Developing linguistic metrics and making linguistic decisions in annotated data

Sunoikisis Digital Classics,
10th Common Session
March 30th, 2017
Eleni Bozia UFL
Anise Ferreira UNESP

Discussing treebank annotations with Perseus Ancient Greek Dependency Treebank (AGDT) and Tündra

Anise Ferreira @ UNESP
March 30th, 2017
10th Common Session

What is Tündra?

- **T**übingen a**N**notated **D**ata **R**etrieval and **A**nalysis tool.

- Web application for searching in treebanks

- Part of CLARIN -D Project - infrastructure supported by the German Federal Ministry for Education and Research (BMBF), integrated with WebLicht language tool.

- AGDT - Ancient Greek Dependency Treebank is searchable on Tündra

- Similar application to Amir Zeldes' ANNIS

How to gain access

- Entry link: Weblicht website -
https://weblicht.sfs.uni-tuebingen.de/Tundra

- Create an account on CLARIN Service Provider Federation

- List of university servers - mine is at clarin.edu
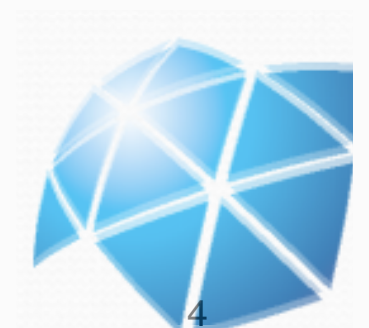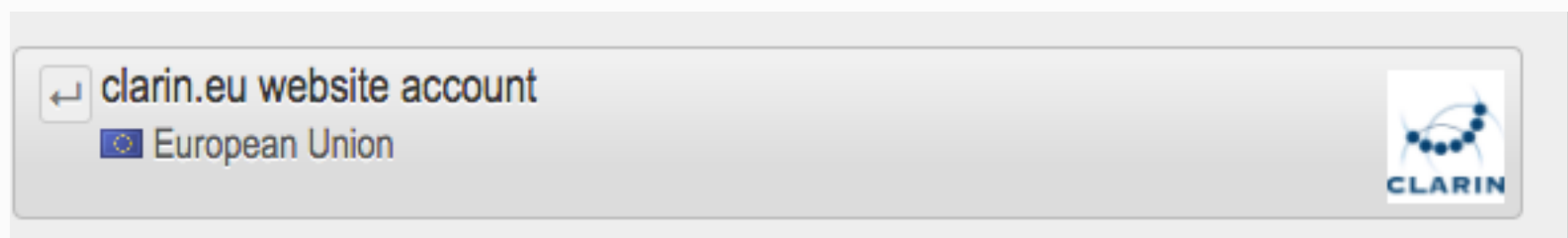
data/Gree… | Ancient Greek - I-p… | Library Genesis | _I-AGLDT - Inter-an… | Dictionaries and Hi… | Greek and Latin in… | Google Keep | » | Outros favoritos

LICHT

Search [ ] Go Search

History | View source | Discussion | Page

started
s and Examples
nual

r
r Manual
ed Applications

n Viewer

Tundra

TüNDRA - the *Tübingen aNnotated Data Retrieval Application* - is a treebank search application based in part on the popular but no longer supported *TIGERsearch* program.

TüNDRA is a web-based application that allows users to search, visualise and generate statistics from treebanks. A number of treebanks are provided, but users may also use the tool to explore resources developed via Weblicht. It uses a lightweight query language inspired by the widely used TIGERSearch application, offering. offering corpus linguists an interface for using corpora with complex annotation and syntactic links.

Using TüNDRA

TüNDRA is available via the WebLicht website at https://weblicht.sfs.uni-tuebingen.de/Tundra 🔒. Login is with any valid WebLicht access credentials.

Available data

The entry link

Sign in via the CLARIN Service Provider Federation

Select your identity provider below. This is usually the institution where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in.

If you cannot find your institution in the list below, please select the *clarin.eu website account* and use your CLARIN website credentials. If you don't have such credentials you can register an account here. For questions please contact spf@clarin.eu.

↵ clarin.eu website account
◻ European Union

CLARIN

CLARIN Identity Provider

You are now trying to access: WebLicht

Please login using your e-mail address by which you are subscribed to the CLARIN website and the corresponding password.

Email:

anise.a@gmail.com

Password:

•••••••••

☐ Don't Remember Login

Login

› Create new account

› Forgot your password?

› Need Help?

The Ancient Greek and Latin Dependency Treebank

Giuseppe G. A. Celano, Gregory Crane, Bridget Almas & al.

View the Project on GitHub
PerseusDL/treebank_data

Download ZIP File | Download TAR Ball | View On GitHub

https://perseusdl.github.io/treebank_data/

Text links point to treebank on Perseids platform

Author	Text	Loci
Aesop	Fables	1.1-1.50
Aeschylus	Agamemnon	
Eumenides		
Coephoroe		
Persae		
Prometheus		
Septem contra Thebas		
Supplices		
Athenaeus	The Deipnosophists	12, 13
Diodorus Siculus	Bibliotheca Historica	11
Herodotus	Historiae	1
Hesiod	Theogonia	
Operae et dies		
Scutum		
Homer	Iliad 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	
	Odyssey 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	
Lysias	De caede Eratosthenis	
In Alcibiadem 1		
In Alcibiadem 2		
In Pancleonem		
Plato	Euthyphro	
Plutarch	Alcibiades	
Lycurgus		
Polybius	Historiae	1
Pseudo Apollodorus	Library	1.1.1-1.4.1
Pseudo Homer	Hymn to Demeter	
Sophocles	Ajax	
Antigone		
Electra		
Oedipus Tyrannus		
Trachinae		
Thucydides	Histories	1.1-1.88, 1.89-1.117, 1.118-1.146

How to search? […]

- word form?　**[word = "καλά "]** or **[**token = …**]**

- lemma?　　　[**lemma**="σοφία"]

- part of speech?　=>　[**pos**="adjective"]

- pos with morphological features

- syntactic function/ tags ?　**[edge**="OBJ"**]**

- AND, OR, NOT ?　[…] **&** […] **|** […] **!** […]

- child　in the tree ?　[…] **>** […]

- distant descendent […]　**>*** […]

- sibling in the tree?　[…] **$** […]

- adjacent word ?　[…] **.** […]

- Distant word?　[…] **.*** […]

How to search?

