

An Ownership Model of Annotation: The Ancient Greek Dependency Treebank

David Bamman, Francesco Mambrini and Gregory Crane
The Perseus Project, Tufts University

Abstract

We describe here the first release of the Ancient Greek Dependency Treebank (AGDT), a 190,903-word syntactically annotated corpus of literary texts including the works of Hesiod, Homer and Aeschylus. While the far larger works of Hesiod and Homer (142,705 words) have been annotated under a standard treebank production method of soliciting annotations from two independent reviewers and then reconciling their differences, we also put forth with Aeschylus (48,198 words) a new model of treebank production that draws on the methods of classical philology to take into account the personal responsibility of the annotator in the publication and ownership of a “scholarly” treebank.

1 Introduction

Data-driven research in linguistics relies on the existence of a large body of texts that have been annotated on several linguistic levels. For modern languages like English, these tend to be comprised of genres like newswire; for Latin, Greek, and other historical languages, our observations are based on a smaller but more heavily studied canon. An article from the *Wall Street Journal* is certainly more representative of how native English speakers actually speak than Homer’s epic *Iliad* is for ancient Greeks, but the *Iliad* has been a focused object of study for almost 3,000 years, with schoolchildren and tenured professors alike scrutinizing its every word, annotating its syntax, semantics and other linguistic levels either privately in the margins of their books or as published commentaries.

Recent scholarship has seen the rise of a number of treebanks for historical languages over the past few years, including Middle English [11], Early Modern English [10], Old English [22], Medieval Portuguese [18], Ugaritic [24], Latin [1, 15] and several Indo-European translations of the New Testament [8]. The long history of philological research on the individual texts that constitute these works highlights what is perhaps the greatest difference between syntactically annotated corpora for modern languages and those for historical ones – while ambiguity is of course present in all language, the individual ad hoc decisions that annotators make

in resolving syntactic ambiguity when creating modern treebanks have, for heavily studied Classical and other historical texts, been debated for centuries; dissertations and entire careers have been made on the study of a single work of a single author. Since over two thousand years separate us from the time when Greek and Roman authors were writing, Classical texts also have additional confounding factors which bring this debate to new levels – not simply the interpretation of the text as it appears to us, but what actually constitutes that text itself.

In creating an annotated corpus of a language for which no native speakers exist (and for which we subsequently cannot rely on native intuitions), we are building on a mountain of prior scholarship that has shaped our fundamental understanding of the text. In order to accommodate this level of scholarly debate on a basic level of annotation, we describe here a new mode of treebank production – what we are terming a “scholarly” treebank. Just as every critical edition and commentary bears the mark and reputation of its author, including the cultural context in which it was written, every act of annotation is here associated with the individual who created it. By stressing such ownership, we hope to transform the act of treebanking from an anonymous practice into a mode of scholarly publication.

The aim of this paper is twofold: we present the first release of the Ancient Greek Dependency Treebank (AGDT), containing 190,903 words of Ancient Greek annotated under a dependency grammar, and describe how it can form the core of scholarly treebanks to come.

2 The Ancient Greek Dependency Treebank

Ancient Greek is a highly inflected language with a considerable degree of variability in its word order. Even the comparatively simpler texts of Homer manifest a high degree of non-projectivity, where constituents themselves are broken up with elements of other constituents, as in the dependency graph shown in figure 1, where an arc drawn from $\mu\tilde{\eta}\nu\iota\nu$ (“rage”) to Ἀχιλλῆος (“Achilles”) crosses that drawn from the root of the sentence to $\tilde{\alpha}\epsilon\iota\delta\epsilon$ (“sing”).¹ This flexibility has encouraged us to base our annotation style on the dependency grammar used by the Prague Dependency Treebank [6] for Czech (another non-projective language), which has since been widely adopted by a number of annotation projects for other languages, including Arabic [7] and Modern Greek [16]. Since Latin and Ancient Greek are so closely related, our specific guidelines have been built as an extension of those used for the Latin Dependency Treebank [1] and the Index Thomisticus [15].

