναν **συντακτικό αναλυτή** οι λέξεις διαβάζονται μία μία και η σημασία βγαίνει από την σύνθεσή τους με βάση τη σύνταξη της δομής.

Όμως οι ΠΛΕ δεν έχουν συνθετική σημασία



# Dependency Grammars

## Γραμματικές Εξαρτήσεων

- 
- Diagram illustrating the dependency parse for the sentence "Κατά καιρούς προτάθηκαν μέθοδοι". The dependencies are shown as arcs connecting the words to their grammatical functions:
- AuxV** (Auxiliary Verb) connects to **προτάθηκαν**.
  - Subj** (Subject) connects to **μέθοδοι**.
  - Adv** (Adverb) connects to **Κατά καιρούς**.
- Κατά καιρούς προτάθηκαν μέθοδοι

- 
- Diagram illustrating the dependency parse for the sentence "Κατά\_καιρούς προτάθηκαν μέθοδοι". The dependencies are shown as arcs connecting the words to their grammatical functions:
- AuxV** (Auxiliary Verb) connects to **προτάθηκαν**.
  - Subj** (Subject) connects to **μέθοδοι**.
  - Adv** (Adverb) connects to **Κατά\_καιρούς**.
- Κατά\_καιρούς προτάθηκαν μέθοδοι

# ΘΥΜΟΣΑΣΤΕ;

## Το παράδειγμα χρήσης

- **ILSP-PAROLE:** Προτυποποιημένες μορφολογικές ετικέτες για την Ελληνική Γλώσσα
- Drop-down menus in IDION

Labropoulou, P., E. Mantzari, and M. Gavrilidou. 1996. Lexicon - Morphosyntactic Specifications: Language Specific Instantiation (Greek), PP-PAROLE, MLAP 63-386 report

(διάλεξη 7)

# ΜΟΡΦΟΛΟΓΙΚΟΣ ΑΝΑΛΥΤΗΣ

## TAGGER



# O tagger του IEA-ILSP tagger

[http://nlp.ilsp.gr/soaplab2-axis/#ilsp.ilsp\\_nlp\\_row](http://nlp.ilsp.gr/soaplab2-axis/#ilsp.ilsp_nlp_row)

token boundaries and generates POS and lemma annotations for each token. The output by default is an XCES document. Standoff alternatives include UIMA and GATE files. For more documentation and ways to cite this work please visit <http://goo.gl/1voev1>.

**Run service** **COMPLETED**

Result	Size	Type
output	4206	text
report	1401	text
detailed_status	1	unknown

**Inputs**

**InputType**

**input**

**InputEncoding**

**OutputType**

**inputsURLlist**  yes  no

**Report**

**Summary:**

Completed: Successfully  
Termination status: 0  
Started: 2016-Apr-20 12:53:37 (EEST)  
Ended: 2016-Apr-20 12:53:43 (EEST)  
Duration: 0:00:06.025

**Report:**

Name: getstarted.ilsp\_nlp  
Job ID: IELST\_getstarted.ilsp\_nlp\_IELEN\_\_24ac10b3.163f7270c81.\_7fed  
Program and parameters:  
/opt/jre1.8.0\_25/bin/java  
-jar  
/usr/local/soaplab2\_execs/bin/ilsp-nlp-asclient.jar  
-b  
tcp://ilspnlp10:61616  
-e  
ilsp-lemmatizer-aggregate-queue  
-it  
txt  
-if  
i\_input  
-il  
el  
-of

# Ο tagger του ΙΕΛ-έξοδος

Ask Home... Inbox (119... Dream Int... PARSEME (74 unread... ILSP NLP Web ... PANACEA ... ILSP NLP Web ... http://...utput x ΕΘΕΓ - Ανζέρε... Dictionary ...

nlp.ilsp.gr/soaplab2-results/IELST\_getstarted.ilsp\_nlp\_IELEN\_24ac10b3.153f7270c81\_7fed\_output

Search easydocmerge Share Files Convert Files Language Tools 75°F Athens, Greece Facebook

