Treebanking Lucian in Arethusa: experiences, problems and considerations

Petar Soldo, Petra Šoštarić

University of Zagreb

Abstract: The article discusses some problems that arose while implementing two sets of annotation rules for treebanking Ancient Greek (the first set by Bamman and Crane, the second by Celano). Educational uses of treebanking are discussed, together with the most common problems that occurred while treebanking the Iudicium vocalium by 2nd-century writer Lucian of Samosata.

Keywords: treebanking, ancient Greek, Lucian

1. Introduction

The aim of this paper is to describe the use of treebanking in Arethusa over a one semester BA level course on Greek syntax and to discuss several problems that occurred frequently during the course. The aim of this paper is not to identify every possible problem that might occur, but to show how applying the sets of rules put forward by Bamman and Crane in 2008 and Celano in 2018 worked in practice in a specific case.

2. Treebanking classical languages

Treebanking is a form of syntactical analysis producing tree-like structures showing the dependency of parts of speech in a sentence on one another, starting from the predicate of the main clause. Apart from being used in linguistic research, treebanking has a notable pedagogical application. In a rapidly increasing world of learning tools, aids and resources, treebanking is a unique way of acquiring and exercising

language skills, and it can be especially useful when it comes to historical languages. A set of rules for treebanking Ancient Greek was put forth in 2008 by David Bamman and Gregory Crane, the *Guidelines for the Syntactic Annotation of the Ancient Greek Dependency Treebank (1.1)* (Bamman, Crane). Giuseppe A. Celano continued their work and his expanded and more detailed set of rules including a semantic layer is available online, with the latest update on March 19th 2018 (Celano). The digital tool used for treebanking is called Arethusa. It has been developed for treebanking classical languages and is freely available at the Perseids platform (Perseids).

A collection of treebanked texts, *The Ancient Greek and Latin Dependency Treebank*, is also available online (Celano, Crane, Almas & al.). It first appeared in 2006 as a Perseus/Tufts University project; later it became a joint effort of Tufts University and Leipzig University, and is currently supervised by Giuseppe G. A. Celano at the Institute of Computer Science, Natural Language Processing, Leipzig University (Celano, Crane, Almas & al., Abteilung Automatische Sprachverarbeitung). The concept of treebanking belongs to the field of Dependency Grammar. The work of Lucien Tesnière marks the beginning of Dependency Grammar in its modern form, although some concepts can be recognized in the work of pre-modern grammarians (Nivre). Several dependency treebanks have been established over the years, with the Prague Dependency Treebank especially influencing the treebanking of Latin (Bamman, Passarotti, Busa & al.) and Greek (Bamman, Crane).

3. Work process

This paper is a direct result of work done during the course titled "Greek Syntactic Period in Classical and Post-Classical Times" in the winter semester of the academic year 2018/2019 at the Department of Classical Philology at the University of Zagreb, Croatia. There were three fourth-year BA students enrolled in the course, one of whom co-authored this paper. The chosen reading for the course was Lucian's *Iudicium vocalium* (*Lis consonantium*) comprising 54 sentences (more in section 4). The four teaching hours per week were divided into two hours of translation seminar and two hours of treebanking, i.e. analysing the treebanked sentences submitted by the students via the Perseids platform to the Perseids group "Zagreb Lucian 2018". The group was established specifically for the course in question and has two board members supervising the students' work, Neven Jovanović and Petra Šoštarić. In the beginning of the course the annotation scheme by Bamman and Crane was used, but as the course progressed, the more detailed scheme by Celano was adopted. The process

of treebanking a sentence starts with identifying the predicate and then determining the dependency of other words in the sentence to the predicate and, if they are not directly dependent on the predicate, identifying the dependency on another word. The examples 1-5 show that the article is usually dependent on the noun it refers to; in example 2 we see a case of coordination (two parts of speech at the same level of dependency); examples 2-4 illustrate the use of prepositions as bridges between the verb and the adverb. Every word in a sentence has to be appropriately labelled (OBJ for object etc.).