2.1 Annotation

The efficient annotation of Ancient Greek is hindered both by the fact that no native speakers exist and that the texts we have available are typically highly stylized

¹See Nivre [13] for a formal definition of projectivity.

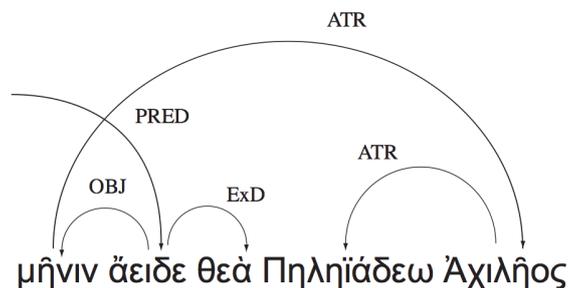


Figure 1: μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος (“Sing, goddess, of the rage of Achilles, the son of Peleus”), Homer, *Iliad* 1.1. Arcs are drawn from heads to their dependents.

in nature. This difficulty and the ability of a sentence to present multiple valid syntactic interpretations has an impact on both annotation speed and inter-annotator agreement. While the Penn Treebank can report a productivity rate of between 750 and 1000 words per hour for their annotators after four months of training [21] and the Penn Chinese treebank can report a rate of 240-480 words per hour [3], our annotation speeds are significantly slower, ranging from 97 words per hour to 211, with an average of 124. Since we preserve the individual streams of annotation from all annotators, we can calculate inter-annotator accuracy (IAA) measures for the treebank in its entirety. Table 1 presents three such measures drawn from [5]: attachment score (ATT), label score (LAB) and labeled attachment score (LABATT), each one being the average annotator accuracy compared to the final corrected data. While our ATT of 87.4% approaches the 91.5% and 89.2% reported by the CATiB Arabic Treebank [5], our LAB and LABATT scores are lower, averaging 85.3% and 80.6%, respectively.

	ATT	LAB	LABATT
Hesiod, W&D	85.1%	85.9%	79.5%
Homer, Iliad	87.1%	83.2%	79.3%
Hesiod, Odyssey	87.5%	85.7%	80.9%
Total	87.4%	85.3%	80.6%

Table 1: Average inter-annotator accuracy in terms of attachment (ATT), label (LAB) and labeled attachment (LABATT) scores.

The backgrounds of the annotators range from advanced undergraduate students to recent PhDs and professors, with the majority being students in graduate programs in Classics. To help provide reading support for more efficient annotations, we have embedded our annotation interface within a larger digital library that presents the Greek source text to be annotated along with contextualizing secondary publications such as translations, commentaries, and references in dictionaries. In addition to an initial training period, annotators are actively engaged

in new learning by means of an online forum in which they can ask questions of each other and of project editors; this allows them to be kept current on the most up-to-date codifications to the annotation guidelines while also helping bring new annotators up to speed. In the “standard” model of production, every sentence is annotated by two independent annotators and the differences are then reconciled by a third. This reconciliation (or “secondary” annotation as it is encoded in the XML release) is undertaken by a more experienced annotator/editor, typically a PhD with specialization in the particular subject area (such as Homer).

As figure 2 illustrates, all annotations are publicly released with the usernames of the primary and secondary annotators (which are then also associated with real names and institutional affiliations). By publicly acknowledging authorship, we are making our first steps toward an ownership model for annotation (more fully discussed below) and hope to provide a means for students, both graduate and undergraduate alike, to engage in the act of scholarly research and produce scientific data that can be useful to the wider Classics community.

```
<sentence id="504" document_id="Perseus:text:1999.01.0135" subdoc="book=2:card=1" span="h)=mos0:0">
  <primary>tovahk</primary>
  <primary>AlexLessie</primary>
  <secondary>jackmitchell</secondary>
  <word id="1" form="h)=mos" lemma="h)=mos" postag="c-----" head="34" relation="AuxC"/>
  <word id="2" form="d" lemma="de/1" postag="g-----" head="34" relation="COORD"/>
  <word id="3" form="h)rige/nea" lemma="h)rige/nea1" postag="n-s---fn-" head="6" relation="ATR"/>
  <word id="4" form="fa/nh" lemma="fai/nw1" postag="v3saip--" head="1" relation="ADV"/>
  <word id="5" form="r(ododa/ktulos" lemma="r(ododa/ktulos1" postag="a-s---fn-" head="6" relation="ATR"/>
  <word id="6" form="*)hw/s" lemma="h)w/s1" postag="n-s---fn-" head="4" relation="SBJ"/>
```

