

# The Syntax of the Heroes? A Treebank-Based Approach to the Language of the Sophoclean Characters

Francesco Mambrini



## 1. Introduction

### 1.1 Computers, Literature, and Language

“It is a truth not generally acknowledged that, in most discussions of works of English fiction, we proceed as if a third, two-fifths, a half of our material were not really there.” [1] With these words, John Burrows opened his book on the language of Jane Austen’s characters—a book that is now rightly considered a classic of the Digital Humanities.

The quoted passage serves as a fitting introduction to Burrows’s work—and, by reflection, to the present paper—not only on account of the witty paraphrase of a famous Austenian *incipit*. More importantly, these words highlight an aspect in the relation between language, literary criticism, and computers that is still very relevant today. In 2021, the notion that computers enable scholars, including linguists and literary critics, to process an amount of data that would be extremely hard to handle for the unaided human eye and brain is bound to sound more like a truism than like a “truth generally acknowledged.” Still, while criticism continues to be mostly concerned with singularities, we often forget that corpus-based studies can aim to investigate the totality of language, even in its most frequent and inconspicuous or “noisy” aspects. And though we do not lack surveys of pervasive facts of language that were compiled without the aid of machines, [2] it is clear

that it is precisely with the investigation of these phenomena that computers may help the most.

In a complex system like language, where redundancy plays a relevant communicative role, some elements are so frequent that they are ordinarily just taken for granted.

“Function words,” such as articles and prepositions, which almost invariably rank at the top of the frequency lists in corpora, are a good example of this. In oral and written communication, frequently repeated features build a sort of general frame of discourse on top of which the most eye-catching elements can be foregrounded. Yet, the fact that they ordinarily escape notice does not mean that their use is deprived of any stylistic value, or that they do not have anything important to reveal. [3]

It is this very linguistic phenomenon—the distribution of the most frequent words in the dialogues of Jane Austen’s novels—that was the object of Burrows’s analysis. The book is a brilliant illustration of how the finesse of the literary critic and the computational power of machines can be combined to produce new insight. Also, it reminds us that quantitative studies in literature do not necessarily have to dig into huge corpora in order to discover new facts, as is customary with those approaches that fall under the general definition of “distant reading” or “macroanalysis.” [4] Even relatively small collections of texts can reveal hitherto-unnoticed aspects when the fabric of their language is studied at the level of granularity of the very frequent words or structures.

It is, thus, fitting that one of the papers presented at a conference entitled *Digital Classics III: Re-thinking Text Analysis* should start with this lesson of Burrows’s *Computation into Criticism*.

## 1.2 The Syntax of Characters

For readers of Greek tragedy, Burrows’s book holds at least one additional element of interest: his investigation of the “idiolects” of the characters is also intriguing because it offers a panoply of statistical instruments, along with a very successful example of interpretation, to investigate a question that has engaged critics since the times of Aristotle. The question is, of course, the problem of the “character.”

A famous passage from the *Poetics* discusses the *ethos* as one of the structural elements of tragedy and argues for its subordination to the *mûthos*, or the structuring of the events. [5] From then, the problem of the degree and the means by which the ancient dramatists

provided their characters with an individual identity has never ceased to engage readers. As Thumiger puts it, in light of modern assumptions about the individual as unique, unrepeatable entity defined by its inner life, “the student of tragedy still finds it difficult not to perceive the representation of tragic character as cold, distant and crude.” [6]

Starting from a perspective that is similar to Aristotle’s, twentieth-century critics have tended to reject the psychologizing approaches that were fashionable in earlier scholarship and were often based on naive assumptions of psychological naturalism. [7] The landmark in this paradigm shift is recognized in the work of Tycho Wilamowitz-Moellendorf (1917). While some of the interpretations advanced by Tycho Wilamowitz would appear too radical nowadays, his book has successfully reset the debate on a perspective that is less preoccupied with psychological nuances and more focused on reading the actions and words of each figure in light of the dramatic needs of the play, or of the relations between the *personae* and the other expressive components of the tragic performance. [8]

Precisely from a viewpoint that takes action, plot, and language into consideration, one aspect of the Sophoclean technique that has come to prominence in postwar studies is the poet’s tendency to build his dramas around figures of “heroic” stature. Though they may differ greatly from one another in term of mythical background, social status, or the type of events that they face in the dramatic arc of their play, Sophocles’ main characters do share several traits that an influential book has discussed under the definition of “heroic temper.” [9]

Questions of unity, typology, and individual characterization of the Sophoclean *ethê* are, therefore, still vital to contemporary criticism. In the present work, we attempt to approach these questions from a point of closer scrutiny of the language, which we study with the help of data from a syntactically annotated corpus. The problem of the linguistic individuality of tragic characters is not one that is too often tackled by critics. Easterling, for one, has tentatively suggested that Sophocles’ use of lexicon and imagery is influenced by his conceptions of the characters. [10] In addition to that, few readers, I believe, would object to the assumption that some meaningful words—such as *mégas* in *Ajax*, [11] or *nómos* and the compounds of *autós* studied by Loraux (1986) in *Antigone*—are used to link some of the play’s most relevant themes to the conflicting perspectives of its characters. Our decision to concentrate on the syntax is motivated by the aspiration to verify whether this same characterization is also recognizable here: is the choice of the most frequent syntactic constructions used by the characters also influenced by, or in any

way linked to, the overall dramatic function that Sophocles assigns to the different *personae*?

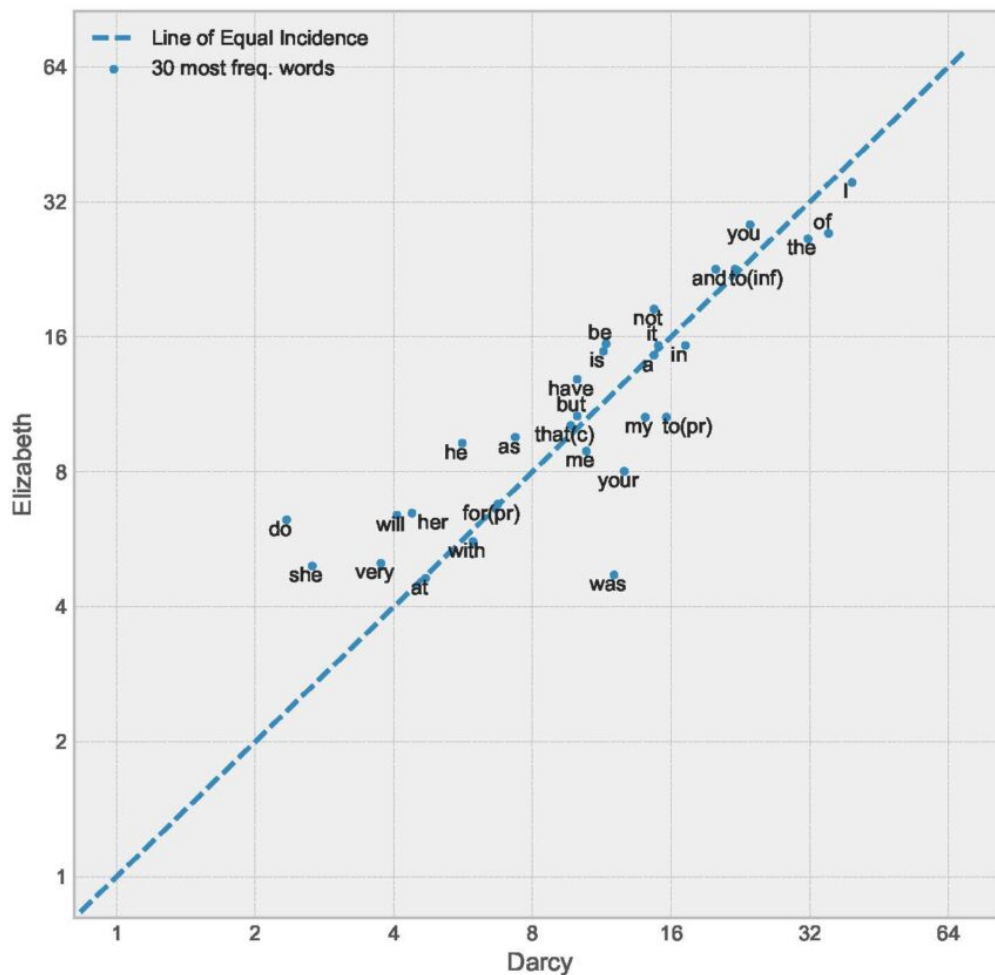


Figure 1. Correlation of the 30 most frequently used words between Darcy and Elizabeth in *Pride and Prejudice*.