All 54 sentences form Lucian's *Iudicium vocalium* were submitted and discussed, either in the course of the lectures and seminars, or via the Perseids comments section. During these discussions certain problems in applying the annotation scheme arose more frequently than others. The most common problems and suggestions on how to deal with them are presented in section 5.

4. The text

The 2nd century author Lucian is somewhat of an enigma. Information on his life is scarce: born around AD 120 in Samosata, a prosperous city on the river Euphrates, most likely he was a native speaker of Aramaic; he travelled extensively throughout the Roman Empire and earned his living by teaching rhetoric and perhaps practicing law (Bricko's introduction to Lukijan, 14-18).[1] Information on his life is sometimes culled from his work, but Richter warns against taking Lucian's "speaking personae" as a realistic representation of the author himself (Richter 327-328). Lucian left behind a substantial body of work comprising around eighty titles, but his contemporaries seem to have been reluctant in discussing his *opus* (Bricko's introduction to Lukijan, 14-16, Richter 327). Bricko cites Strohmaier's suggestion that Lucian's tendency to mock his contemporaries made them unwilling to write about him (Bricko's introduction to Lukijan, 16). Nevertheless, he later influenced the likes of Alberti, Erasmus, Thomas More, Jonathan Swift and Voltaire (Costa's introduction to Lucian, *Selected Dialogues*, xiii-xiv).

The exact number of Lucian's works is disputed, and so are their dates (Bricko's introduction to Lukijan, 19). Bricko deems the *True Histories* "his best known and most read text" (Bricko's introduction to Lukijan, 29); together with his *Icaromenippus*, they

---

[1] For basic information in English see Costa's introduction to Lucian (*Selected Dialogues*) or J. Roberts s. v. "Lucian"; a more detailed discussion can be found in Richter.

are considered a piece of science-fiction *avant la lettre* (A. Roberts 31). Another popular work is the *Dream*, usually considered Lucian's autobiography, although Richter suggests the autobiographical presentation should not be taken at face value (Richter 334-336). The text relevant for this paper, *Δίκη φωνηέντων* or *Iudicium vocalium*, also known as *Lis consonantium*, is a judicial oration by the consonant Sigma which is accusing his fellow consonant Tau of usurping many words originally belonging to Sigma (i.e. pronounced with an /s/, not a /t/). The indignant Sigma also lists the instances of Tau's violent behaviour against other consonants, and even his malicious intrusions into the human world. The oration is addressed to a jury of vowels, hence the title. It remains to be seen to what extent this text might represent Lucian's reaction to hyper-Atticist language practice and a comment on the phonological changes that have been taking place since Hellenistic times.

The version of the text used for the purposes of the "Greek Syntactic Period in Classical and Post-Classical Times" course and consequently this paper is the one available at the Perseus Digital Library (Crane). An English translation is to be found in the relevant section of Lucian (*Works*).

5. Problems identified

It takes some time to adjust to the change from traditional grammar to the treebanking annotation schemes by Bamman and Crane and Celano (more on this in section 6), but even after a certain period of adjustment a few problems still persist. Typically, they belong to those areas of grammar already known not to be completely straightforward.