Figure 2: XML fragment from the AGDT (Homer, *Odyssey* 2.1).

2.2 AGDT 1.0

Using this model, we have annotated a total of 190,903 words from three different authors (Hesiod, Homer and Aeschylus), as distributed in table 2.

In addition to the index of its syntactic head and the type of relation to it, each word is also annotated with the lemma from which it is inflected and its morphological code (a composite of nine different morphological features: part of speech, person, number, tense, mood, voice, gender, case and degree). All of the files have been freely released under a Creative Commons license.²

For the works of Homer and Hesiod, we have followed the standard production method of soliciting annotations from two different annotators and then reconciling the differences between them. Aeschylus, whose textual tradition is much more fragmentary, has presented an ideal case for annotation as a scholarly treebank.

²All treebank data can be found at: <http://nlp.perseus.tufts.edu/syntax/treebank/>.

Method	Author	Work	Sentences	Words	
Standard	Hesiod	Works and Days	446	6,214	
	Homer	Iliad	2,470	37,223	
		Odyssey	6,417	99,268	
Scholarly	Aeschylus	Agamemnon	809	9,796	
		Eumenides	521	6,376	
		Libation Bearers	572	6,563	
		Persians	478	6,223	
		Prometheus Bound	589	7,045	
		Seven Against Thebes	478	6,206	
		Suppliants	518	5,989	
		Total:		13,298	190,903

Table 2: AGDT 1.0 composition by work.

3 Scholarly Treebanks

Linguistic annotation projects have, of necessity, long focused on the creation of the single-best annotation, enforcing agreement between annotators even in cases of ambiguity. This approach works well for generic text such as newswire (where the value lies not in any individual sentence but rather in the aggregation of many) but breaks down when the objects of annotation are themselves the focus of scholarly debate. In these cases we must provide a means for encoding multiple annotations for a text and allowing scholars who disagree with a specific annotation to encode their disagreement in a quantifiable form.

For historical texts especially, scholarly disagreement can be found not only on the level of the correct syntactic parse, but also on the form of the text itself. These two levels are not completely isolated from each other, since it is often a scholar’s understanding of the meaning of the text – i.e., what it *should* say – that informs their decisions about its reconstruction (i.e., what it actually *did* say). The need for this reconstruction is due to the process of textual transmission. We do not have a copy of Plato’s *Apology* in his own hand; what we have instead is a series of manuscripts, one copied from the other, with errors introduced into each generation by the process of hand-copying by medieval scribes. This manuscript transmission allowed the work of the author to survive, but resulted in a considerable alteration of the text. Modern critical editions attempt to reconstruct the original text by a systematic comparison of that manuscript tradition.

As the product of scholarly labor, a critical edition displays the text as it is reconstructed by an editor; it is thus an interpretative hypothesis whose foundations lie on the methods of textual criticism. A scholarly treebank may be defined by analogy as a syntactically annotated corpus that again reflects an interpretation of a single scholar, based not only on the scholar’s philological acumen but also on an inevitable degree of personal taste and opinions that are culturally and historically

determined. A scholarly treebank thus distances itself from the notion that linguistic annotations can be absolute; when dealing with non-native historical languages especially, a syntactic interpretation of a sentence is always the interpretation of an individual and therefore subject to debate.