In what follows, we will approach this question starting with some of the most basic exploratory analyses that Borrow (1987) applied to the lexicon of Jane Austen. We will leave for further studies other techniques of multivariate statistics, such as Principal Component Analysis, which was popularized by *Computation into Criticism* and widely applied and discussed ever since. [12]

One example of the type of investigations that we will replicate on Sophocles is reproduced in Figure 1. The chart illustrates the distribution of the thirty most frequent words in Austen's dialogue, as computed by Burrows, relative to the two main figures of

*Pride and Prejudice* with Elizabeth Bennet on the y- and Darcy on the x-axis; the raw occurrences are normalized to the per-thousand ratio, and the plot is shown on a logarithmic scale—so that the distances on the top portion of the axes appear less dramatic than they actually are. [13] The diagonal line connects the points that have equal values on the two axes. The chart displays “the close resemblance between the idiolect of two strong-minded, intelligent and essentially well-mannered characters.” [14] The “resemblance” is evidenced by the fact that the words are mostly distributed along the line of equal incidence and so the frequency (per 1000 words) of one word in Elizabeth is met with a more or less identical frequency in Darcy.

The correlation can be measured with statistical methods such as the Pearson Correlation Coefficient, which yields values ranging from 0 (no correlation) to 1 or -1 (positive or negative perfect correlation). [15] The r-value of this couple of characters is 0.930—a very high score and yet not among the highest in the corpus Austen’s dialogues, or indeed in the novel. [16] One may also note that one clear advantage of this visualization is that it points out the elements of similarity and difference, which can then become the object of closer scrutiny: the “was” that is conspicuously on Darcy’s side, for instance, is easily explained on account of the role that the past tense plays in Darcy’s recollections and explanations, especially in the long letter that Elizabeth reads in chapter thirty-five.

In the following sections, we start with similar visualizations (Section 3.1) and then attempt to provide a bird’s-eye view of the correlation scores among the characters (Section 3.2). As said, the present survey is intended as a preliminary exploration only. Yet, we hope that the paper succeeds in confirming the interest of the methodology that we propose and the fruitfulness of this peculiar kind of computational approach to Greek literature.

### 1.3 A Treebank of Sophocles

Since 1987, much has changed in the world of computing. One of the most relevant innovations for the study of language is the diffusion and the widespread recognition of linguistically annotated corpora known as “treebanks.” [17] In the parlance of corpus and computational linguistics, a treebank is a corpus of sentences with at least two layers of word-by-word annotation: part-of-speech tagging, and a representation of the syntactic structure codified according to some form of a theory of syntax.

Since 2009, the Perseus Digital Library has maintained a treebank of Ancient Greek and Latin literary texts: The Ancient Greek and Latin Dependency Treebank (AGLDT, or AGDT for the Greek section only). [18] Five complete tragedies of Sophocles (*Ajax*, *Electra*, *Oedipus Rex*, *Antigone*, and *Women of Trachis*) are published in this collection and are freely available for browsing and downloading. Each word of these texts is annotated with the lemma and full morphological analysis, along with other metadata whose importance will be discussed later. [19] The syntax of the sentence is codified according to a form of Dependency Grammar, where words are attached directly to one another with no intermediate constituent, such as a noun or verb phrase, introduced; the relation between head and dependent is described with a label from a set of about twenty tags. A definition of the labels and the rules for syntactic annotations are defined in the project guidelines. [20] To simplify the matter, in many prototypical cases the formalism adopted reproduces some “intuitions” about the syntax that should be familiar to most classicists: a verb governs its satellites, like subjects, direct objects, and complements; equally, nouns are licensed to govern definite articles and adjectives that agree with them. [21]

Work	Annotator	Nr. of Treebank Tokens	
		TOT	Artificial nodes
<i>Ajax</i>	D. Libatique	9693	275
<i>Electra</i>	F. Mambrini	10808	316
<i>Oedipus Rex</i>	F. Mambrini	11523	337
<i>Trachiniae</i>	F. Mambrini	9028	206
<i>Antigone</i>	F. Mambrini	8993	240
TOT		50045	1374

Table 1. Sophoclean tragedies included in the AGDT

It is important to note that four of the five tragedies (with the exception of *Ajax*) have been curated by one single annotator, according to a workflow called “scholarly annotation,” [22] which aims to transfer the model of the critical commentary to the treebanking format. [23] In this framework, the level of text reconstruction is also part of the annotation work. While the treebanks of other texts of the AGDT, including also *Ajax*, reproduce the same edition that is used in the Perseus Digital Library (Storr 1912, for *Ajax*), as the editor of the other four Sophoclean tragedies, I modified the text whenever I

thought that Storr's edition needed to be updated in light of the results of the most recent scholarship. [24]

Table 1 lists the tragedies of Sophocles that are currently distributed with the AGDT. Note that the “tokens,” which are the unit of analysis of a treebank, do not necessarily correspond to words, although they tend to: single words like the coordinating negative conjunctions, such as οὔτε, are split into two tokens in order to annotate the peculiar syntactic function of the two components. The AGDT also includes artificial nodes—that is, tokens that do not correspond to words in the text but represent elements that are elided in elliptical constructions and are inserted to account for the syntactic structure of the sentence. The total number of artificial nodes, which will not be the focus of this study, is also reported in Table 1.

## 2 Methodology

### 2.1 In Search of the Speaking Character

The plan to investigate the syntax of the Sophoclean characters, following the example of Burrows, is fraught with a multitude of problems, both on the technical and on the methodological side. We will start with the more concrete issues.

Although the treebank is an invaluable source of knowledge on many aspects of language, [25] there is one crucial piece of information that is conspicuously absent for our present needs: namely, the identity of the speaker of each sentence is not recorded anywhere.

What the treebank does register, on the other hand, is the canonical reference attached to every token of the collection, expressed in the form of a CTS-URN. This string informs us that, for instance, the word τέκνα in the first sentence of *Oedipus Rex* is located at line 1 of a digital edition of the play. [26] As the digital versions that the AGDT is based upon are compliant with the standards defined by the Text Encoding Initiative (TEI), they should (and in fact, they do) contain the information that we are trying to retrieve. In other words, a CTS-URN acts as a sort of entry in an “address book” that enables users to cross the linguistic information of the treebank with the editorial metadata stored in the digital edition.

The TEI files of the Perseus Digital Library make use of the “sp” (speech) element to group the lines (or line sections, in the case of *antilabé*) in a speech, [27] so that a change of

speaker marks the beginning of a new “sp” section. The label of the speaking character is also recorded in a child element of “sp” named “speaker.” [28] By programmatically querying this structure, starting from the line number in the CTS-URN that is attached to the treebank tokens, it is possible to ascertain how to attribute each word to their speaker. [29] That combined information provides the backbone data of this investigation.

I have lengthened the technical discussion not only to document the methodology in full but because I believe that the present case provides a good illustration of a best practice in designing (and exploiting) digital resources for classical philology. Instead of confusing information pertaining to different conceptual layers, the AGDT wisely keeps the two aspects—the linguistic interpretation stored in treebanks and the editorial and philological features—distinct. A standardized unique identifier, such as a CTS-URN, acts as a link between the two. Conversely, users can profit from this synergy to obtain the data that suits their research needs.

## 2.2 Defining the Syntax

The most consequential methodological issue for this investigation is deciding what information we want to extract from the annotation so as to be able to study the syntax of the characters.