- search for inflection categories: number, person, gender, tense, person, mood, voice, case and degree Ex. [case="dative"]

- [pos="…"] parts of speech: noun, verb, adjective, adverb, article, particle, conjunction, adposition, pronoun, interjection, numeral, punctuation.

- morphological features: first letter of each part of speech or category of inflection (s p d for numbers, etc).

Tagset descriptions or "edges" on Tündra

- *AGDT Guidelines:*

- Celano's AGDT 2.0 -https://github.com/PerseusDL/ treebank_data/blob/master/AGDT2/guidelines/ Greek_guidelines.md

- Bamman & Crane's AGDT 1.7 http:// nlp.perseus.tufts.edu/syntax/treebank/agdt/1.7/docs/ guidelines.pdf

Combination of particles

- ἀλλ᾽ ἤ … ; which one is the dependent in questions

- [token="ἀλλ᾽"]. [word="ἤ"] . * [lemma=";"]

- ἤ = AuxZ dependent on ἀλλά as AuxY

- ἀλλά = AuxY   and ἤ = AuxY   both dependent on the verb

- AuxY - AGDT : sentence adverbs

- AuxZ - "AuxZ label is reserved for logical operators that are adverbs, such as those meaning "not", "even", and "also". "is attached to the word which it modifies" (AGDT 2.0) "…emphasizes one specific word in the sentence"(AGDT 1.7)

```
[token="ἀλλ" ] . [word="ἦ"] . * [lemma=";"]
```
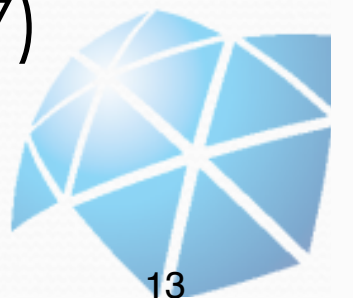
Tree

#31198: ἀλλ ἦ φρονεῖς εὖ τοῖσι νῦν ἠγγελμένοις ;
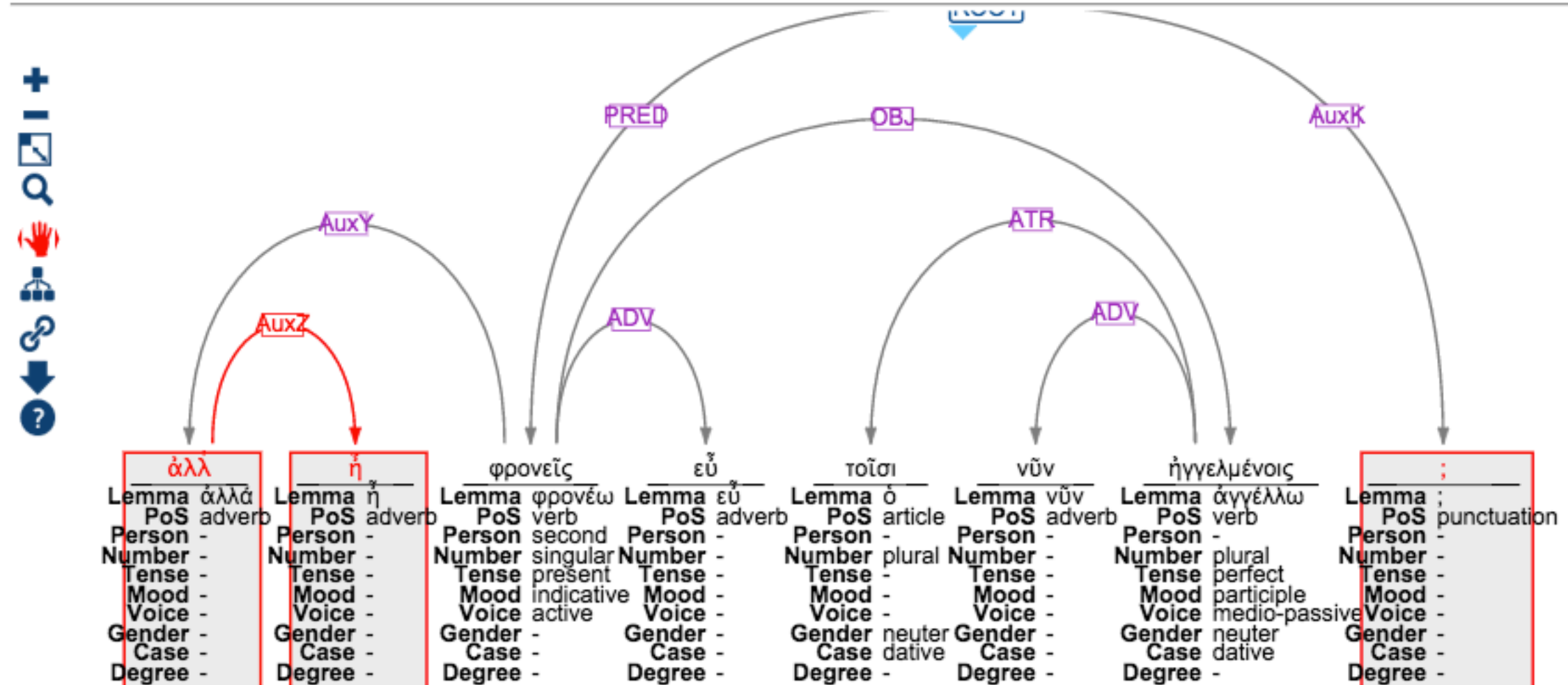
AESCHYLUS Trag. Choephoroe {0085.016} Line 774 - http://data.perseus.org/citations/urn:cts:greekLit:tlg0085.tlg006.perseus-grc1:766-782

