Nominal vs copular clauses in a diachronic corpus of Ancient Greek historians

A treebank-based analysis

Francesco Mambrini
Università Cattolica del Sacro Cuore, Milan
francesco.mambrini@unicatt.it

Abstract

We study the distribution of the nominal and copular construction of predicate nominals in a subset of authors from the Ancient Greek Dependency Treebank (AGDT). We concentrate on the texts of the historians Herodotus, Thucydides (both 5th century BCE) and Polybius (2nd century BCE). The data comprise a sample of 440 sentences (Hdt = 175, Thuc = 91, Pol = 174). We analyze the impact of four features that have been discussed in the literature and can be observed in the annotation of AGDT: (1) order of constituents, (2) part of speech of the subjects, (3) type of clause and (4) length of the clause. Furthermore, we test how the predictive power of these factors varies in time from Herodotus and Thucydides to Polybius with the help of a logistic-regression model. The analysis shows that, contrary to a simplistic opinion, the nominal construction does not drop into irrelevance in Hellenistic Greek. Moreover, an analysis of the distributions in the authors highlights a remarkable continuity in the usage patterns. Further work is needed to improve the predictive power of our logistic-regression model and to integrate more data in view of a more comprehensive quantitative diachronic study.

Keywords

Ancient Greek syntax – treebanks – nominal clauses – copula – Greek historians

1   Introduction

Syntactically annotated corpora known as treebanks are doubtlessly a very powerful tool for studying the syntax of a language. With the help of query software that are capable of extracting information from annotation, researchers can quickly obtain data about the distribution of different constructions and the co-occurring factors that correlate positively or negatively with them. However, how linguistic annotation can help research on syntactic change and on the history of ancient languages is a question that linguists have begun to explore only recently (Eckhoff, Luraghi, and Passarotti 2018). In this work, we would like to attempt to contribute to this discussion with a limited example and a preliminary investigation on a well-known syntactic phenomenon of Ancient Greek.

The nominal clauses found in some of the oldest surviving Indo-European texts are probably among the most debated topics in the syntax of ancient languages. This long-standing interest probably owes something to the centrality of equational and ascriptive sentences and, more generally, of the verb "to be" in the history of Western philosophy (see Moro 1997, 248–261, for an overview), or to the long-lasting influence of Latin maxims such as *omnia praeclara rara* ("all excellent things [are] rare") in European culture.

Perhaps one of the most intriguing aspects of the phenomenon is the pervasive presence of the clauses constructed without copula in some of the most archaic surviving texts, like the Homeric poems, the *Avesta* or the *Vedas*. In Ancient Greek, the nominal construction is attested in competition with the overt copula as early as in the 8th century BCE *Iliad* and *Odyssey*, and the omission of the copula eventually became non-grammatical in Greek. On account of those facts, it is tempting to conclude that, while both constructions can be ascribed to the Proto Indo-European syntax, the nominal sentence is the older of the two and survives in historical times mainly as a poetic archaism (Schwyzer 1950, 623–624). The diachronic dimension of the phenomenon has thus been crucial at least since the work of Delbrück (1900, 117–121) and Meillet (1906). However, the stress placed on the reconstruction of the Indo-European origins and prehistory of the two constructions has had the effect of concentrating the scholarly attention on the earliest documents; only sporadic attention has been paid to the later developments in the history of Greek.

In this study we focus on a set of authors, the historians Herodotus (ca. 485–424 BCE), Thucydides (ca. 460-after 404 BCE) and Polybius (ca. 199–120 BCE), that, for reasons that we see (Section 1.2), have been relatively overlooked. Moreover, instead of discussing one of the many general theories proposed on the subject, we make use of the corpus data to discuss the explicatory power of

a set of predictors whose influence on the choice of the construction is known or may be hypothesized. In particular, we see whether the explanatory force of some of them (which can be tested on the corpus as a whole) varies among the three authors that we consider.

The work is organized as follows: Section 1.1 introduces the constructions under consideration more formally; Section 1.2 provides a brief sketch of the main contributions and open problems. Section 2.1 describes the corpus and the authors that we will investigate, while Section 2.2 discusses the methodology of this study. An overview of the predictors that are expected to influence the choice between the two constructions is offered in Section 3. Section 4 discusses the regression model that we use to assess the influence of the different predictors and their interactions. Section 5 offers conclusions and plans for future work.