The principal difficulty that a treebank allows us to bypass is the fact that, as Burrows himself noted, bigrams of lexical items do not recur frequently enough in a corpus of the size of Austen’s novels to support quantitative analysis. [30] Treebanks, on the other hand, encapsulate a description of words and sentences that is formalized and abstract enough to get away altogether with the scarcity of lexical attestations. The phrase τέκνον Λαρτίου, for instance, is attested only once in our five Sophoclean tragedies and so is the combination of the two lemmata (*Ajax*, 380). However, the lemma τέκνον plus a genitive noun is found in three other passages (*Trachiniae* 598, 665, *Oedipus Rex* 157). To a higher level of abstraction, the couplet formed by a noun head and a noun dependent is attested 981 times, while the combination between noun and a noun in genitive occurs 732 times. Finally, the syntactic label that is attributed to the phrase (ATR) yields 7688 results.

All these descriptions are perfectly legitimate ways to represent this syntactic unit, but they are not equally useful to support a data-driven, quantitative study of the syntax of the characters. The first and the second options provide still too few occurrences for statistical analysis. The approach that relies on the syntactic labels adopted in the AGLDT, on the



other hand, yields numerically relevant results but leaves too much room for ambiguity. As said, the label ATR conflates several constructions—including couplets of name–adjective or noun and main verb of a relative clause—that, although functionally similar, are simply too different from the standpoint of stylistic analysis.

Readers that are interested in finding parallels for the phrase τέκνον Λαρτίου would most likely adopt an abstract representation of the structure as “noun + genitive noun.” This solution, however, would scale poorly at a level of a systematic survey of the whole work of Sophocles and is also rather impractical. The morphology of a noun is described by three features (number, gender, and case) and the choice to isolate just one of the possible values (the genitive) from one of the features (the case) for the dependent word, though based on a sensible assumption, might be rather arbitrary. The risk of subjectivity is amplified by the fact that a similar work of the selection of relevant features for head and dependent must then be extended to all the other parts of speech, some of which (like the verbs) are morphologically quite complex. Also, whenever head and dependent are both in the genitive, as in πατὴρ Ὀϊνέως, [31] this representation would not even be enough to distinguish between nouns used attributively and genitives of possession or origin with nouns. Those cases would still require manual disambiguation.

The use of couplets formed by the part of speech tag (POS) of the head and the POS of the dependent strikes a good compromise between stylistic meaningfulness, quantitative relevance, and scalability to the whole corpus. In fact, the data can be downloaded from the treebank as is, and no further pre-processing—apart from the identification of the speakers—or arbitrary selection of features is needed in order to extract those constructions.

The AGDT guidelines define 11 POS that are reasonably fine-grained to describe the morphological class of the words: article, noun, adjective, pronoun, verb, adverb, preposition, conjunction, numeral, interjection, and punctuation. Since punctuation is the product of the editorial work of modern editors, we will exclude it while we keep all the other ten classes.

The choice of POS as the only relevant feature for head and dependent is not exempt from problems. Some POS are, if not as polysemous as the syntactic labels, still ambiguous. As the POS tagging is based on the lemma, some words, such as the different forms of ὄδε, τις, or αὐτός are (or should be) consistently tagged as pronouns, even when they are used adjectivally. Moreover, the same class of “pronouns” holds together a conglomerate of words with quite different syntactic functions, such as personal, relative, and demonstrative

pronouns—a classification that might obscure some important nuances, as we will have the opportunity to see. A more fine-grained classification that accounts for these differences, however, would again require a considerable amount of manual disambiguation. Finally, some couplets, like the noun-noun unit discussed above, are also ambiguous as they conflate different constructions, such as nouns used as attributes and nouns with genitives. All this considered, however, the ten parts of speech make for a reasonable starting point to verify the methods and obtain preliminary results. It will be the critic's job, then, to investigate cases of potential ambiguity and bring about the nuances that are obliterated in the bird's-eye view. [32]

In the five tragedies, 105 couplets are attested with a skewed distribution ranging from the 4443 occurrences of the verb-adverb structure to the single instance of 15 couplets. The choice of the threshold after which a structure may be considered “frequent” is somewhat arbitrary. Like Burrow's, the present study is based upon the 30 most frequent units, which include all the syntactic couplets that occur at least 100 times.

The full list of the POS-POS syntactic arcs that were considered in this work is reported in Appendix A.1.

## 2.3 What Characters?

Defining the list of the characters that meet a minimum required number of words raises similar problems. As shown in Table 1, our five tragedies of Sophocles vary sensibly in length as well as in the number of speaking characters. The distribution of the words is obviously very uneven between the different roles, with the protagonist and the choruses dominating the scene in each play; for some characters, such as Euridice in the *Antigone* or the old leader of the Thebans in the prologue of the *Oedipus Rex*, the number of attributed treebank tokens is as low as 50 and 70 respectively.

Once again, the decision of setting a threshold may be arbitrary. By looking at the list of characters and their total number of tokens (excluding punctuation marks and reconstructed nodes), 700 tokens prove a suitable limit because this threshold allows us to include between four and six characters per play and to maintain some interesting figures like the Guard of the *Antigone* or Clytemnestra in *Electra*.

The most difficult decision, however, is whether to include the choruses in the list of characters. The chorus is not only a fundamental component of the tragic performance but

its involvement in the dramatic action, as well as its characterization, is also a crucial aspect of the art of an ancient dramatist, an area in which Sophocles excelled. [33] At the same time, however, it is clear to every reader that the language of a *stasimon* is very different from that of dialogue. The risk implied here is that the statistical signal of a difference between the language of choral and iambic parts is so strong that it will obfuscate all other contrasts. The choruses, however—or better, the chorus leaders—also take part in the iambic dialogues, while actors often engage in lyric exchanges with choruses too. Given the current state of both the treebank and the digital editions, none of which provides metrical annotation, it would be hard to disentangle those sections. For these reasons and at this current stage in our work, we decided to treat the choruses as normal characters. Our preliminary investigation will allow us to verify whether the distribution of the constructions points to a differentiation between the language of the choruses and actors that is stronger than that between character and character and chorus and chorus.

The list of characters is reported in Appendix A.2. It is particularly notable that for one tragedy (*Electra*) our list includes all the main speaking characters with the sole exception of Aegisthus.

### 3. Patterns of Correlation

#### 3.1 Pronouns and Prepositions by *Electra* and *Oedipus*

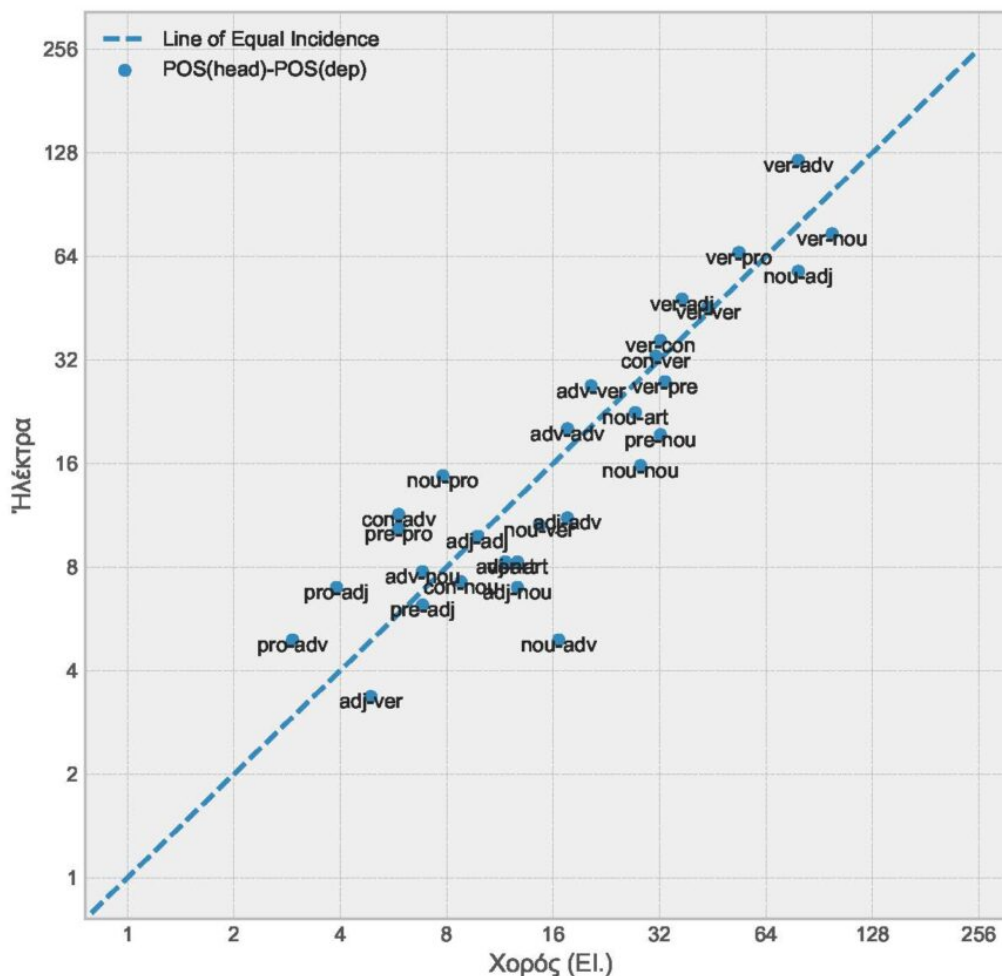


Figure 2. Correlation between Electra and the Chorus of *Electra*. Correlation coefficient: 0.899. Occurrences x 1000 tokens.