en-What! Are you gladdened at heart by the present news? (Smyth, 1926)
por -Mas, estás satisfeita com as novidades? (MGK 2000)
fr - Mais peux-tu espérer après cette nouvelle? (Mazon, 1949)
Smyth Grammar: 2654: objection (with surprise or astonishment)

#7878: ἀλλ ἦ μέμηνας , ὦ τάλαινα , κἀπὶ τοῖς σαυτῆς κακοῖσι κἀπὶ τοῖς ἐμοῖς γελᾷς ; --
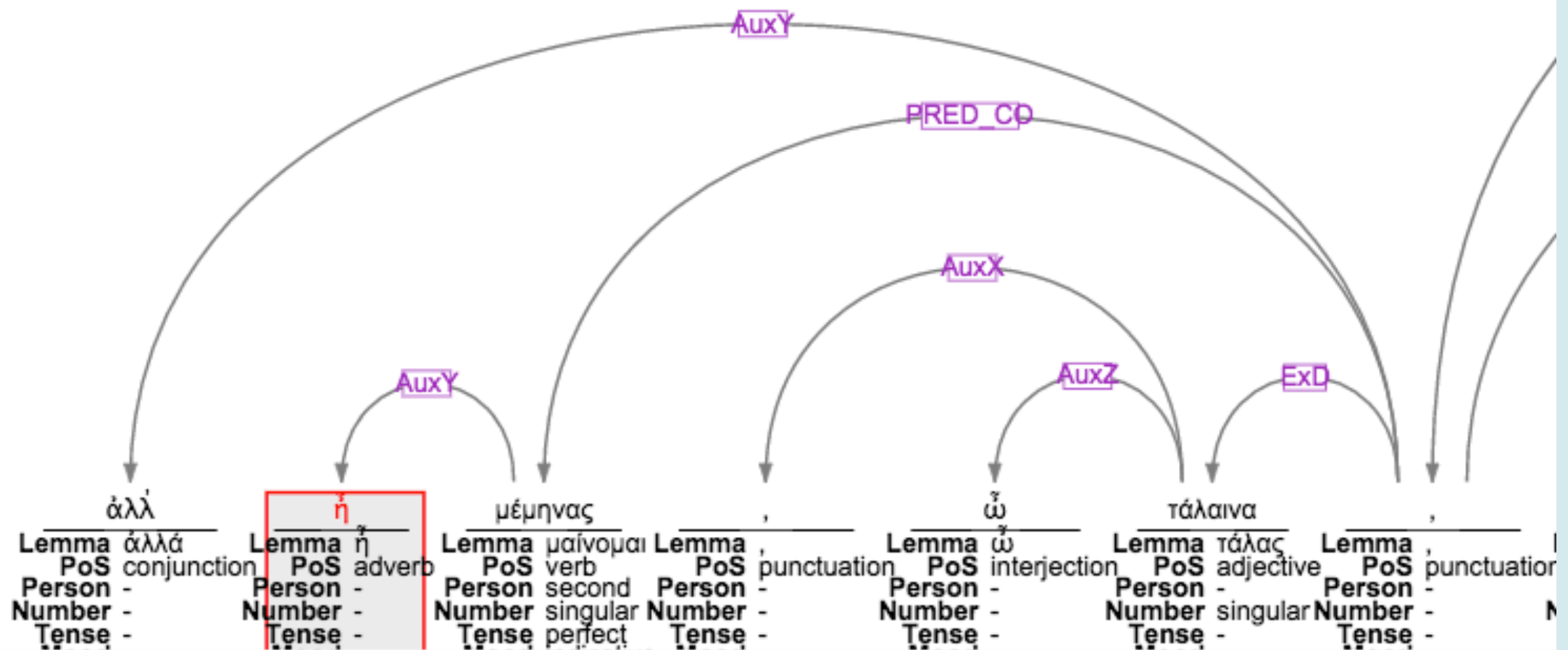
SOPHOCLES Trag. Electra {0011.012} Line 879 - http://data.perseus.org/citations/urn:cts:greekLit:tlg0011.tlg005.perseus-grc1:871-891

En: What, are you insane, poor girl? (…) (Jebb, 1894); What, are you mad, poor girl? (Kitto, 1994)
=> meaning different emphasis? "What? are you really mad?"
ἦ asseverative (Smyth 2864)
ἦ just a question indicator? (Smyth 2650)
surprise (ἀλλά) + question (ἦ) (Smyth 2786)

Combination of particles

- ἀλλὰ καὶ... [word="ἀλλὰ"] . [word="καὶ"]

  - with the expression οὐ μόνον... (not only) [token="οὐ"].[token="μόνον"] .* [word="ἀλλὰ"] . [word="καὶ"]

    - ἀλλά is a COORD (but...)