3.1 Aeschylus

We have decided to treat the corpus of Aeschylus' plays as the first example of a scholarly treebank due to the difficulty (even by Classical standards) of its textual tradition. The historical position of this author (ca. 525 BCE – ca. 456 BCE) may partly account for this complexity. Classical authors established him as the true founder of tragedy,³ the poet who took a genre that was already characterized by a high degree of linguistic diversity and complexity and transformed it from its humble and rustic origins into a sublime form of poetry.

When Ancient Greek literature was rediscovered in Western Europe in the Renaissance, the difficulty of reading Aeschylus (along with all other Classical texts) was increased by the errors that inevitably intruded into the text during the process of copying. Out of a whole corpus that included between 70 and 90 tragedies, a canon of seven plays traditionally attributed to Aeschylus was chosen most likely in late antiquity to be copied integrally (a number of fragments of other works also survived independently).⁴ Of these, only three (*Prometheus Bound*, *Persians* and *Seven Against Thebes*) have been preserved by a group of manuscripts large enough to assure a good transmission, and two (*Libation Bearers* and *Suppliants*) survive only in one single manuscript (and its copies), the Laurentianus 32.9. Starting from this controversial evidence, a vast number of scholars, beginning from the first printed edition of 1518, have undertaken the enterprise of giving justice to the complex poetry of the author and amending the text of all errors [12]. The main bibliographic catalogue for Aeschylus lists no less than 127 editions of the seven plays for the years 1518-1974, counting also the major reissues [23, 633-35]; if we include the separate editions of the single tragedies or of the trilogy (the *Oresteia*), the count rises exponentially.

3.2 Example: *Agamemnon* 176-8

One example (*Ag.* 176-8) may give an idea of how this complex history affects the practical task of treebanking. In a pivotal passage of the so-called “Hymn to Zeus” in the *Agamemnon*, the chorus voices for the first time in the play a theological vision that will dominate the whole *Oresteia*: the rule of “learning through suffering” as the means by which Zeus instructs the mortals to wisdom. Smyth's

³Cf. Dioscorides (3rd century BCE), *Tragicorum Graecorum Fragmenta*, Testimonium 163 [17, 107-8].

⁴The conflicting ancient evidence on the number of plays attributed to Aeschylus in antiquity is collected by Radt [17].

[20] edition⁵ of the Greek text reads:

τὸν φρονεῖν βροτοὺς ὀδῶ-
[the . to be wise . mortals . putting on ...]
σαντα, τὸν πάθει μάθος
[... the way . the . through suffering . learning]
θέντα κυρίως ἔχειν.
[establishing . authoritatively . hold]

The precise meaning of the passage is subject to debate (see below), but a basic translation is: “[Zeus] ... who put men on the path of wisdom, who established that the law ‘learning through suffering’ shall be in force.”

Though the formula *πάθει μάθος* (“learning through suffering”) is both quoted and commented upon in many general introductions to the theater of Aeschylus (it was even quoted by Robert F. Kennedy in his speech on the assassination of Martin Luther King Jr. [9]), both the text and syntactic interpretation of the sentence are highly controversial.

For instance, we may note at once that the second masculine accusative article τὸν (l. 177: τὸν ... θέντα, “the one establishing”) is a modern conjecture proposed by Schütz [19] and subsequently accepted by many followers, including Smyth above [20]. In contrast, all the manuscripts of the *Agamemnon* unanimously read a dative neuter article (τῷ) that is morphologically licensed to modify the dative noun *πάθει* instead (“the suffering”).

This conjecture of Schütz is directly related to a question of syntactic interpretation. There is a fundamental ambiguity in the relationship between the two participles *ὀδῶσαντα* (“put on the way”) and *θέντα* (“establish”). Is it apposition (Zeus is the god that “put the men on the path of wisdom, *i.e.*, the one who established the law”) or subordination (“Zeus gave wisdom to men *by establishing* the law”)? Whichever interpretation we choose, the article with *πάθει* is rather difficult to understand. This inherent ambiguity on several levels has led the three most recent commentaries on the play – Fraenkel [4], Denniston-Page [14] and Bollack [2] – to adopt three very different solutions based on their own weighing of the philological evidence, each resulting in a markedly different syntactic tree. Figures 3 and 4 present these three trees annotated under a dependency grammar, and illustrate the variety of interpretations that have been argued in print for just this one sentence alone.⁶

⁵A digital version of this edition is available at: <http://www.perseus.tufts.edu/hopper/>.