Our investigation starts with one example of a comparison that brings us back to the question raised at the end of Section 2: How different is the syntax of a character from that of a chorus? Figure 2, which follows the model of Figure 1 closely, illustrates the correlation between the 30 most frequent syntactic constructions in the five Sophoclean tragedies of the AGDT as used by Electra and the Chorus of the same tragedy. With a coefficient of 0.899, the correlation is low but not the lowest for Electra, not even among the characters of the same play. [34]

In the upper part of the chart, we see that some constructions that should intuitively be extremely common—like noun-noun, verb-adverb, or noun-adjective—are, in fact, rather polarized towards the one or the other.

As with pronouns, “adverb” is a category that holds different words under its roof; discursive particles (like δέ, μέν, γάρ or γε) are classified as adverbs, as well as the negative conjunctions (οὐ, οὐδέ, μηδέ) together with the forms in -ως that are derived from the adjective. With 73 negative particles out of her 471 adverbs—and with οὐ the most frequent lemma with 50 occurrences, followed by γάρ (49) and δέ (40)—Electra seems to resort to negation more often than the Chorus, which joins a verb with οὐ (6 cases) less frequently than with γάρ (12 cases). However, Electra’s numbers for the negative adverbials are in line with that of the other characters and do not seem to point to a significant role of negative forms in her idiolect.

As for the verb–noun couplet, we may observe that while Electra tends to use almost the same number of nouns and pronouns as arguments of the verbs (288 vs 254), the Chorus resorts to the former more frequently (101 vs 55). Most of Electra’s pronouns that are constructed with verbs are second- and first-person pronouns—a point to which we will return. The fact that in the same tragedy, the Paedagogus displays a ratio between the two types of verbal constructions that tends even more dramatically towards verb–noun than verb–pronoun (91 vs 35), as do Lichas in the *Women of Trachis* (75 vs 34) and the Guard in the *Antigone* (85 vs 36), is not surprising. What these three characters have in common is the fact that a very long stretch of their role is occupied with lengthy narratives, which report off-stage events in a messenger-like fashion. Long tales about off-stage actions, where the attention shifts from the *hic et nunc* on the scene to absent places and persons, entail a predictable drop in the use of pronouns—especially personal and deictic pronouns.

In light of these facts, it is tempting to conclude that the distribution of verbal arguments is entirely coherent with a dichotomy between characters, who tend to engage in dialogue, and choruses, whose long *stasima* are delivered to an empty stage. Those characters, like the Paedagogus, who play the role of messengers and whose long narratives mark a suspension in the flow of dialogues that is similar to what happens with choral odes, follow the same trend as the Choruses. Though the observation is valid, the situation is in fact more complex. If some main characters like Antigone or Oedipus display a more or less 1:1 ratio between verb–pronoun and verb–noun constructions that is comparable to Electra’s, the same does not hold true for others who seem to side with the chorus in their preference for nouns.

With 150 nominal arguments of verbs vs 85 pronouns, one such character is Ajax, whose strong penchant for “monologic” expression has been stressed by scholars like Schadewaldt (1926) or Di Benedetto (1988). With a distribution of 182 nouns and 105 pronouns

constructed with a verb, Creon (*Antigone*) is another one that clearly belongs to this group. His tendency to generalize, to rely on aphorisms, or speak of (often political and social) abstract entities is well reflected in the lexicon: the most frequent nouns that show up in his part as arguments or satellites of verbs are πόλις (12x), [35] ἀνὴρ (10x), [36] παῖς (6x), and νόμος (5x). But, ultimately, this aspect of Creon's language is clear in his preference for nouns as verbal arguments, which sets him in a very apt contrast with Antigone.

In the first scene of the second episode (see lines 450–496, in particular) and their most famous confrontation, we can argue that that syntactic feature is—together with the semantics of such loaded words as νόμοι, κήρυγμα, or νόμιμα—one of the linguistic elements that is widening the gulf between the two characters. In the 20 iambic trimeters of her speech (ll. 450–470), Antigone utters 7 first- and second-person personal pronouns, including one in her famous *incipit*: οὐ γὰρ τί μοι Ζεὺς ἦν ὁ κηρύξας τάδε, “for it was not Zeus that had published me that edict,” (line 450 – transl. Jebb). The pronouns, especially of first-person, become increasingly frequent towards the center of the speech:

460 θανουμένη γὰρ ἐξήδη, τί δ' οὐ;  
 κεί μὴ σὺ προῦκήρυξας. εἰ δὲ τοῦ χρόνου  
 πρόσθεν θανοῦμαι, κέρδος αὐτ' ἐγὼ λέγω.  
 ὅστις γὰρ ἐν πολλοῖσιν ὡς ἐγὼ κακοῖς  
 ζῆ, πῶς ὄδ' οὐχὶ κατθανῶν κέρδος φέρει;  
 465 οὕτως ἔμοιγε τοῦδε τοῦ μόρου τυχεῖν  
 παρ' οὐδὲν ἄλγος.

I knew that I would die, of course I knew, even if you had made no proclamation. But if I die before my time, I account that gain. For does not whoever lives among many troubles, as I do, gain by death? So it is in no way painful for me to meet with this death.

Sophocles, *Antigone* 460–465 (transl. Lloyd-Jones)

In comparison, only one ἐγὼ is found in Creon's answering speech (ἦ νῦν ἐγὼ μὲν οὐκ ἀνὴρ, αὕτη δ' ἀνὴρ, “indeed, now I am no man, but she is a man,” line 484, transl. Lloyd-Jones), while the new king refers to Antigone only with the third-person pronoun αὐτή (also lines 480, 488). Nouns, on the other hands, are particularly abundant in his aphorisms, as is exemplified in the following passage:

473 ἄλλ' ἴσθι τοι τὰ σκλήρ' ἄγαν φρονήματα  
πίπτειν μάλιστα, καὶ τὸν ἐγκρατέστατον  
475 σίδηρον ὀπτὸν ἐκ πυρὸς περισκελῆ  
θραυσθέντα καὶ ραγέντα πλεῖστ' ἂν εἰσίδοις:  
σμικρῶ χαλινῶ δ' οἶδα τοὺς θυμουμένους  
ἵππους καταρτυθέντας: οὐ γὰρ ἐκπέλει  
φρονεῖν μέγ' ὅστις δοῦλός ἐστι τῶν πέλας.

Why, know that over-stubborn wills are the most apt to fall, and the toughest iron, baked in the fire till it is hard, is most often, you will see, cracked and shattered! I know that spirited horses are controlled by a small bridle; for pride is impossible for anyone who is another's slave.

Sophocles, *Antigone* 473–479 (transl. Lloyd-Jones)

Returning to Figure 2 and the *Electra*, an interesting couplet that is easy to overlook is the one formed by prepositions governing pronouns. That particular construction is well placed in *Electra*'s side, although, in the comparison with the Chorus of the same tragedy (where the per-1000 occurrences are 10.34 vs 5.85), the distance is less pronounced than it would have been had we selected the Chorus of *Antigone* (1.98 per 1000 tokens) as her counterpart.