    - καί (also) is a AuxZ emphasizes the element being coordinated.

#1261: ἦν δὲ [οὐ μόνον καλὸν τοῦτο καὶ μεγαλόψυχον , ἀλλὰ καὶ χρήσιμον] .

PLUTARCHUS Biogr. et Phil. Lycurgus {0007.004} Chapter 22 section 5 line 7

And this was not only a noble and magnanimous policy, but it was also useful (Perrin, 1914)

ἄν modal particle

- AuxZ or AuxY?

  - [lemma="ἄν"] . * [mood="optative"]

  - [lemma="ἄν" & edge="AuxZ"] . * [mood="optative"]

    - (> 500 in the corpus)

  - [lemma="ἄν" & edge="AuxY"] . * [mood="optative"]

    - (>700 in the corpus) AuxZ - courteous formula: used to soften a statement, or request
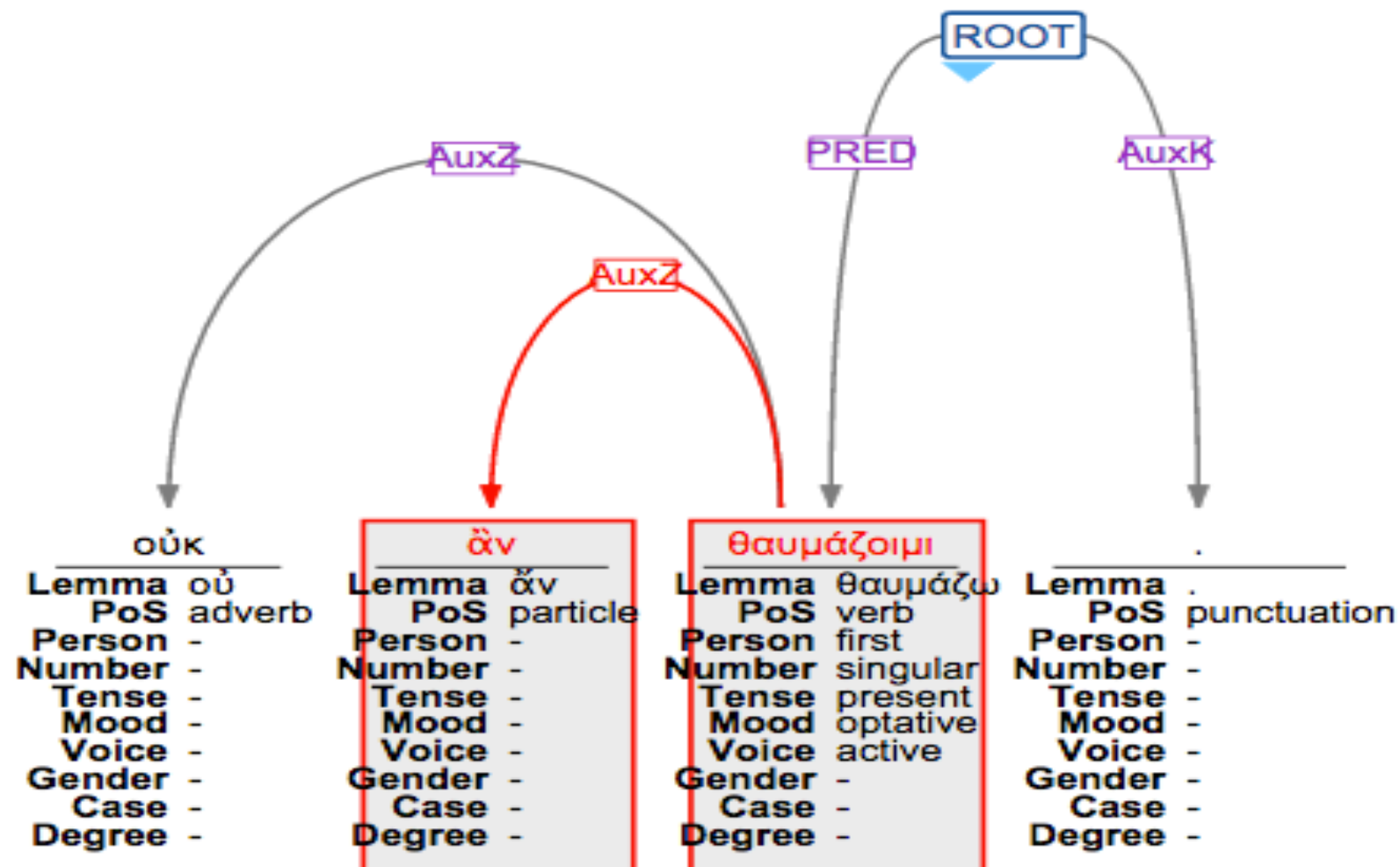
- AuxY- potencial - possibility/likelihood

#26561: [οὐκ ἂν] θαυμάζοιμι .