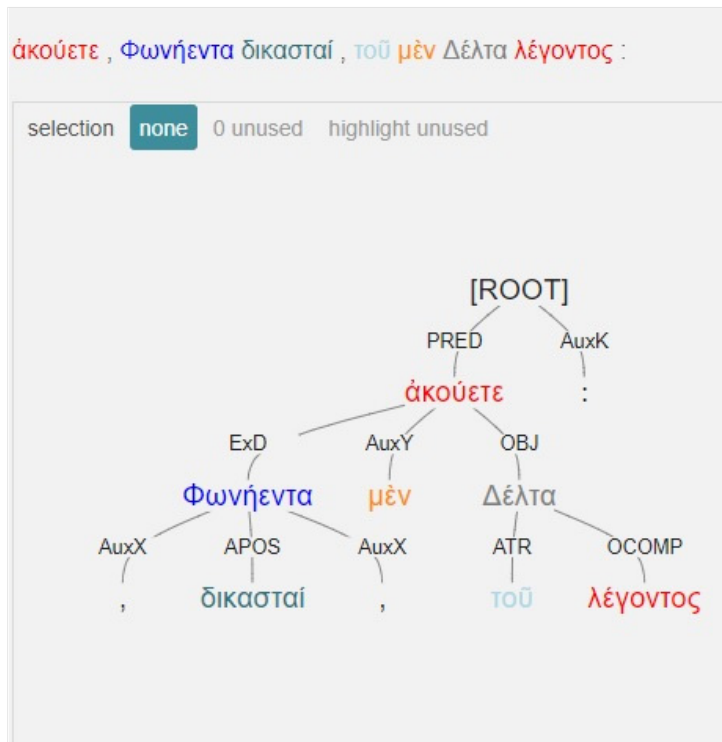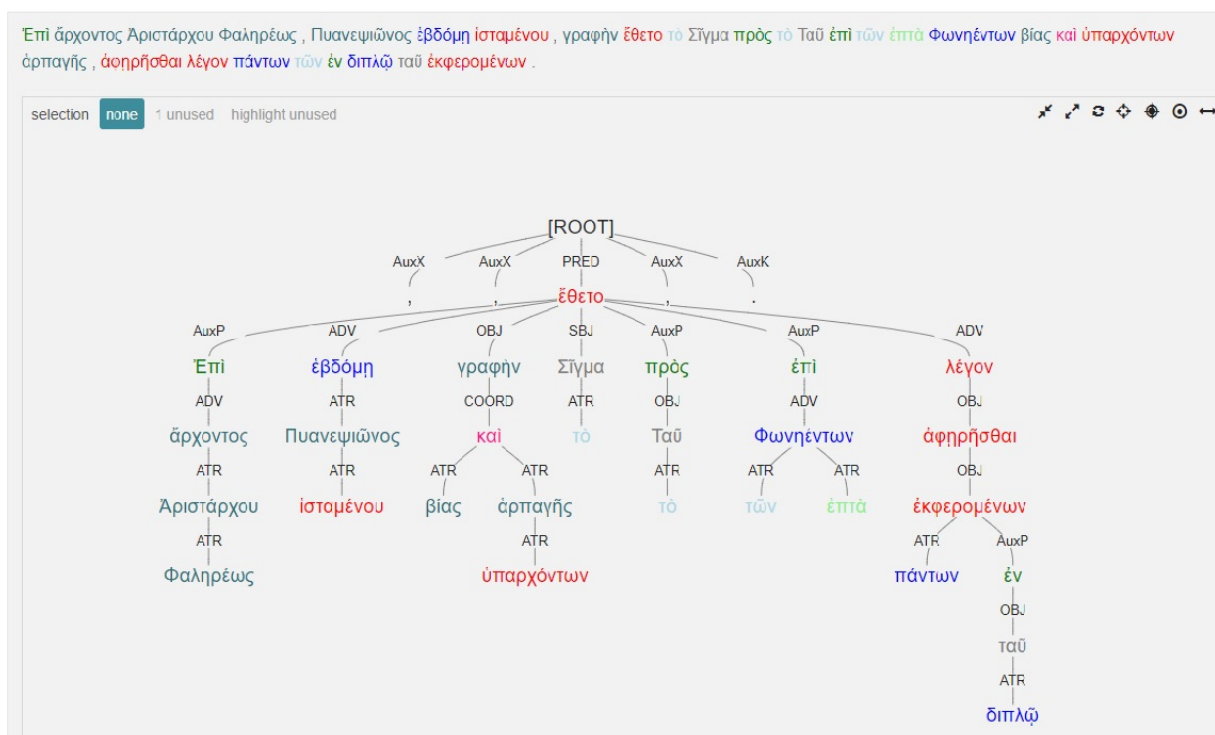### 5.1. Labelling verbal complements: OBJ or ADV?