Search onlinemapfinder Map Search Map Links Traffic Yellow Pages Tools Facebook Listen to the Radio 75°F Athens, Greece

```
<t id="t19" word="αξέχαστων" tag="AjBaMaPIGe" lemma="αξέχαστος"/>
<t id="t20" word="«" tag="OPUNCT" lemma="«"/>
<t id="t21" word="εκουγχρονιστών" tag="NoCmMaPIGe" lemma="εκουγχρονιστής"/>
<t id="t22" word="»" tag="CPUNCT" lemma="»"/>
<t id="t23" word="," tag="PUNCT" lemma=","/>
<t id="t24" word="εκδηλώσε" tag="VbMnIdPa03SgXxPeAvXx" lemma="εκδηλώνω"/>
<t id="t25" word="από" tag="AsPpSp" lemma="από"/>
<t id="t26" word="ένα" tag="AtIdNeSgAc" lemma="ένας"/>
<t id="t27" word="βήμα" tag="NoCmNeSgAc" lemma="βήμα"/>
<t id="t28" word="της" tag="PnPofe03SgGeXx" lemma="μου"/>
<t id="t29" word="," tag="PUNCT" lemma=","/>
<t id="t30" word="προ" tag="AsPpSp" lemma="προ"/>
<t id="t31" word="ημερών" tag="NoCmFePIGe" lemma="ημέρα"/>
<t id="t32" word="," tag="PUNCT" lemma=","/>
<t id="t33" word="την" tag="AtDfFeSgAc" lemma="ο"/>
<t id="t34" word="οργή" tag="NoCmFeSgAc" lemma="οργή"/>
<t id="t35" word="του" tag="PnPoma03SgGeXx" lemma="μου"/>
<t id="t36" word="," tag="PUNCT" lemma=","/>
<t id="t37" word="για" tag="AsPpSp" lemma="για"/>
<t id="t38" word="το" tag="AtDfNeSgAc" lemma="ο"/>
<t id="t39" word="γεγονός" tag="NoCmNeSgAc" lemma="γεγονός"/>
<t id="t40" word="ότι" tag="CjSb" lemma="ότι"/>
<t id="t41" word="την" tag="AtDfFeSgAc" lemma="ο"/>
<t id="t42" word="περασμένη" tag="AjBaFeSgAc" lemma="περασμένος"/>
<t id="t43" word="Κυριακή" tag="NoPrFeSgAc" lemma="Κυριακή"/>
<t id="t44" word="ο" tag="AtDfMaSgNm" lemma="ο"/>
<t id="t45" word="πρωθυπουργός" tag="NoCmMaSgNm" lemma="πρωθυπουργός"/>
<t id="t46" word="έδωσε" tag="VbMnIdPa03SgXxPeAvXx" lemma="δίνω"/>
<t id="t47" word="συνέντευξη" tag="NoCmFeSgAc" lemma="συνέντευξη"/>
<t id="t48" word="στην" tag="AsPpPaFeSgAc" lemma="στον"/>
```

EL 12:58 μμ 20/4/2016

# ΕΚΓΥΜΝΑΣΗ ΜΗΧΑΝΩΝ

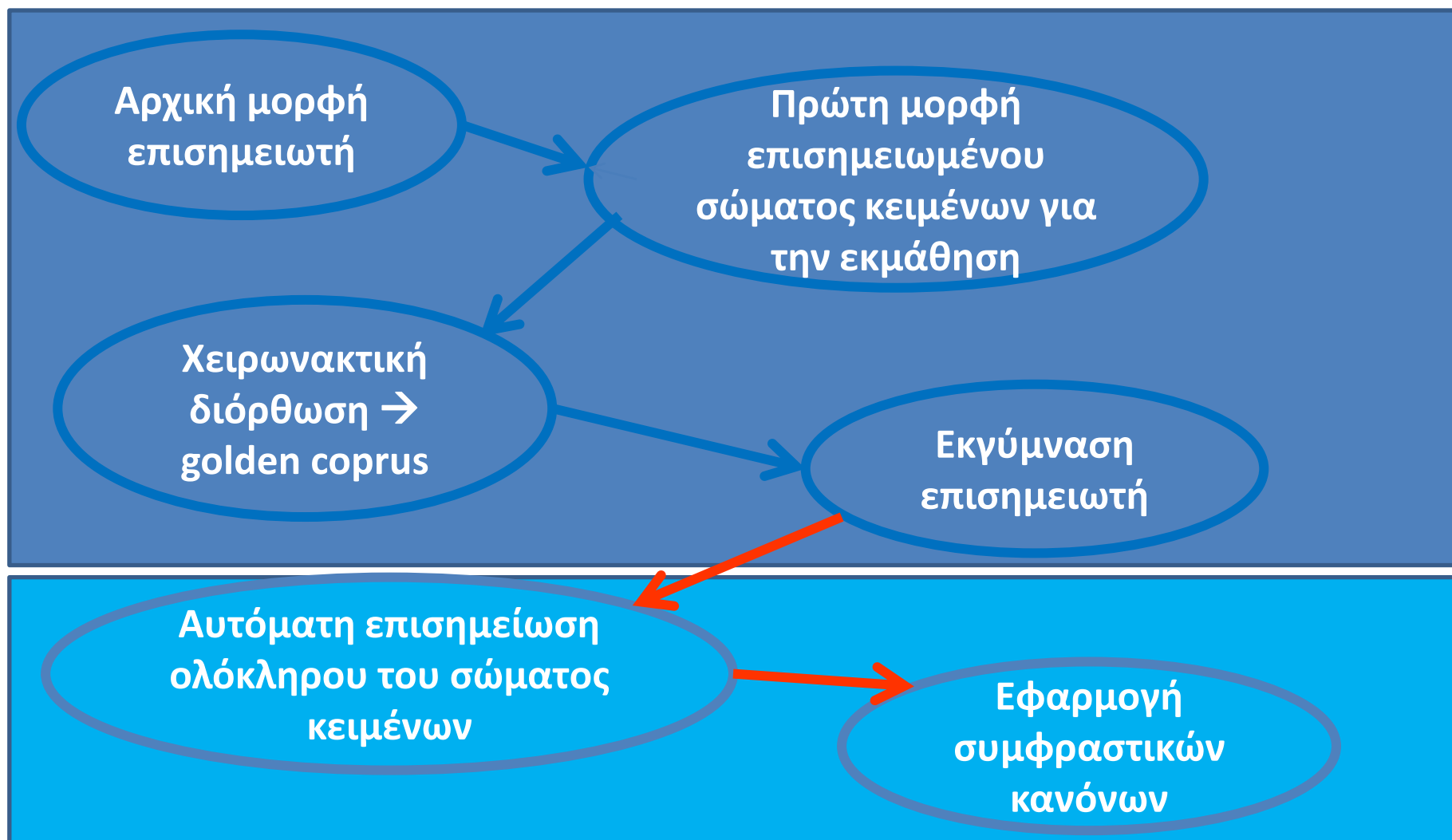
## MACHINE LEARNING

Ή αλλιώς, Μηχανική Μάθηση:

Προγράμματα τρέχουν πάνω στα δεδομένα που μας ενδιαφέρουν και **ΜΑΘΑΙΝΟΥΝ** να κάνουν κάτι, π.χ. να αποδίδουν μορφολογικές ιδιότητες

Δηλαδή, να κολλάνε στις λέξεις ετικέτες, συνήθως σε XML, που γράφουν ενικός, πληθυντικός, ουδέτερο, αόριστος κλπ.

# Πώς αναπτύχθηκε ο tagger του ΙΕΛ



# Η ημι-αυτόματη επισημείωση του σώματος κειμένων για την εκμάθηση

Prokopis Prokopidis, Byron Geograntopoulos, Haris Papageorgiou. 2001.  
A suite of NLP tools for Greek. *Proceedings of the 10th International  
Conference of Greek Linguistics*,  
Komotini, Greece [http://nlp.ilsp.gr/nlp/ICGL2011\\_Prokopidis\\_etal.pdf](http://nlp.ilsp.gr/nlp/ICGL2011_Prokopidis_etal.pdf)

- «...**part of speech tagger** called FBT. FBT is an adaptation of the Brill tagger (Brill, 1992) **trained on a manually annotated corpus of Greek texts** amounting to 455K tokens. During manual and automatic annotation, we use a tagset of **584 combinations of basic POS tags** and morphosyntactic features, which capture the rich morphology of the Greek language...»
- «For the construction of the corpus, **linguists had to correct** automatically assigned tags from an initial version of the tagger.»

# Διεπιφάνεια για την χειρωνακτική φάση της επισημείωσης

The screenshot shows the TRee Editor interface. The main window displays a text document with the following content: "#41 Το στρατόπεδο στο οποίο βρισκόμουν , στα Ουράλια , ήταν πολύ μικρό , υπήρχαν μόλις 60 κρατούμενοι και 100". The word "στρατόπεδο" is highlighted in yellow. An "Edit Tag" dialog box is open, showing a list of tag suggestions from a lexicon database. The current tag is "NoCmNeSg". The dialog box has a list of tag suggestions on the left and a list of tag suggestions from the lexicon DB on the right. The tag "NoCmNeSgNm" is selected in the list of suggestions from the lexicon DB. The dialog box also has "OK" and "Cancel" buttons.

TRee Editor Default(1/1): http://nlp.ilsp.gr/nlp/examples/OP\_PRIS.fs  
Mode: GDT\_M\_Edit

File Node Tree View Macros Setup Help

#41 Το **στρατόπεδο** στο οποίο βρισκόμουν , στα Ουράλια , ήταν πολύ μικρό , υπήρχαν μόλις 60 κρατούμενοι και 100 41/115