ROOT

AuxZ   PRED   AuxK

AuxZ

	οὐκ		ἂν		θαυμάζοιμι		.
Lemma	οὐ	**Lemma**	ἂν	**Lemma**	θαυμάζω	**Lemma**	.
PoS	adverb	**PoS**	particle	**PoS**	verb	**PoS**	punctuation
Person	-	**Person**	-	**Person**	first	**Person**	-
Number	-	**Number**	-	**Number**	singular	**Number**	-
Tense	-	**Tense**	-	**Tense**	present	**Tense**	-
Mood	-	**Mood**	-	**Mood**	optative	**Mood**	-
Voice	-	**Voice**	-	**Voice**	active	**Voice**	-
Gender	-	**Gender**	-	**Gender**	-	**Gender**	-
Case	-	**Case**	-	**Case**	-	**Case**	-
Degree	-	**Degree**	-	**Degree**	-	**Degree**	-

PLATO Phil. Euthyphro {0059.001} Stephanus page 6 section c line 8
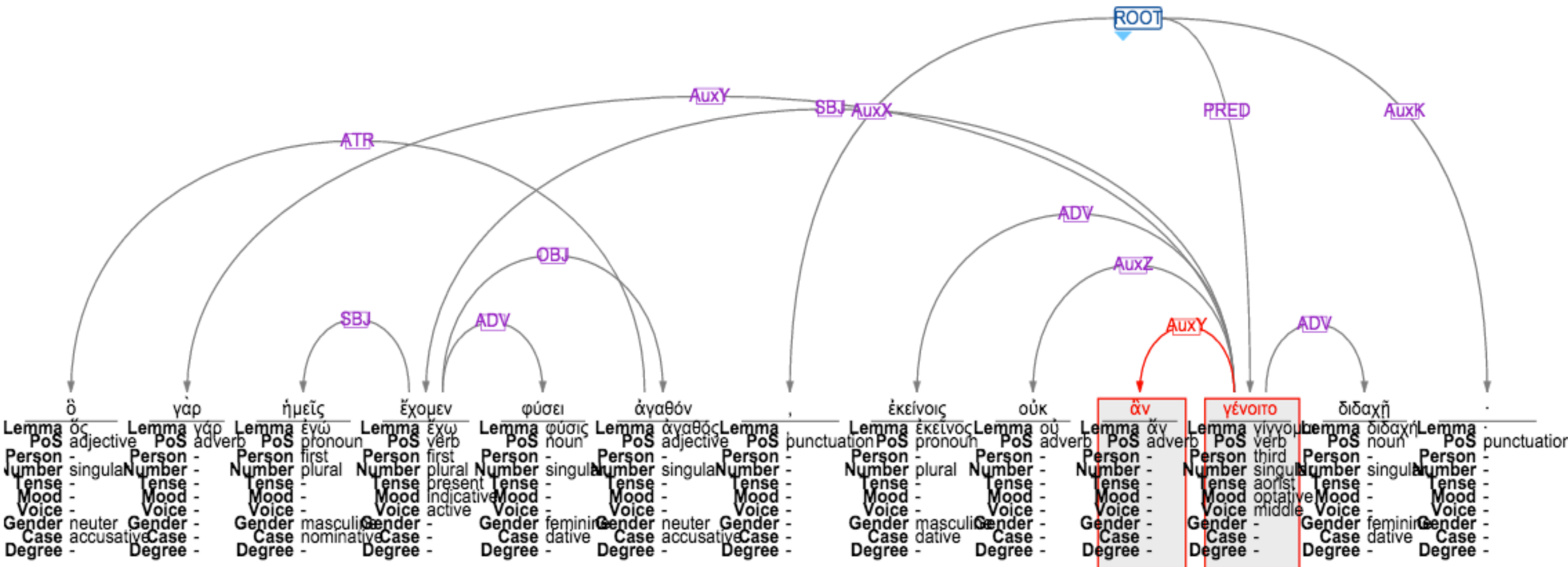**http://data.perseus.org/citations/urn:cts:greekLit:tlg0059.tlg001.perseus-grc1:6c**

En: S.: " I dare say". (Burnet, 1903; Fowler, 1914)
Fr: "Je le crois."
Por: Não me espantaria! (Malta, 2007)

#752: [ὃ γὰρ ἡμεῖς ἔχομεν φύσει ἀγαθόν , ἐκείνοις οὐκ ἂν γένοιτο διδαχῇ] ·

THUCYDIDES Hist. Historiae {0003.001} Book 1 chapter 121 section 4 line 6
http://data.perseus.org/citations/urn:cts:greekLit:tlg0003.tlg001.perseus-grc1:1.121.4

En:For the advantages that we have by nature they cannot acquire by education (Dent, 1910)
Por: A coragem que temos por natureza, eles não poderiam ter com a instrução (Prado, 1999) "The courage we have by nature they couldn't have with instruction.

Word repetitions

- Syntax - word or token = #number & other category

- [word = #1 & case="dative"]. [word=#1 & case="dative"]

- [word= #1 & pos="noun"]. [word=#1 & pos="noun"]

AESCHYLUS Trag. Persae {0085.012} Line 1010
http://data.perseus.org/citations/urn:cts:greekLit:tlg0085.tlg002.perseus-grc1:1008-1013

En: By strange woe, strange woe! (Smyth, 1926)
Por: Com novas novas dores dores (Torrano, 2009)

AESCHYLUS Trag. Agamemnon {0085.015} Line 121

En: Sing the song of woe, the song of woe… (Smyth, 1926)
Por:Tristezas, canta tristezas…