In Celano's scheme, "The OBJ ('object') label is attached to any dependent which is taken to be an argument of the verb, adjective, or adverb (excluding, of course, arguments which are captured by other labels, such as SBJ)" (Celano). Probably the most frequent discussion held during the course was when to assign the OBJ label, and when the ADV label to a verb complement. In Celano's own words, "The concept of argument is notoriously difficult to define" (Celano). Celano states that "verbs of motion take an argument expressing direction," and suggests consulting Verbnet, a lexicon of English verbs, in case of other classes of verbs (Celano). We would hereby

like to suggest additional classes of verbs requiring OBJ complements. These classes of verbs can be found in Smyth's *Greek Grammar* (Smyth).[2] Traditional case syntax is helpful in this regard since it identifies classes of verbs that typically have an argument in a certain case.

In paragraphs 1343-1411 Smyth lists the following classes of verbs taking a genitive complement: "verbs of sharing" (Smyth 320; par. 1343), "verbs signifying *to touch*, *take hold of*, *make trial of*" (Smyth 321; par. 1345), "verbs of beginning" (Smyth 321; par. 1348), "verbs signifying *to aim at*, *strive after*, *desire*" (Smyth 321; par. 1349), "verbs signifying *to reach*, *obtain*" (Smyth 321; par. 1350), "verbs of approaching and meeting" (Smyth 322; par. 1353), "verbs of smelling" (Smyth 322; par. 1354), "verbs signifying *to remember*, *remind*, *forget*, *care for* and *neglect*" (Smyth 322; par. 1356), "verbs signifying *to hear* and *perceive*" (Smyth 323; par. 1361) (example 1), "verbs signifying *to fill*, *to be full of*" (Smyth 324; par. 1369), "verbs signifying *to rule*, *command*, *lead*" (Smyth 324; par. 1370). Further types of genitive we consider requiring an OBJ labels are the "genitive of price and value" (Smyth 325; par. 1372-4), "genitive of crime and accountability" (Smyth 325; par. 1375-9), "genitive of connection… with verbs of saying and thinking" (Smyth 326; par. 1380) and the types of ablatival genitive: "genitive of separation with verbs signifying *to cease*, *release*, *remove*, *restrain*, *give up*, *fail*, *be distant from*" (Smyth 328; par. 1392) (example 2), "*to want*, *lack*, *empty*" (Smyth 329; par. 1396). Finally, there are the "genitive of distinction and of comparison" (Smyth 330; par. 1401-4), "genitive of cause" (Smyth 330; par. 1405) and "genitive of source" (Smyth 331; par. 1410-1).

Things are much clearer with the dative and accusative since they are traditionally interpreted as indirect and direct object respectively. Still, classes of verbs requiring an OBJ are those complemented by the "dative of association" (Smyth 349; par. 1523) and "dative of accompaniment" (Smyth 349; par. 1524).

---

[2] Grammatical concepts and nomenclature are understood as in Smyth's grammar which is the standard English-language Greek grammar and also used by Celano.

Example 1: Genitive as OBJ with verbs of hearing

ἀκούετε , Φωνήεντα δικασταί , τοῦ μὲν Δέλτα λέγοντος :

selection none 0 unused highlight unused

[ROOT]
PRED   AuxK
ἀκούετε   :
ExD   AuxY   OBJ
Φωνήεντα   μὲν   Δέλτα
AuxX   APOS   AuxX   ATR   OCOMP
,   δικασταί   ,   τοῦ   λέγοντος

Example 2: Genitive of separation as OBJ

Ἐπὶ ἄρχοντος Ἀριστάρχου Φαληρέως , Πυανεψιῶνος ἑβδόμη ἱσταμένου , γραφὴν ἔθετο τὸ Σῖγμα πρὸς τὸ Ταῦ ἐπὶ τῶν ἑπτὰ Φωνηέντων βίας καὶ ὑπαρχόντων ἁρπαγῆς , ἀφῃρῆσθαι λέγον πάντων τῶν ἐν διπλῷ ταῦ ἐκφερομένων .