1.1   *Nominal vs verbal clauses in Ancient Greek*

In Ancient Greek, subjects and predicate nominals can be constructed in two different ways:
1.   they can be juxtaposed without a governing verb;
2.   a "copular" verb (analogous to English "to be") is used to connect them.[1]
The first construction is often seen as inherited from common Indo-European (Lehmann 1974, 116); typologically, however, it is by no means limited to Indo-European languages, but is admissible or mandatory in a wide range of languages and even diastratic varieties of languages (like English) that would not admit the "zero copula" in their standard version.[2]

In Ancient Greek, both constructions are attested since the earliest surviving texts. Ultimately, the nominal construction disappears from the language and it is not allowed in modern Greek. Examples 1 and 2 show how the two struc-

---

1   Though other types of copular verbs may fulfil this syntactic function (such as the equivalent of English "become", which adds to the linking function the semantic notion of transformation in time), we will limit our attention to the (semantically emptier) copula *par excellence*, the Greek verb *eînai* 'to be'. In Greek, as in English and in many other languages (cfr. French *être*, Italian *essere* etc., but not in Spanish, where the two notions are distinguished), this function coexists with a fuller existential meaning. Note that the notion of "copula" is not a neutral nor an uncontroversial one; for its history, see Moro (1997, 248–261); on the Greek verb *eînai* see especially Kahn (1973), with the remarks now reprinted in Ruijgh (1991, 775–815).

2   On the geographical distribution of the languages that allow or do not allow zero-copula constructions (and for a discussion of the problems in the classification) see WALS (Stassen 2013); much work has been devoted to the omission of the copula in the African American Vernacular English: see Walker (2006) for an introduction and Rickford et al. (1991) for a useful discussion on the methodological aspects of how to count and compare the two constructions.

tures can be employed in very similar contexts (and, in this case, how they can be accommodated to the very same metrical position); these two texts from the corpus of the tragedies attributed to Aeschylus (5th century BCE)[3] give a good idea of how difficult it is to account for the alternation between the structures.

(1) *toiáuta   mèn tád'   estín*
    such:NOM.PL PTC DEM.NOM.PL is:PRS.3SG
    'So stands the case' (Aesch. *Eum.* 480)

(2) *toiáuta   mèn dē   taût'*
    such:NOMPL PTC PTC DEM.NOM.PL
    'This [is] how things are' ([Aesch.] *PV.* 500)

Example 3 and 4 show how the two constructions can even coexist in the same sentence, in two comparable contexts (a proverb and its explanation), but with opposite distribution.

(3) *all' est'   alēthḕ   he   brotôn*
    PTC is:PRS.3SG true:NOM.SG DET:NOM mortals:GEN.PL
    paroimía,   ekhthrôn   adōra   dôra
    proverb:NOM.SG, enemies:GEN.PL no-gift:NOM.PL gifts:NOM.PL
    'But it is true what mortals say: the gifts of the enemy [are] no gifts'
    (Sophocles, *Ajax* 664–665)

(4) *palaià   paroimía   hóti chalepà   tà*
    old:NOM.SG proverb:NOM.SG that hard:NOM.PL DET
    kalá   estin
    good-things:NOM.PL is:PRS.3SG
    'It [is] an old saying that good things are hard to gain' (Plato, *Cratylus* 384, a-b)

The terminology is at times confusing. While in Greek grammars and in Indo-European studies it is customary to talk about the "nominal sentence" (*phrase nominale*, *Nominalsatz*) for constructions such as Example 2, "zero copula" is

---

3 The disputed authenticity of the *Prometheus Bound* (*PV*) is not relevant here: all critics agree that the tragedy, even if not by Aeschylus, was written in the classical period, most likely around the middle of the 5th Century BCE.

often employed, especially for other languages that allow constructions comparable to the one in point 1 above, where subject and nominal predicate are juxtaposed without a governing verb. While the latter definition can be criticized for suggesting the underlying assumption that the use of a copula is the norm and its absence a deviation (Stassen 2013), the former potentially covers (and in fact does cover in many studies) every form of non-verbal predication attested in Greek, even those that are not, strictly speaking, in competition with copular constructions (e.g., Lanérès 1994, 150–179).

As the present work is dedicated to the factors influencing the choice between the two competing structures, we will limit our attention to the two constructions listed above in points 1 and 2. We will survey the attestations in our reference corpus where a word or a phrase whose syntactic function is labelled as predicate nominal is constructed with a form of *eînai* 'to be', or is constructed without a governing verb; in Section 2.2, further restrictions are introduced in order to have a more controlled sample to review the impact of influencing factors. The choice of authors and corpus is discussed in Section 2.1.