A common repetition: the article

- … τοῦ πολέμου τοῦ πρὸς τοὺς Ἀθηναίους…

; δὲ τῆς γερουσίας ἐβουλεύοντο περὶ [τοῦ πολέμου τοῦ πρὸς [τοὺς Ἀθηναίους]] ὑπὲρ τῆς κατὰ θά

artigos ATR ligados a πολέμου
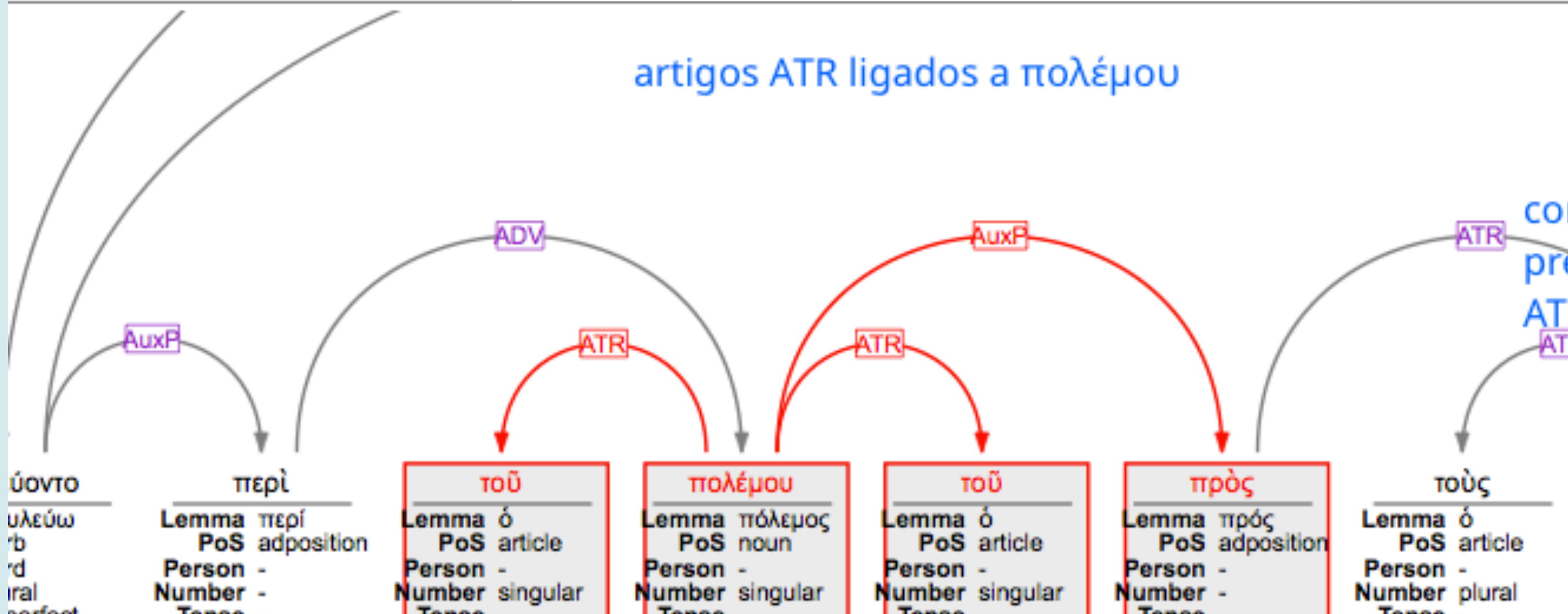
DIODORUS SICULUS Hist. Bibliotheca historica (lib. 1-20) {0060.001} 11. 50.2

En: "… war upon the Athenians…")
Fr: "… guerre aux Atheniens…"

MWE - Multiword expression

- *…κἂν τὰ μάλιστα χρηστότητα ὑποκρίνονται…* (. AESOPUS et AESOPICA Scr. Fab. Fabulae {0096.002} Fable 7 version 1 line Aesopica, 7.1. The cat doctor and the chicken)

- … although they perform honesty to the maximum

#31948: Οὕτως καὶ τῶν ἀνθρώπων οἱ πονηροὶ τοὺς φρονίμους οὐ λανθάνουσι , κἂν τὰ μάλιστα χρηστότητα ὑποκρίνωνται .
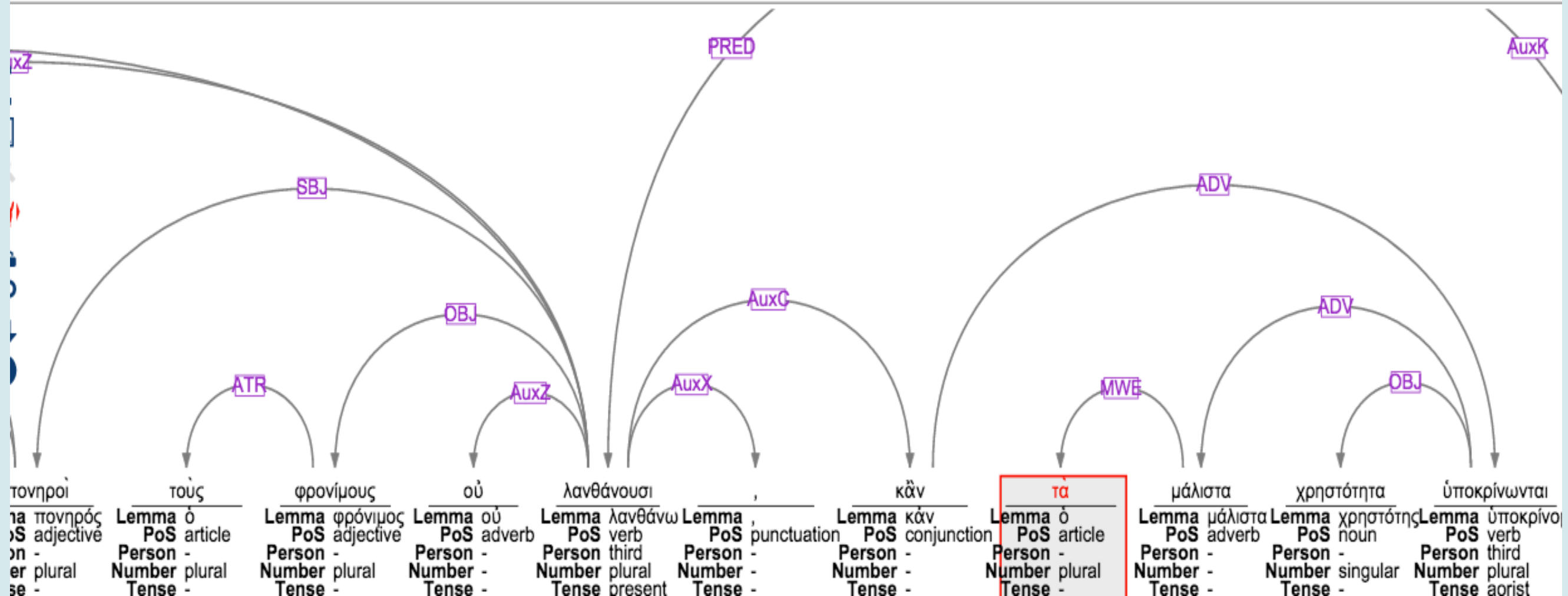
AESOPUS et AESOPICA Scr. Fab. Fabulae {0096.002} Fable 7 version 1 line Aesopica, 7.1.