⁶The different interpretations are, of course, reflected also in different translations or paraphrases. Fraenkel (1950): “it is Zeus who has put men on the way to wisdom by establishing as a valid law *By suffering they shall win understanding*” [4]; Denniston-Page (1957): “he who set men on the path to understanding, who laid down the law, ‘learning through suffering’, to hold good” [14]; Bollack (1981): “de celui qui a ouvert aux mortels le penser, posant qu’ils tiendraient principalement leur savoir par la souffrance” (“of the one who opened the way of thinking for mortals, by establishing that chiefly by their suffering they will have their knowledge”) [2].

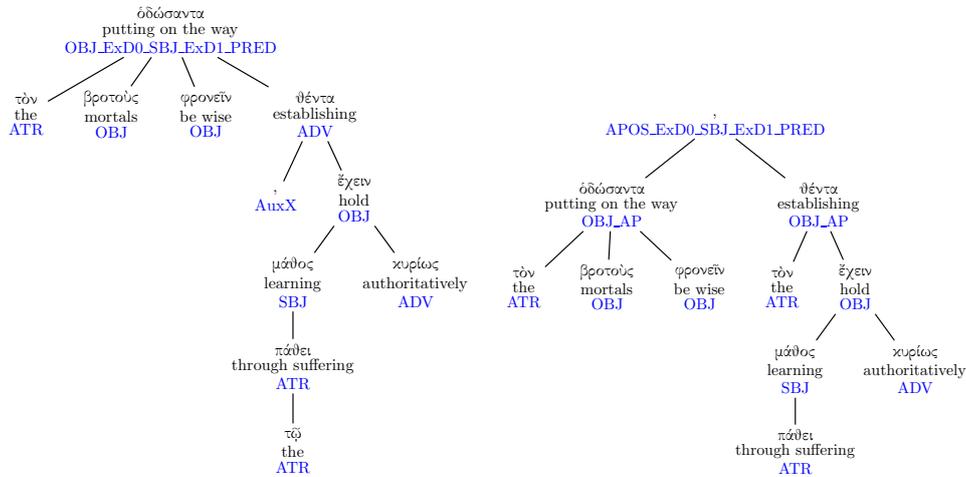


Figure 3: Trees of Fraenkel (left) and Denniston-Page (right) for Ag. 176-8.

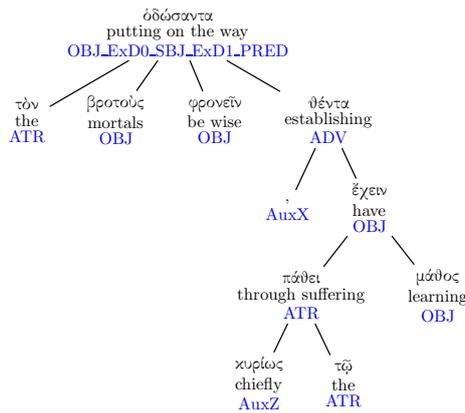


Figure 4: Bollack's tree for Ag. 176-8.

3.3 A Critical Treebank of Aeschylus

The variety of textual and syntactic interpretations for just these three lines of Aeschylus begins to point out the shortcomings of a standard treebank production model for texts of ongoing scholarly debate. While distributing the task of annotation across two independent annotators and then reconciling their differences does help remove any single annotator's personal bias from the final annotated corpus, for these texts what we want is exactly that – the quantified decisions of a single individual (whether Fraenkel, Denniston-Page, Bollack, or some other scholar), along with the sense of ownership and personal responsibility that attend such work. In this, the scholarly practice of annotation is practically indistinguishable from the creation of a critical edition of a text and attendant commentary.