Edit Tag

NoCmNeSg

.. Ac Da Ge Nm Vo

Ac Da Ge Nm Vo

Tag suggestions from lexicon DB

NoCmNeSgNm

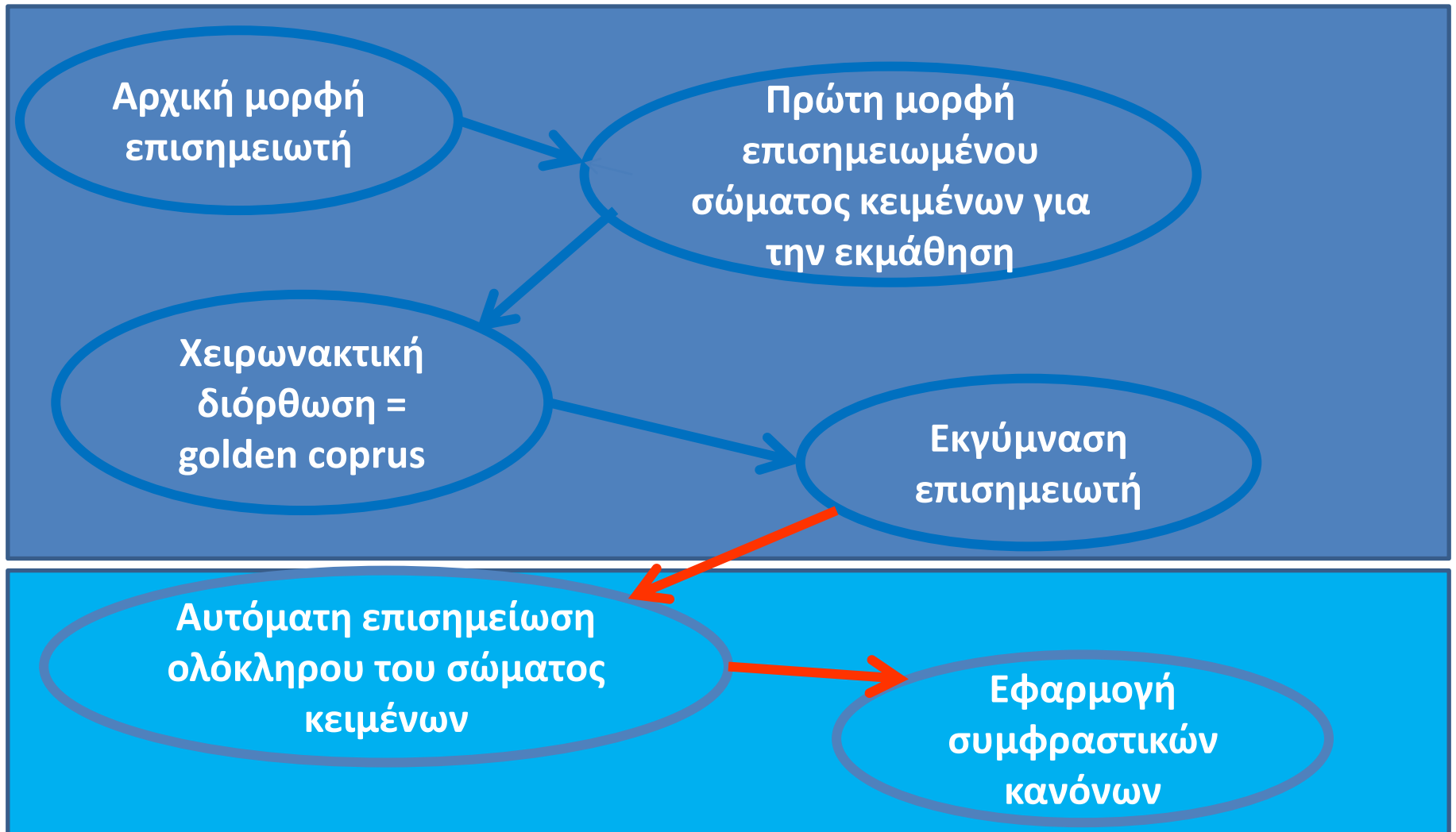
NoCmNeSgVo

OK Cancel

#41	Z#-----
Το	AtDfNeSgNm
στρατόπεδο	NoCmNeSgAc
στο	AsPpPaNeSgAc
οποίο	PnReNe03SgAcXx
βρισκόμουν	VbMnIdPa01SgXxlp
,	PUNCT
στα	AsPpPaNePIAc

Scale: 100%

# Πώς αναπτύχθηκε ο tagger



# Η αυτόματη επισημείωση ολόκληρου του σώματος κειμένων

- «The tagger assigns to each token the most frequent tag in a lexicon compiled from the training corpus and augmented with entries from ILSP's Morphological Lexicon.
- ... a set of about 800 contextual rules is applied to correct initial tags. The rules were automatically learned from the training corpus...».

**Input** μνημονιακού/AjBaNeSgGe χειμώνα/NoCmMaSgGe  
**Rule** AjBaNeSgGe ->AjBaMaSgGe NEXTTAG NoCmMaSgGe  
**Output** μνημονιακού/AjBaMaSgGe χειμώνα/NoCmMaSgGe

**Input** απαιτείται διαρκής /AjBaFeSgNm επαγρύπνηση/NoCmFeSgAc  
**Rule** NoCmFeSgAc ->NoCmFeSgNm PREVTAG AjBaFeSgNm  
**Output** απαιτείται διαρκής AjBaFeSgNm επαγρύπνηση/NoCmFeSgNm



# Ακρίβεια του tagger

- Ο επισημειωτής επιτυγχάνει ακρίβεια
  - 97.49% για τα βασικά μέρη του λόγου (POS)
  - 92.54% όταν αξιολογούνται όλα τα μορφολογικά χαρακτηριστικά

## ΑΠΟΔΟΣΗ ΛΗΜΜΑΤΟΣ

- Οι λεξικοί τύποι αντιπαραβάλλονται με ένα μορφολογικό λεξικό που περιέχει 66K λήμματα (με τις αναπτυγμένες μορφές, σύνολο 2M εγγραφές). Οι αμφισημίες επιλύονται με χρήση των ετικετών που έχει αποδώσει ο επισημειωτής.
- π.χ. Ο τύπος *ενοχλήσεις* θα λάβει ως λήμμα το *ενοχλώ* εάν ο επισημειωτής τον έχει χαρακτηρίσει «ρήμα». Εάν τον έχει χαρακτηρίσει «όνομα» θα λάβει ως λήμμα την *ενόχληση*.

# Ακρίβεια tagger και αναζητήσεις

- Ο ΕΘΕΓ είναι επισημειωμένος αυτόματα
- Όταν ρωτάμε για μια λέξη, αυτόματα γίνεται η επισημείωσή της και έτσι η αναζήτηση μετατρέπεται σε ένα απλό πρόβλημα ταιριάσματος ετικετών.
- Άρα, ο ΕΘΕΓ επιστρέφει και λάθη!!! (λίγα, αλλά υπάρχουν)
- Τα λάθη τα λέμε «ΘΟΡΥΒΟ»
- **ΕΡΩΤΗΣΗ: πώς θα τα διορθώναμε αυτά τα λάθη;**