As we said, pronoun is a broad category that groups together different words. If one considers the lemmata of the pronouns governed by a preposition in *Electra*'s part, we see that the first- (8) and second-person (13) [37] pronouns are the most represented. This is not always the case. Iocaste, who at 10.05 ranks close to *Electra*'s 10.34 for the number of per-1000 occurrences of the construction, perhaps as an effect of her much lower number of total tokens (796 vs the 3870 of *Electra*), has a few second-person pronouns (3) but no first persons; the remaining five occurrences are divided between the relative ὅς (2), τις (2) and ὅδε (1).

Character	Tot	Freq. x 1000
<i>Electra</i>	21	5.43
Deianeira	9	3.86
Tecmessa	5	3.97
Chorus (OT)	5	3.05
Oedipus	11	2.60

Table 2. Frequency of construction preposition-1st/2nd pers. pronoun for characters with at least 5 total occurrences of the unit.

The numbers of occurrences of prepositions governing a personal pronoun of first- and second-person are given in Table 2—this is limited to characters that have at least 5 occurrences of personal pronouns in that construction. The Thebans of *Oedipus Rex* are the only chorus to figure here, while the others tend rather to shy away not just from the construction with personal pronouns but from the couplet preposition-pronoun in general. Although other characters often employ the structure, which seems after all to belong to a fairly common register of expression for a language like Ancient Greek, it is clear that this construction features rather prominently in the language of Electra.

In fact, one can almost perfectly trace the progress of the dramatic action of the tragedy by simply reading the sentences where the main character utters prepositions that govern a first- and, in particular, a second-person pronoun. In the first half of the play, as Electra engages in quarrels with Chrysothemis and Clytemnestra, she resorts to the second-person to stress the pain that she is suffering and her isolation from the rest of the family. When Chrysothemis asks why she desires death so much, the answer is, “to fly as far away as possible from you” (ὅπως ἀφ’ ὑμῶν ὡς προσωτάτω φύγω, line 391). As she later retorts to her mother, “torments that come from you” are what she lives with (ἔκ τε σοῦ κακοῖς / πολλοῖς ἀεὶ ξυνοῦσα). [38]

The report of Orestes’ death will not only shift the center of Electra’s attention, it will also change how she employs this construction. At first, she will turn to her sister again, hoping to win her over for a desperate plot—“but now that he is no more, I look to you,” εἰς σὲ δὴ βλέπω, line 954—only to conclude that, “there is no help in her” (σοὶ γὰρ ὠφέλησις οὐκ ἔνι, line 1031).

It is, however, in the famous speech to the urn, when she addresses what she believes to be the ashes of her brother, [39] that the highest concentration of the structure is attested. Her impassioned speech mentions the care that she had for Orestes (τροφῆς [...] ἀμφὶ σοί, line 1143–1144) and how all love and hope have vanished with his death (θανόντι σὺν σοί, line 1150); later, she expresses the wish to go live below the earth with him (ὡς σὺν σοὶ κάτω / ναίω, line 1166–1167), just like the two shared their lot when he was alive (ξὺν σοὶ μετεῖχον τῶν ἴσων, line 1168). After this moving lamentation, Electra will resort to this construction again only twice (1303, 1411).



Although Oedipus ranks last in the per-1000 occurrences reported in Table 2, an investigation of his 11 sentences where the structure is attested proves no less rewarding than those of Electra. For whereas Electra tended to construct prepositions with the second-person pronouns, Oedipus is, not surprisingly, the master of the construction with ἐγώ, which covers 8 of his total 11 attestations.

What is even more striking is that Oedipus seems to transition gradually from one construction to the other as the plot advances and the center of the drama moves from the main character's place within the Theban society to the question of his identity. All three instances of a preposition with a second-person pronoun are attested in the first part of the play, where the question at hand is how to find Laius' murderer and, or so the king thinks, how to fend off a traitorous plot against him. [40] But as the play progresses, the use of the construction shifts toward the first-person as the problem of Oedipus' true identity (his true "me") becomes more and more dominant. Oedipus' distress erupts in the question, "Oh Zeus, what do you want to do about me?" (ὦ Ζεῦ, τί μου δρᾶσαι βεβούλευσαι πέρι, line 738). Again, if the only witness confirms that Laius was killed by a single person, that would tilt the scales against him (εἰς ἐμέ, line 847). It is to none other than against himself that he unknowingly threw his curse (καὶ τάδ' οὔτις ἄλλος ἦν ἢ ἔγωγε 'π' ἐμαυτῷ τάσδ' ἀρὰς ὁ προστιθείς, 819–820). And when the truth comes to light, Oedipus will be forced to conclude that no other man except him will be able to bear his woes (τάμα γὰρ κακὰ / οὔδεις οἴός τε πλὴν ἐμοῦ φέρειν βροτῶν, line 1414–1415).

Though we are probably far from identifying the "syntax" that is peculiar to each character, an exploration guided by data like those visualized in Figure 2 has allowed us to isolate a number of syntactical constructions whose distribution, just like for the lexical facts discussed in the previous sections, seems both to be skewed to some characters and to serve some neatly-identifiable dramatic functions: Creon's preference for nominal arguments of verbs, and Electra's and Oedipus' use of the personal pronouns with prepositions, highlight some important themes within the structure of the scenes where the constructions are found. Even more interestingly, at least for Electra and Oedipus, the use of these constructions seems adapted to the evolving dramatic tension within the plays.

### 3.2 Lowest and Highest Correlations

In the previous section, we started from a single example to familiarize ourselves with the patterns of distributions of the POS-POS dependency structures between the Sophoclean

characters. Still following Burrows' example, the next step is to collect and compare the complete set of the highest and lowest correlation coefficients for each character. In the present paper, we will discuss the data only briefly; for the most part, we will not do more than confront the reader with a list of open problems.

Appendix A.3 presents the three highest and the three lowest correlation scores for each character on our list. Even with a cursory reading, several elements stand out.

Looking at the top associations, one can identify a series of familiar patterns. The second highest coefficient of all is the correlation between Antigone and Electra (0.987). As is well known, there is much that these two figures have in common. Both are confronted with a sister (and a Chorus) who is sympathetic, and yet, to the character's taste, way too accommodating to those that are in power. Both are pitted against a harsh and embittered relative. Finally, each of their parts is made of lengthy speeches, edgy dialogues, and moving lyric exchanges.

Slightly higher and, thus, the highest of all the coefficients is the correlation between Teucros and Tecmessa in *Ajax*. This parallelism is certainly interesting, as these two figures share an intense attachment to the main hero, which is also strengthened by the fact that the arrival of the former marks the substantial disappearance of the latter from the play; Tecmessa will not speak again in the tragedy after Teucros' entrance at line 974. However, the high coefficient may result from the fact that these two characters appear in the only tragedy that was annotated by a different annotator; we cannot exclude that some differences in the application of the annotation guidelines or in the way of interpreting the syntax of Sophocles between the two annotators are setting the characters of *Ajax* apart from the others. Indeed, by looking at the correlation scores, *Ajax's* characters seem to cluster together to a degree that provokes suspicion.



What then do Chrysothemis and Creon have in common? And what makes the Chorus of *Antigone* seem so characteristically “choral?” One hypothesis is that both Creon and Chrysothemis’ parts are rather limited in terms of forms of expressions; Creon and Chrysothemis do not engage in lyric dialogue, although they are by no means the only ones in the list that do not sing. On the other hand, *Antigone* is the tragedy with the highest number of *stasima* of the five, so that the language of that Chorus is more shaped by the expressive features of choral odes than that of the rest of the choruses. Such questions, however, must remain unresolved for the present.

There are some notable exceptions to the clustering rule that we just discussed. Ajax’s patterns of associations are remarkable, since his highest correlation is with a Chorus (that of *Oedipus Rex*) and his lowest is not with the ever-present Chorus of *Antigone* but with the less “lyric” Creon (*Oedipus Rex*) and Chrysothemis. On the one hand, there may be some peculiar differences in the annotation style between *Ajax* and the other tragedies, as it was noted above, again at play. On the other hand, however, Heracles shows a remarkably similar trend. The fact that these two heroes are associated and display some traits that accentuate the lyric dimension of their expression—an aspect in the tragic portrait of these two traditionally indomitable heroes that was well commented by Di Benedetto (1988)—is perhaps not so surprising.