En: "Although they play the role of very generous characters" (from Port. transl.)
Por: …ainda que representem o papel de personagens muito generosas (Dezotti, 2013).
Fr. … malgré toutes leurs affectations d'honnêteté

Article & participle

- [pos="article" & edge="SBJ"] > [mood="participle"]

- [mood="participle" & edge="SBJ"] > [pos="article" & edge="ATR"]

Tree

Query Match: **6** / 17
Sent match: **6** / 16
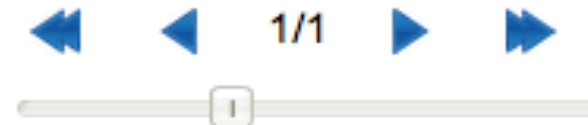
◄◄ ◄ 1/1 ► ►►

#1239: ὁ δὲ [τῶν ἀκμαζόντων] ἀμειβόμενος ἔλεγεν ·

	ὁ	δὲ	τῶν	ἀκμαζόντων	ἀμειβόμενος	ἔλεγεν	·
Lemma	ὁ	δέ	ὁ	ἀκμάζω	ἀμείβω	λέγω	·
PoS	article	adverb	article	verb	verb	verb	punctuation
Person	-	-	-	-	-	third	-
Number	singular	-	plural	plural	singular	singular	-
Tense	-	-	-	present	present	imperfect	-
Mood	-	-	-	participle	participle	indicative	-
Voice	-	-	-	active	medio-passive	active	-
Gender	masculine	-	masculine	masculine	masculine	-	-
Case	nominative	-	genitive	genitive	nominative	-	-
Degree	-	-	-	-	-	-	-
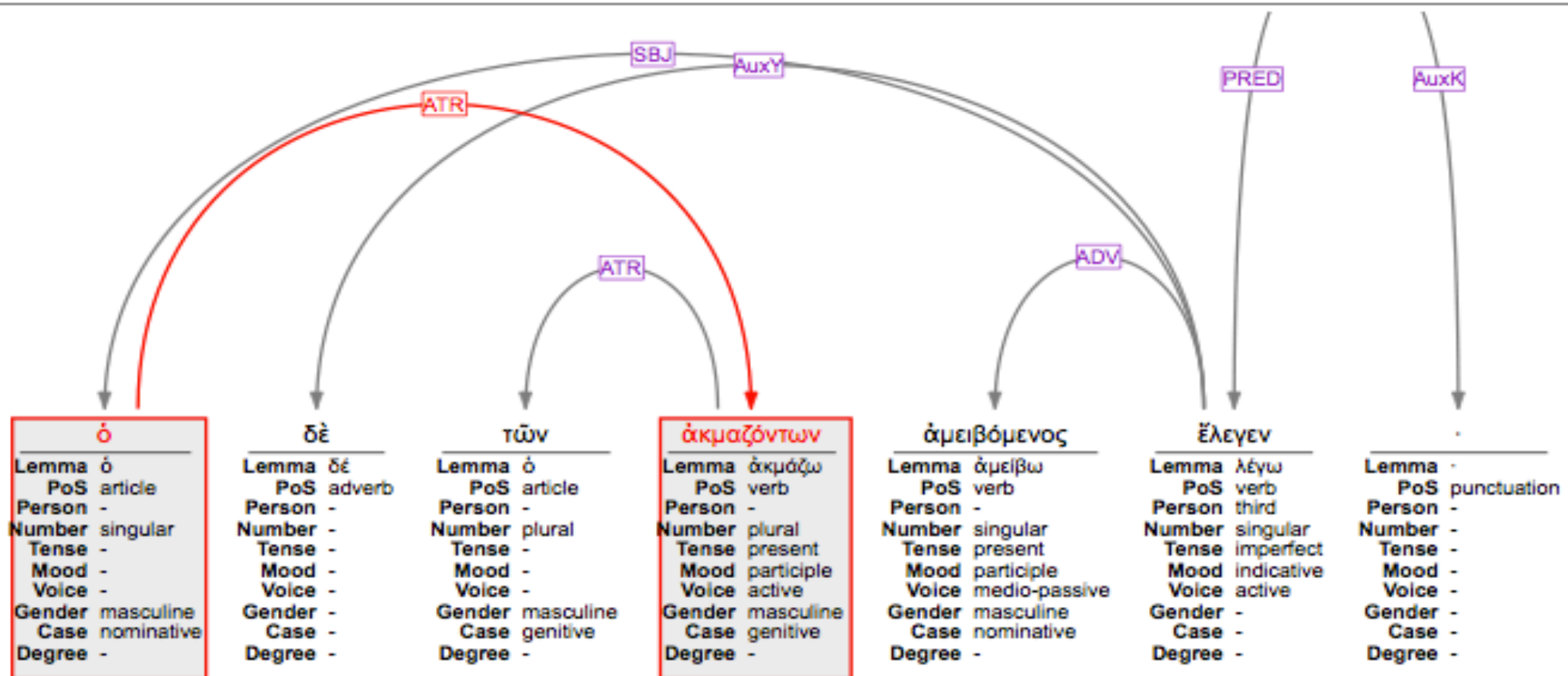
PLUTARCHUS Biogr. et Phil. Lycurgus {0007.004} Chapter 21 section 2 line 6