# Θόρυβος στον ΕΘΕΓ

hnc.ilsp.gr/results.asp

αρελθόντος, περιορίζονται απλώς στο να τους γι... Search

αρελθόντος, περιορίζονται απλώς στο να τους γι... Search

155	Ο 25χρονος Σέρβος, μολονότι έπαιζε ιώθοντας <b>ενοχλήσεις</b> στο αριστερό του πόδι, ήταν και πάλι επιθετικά η κύρια πηγή όχι μόνο κινδύνου για την εστία του Εδεσσαϊκού, αλλά και δημιουργίας.
156	Ο 21χρονος Δημήτρης Πολυμέρου έφθασε το ακόντιό του στα 79, 86 αλλά δεν είδαμε την κόντρα με Κώστα Γκατσιούδη που αισθάνθηκε <b>ενοχλήσεις</b> στους προσαγωγούς κατά το ζέσταμα και δεν αγωνίστηκε.
157	Θα φταίνε οι συχνές και ξαφνικές <b>ενοχλήσεις</b> του στην πλάτη, ίσως και το πολύ πρόσφατο σοκ από το θάνατο του προπονητή του Τιμ Γκούλικσον.
158	Δεν υποβλήθηκε σε <b>ενοχλήσεις</b> , βασανιστήρια κ.ο.κ.
159	Χρόνιοι στομαχόπονοι, <b>ενοχλήσεις</b> κατά την κατάποση, κυρίως όμως τις ασυνήθιστα συχνές κενώσεις του εντέρου, ταλαιπωρούσαν από πολύ παλαιά αρκετά άτομα.
160	Ένα δεύτερο νευρικό κέντρο, που βρίσκεται στην κοιλιακή χώρα, ευθύνεται για τις <b>ενοχλήσεις</b> και παρουσιάζει στη δομή του ομοιότητες εκπληκτικές με το κεντρικό ανθρώπινο όργανο:
161	Τις λιγότερες πιθανότητες έχει ο Μαρίνος Ουζουνίδης που έκανε και χθες ατομικό πρόγραμμα, αφού οι <b>ενοχλήσεις</b> στη μέση του δεν έχουν υποχωρήσει.
162	Μοναδικό ερωτηματικό παραμένει για τον Παναθηναϊκό ο Μαρίνος Ουζουνίδης, που συνεχίζει να έχει <b>ενοχλήσεις</b> στη μέση.
163	Είτε οι νογείς σπονδυλίες οποιαδήποτε αγέννημα μαζί τους και τους απαγορεύουν τέτοιου είδους <b>ενοχλήσεις</b> ή τα ίδια τα παιδιά σπονδυλίες να τους... σπαστηθούν ως αισθητοί κροσσι του παρελθόντος, περιορίζονται απλώς στο να τους γνωρίσουν από μακριά και μετά κλείνουν οριστικά το μεγάλο φάκελο της ζωής τους που αφορά την ανεύρεση της χαμένης τους οικογένειας...
164	Για την επίτευξη της μέγιστης δυνατής ταχύτητας με μια δεδομένη ισχύ, πρέπει να <b>ενοχλήσεις</b> την ατμόσφαιρα όσο το δυνατόν λιγότερο με το πέρασμά σου.
165	Ο Στάνιεκ, πάντως, που είχε <b>ενοχλήσεις</b> στο πόδι, προπονήθηκε χθες και θα παίξει.
166	Προτείνουμε να συζητηθούν όλα τα θέματα που ενδεχομένως θα δημιουργούσαν <b>ενοχλήσεις</b> ή και προβλήματα μεταξύ των δύο κρατών.
167	Όμως αν επιστρέψουν μελλοντικά τα F16 σ' εκείνο το αεροδρόμιο, γιατί αν θα γίνει αποκλειστικά στρατιωτικό, δεν φαντάζομαι ότι θα ξεκουραστεί από τις <b>ενοχλήσεις</b> , πρέπει να βρεθεί μία λύση, ώστε κανείς να μην είναι χαμένος.
168	Πρώτα- πρώτα είναι Βουλευτής ενός νησιού κατεξοχήν τουριστικού και είναι φυσικό να έχει τέτοια μηνύματα, <b>ενοχλήσεις</b> και παρεμβάσεις για να επιλύσει σχετικά θέματα, όπως για παράδειγμα η Τουριστική Σχολή.
169	Δέχονταν <b>ενοχλήσεις</b> ακόμα και όσοι έχαιραν εκτιμήσεως ανάμεσα στους κακοποιούς, έστω κι αν δεν είχαν πάρει μέρος στις συμπλοκές των συμμοριών κανένας, όμως, από αυτούς δε σκοτώθηκε.

Σελίδα 1 από 1

Προηγούμενη Επόμενη

Screenshot Added  
A screenshot was added to your Dropbox.