The exception that is easier to explain—regarding the clustering of choruses with choruses and actors with actors—is the one that appears to be consistent throughout many correlations: the Paedagogus of *Electra*. We have already seen that in the peculiar feature of the choice of verbal arguments and in particular in the opposition between nouns and pronouns, the Paedagogus sides with the choruses, together with other messenger-like characters. However, the fact that Orestes’ old companion figures in the top three list of correlated characters for all choruses hints at something deeper than the old servant’s role as the messenger in the second scene of the second episode. Neither Lichas nor the Guard feature anywhere near the Paedagogus in these rankings.

If we compare the distribution of the syntactic constructions of the Paedagogus and the Chorus of *Electra* using the usual methodology (Figure 3), we notice that there is at least one important couplet that appears very close to the line of equal incidence: the noun-adjective structure.

The importance of noun-epithet couplets in the language of the Paedagogus is once again clear to everyone who reads the first lines of his (fake) report on Orestes’ death:

κείνος γὰρ ἔλθων εἰς τὸ κλεινὸν Ἑλλάδος  
 πρόσχημ' ἀγῶνος Δελφικῶν ἄθλων χάριν,  
 ὅτ' ἦσθετ' ἀνδρὸς ὀρθίων κηρυγμάτων  
 δρόμον προκηρύξαντος, οὗ πρώτη κρίσις,  
 685 εἰσῆλθε λαμπρός, πᾶσι τοῖς ἐκεῖ σέβας:  
 δρόμου δ' ἰσώσας τῇ φύσει τὰ τέρματα  
 νίκης ἔχων ἐξῆλθε πάντιμον γέρας.

He came to the pride of Greece, the contest, for the sake of Delphic prizes, and when he heard the loud pronouncement of the man who proclaimed the race, which is decided first, he entered the course a brilliant figure, admired by all. He made the results of the race correspond with his appearance, and emerged holding the greatly honoured prize of victory

Sophocles, *Electra* 681–687 (transl. Lloyd-Jones)

The style of this beginning is certainly very much in tune with the tale of the glorious but tragic death of a young athlete during a chariot race, which immediately evokes the precedents of epic and epinikian poetry. However, the long narrative of lines 680–764 is not the only passage where the Paedagogus displays a high number of “exornative epithets” coupled with nouns: the very beginning of the tragedy, where the Paedagogus addresses Orestes and describes the landscape of the Argolid for him, makes a no less prominent and abundant exhibition of noun–epithet couplets:

ὦ τοῦ στρατηγήσαντος ἐν Τροίᾳ ποτὲ  
 Ἀγαμέμνονος παῖ, νῦν ἐκεῖν' ἔξεστί σοι  
 παρόντι λεύσσειν, ὧν πρόθυμος ἦσθ' ἀεί.  
 τὸ γὰρ παλαιὸν Ἄργος οὐπόθεις τόδε,  
 5 τῆς οἰστροπλήγος ἄλσος Ἰνάχου κόρης:  
 αὕτη δ', Ὀρέστα, τοῦ λυκοκτόνου θεοῦ  
 ἀγορὰ Λύκειος: οὐξ ἀριστερᾶς δ' ὄδε  
 Ἦρας ὁ κλεινὸς ναός: οἱ δ' ἰκάνομεν,  
 φάσκειν Μυκίνας τὰς πολυχρύσους ὄρᾶν  
 10 πολύφθορόν τε δῶμα Πελοπιδῶν τόδε

Son of Agamemnon, who once led the army before Troy, now you can gaze with your own eyes on what you have always longed to see! This is ancient Argos for which you used to long, the precinct of the daughter of Inachus whom the gadfly

stung; and this, Orestes, is the Lycean marketplace of the wolf-killing god; this to the left is the famous temple of Hera; and at the palace where we have arrived, you may say that we reached Mycenae, rich in gold, and the house of the sons of Pelops, rich in disasters

Sophocles, *Electra* 1–10 (transl. Lloyd-Jones)

### 3.3 Some Conclusions (and Future Plans)

The data presented in the previous two sections have confronted us with a series of syntactic constructions whose patterns of distribution among characters seem highly revealing in light of the thematic structure and tensions in the play. Some—like the correlation patterns of Heracles and Ajax, the Paedagogus and the Choruses, or Electra and Antigone—suggest some possible directions in the investigation of Sophocles' dramatic technique.

The significant role played by the noun-epithet constructions in the two most important scenes where the Paedagogus is involved conveys, if not a sense of individuality, certainly a sense of stylistic unity. This feature is, as we saw, carried over from one scene to the other—from the prologue to the second episode. His language helps define the old man's function as Orestes' mentor, in view of the vengeance plan that the two must plot, and as the herald of the glorious and lamentable fate of the young son of Agamemnon. The fact that this report is entirely fictional, notwithstanding its elaborate language, strikes a sinister note—a note that is perhaps in tune with the general somber atmosphere of the vengeance and of the play's finale. Be it as it may, the syntax is certainly an important element in this elaborated dramatic construction.

Up until now, we have proceeded mainly by isolating prominent syntactic features, such as the verb-pronoun or noun-adjective couplets. The whole picture, however, together with the many unanswered questions that we posed, is still in want of an explanation.

## References

Abeillé, Anne, ed. 2003. *Treebanks. Building and Using Parsed Corpora*. Dordrecht.

Bamman, David, Marco Passarotti, Gregory Crane, and Savina Raynaud. 2007. "A Collaborative Model of Treebank Development." *Proceedings of the Sixth Workshop on Treebanks and Linguistic Theories (TLT 6)* 1–6. Bergen

Bamman, David, Francesco Mambrini, and Gregory Crane. 2009. "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank." *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT 9)*, eds. M. Passarotti, A. Przepiorkowski, S. Raynaud, and F. Van Eynde, 5–15. Milan.

Burrows, John F. 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford.

Denniston, John D. 1934. *Greek Particles*. Second edition revised by K. Dover, 1954. Oxford.

Di Benedetto, Vincenzo. *Euripide. Teatro e società*. Torino.

———. 1988. *Sofocle*. Firenze.

Du , Casey, Blackwell, Christophe, and Smith, D. Neel. 2012. "A Gentle Introduction to CTS & CITE URNs." <http://www.homermultitext.org/hmt-doc/guides/urn-gentle-intro.html>.

Easterling, Patricia E. 1977. "Character in Sophocles." *Greece & Rome* 24(2):121–129.

Finglass, Patrick J., ed. 2007. *Sophocles: Electra*. Cambridge.

Gardiner, Cynthia P. 1987. *The Sophoclean Chorus*. Iowa City.

Garvie, Alex. 1998. *Sophocles. Ajax*. Warminster.

Goldhill, Simon. "Modern Critical Approaches to Greek Tragedy." *The Cambridge Companion to Greek Tragedy*, ed. P. Easterling, 324–347. Cambridge.

Gries, Stefan Th. 2009. *Statistics for Linguists with R*. 2nd edition. Berlin.

- Hoover, David L. 2008. "Quantitative Analysis and Literary Studies." *A Companion to Digital Literary Studies*, eds R. Siemens and S. Schreibman. Oxford.  
<http://www.digitalhumanities.org/companionDLS/>.
- Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. Chicago.
- Kestemont, Mike. 2014. "Function Words in Authorship Attribution. From Black Magic to Theory?" Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL) 55–69. Gothenburg.
- Knox, Bernard. 1964. *The Heroic Temper: Studies in Sophoclean Tragedy*. Berkeley.
- Leech, Geoffrey. "Adding Linguistic Annotation." *Developing Linguistic Corpora: A Guide to Good Practice*, ed. M. Wynne, 17–29. Oxford.
- Lloyd-Jones, Hugh. 1972. "Tycho von Wilamowitz-Moellendorff on the Dramatic Technique of Sophocles." *The Classical Quarterly* 22(2):214–228.
- Loroux, Nicole. 1986. "La main d'Antigone." *Métis* 1:165–196.
- Mambrini, Francesco, and Marco Passarotti. 2016. "Subject-Verb Agreement with Coordinated Subjects in Ancient Greek. A Treebank-Based Study." *Journal of Greek Linguistics* 16:87–116.
- Rosenbloom, David. 2001. "Ajax is *megas*. Is that all we can say?" *Prudentia* 33:109–130.
- Schadewaldt, Wolfgang. 1926. *Monolog und Selbstgespräch. Untersuchungen zur Formgeschichte der griechischen Tragödie*. Berlin.
- Storr, Francis, ed. 1912. *Sophocles, With an English Translation*. London and New York.
- Thumiger, Chiara. 2007. *Hidden paths. Self and characterization in Greek tragedy. Euripides' Bacchae*. BICS Supplements, 99. London.



von Wilamowitz-Moellendorf, Tycho. 1917. *Die dramatische Technik des Sophocles*. Berlin.