selection none 1 unused highlight unused

[ROOT]
AuxX   AuxX   PRED   AuxX   AuxK
,   ,   ἔθετο   ,   .
AuxP   ADV   OBJ   SBJ   AuxP   AuxP   ADV
Ἐπὶ   ἑβδόμη γραφὴν Σῖγμα πρὸς   ἐπὶ   λέγον
ADV   ATR   COORD   ATR   OBJ   ADV   OBJ
ἄρχοντος Πυανεψιῶνος καὶ   τὸ   Ταῦ   Φωνηέντων   ἀφῃρῆσθαι
ATR   ATR   ATR   ATR   ATR   ATR   ATR   OBJ
Ἀριστάρχου ἱσταμένου βίας ἁρπαγῆς τὸ   τῶν   ἑπτὰ   ἐκφερομένων
ATR   ATR   ATR   AuxP
Φαληρέως   ὑπαρχόντων   πάντων   ἐν
                                                         OBJ
                                                         ταῦ
                                                         ATR
                                                         διπλῷ

When in doubt, it is useful to consult the dictionary entry and consider the following: is there a typical complement of the verb in question? Are there different meanings listed, and do they require different complements? If the answer to one or both of these questions is yes, it is worth considering the OBJ label. Additionally, a search of the *corpus* can be performed.

In case of adjectives, it is once again helpful to follow Smyth. In the first paragraph on genitive with adjectives he states that the adjectives requiring a genitive complement are "corresponding in derivation or meaning to verbs taking the genitive" (Smyth 332; par. 1414). The types of adjectives with genitive complements requiring an OBJ label are, in Smyth's nomenclature, the following: those expressing "possession and belonging" (Smyth 332; par. 1297), "sharing" (Smyth 332; par. 1415), "touching, desiring, attaining, tasting" (Smyth 332; par. 1416), "connection" (Smyth 332; par. 1417), "capacity and fitness" (Smyth 332; par. 1418), "experience" (Smyth 332; par. 1419), "remembering, caring for" (Smyth 333; par. 1420), "perception" (Smyth 333; par. 1421), "fullness" (Smyth 333; par. 1422), "ruling" (Smyth 333; par. 1423), "value" (Smyth 333; par. 1424) (example 3), "accountability" (Smyth 333; par. 1425), "place" (Smyth 333; par. 1426), "separation" (Smyth 333; par. 1427), "want" (Smyth 334; par. 1429), "distinction" (Smyth 334; par. 1430), "comparison" (Smyth 334; par. 1431) and "cause" (Smyth 334; par. 1405).

Example 3: adjective as OBJ
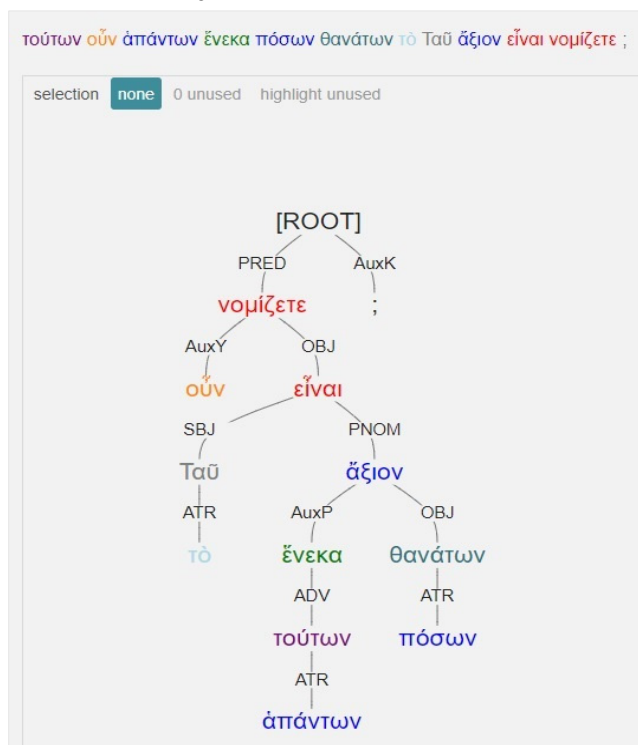
Verbal adjectives ending in -τος/τεος were another problematic point. We concluded they should have an OBJ complement in the dative case attached to them, not to the verb *to be* to which these adjectives are connected, because the adjectives bear the passive meaning which requires the *dativus auctoris*. We came to this conclusion following Bamman and Crane who state that "actors in passive constructions should be annotated as objects" and provide a genitive example (Bamman and Crane 13-14).