να ενοχλήσεις Highlight All Match Case 1 of 1 match

EL 3:20 μμ  
20/4/2016

Greek biblical texts

**UNIVERSITY OF PENNSYLVANIA (UP), CENTER  
FOR COMPUTER ANALYSIS OF TEXTS (CCAT)**

<http://ccat.sas.upenn.edu/gopher/text/religion/biblical/lxxmorph/>

JudgA 1:1

KAI\	C	KAI/	
E) GE/NETO	VBI AMI3S	GI/GNOMAI	
META\	P	META/	
TH\N	RA ASF	O (	
TELEUTH\N	N1 ASF	TELEUTH/	
*I) HSOU=	N GSM	*I) HSOU=S	
KAI\	C	KAI/	
E) PHRW/TWN	V3I IAI3P	E) RWTA/W	E) PI
OI (	RA NPM	O (	
UI (OI\	N2 NPM	UI (O/S	
*ISRAHL	N GSM	*I) SRAH/L	
E)N	P	E)N	
KURI/W	N2 DSM	KU/RIOS	
LE/GONTES	V1 PAPNPM	LE/GW	
TI/S	RI NSM	TI/S	
A) NABH/SETAI	VF FMI3S	BAI/NW	A) NA
H (MI=N	RP DP	E) GW/	
PRO\S	P	PRO/S	
TO\N	RA ASM	O (	
*XANANAI=ON	N2 ASM	*XANANAI=OS	
A) FHGOU/MENOS	V2 PMPNSM	H (GE/OMAI	A) PO
TOU=	RA GSN	O (	
POLEMH=SAI	VA AAN	POLEME/W	
E)N	P	E)N	
AU) TW=	RD DSM	AU) TO/S	

JudgA 1:2

KAI\	C	KAI/	
EI) =PEN	VBI AAI3S	EI) =PON	
KU/RIOS	N2 NSM	KU/RIOS	
*IOUDAS	N1T NSM	*I) OU/DAS	
A) NABH/SETAI	VF FMI3S	BAI/NW	A) NA
I) DOU\	I	I) DOU/	
DE/DWKA	VX XAI1S	DI/DWMI	
TH\N	RA ASF	O (	
GH=N	N1 ASF	GH=	
E)N	P	E)N	
XEIRI\	N3 DSF	XEI/R	
AU) TOU=	RD GSM	AU) TO/S	

JudgA 1:3

# Chunker

```
<types:Lemma xmi:id="529" sofa="1" begin="0" end="5" value="απορώ"/>
<types:Lemma xmi:id="537" sofa="1" begin="6" end="9" value="πώς"/>
<types:Lemma xmi:id="545" sofa="1" begin="10" end="17" value="κάποιος"/>
<types:Lemma xmi:id="553" sofa="1" begin="18" end="25" value="ανόητος"/>
<types:Lemma xmi:id="561" sofa="1" begin="26" end="32" value="θέλω"/>
<types:Lemma xmi:id="569" sofa="1" begin="33" end="35" value="να"/>
<types:Lemma xmi:id="577" sofa="1" begin="36" end="43" value="κάμπτω"/>
<types:Lemma xmi:id="585" sofa="1" begin="44" end="46" value="ο"/>
<types:Lemma xmi:id="593" sofa="1" begin="47" end="52" value="ηθικός"/>
<types:Lemma xmi:id="601" sofa="1" begin="53" end="56" value="μου"/>
<types:Lemma xmi:id="609" sofa="1" begin="57" end="58" value=","/>
<types:Lemma xmi:id="617" sofa="1" begin="58" end="61" value="που"/>
<types:Lemma xmi:id="625" sofa="1" begin="62" end="64" value="εγώ"/>
<types:Lemma xmi:id="633" sofa="1" begin="65" end="73" value="ανεβάζω"/>
<types:Lemma xmi:id="641" sofa="1" begin="74" end="78" value="ψηλά"/>
<types:Lemma xmi:id="649" sofa="1" begin="79" end="81" value="ο"/>
<types:Lemma xmi:id="657" sofa="1" begin="82" end="88" value="παιδί"/>
<types:Lemma xmi:id="665" sofa="1" begin="89" end="92" value="ο"/>
<types:Lemma xmi:id="673" sofa="1" begin="93" end="96" value="Ρίο"/>
<types:Lemma xmi:id="681" sofa="1" begin="96" end="97" value="!"/>
<types:Chunk xmi:id="689" sofa="1" begin="0" end="5" label="vg"/>
<types:Chunk xmi:id="697" sofa="1" begin="10" end="25" label="np_nm"/>
<types:Chunk xmi:id="705" sofa="1" begin="18" end="25" label="adjp_nm"/>
<types:Chunk xmi:id="713" sofa="1" begin="26" end="32" label="vg"/>
<types:Chunk xmi:id="721" sofa="1" begin="33" end="43" label="vg_s"/>
<types:Chunk xmi:id="729" sofa="1" begin="44" end="56" label="np_ac"/>
<types:Chunk xmi:id="737" sofa="1" begin="47" end="56" label="adjp_ac"/>
<types:Chunk xmi:id="745" sofa="1" begin="58" end="61" label="pou_np_nm"/>
<types:Chunk xmi:id="753" sofa="1" begin="62" end="64" label="np_ac"/>
<types:Chunk xmi:id="761" sofa="1" begin="65" end="73" label="vg"/>
<types:Chunk xmi:id="769" sofa="1" begin="74" end="78" label="advp"/>
<types:Chunk xmi:id="777" sofa="1" begin="79" end="88" label="np_nm"/>
<types:Chunk xmi:id="785" sofa="1" begin="89" end="96" label="np_ge"/>
<types:Clause xmi:id="793" sofa="1" begin="0" end="5" label="cl"/>
<types:Clause xmi:id="801" sofa="1" begin="6" end="32" label="cl o"/>
```

Απορώ πως κάποιος ανόητος θέλουν να κάμψουν το ηθικό μας ,που το ανέβασαν ψηλά τα παιδιά του Ρίο!