For the complete works of Aeschylus, we have created a treebank based on the

work of a single scholar following these philological principles. As in the creation of critical editions of texts, each syntactic annotation is created in consultation with the current state of Aeschlyean criticism; the resulting work stands as a contribution to that ongoing body of research. In total, the scholarly treebank amounts to 48,198 words (3,965 sentences) from 7 different plays and is included in the public release of AGDT 1.0 (see Figure 2). Figure 5 displays a fragment of that data – unlike the canonically produced texts of Homer and Hesiod, where a consensus is established among three individuals, this work here is the sole responsibility of the scholar who created it and remains that scholar’s published interpretation of the text.

```
<sentence id="31" document_id="Perseus:text:1999.01.0003" subdoc="card=104" span="ai)/linon0:1">
  <annotator>FrancescoM</annotator>
  <word id="1" form="ai)/linon" lemma="ai)/linon1" postag="n-s---na-" head="2" relation="AuxY"/>
  <word id="2" form="ai)/linon" lemma="ai)/linon1" postag="n-s---na-" head="3" relation="OBJ"/>
  <word id="3" form="ei)pe/" lemma="ei)=pon1" postag="v2sama---" head="6" relation="PRED_CO"/>
  <word id="4" form="," lemma="comma1" postag="u-----" head="6" relation="AuxX"/>
  <word id="5" form="to\" lemma="o(1" postag="l-s---nn-" head="7" relation="ATR"/>
  <word id="6" form="d" lemma="de/1" postag="g-----" head="0" relation="COORD"/>
  <word id="7" form="eu)=" lemma="e)u/s" postag="a-s---nn-" head="8" relation="SBJ"/>
  <word id="8" form="nika/tw" lemma="nika/w1" postag="v3spma---" head="6" relation="PRED_CO"/>
  <word id="9" form="." lemma="period1" postag="u-----" head="0" relation="AuxK"/>
</sentence>
```

Figure 5: XML fragment from the AGDT (Aeschylus, *Ag.* 121).

4 Conclusion

By focusing on authorship in the release of the AGDT, we hope to drive future research in two directions. First, by publicly releasing the data with citable attributions of ownership, we hope to provide the core around which other interpretations of the data can be layered – a scholar who disagrees with a single annotation decision need not start from scratch to contribute a new annotation, but can simply build on the existing data and change only the elements subject to debate. As the example from *Agamemnon* 176-8 from above clearly shows, Classical texts very often license multiple syntactic interpretations, and providing a quantified record of how these multiple interpretations differ can only help drive future research.

Second, by publicly acknowledging the creator of the annotation, we hope to promote the act of treebanking as a scholarly publication no different than a critical edition or commentary. In so doing, we hope to engage a much wider audience in the creation of syntactically annotated data for historical languages – not only the corpus and computational linguists who have typically promoted them, but Classicists as well, for whom treebanking is simply a quantified form of the traditional scholarship that has been conducted for centuries.

5 Acknowledgments

Grants from the Alpheios Project (“Building a Greek Treebank”), the Andrew W. Mellon Foundation (“The CyberEdition Project: Workflow for Textual Data in Cyberinfrastructure”) and the National Endowment for the Humanities (PR-50013-08, “The Dynamic Lexicon: Cyberinfrastructure and the Automated Analysis of Historical Languages”) provided support for this work.