5.2. Particles

In the annotation schemes by Bamman and Crane and Celano particles can be labelled as COORD, AuxY and AuxZ. In Celano's annotation scheme, "The function COORD ('coordination') is reserved for coordinate conjunctions (the list can be found in SG 2163, excluding the inferential, causal, and some of the adversative conjunctions (such as *μέντοι*); these are all treated here as (sentence) adverbs)…" (Celano). The adversative conjunctions taking the COORD function are *ἀλλά, ἀτάρ, δέ, καίτοι* (Celano). The particle *δέ* was designated as a COORD already in Bamman and Crane (25, 38). Celano discusses particles in more detail: "As is known, the meaning and function of AG particles is very complex to capture. In the following guidelines there is no attempt to provide detailed criteria to annotate particles. The annotator is only required to be able to identify those particles (in the traditional large definition of terms) which are conjunctions, such as: *ἀλλά, ἀτάρ, δέ, εἴτε, ἤ, ἠδέ, καί, καίτοι, οὐδέ, οὔτε, μηδέ, μήτε, τε*. … Any other particle which is not one of the above conjunctions is attached to the verb of the main clause and receives the AuxY label, if the particle is a sentence adverb, or is attached to the word which it modifies and receives the label AuxZ. In the current version of the treebank it is left to the annotator to decide which of these two options is the right one for each of the non-conjunction particles" (Celano).

The particles listed by Smyth (Smyth 631-671; par. 2769-3003) that are not allowed the COORD function can therefore take both the AuxY and AuxZ label, although, in our experience, usually one more often than the other: *ἄρα* and *μέν* tend to be sentence adverbs (AuxY) more often than AuxZ, but Smyth's grammar allows the latter, albeit in more restricted circumstances (Smyth 635, 654; par. 2787, 2895). Perhaps it is possible that *ἄρα* receives the AuxZ label when next to a pronoun, and AuxY in other cases, but a rule of this kind could be formulated only after more research is done. Of all the particles listed in Smyth (Smyth 631-671; par. 2769-3003) we would suggest that some necessarily take the label AuxZ: *καίπερ* with participles, *μά, ναι* and *νή* with accusative, and *πέρ*.

5.3. ExD

In Celano's annotation scheme, "The label ExD is used for those constituents that do not syntactically belong to a sentence, such as vocatives and the head of a parenthetical clause. They are appended to the PRED node" (Celano). In the case of example 4, we find the particle *δέ* within the parenthetical clause. If it is to be labelled as COORD, the head of the parenthetical clause (the verb *προσάπτουσι*) cannot be appended directly to the PRED, therefore violating the relevant rule. If the predicate of the parenthetical clause (*προσάπτουσι*) is appended to the PRED, the particle must then be labelled AuxY, which is inappropriate since it is considered a conjunction (see 5.2). The former option was chosen.