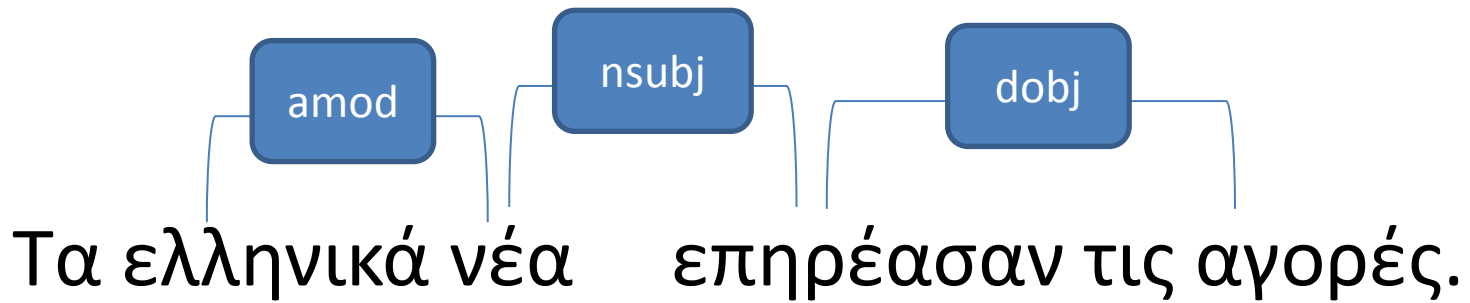
# ΓΡΑΜΜΑΤΙΚΗ ΤΩΝ ΕΞΑΡΤΗΣΕΩΝ

## DEPENDENCY GRAMMAR

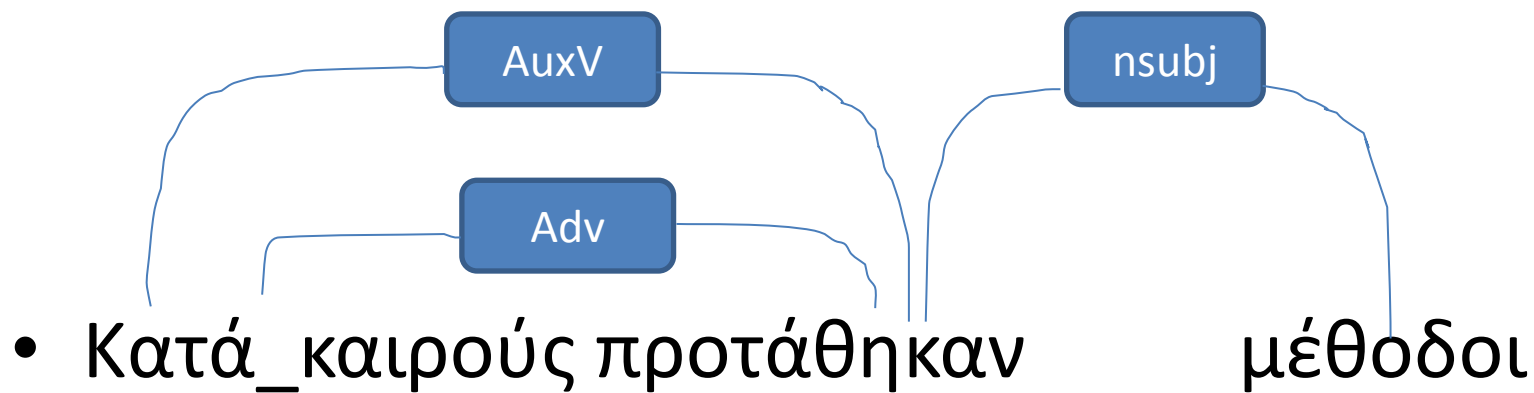
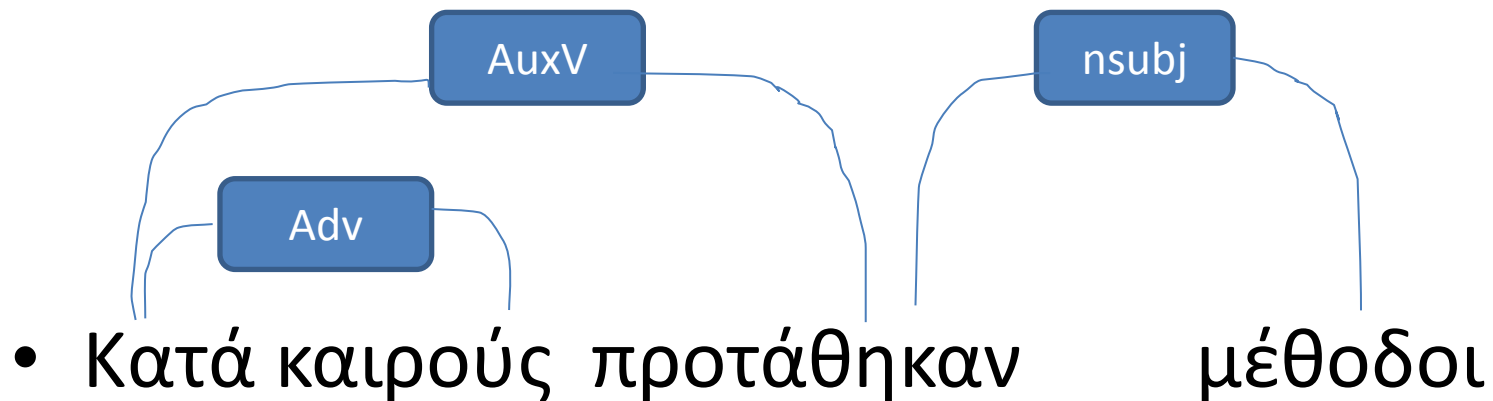
Η συντακτική δομή αποτελείται από λέξεις που συνδέονται μεταξύ τους με δυαδικές σχέσεις τις οποίες αποκαλούμε εξαρτήσεις.