References

- [1] David Bamman and Gregory Crane. The Latin Dependency Treebank in a cultural heritage digital library. In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*, pages 33–40, Prague, 2007. Association for Computational Linguistics.
- [2] Jean Bollack and Pierre Judet de La Combe. *L’Agamemnon d’Eschyle: le texte et ses interprétations*. Presses universitaires de Lille, Lille, 1981.
- [3] Fu-Dong Chiou, David Chiang, and Martha Palmer. Facilitating treebank annotation using a statistical parser. In *Proceedings of the First International Conference on Human Language Technology Research HLT ’01*, 2001.
- [4] Eduard Fraenkel. *Aeschylus. Agamemnon*. Clarendon Press, Oxford, 1950.
- [5] Nizar Habash and Ryan Roth. CATiB: The Columbia Arabic Treebank. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 221–224, Suntec, Singapore, 2009. Association for Computational Linguistics.
- [6] J. Hajič. Building a syntactically annotated corpus: The Prague Dependency Treebank. In E. Hajičová, editor, *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*. Prague, Charles University Press, 1998.
- [7] J. Hajič, O. Smrž, P. Zemánek, J. Šnaidauf, and E. Beška. Prague Arabic dependency treebank: Development in data and tools. In *Proc. of the NEMLAR Intern. Conf. on Arabic Language Resources and Tools*, 2004.
- [8] D.T.T. Haug and M.L. Jøhndal. Creating a Parallel Treebank of the Old Indo-European Bible Translations. In *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, 2008.
- [9] Robert F. Kennedy. Statement on the assassination of Martin Luther King, Indianapolis, Indiana, April 4, 1968.
- [10] A. Kroch, B. Santorini, and L. Delfs. Penn-Helsinki Parsed Corpus of Early Modern English. <http://www.ling.upenn.edu/hist-corpora/ppceme-release-1>, 2004.

- [11] A. Kroch and A. Taylor. Penn-Helsinki Parsed Corpus of Middle English, 2nd edition. <http://www.ling.upenn.edu/hist-corpora/ppcme2-release-2/>, 2000.
- [12] Monique Mund-Dopchie. *La survie d'Eschyle à la Renaissance. Éditions, traductions, commentaires et imitations*. Peeters, Louvain, 1984.
- [13] Joakim Nivre. Constraints on non-projective dependency parsing. In *EACL*. The Association for Computer Linguistics, 2006.
- [14] Denys Page. *Aeschylus Agamemnon. Edited by the late John Dewar Denniston and Denys Page*. Clarendon Press, Oxford, 1957.
- [15] Marco Passarotti. Verso il Lessico Tomistico Biculturale. La treebank dell'Index Thomisticus. In P. Raffaella and F. Diego, editors, *Il filo del discorso. Intrecci testuali, articolazioni linguistiche, composizioni logiche*, pages 187–205. Roma, Aracne Editrice, 2007.
- [16] Prokopis Prokopidis, Elina Desipri, Maria Koutsombogera, Harris Papageorgiou, and Stelios Piperidis. Theoretical and practical issues in the construction of a Greek dependency treebank. In *In Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT)*, pages 149–160, 2005.
- [17] Stefan Radt. *Tragicorum Graecorum Fragmenta. Vol. III: Aeschylus*. Vandenhoeck und Ruprecht, Göttingen, 1985.
- [18] Vitor Rocio, Mário Amado Alves, J. Gabriel Lopes, Maria Francisca Xavier, and Graça Vicente. Automated creation of a Medieval Portuguese partial treebank. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, pages 211–227. Kluwer Academic Publishers, 2003.
- [19] Christian Gottfried Schütz. *Commentationum in Aeschyli Agamemnon libellus I*. Jena, 1780.
- [20] Herbert Weir Smyth. *Aeschylus. With an English Translation*. Loeb Classical Library. Harvard University Press, Cambridge, 1922.
- [21] Ann Taylor, Mitchell Marcus, and Beatrice Santorini. The Penn Treebank: An overview. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, pages 5–22. Kluwer Academic Publishers, 2003.
- [22] Ann Taylor, Anthony Warner, Susan Pintzuk, and Frank Beths. York-Toronto-Helsinki Parsed Corpus of Old English Prose, 2003.
- [23] André Wartelle. *Bibliographie d'Eschyle et de la tragédie grecque 1518-1974*. Les Belles Lettres, Paris, 1978.
- [24] Petr Zemánek. A treebank of Ugaritic: Annotating fragmentary attested languages. In *Proceedings of the Sixth Workshop on Treebanks and Linguistic Theories (TLT2007)*, pages 213–218, Bergen, 2007.