Example 4: particle *δέ*
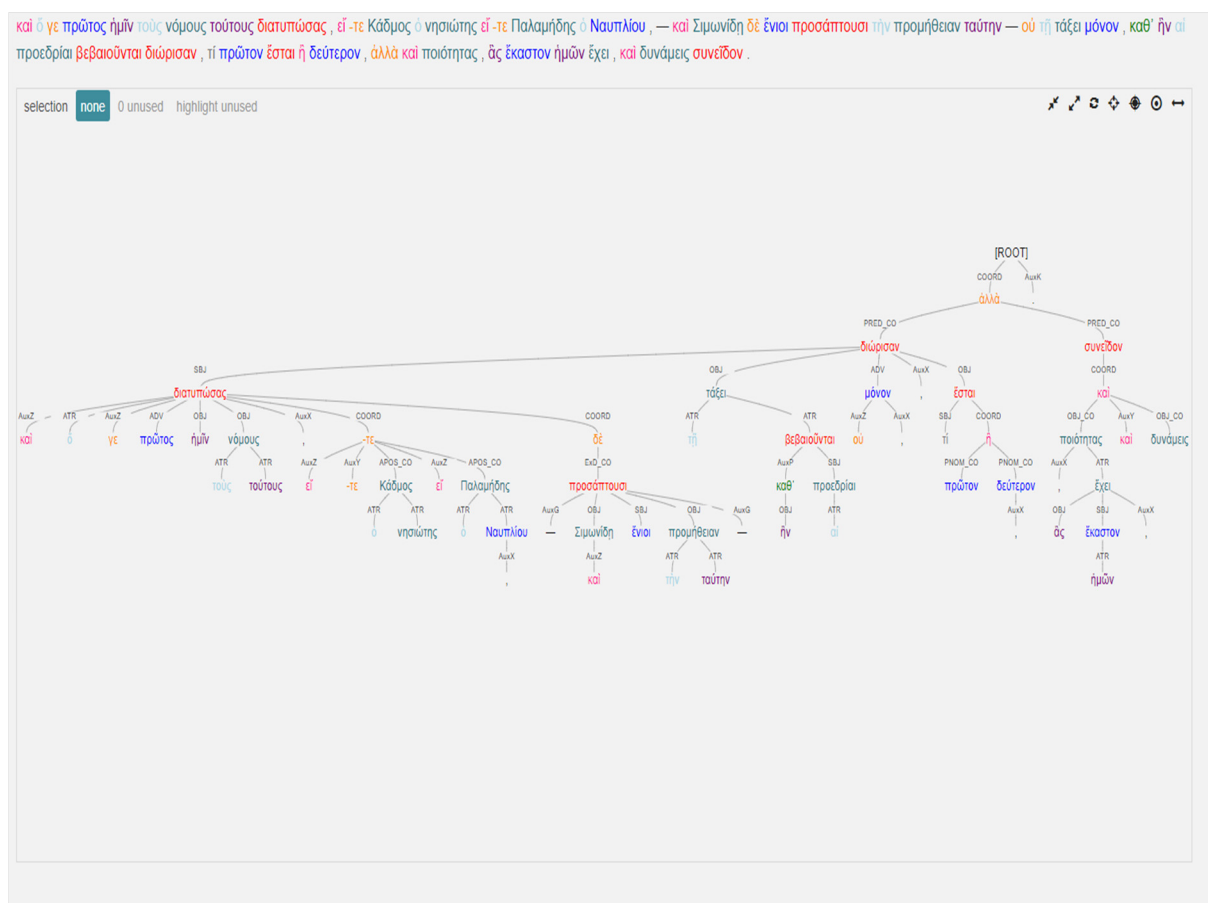
6. Treebanking in teaching Ancient Greek:
advantages and problems

The use of treebanking in teaching Ancient Greek leads to a greater understanding of Greek syntax and relationships between parts of speech in a sentence, but it's not all smooth sailing since there are some gray areas and things that can be interpreted in more than one way. To start with, the student is required to think and to deeply understand the sentence. Instead of just using morphological and syntactic analysis to break down a period, treebanking introduces the semantic layer. Having to connect each word to exactly one other word and having to determine their relation makes the student think about the usage and precise meaning of a word or an expression. Sometimes it is easier for the students to explain and discuss something in metalanguage than translating into their mother tongue. Treebanking also helps notice the valence of verbs, certain sentence structures, parallelism etc. It should also be mentioned that the straightforward and intuitive tree structure visualization and morphological colours (at least when using Perseids) help a great deal to make the sentence clearer, especially when analysing highly complex periods. Some students also experience treebanking as more fun than a purely translational approach and enjoy the opportunity to have a linguistic discussion in class.