Joakim Nivre. Introduction to Dependency Grammar and Dependency Parsing.

<http://ufal.mff.cuni.cz/~bejcek/parseme/prague/nivre1.pdf>







```

# sent_id = o_outputout1
# text = "Απορώ πως κάποιοι ανόητοι θέλουν να κάψουν το ηθικό μας ,που το ανέβασαν ψηλά τα παιδιά του Ρίο!"
1      "      "      PUNCT  OPUNCT  PunctType=OPUNCT      2      punct      SpaceAfter=No
2      Απορώ  απορώ  VERB    VbMn    Aspect=Imp|Mood=Ind|Number=Sing|Person=1|Tense=Pres|Voice=Act  0      root      _      _
3      πως    πως    SCONJ   CjSb    _      6      mark      _      _
4      κάποιοι  κάποιος PRON    PnId    Case=Nom|Gender=Masc|Number=Plur|Person=3|PronType=Ind  5      det      _      _
5      ανόητοι  ανόητος ADJ     Aj      Case=Nom|Gender=Masc|Number=Plur      6      nsubj    _      _
6      θέλουν  θέλω    VERB    VbMn    Aspect=Imp|Mood=Ind|Number=Plur|Person=3|Tense=Pres|Voice=Act  2      ccomp    _      _
7      να      να      PART    PtSj    ParticleType=Sub      8      aux      _      _
8      κάψουν  κάμπτω VERB    VbMn    Aspect=Perf|Mood=Ind|Number=Plur|Person=3|Voice=Act  6      ccomp    _      _
9      το      ο      DET     AtDf    Case=Acc|Definite=Def|Gender=Neut|Number=Sing|PronType=Art  10     det      _      _
10     ηθικό   ηθικός  ADJ     Aj      Case=Acc|Gender=Neut|Number=Sing      8      obj      _      _
11     μας    μου    PRON    PnPp    Case=Gen|Gender=Masc|Number=Plur|Person=1|Poss=Yes|PronType=Prs  10     nmod     _      _
12     ,      ,      PUNCT   PUNCT   _      15     punct    SpaceAfter=No
13     που    που    PRON    PnRe    Case=Nom|Gender=Fem|Number=Plur|Person=3|PronType=Rel  15     advmod   _      _
14     το     εγώ    PRON    PnPp    Case=Acc|Gender=Neut|Number=Sing|Person=3|PersPronType=We|PronType=Prs  15     nsubj    _      _
15     ανέβασαν  ανεβάζω VERB    VbMn    Aspect=Perf|Mood=Ind|Number=Plur|Person=3|Tense=Past|Voice=Act  10     acl:relcl _      _
16     ψηλά   ψηλά   ADV     Ad      _      15     advmod   _      _
17     τα     ο      DET     AtDf    Case=Nom|Definite=Def|Gender=Neut|Number=Plur|PronType=Art  18     det      _      _
18     παιδιά  παιδί  NOUN    NoCm    Case=Nom|Gender=Neut|Number=Plur      15     nsubj    _      _
19     του    ο      DET     AtDf    Case=Gen|Definite=Def|Gender=Neut|Number=Sing|PronType=Art  20     det      _      _
20     Ρίο    Ρίο    PROPN   NoPr    Case=Gen|Gender=Neut|Number=Sing      18     nmod     _      SpaceAfter=No
21     !      !      PUNCT   PTERM_P PunctType=PTERM_P      2      punct    _      _

```

```

# sent_id = o_outputout2
# text = "/"
1      "      "      PUNCT  OPUNCT  PunctType=OPUNCT      _      _      _      SpaceAfter=No
2      /      /      PUNCT  PUNCT   _      1      det      _      _

```

# ΜΕΡΙΚΕΣ ΤΡΑΠΕΖΕΣ ΔΕΝΔΡΩΝ ΜΕ ΚΑΘΟΛΙΚΕΣ ΕΞΑΡΤΗΣΕΙΣ

## UNIVERSAL DEPENDENCY TREEBANKS

Michael Colburn-MT για Χριστιανικούς ψαλμούς

<https://www.dh.uni-leipzig.de/wo/projects/ancient-greek-and-latin-dependency-treebank-2-0/> για τον Αίσωπο

- για μετάφραση
- εξαγωγή λεξιλογίων
- συνδέσεις μεταξύ κειμένων
- κλπ

**Για το επόμενο μάθημα:**

**Basic creation of name and title  
authorities (module 1)**

[https://www.loc.gov/catworkshop/courses/name  
titleauth/pdf/Name-Title Instr Manual.pdf](https://www.loc.gov/catworkshop/courses/name<br/>titleauth/pdf/Name-Title Instr Manual.pdf)

**Για το επόμενο μάθημα:**

**About AAT (Information in the record)**

<http://www.getty.edu/research/tools/vocabularies/aat/about.html>