En. Then the choir of young men would respond:— (Perrin, 1914)
"the one of young men / (the young man) said responding"

#569: [οἱ δὲ ἀκούοντες τῷ μὲν Θεμιστοκλεῖ] ἐπείθοντο διὰ φιλίαν αὐτοῦ , τῶν δὲ ἄλλων ἀφικνουμένων καὶ σαφῶς κατηγορούντων ὅτι τειχίζεταί τε καὶ ἤδη ὕψος
ὅπως χρὴ ἀπιστῆσαι .

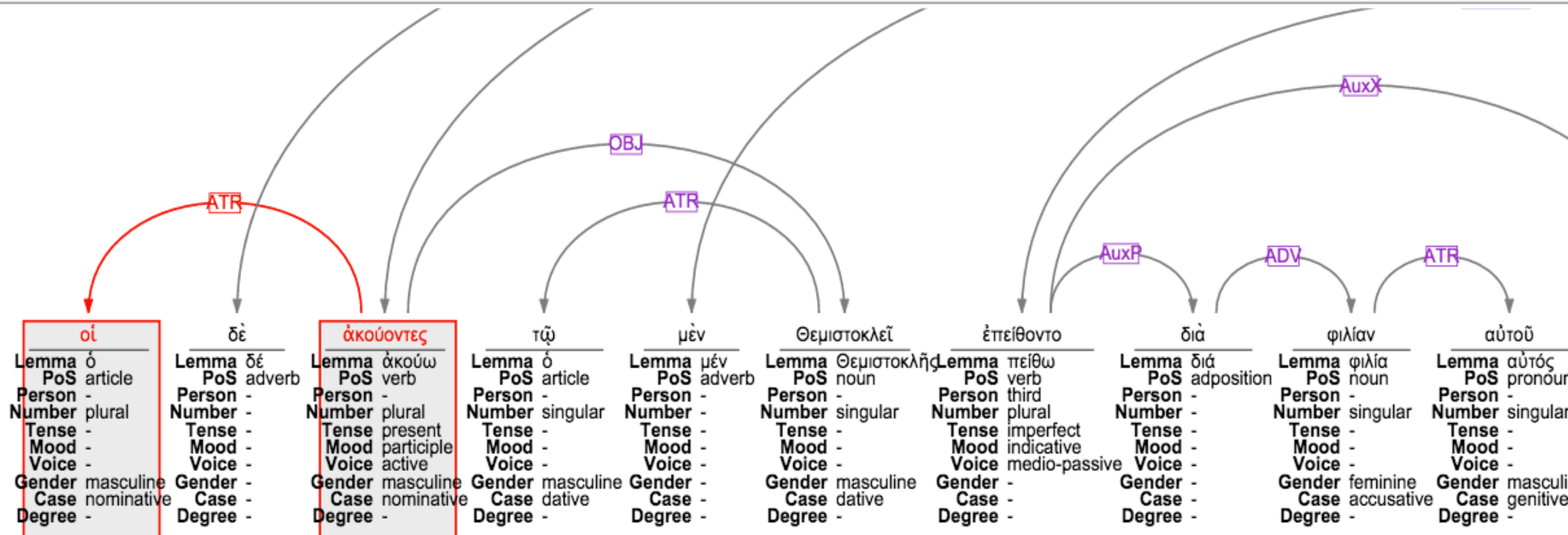|	οἱ	δὲ	ἀκούοντες	τῷ	μὲν	Θεμιστοκλεῖ	ἐπείθοντο	διὰ	φιλίαν	αὐτοῦ
Lemma	ὁ	δέ	ἀκούω	ὁ	μέν	Θεμιστοκλῆς	πείθω	διά	φιλία	αὐτός
PoS	article	adverb	verb	article	adverb	noun	verb	adposition	noun	pronoun
Person	-	-	-	-	-	-	third	-	-	-
Number	plural	-	plural	singular	-	singular	plural	-	singular	singular
Tense	-	-	present	-	-	-	imperfect	-	-	-
Mood	-	-	participle	-	-	-	indicative	-	-	-
Voice	-	-	active	-	-	-	medio-passive	-	-	-
Gender	masculine	-	masculine	masculine	-	masculine	-	-	feminine	masculine
Case	nominative	-	nominative	dative	-	dative	-	-	accusative	genitive
Degree	-	-	-	-	-	-	-	-	-	-

THUCYDIDES Hist. Historiae {0003.001} Book 1 chapter 91 section 1 line 1
http://data.perseus.org/citations/urn:cts:greekLit:tlg0003.tlg001.perseus-eng3:1.91.1

Por. E eles, ouvindo Temístocles, pela amizade que lhe tinham, deixavam-se persuadir (Prado, 1999). And they listening to Themistocles, through their friendship for him, let being persuaded.
En: At first the Lacedaemonians trusted the words of Themistocles, through their friendship for him; (Dent, 1910)

The end