Winnington-Ingram, Reginald P. 1980. *Sophocles. An Interpretation*. Oxford.

## Appendix

### A.1 The top thirty POS-POS syntactic relations in the AGDT's Sophocles

The raw frequency of the construction is reported in parenthesis:

verb-adverb (4443); verb-noun (3593); noun-adjective (2585); verb-pron (2464); verb-verb (2297); verb-adjective (1874); conjunction-verb (1514); verb-conjunction (1332); verb-preposition (1172); adverb-verb (1069); noun-article (1005); noun-noun (952); preposition-noun (947); adverb-adverb (815); noun-pron (693); noun-verb (444); adjective-adverb (394); adverb-noun (384); conjunction-noun (373); adjective-adjective (367); adjective-article (346); noun-adverb (313); adjective-noun (311); conjunction-adverb (304); verb-article (293); preposition-adjective (256); preposition-pron (253); pron-adjective (220); pron-adverb (203); adjective-verb (203).

### A.2 Characters in Sophocles that speak more than 700 treebank tokens

The first number in parentheses reports the raw total; the second the total normalized per 1000 tokens.

*Ajax*

Ajax (1629; 206.18), Chorus (1494; 189.09), Teucros (1319; 166.94), Tecmessa (1259; 159.35).

*Electra*

Electra (3870; 441.23), Chorus (1025; 116.86), Chrysothemis (1014; 115.61), Orestes (995; 113.44), Paedagogus (873; 99.53), Clytemestra (766; 87.33).

*Oedipus Rex*

Oedipus (4233; 452.10), Chorus (1642; 175.37), Creon (849; 90.68), Iocaste (796; 85.02).

*Antigone*

Creon (2100; 285.52), Chorus (1518; 206.39), Antigone (1231; 167.37), Guard (712, 96.80).

*Women of Trachis*

Deianeira (2330; 313.59), Heracles (1194; 160.70), Chorus (1042; 140.24), Hyllos (1071; 144.15), Lichas (719; 96.77).

### A.3 Full prospect of the correlation among characters

Stars (\*) mark characters from the same play. Characters that recur in more plays (and choruses) are marked with an underscore and an abbreviation for the play title: e.g., Creon\_OT = Creon in *Oedipus Rex*.

*Ajax*

Ajax:

TOP: Chorus\_OT (0.979); Tecmessa (0.974)\*; Hyllos (0.968).

LOW: Creon\_OT (0.863); Chorus\_Ant. (0.867); Chrysothemis (0.880).

## Teucros:

TOP: Tecmessa (0.989)\*; Ajax (0.965)\*, Electra (0.961).

LOW: Chorus\_Ant (0.796); Chorus\_Tr (0.897); Paedagogus (0.902).

## Tecmessa:

TOP: Teucros (0.989)\*; Ajax (0.974)\*; Hyllos (0.973).

LOW: Chorus\_Ant (0.806); Paedagogus (0.908); Chorus\_Tr (0.909).

## Chorus\_Aj:

TOP: Paedagogus (0.967); Ajax\* (0.964); Chorus\_Tr (0.957).

LOW: Creon\_OT (0.756); Chrysothemis (0.776); Clytaemestra (0.803).

*Electra*

## Electra:

TOP: Antigone (0.987); Oedipus (0.981); Orestes\* (0.977).

LOW: Chorus\_Ant. (0.730); Paedagogus\* (0.860); Chorus\_Tr. (0.864).

## Chorus\_El:

TOP: Chorus\_OT (0.968); Paedagogus\* (0.967); Ajax (0.958).

LOW: Creon\_OT (0.810); Clytaemestra\* (0.824); Chrysothemis\* (0.833).

## Chrysothemis:

TOP: Orestes\* (0.970); Clytaemestra\* (0.968); Oedipus (0.967).

LOW: Chorus\_Ant. (0.635); Chorus\_Aj (0.776); Paedagogus\* (0.787).

## Orestes:

TOP:Oedipus (0.983); Electra\* (0.977); Clytaemestra (0.973).

LOW:Chorus\_Ant. (0.692); Chorus\_Tr. (0.836); Chrous\_Aj. (0.847).

## Paedagogus:

TOP:Chorus\_Aj. (0.967); Chorus\_El.\* (0.967); Chorus\_Tr. (0.959).

LOW:Creon\_OT (0.760); Clytaemestra\* (0.786); Chrysothemis\* (0.787).

## Clytaemestra:

TOP:Oedipus (0.974); Orestes\* (0.973); Chrysothemis\* (0.968).

LOW:Chorus\_Ant. (0.631); Paedagogus\* (0.786); Chorus\_Tr. (0.795).

*Oedipus Rex*

## Oedipus:

TOP:Orestes (0.983); Electra (0.981); Antigone (0.978).

LOW:Chorus\_Ant. (0.662); Paedagogus (0.808); Chorus\_Aj. (0.815).

## Chorus\_OT:

TOP:Ajax (0.977); Chorus\_El. (0.968); Chorus\_Tr. (0.968).

LOW:Creon\_OT\* (0.861); Chrysothemis (0.865); Clytaemestra (0.877).

## Creon\_OT:

TOP:Oedipus (0.975); Clytaemestra (0.968); Chrysothemis (0.964).

LOW:Chorus\_Ant. (0.611); Chorus\_Aj. (0.756); Chorus\_Tr. (0.758).

Iocaste:

TOP:Orestes (0.972); Deianeira (0.971); Oedipus\* (0.969).

LOW:Chorus\_Ant. (0.726); Chorus\_Tr. (0.839); Chorus\_Aj. (0.851).

### *Antigone*

Creon\_Ant.

TOP:Deianeira (0.978); Teucros (0.960); Orestes (0.959).

LOW:Chorus\_Ant. (0.752); Chorus\_Tr. (0.864); Chorus\_Aj. (0.884).

Chorus\_Ant.:

TOP:Chorus\_Tr. (0.947); Paedagogus (0.935); Chorus\_Aj. (0.912).

LOW:Creon\_OT (0.611); Clytaemestra (0.631); Chrysothemis (0.635).

Antigone:

TOP:Electra (0.987); Oedipus (0.978); Deianeira (0.973).

LOW:Chorus\_Ant.\* (0.742); Chorus\_Aj. (0.855); Paedagogus (0.861).

Guard:

TOP:Lichas (0.974); Creon\_Ant.\* (0.953); Deianeira (0.949).

LOW:Chorus\_Ant. (0.768); Clytaemestra (0.873); Chorus\_Aj. (0.878).

### *Women of Trachis*

Deianeira:

TOP:Creon\_Ant. (0.978); Electra (0.975); Antigone (0.972).

LOW:Chorus\_Ant. (0.788); Chorus\_Aj. (0.896); Chorus\_Tr.\* (0.900).

Heracles:

TOP:Chorus\_OT (0.967); Chorus\_Tr.\* (0.965); Ajax (0.958).

LOW:Creon\_OT (0.853); Chorus\_Ant. (0.871); Chrysothemis (0.874).

Hyllos:

TOP:Tecmessa (0.973); Antigone (0.972); Deianeira\* (0.971).

LOW:Chorus\_Ant. (0.787); Paedagogus (0.899); Chorus\_Aj. (0.903).

Lichas:

TOP:Guard (0.974); Deianeira\* (0.965); Hyllos\* (0.963).