On the other hand, dependency treebanking is based on dependency syntax theory, which differs from what most teachers and professors use in everyday classes. Students are traditionally taught that the participle is necessarily an attribute, objects are either in accusative (direct objects) or dative (indirect objects) without prepositions, prepositional phrases cannot be attributes, and not much attention is paid to particles, which are a problematic part of speech in themselves. The adverb-or-object debate can be especially frustrating, and the analysis of relative clauses can seem counter-intuitive.

Conclusion

The differences between the traditional approach and the guidelines proposed by Bamman and Crane and Celano can be confusing and counter-intuitive and it takes some time for the students to adjust. Therefore, this approach is appropriate for higher education levels (the final year of a BA program, MA program), while at a lower level (early BA, even high school) simplified sentences should be analysed. Although there are some downsides to using treebanking as an educational tool, especially in the areas

where there is some ambiguity or lack of clarity (one could argue, though, that this is true for any kind of linguistic description), we are certain that treebanking has not only a great research potential, but also a bright didactic future.

Works cited:

Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig. DFG-Projekt: Überarbeitung, Standardisierung und Erweiterung der "Ancient Greek and Latin Dependency Treebank" http://asv.informatik.uni-leipzig.de/projects/41. 2019. Accessed on 15 January 2019.

Bamman, David, Gregory Crane. *Guidelines for the Syntactic Annotation of the Ancient Greek Dependency Treebank (1.1)*. The Perseus Project, Tufts University. http://nlp.perseus.tufts.edu/syntax/treebank/greek.html. September 1, 2008. Accessed on 15 January 2019.

Bamman, David, Marco Passarotti, Roberto Busa, Gregory Crane*. The Annotation Guidelines of the Latin Dependency Treebank and Index Thomisticus Treebank: The Treatment of Some Specific Syntactic Constructions in Latin*. 2008. www.lrec-conf.org/proceedings/ lrec2008/summaries/25.html. Accessed on 15 February 2019.

Lukijan. *Djela*. Translation and introduction by Marina Bricko. Matica hrvatska 2002.

Celano, Giuseppe G. A., Gregory Crane, Bridget Almas & al. *The Ancient Greek and Latin Dependency Treebank v.2.1.*, http://perseusdl.github.io/treebank_data/. 2018. Accessed on 15 January 2019.

Celano, Giuseppe G. A. *Guidelines for the Ancient Greek Dependency Treebank 2.0*. https://github.com/PerseusDL/treebank_data/blob/master/AGDT2/guidelines/Greek_ guidelines.md. March 19, 2018. Accessed on 15 January 2019.

Crane, Gregory (editor in chief). *Perseus Digital Library*. http://www.perseus.tufts.edu/hopper/ 2007. Accessed on January 15th 2019.

Lucian. *Works*. Translated by A. M. Harmon, Harvard University Press and William Heinemann Ltd. Vol. 1. 1st ed. 1913. Reprinted 1961.

Lucian. *Selected Dialogues*. Translated with an Introduction and Notes by C. D. Costa, Oxford University Press, 2005.

Nivre, Joakim. *Dependency Grammar and Dependency Parsing*. 2005. https://cl.lingfil.uu.se Accessed on 15 February 2019.

Perseids. http://sites.tufts.edu/perseids/. 2017. Accessed on 15 January 2019.

Richter, Daniel S. "Lucian of Samosata." In: *The Oxford Handbook of Second Sophistic*. Edited by Daniel S. Richter and William A. Johnson. Oxford University Press, 2017, pp. 327-344.

Roberts, Adam. *The History of Science Fiction*, 2nd ed. Palgrave Macmillan, 2016.

Roberts, John (ed.). *The Oxford Dictionary of the Classical World*. Oxford University Press, 2005.

Smyth, Herbert Weir. *Greek Grammar*. Copyright 1920 by Herbert Weir Smyth. Harvard University Press, 1984.