LOW:Chorus\_Ant. (0.804); Creon\_OT (0.883); Clytaemestra (0.885).

## Footnotes

[ [back](#) ] 1. Burrows 1987:1.

[ [back](#) ] 2. For Ancient Greek literature, it is sufficient to quote Denniston (1934).

[ [back](#) ] 3. In stylometry, frequent words are often used to tackle concrete problems of authorship attribution (Kestemont 2014). However, linguistic and stylistic interpretations of their patterns of use that are comparable to those advanced by Burrows (1987) are, unfortunately, rare.

[ [back](#) ] 4. Jockers 2013.

[ [back](#) ] 5. Aristotle, *Poetics* 1450a12–13.

[ [back](#) ] 6. Thumiger 2007:19.

[ [back](#) ] 7. For a brief overview of the history of modern critical approaches to Greek tragedy see the synthesis of Goldhill (1997).

[ [back](#) ] 8. For an assessment of Tycho Wilamowitz–Moellendorf’s book, and for a discussion of the instances in which the author has pushed his thesis too far, see Lloyd–Jones 1972 (with notes from an unpublished essay by E. Fraenkel). For Euripides, Di Benedetto (1971) has analyzed the instances of transition, often quite abrupt, from monodies, or lyrical dialogues to *rheseis*. Di Benedetto (1988:45–48 in particular) also discusses the original way in which Sophocles used the same technique of transition from the lyric to iambic meter in order to explore new dimensions in his characters.

[ [back](#) ] 9. Knox 1964.

[ [back](#) ] 10. Easterling 1977:127–128.

[ [back](#) ] 11. The importance of *mégas* in Ajax was noted most prominently by Winnington–Ingram (1980:22n35). Garvie, in his commentary (1998), placed much emphasis on the “greatness” of Ajax, and his interpretation has been debated—e.g., by Rosenbloom 2001. On the conflicting perspectives in the employment of the words, it is sufficient here to point how *mégas* is constantly used in the parodos to illuminate both the moral and social superiority of Ajax (e.g. lines 154, 169) and the malignant force of the slanders addressed to him (142, 173, 226); Tecmessa (205) uses the adjective to highlight the contrast between the heroic greatness that was and the current misery. Finally, the sense of the adjective is diminished to one of physical hugeness and bodily strength in the veiled threats of Menelaus (1076–1077) and Agamemnon (1253–1254), which remind Teucros that the smallest means are sufficient to those in power to rein in the ambitions of the “big men.”

[ [back](#) ] 12. Hoover 2008.

[ [back](#) ] 13. On the other hand, the log scale allows for a much clearer visualization of the less frequent words. For this reason, it will be adopted in our experiments as well. The chart is a personal reconstruction of the one published by Burrows (1987:83), based on the numbers documented in the Appendixes of the book.

[ [back](#) ] 14. Burrows 1987:83.

[ [back](#) ] 15. Literary scholars and linguists can find a very good introduction to correlation in Gries (2013).

[ [back](#) ] 16. Within the novel, the highest correlation for Elizabeth Bennet is with her father (0.938). The highest correlation of all characters in Jane Austen is that between Elizabeth Bennet and Emma Woodhouse (0.973); the lowest is that between Mr. Collins (*Pride and Prejudice*) and Harriet Smith (*Emma*) at 0.408 (Burrows 1987:121).

[ [back](#) ] 17. Though the field has obviously evolved, the essays in Abeillé (2003) and the discussion of the principles of linguistic annotation by Leech (2004) still provide an excellent introduction to the subject.

[ [back](#) ] 18. [https://perseusdl.github.io/treebank\\_data/](https://perseusdl.github.io/treebank_data/).

[ [back](#) ] 19. See Section 2.1, with some examples.

[ [back](#) ] 20. The current annotation guidelines (v. 2.1) can be found on the project's website. Note, however, that the annotation of Sophocles conformed to a previous and slightly different version of the guidelines, which were directly derived from those of the Latin treebank (Bamman et al. 2007).

[ [back](#) ] 21. In other cases, such as with coordination, apposition, and even (at least according to some specialists with whom I have discussed this) participles and relative clauses, the level of abstraction is higher and the distance with the notions of ordinary grammar is wider. Though it is convenient to appeal to the notions of school grammars in short introductions like this one, readers and users of the AGLDT should never forget that the annotation is based on a formalized meta-language. To make this point clear with one example: although the syntactic label ATR is indeed used to tag (among other relations) the syntactic arc that connects a noun to its attribute, readers should refrain from making the immediate equivalence between the tag and the ordinary notion of “attribute,” even if this equivalence seems to be implied by the name of the label itself. In the AGLDT, ATR means primarily ATR—i.e., a set of relations that are defined as such in the formalized language.

[ [back](#) ] 22. Bamman et al. 2009.

[ [back](#) ] 23. *Ajax* was curated by D. L. Libatique, who was at the time a student at the College of Holy Cross, Worcester, MA. The other four tragedies were annotated by



me.

[ [back](#) ] 24. See for instance *Electra* 533, where my treebank has the transmitted ὄτ' in place of the ὄς printed by Storr; or *Electra* 742, where I write ὠρθοῦθ' with the manuscripts in place of the conjecture ὠρμᾶθ' that was printed by Storr.

[ [back](#) ] 25. For one example of a treebank-based analysis of a syntactical construction in the AGDT (number agreement with coordinated subjects), see Mambrini and Passarotti (2016).

[ [back](#) ] 26. The CTS-URN attached to the cited word is: <urn:cts:greekLit:tlg0011.tlg004:1>. For a “gentle introduction” to the CTS-URN notation see Dué et al. (2012).

[ [back](#) ] 27. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sp.html>.

[ [back](#) ] 28. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-speaker.html>.

[ [back](#) ] 29. An example of a Python script that implements those steps can be found here:

[https://github.com/francescomambrini/gAGDT/blob/master/scr/extract\\_speaker.py](https://github.com/francescomambrini/gAGDT/blob/master/scr/extract_speaker.py).

[ [back](#) ] 30. Burrows 1987:100.

[ [back](#) ] 31. *Women of Trachis*, 6.

[ [back](#) ] 32. Another important aspect that is lost in the current survey is word order. Our POS-POS couplets are always ordered from the head to the dependent, regardless of whether, in the text, the governing word precedes or follows its dependent. A study that also factors the effect of word order will be left for future work.

[ [back](#) ] 33. The point has been vigorously made in particular by Gardiner (1987).

[ [back](#) ] 34. The correlation with the syntax of Paedagogus (0.860) is *Electra*'s lowest among the characters of the play. See below Section 3.2 and the data reported in Appendix A.3.

[ [back](#) ] 35. See, for instance, εὐθύνων πόλιν (“guiding the city”) *Antigone* 178 in the context of a statement about the duties of the governor to do what is right. Other

instances involve sentences where Creon presents his program of government:

τοιιοῖσδ' ἐγὼ νόμοισι τῆδ' αὖξω πόλιν (“these are the rules by which I make our city great,” 191); ψευδῆ γ' ἐμαυτὸν οὐ καταστήσω πόλει (“I shall not show myself false to the city,” 657).

[ [back](#) ] 36. E.g., line 221–222: ἀλλ' ὑπ' ἐλπίδων ἄνδρας τὸ κέρδος πολλάκις διώλεσεν (“but hope has often caused the love of gain to ruin men”). Line 661–662: ἐν τοῖς γὰρ οἰκείοισιν ὅστις ἔστ' ἀνὴρ χρηστός, φανεῖται κἀν πόλει δίκαιος ὢν (“the man who acts rightly in family matters will be seen to be righteous in the city also”). In line 162, ἄνδρες is the vocative used by Creon to address the Chorus.

[ [back](#) ] 37. Note that the lemmata σύ and ἐγώ are assigned to both singular and plural pronouns; thus, σύ might include forms of ὑμεῖς as well.

[ [back](#) ] 38. See also: ὑπο σοῦ, line 553; and ἐκ σέθεν, line 579.

[ [back](#) ] 39. See Finglass (2007:443–446) for an introduction to the significance of this scene.

[ [back](#) ] 40. See lines 314, 370, 382.

## **VOLUME 20**