1.2 *Previous literature and open problems*

The first Greek grammarians interpreted the nominal constructions as cases of verbal ellipsis (Lanérès 1994, 14–16). This point of view is generally replicated by modern grammars, where constructions such as Example 2 are regarded as a variation of 1 with elision of the main verb (Kühner and Gerth 1898; Schiefer 1974, 1, 40). Hjelmslev (1948), Meillet 1906 and Émile Benveniste (1950, reprinted in Benveniste 1966, from which we cite) have made clear that the two constructions, instead of being mere variants with elision, embody two different types of predication, with specific characters, nuances of meaning and partially overlapping features.

In particular, Benveniste (1950) has drawn attention to the fact that in nominal constructions the "verbal function" (i.e., the role of providing unity and cohesion to the elements of the sentence) is performed by a word that, because of its morphology, belongs to the class of nouns. Thus the nominal sentence does not make any reference to time, place or person, precisely because it lacks a word that, by its morphology, could mark person, tense and modality. For this reason, the nominal construction is unfit to describe or point to contingent realities, and it is rather employed to utter general statements about universal truths, often in the form of proverbs (as it is the case with Example 3, but, one may note, not with Example 4).

Benveniste (1950) supported his argument with a thorough discussion of sample data from the Ancient Greek literature; he inspected sentences from

two poetic texts from the 8th and 5th century (the *Iliad* and Pindar's *Pythians*), and used a prose author (Herodotus) as a counter-example. In Homer and Herodotus (where the nominal construction is in any case much rarer), the nominal sentence was practically absent outside direct speech. In dialogue (and in the pedagogic tone of Pindar's poetry), the construction is mostly used to express a universal truth, often introduced with a sententious tone or to add vigour to a point or argument.

A more systematic inspection of a corpus of texts, however, complicates the picture drawn by Benveniste. To begin with, the pattern of competition between nominal and copular structures is observed also outside the domain of assertions, to which Benveniste's analysis is limited, for example in interrogative sentences.[4] Also, nominal constructions with deictic pronouns, which clearly point to the context of the enunciation in a way that could hardly be called general or 'unbound to time and place', abound in the texts.[5] In a complete survey of Greek poetry and prose down to Euripides (late 5th Century BCE), Guiraud (1962) attempted to shield Benveniste's theory from these objections by postulating a difference between genuine and spurious nominal sentences. While Guiraud's book provides a great amount of useful data and analyses, this dichotomy is clearly artificial and untenable (see Lanérès 1994, 45–58, for a detailed criticism); besides, distinguishing between two groups of nominal sentences would not clarify why and how both classes were in competition with the copular construction.

More recently, Lanérès (1994) produced a two-volume analysis of the nominal constructions in the *Iliad*, where nominal and copular predications are investigated in the light of the opposition between narrative and discourse (*récit* and *discours*) and of the different pragmatic values of the two constructions. Lanérès starts from the hypothesis, made by Émile Benveniste (1960, reprinted in Benveniste 1966, 187–207) in a successive revision of his previous arguments, that subject and nominal predicates in a nominal sentence are joined by a null morpheme, which can be perceived as a "pause" in the phonetic realization of the sentence. For Lanérès, this is an important hint of the primarily discursive nature of the nominal construction, which is then more at home in oral communication than it is in writing (Lanérès 1994, 33–43 and 681). The

---

4   A treebank query on the tragedies traditionally attributed to Aeschylus in the AGDT, for instance, yields 33 interrogatives with nominal construction vs 37 questions with copula.

5   Tragedies, for instance, often begin with a description of the scene or a reference to the current time with the deictic pronoun *hóde* 'this' construed nominally, in the form: "this (that you see) [is] …". See e.g. Aesch. *Ag.* 40, Soph. *El.* 4–5, *Phil.* 1–2, Eur. *Hel.* 1–3; of course, the copula is always possible in similar constructions.

author's conclusions are that, while *eînai* is used to anchor the predication to reality or point to a nexus between subject and predicate that, for the speaker, exists outside the act of communication, in the nominal construction the predication is left only at the level of the linguistic act. For this reason, the nominal sentence is employed frequently to express value judgements or to mark the speaker's point of view (Lanérès 1994, 674). While the link between nominal construction, oral communication and discourse is persuasive, it remains difficult to operationalize it for quantitative assessment.

Nominal constructions are frequently attested in all the possible types of predications where the verb *eînai* is used, although the taxonomy adopted by the linguists vary significantly. All authors agree that, in Greek and in several other languages, both copular and nominal constructions can be employed in the following classes of sentences: existential statements (Lanérès 1994, 150–179; 668–669); predications that express the identity of subjects and predicate nominals or assign a quality to a subject (Benveniste 1966, 187–188, Guiraud 1962, 63–160, Lanérès 1994, 74–79); locative sentences (Lanérès 1994, 79–81, Kahn 1973, 156–167); possessive constructions, where the predicate nominal in Greek is in genitive or dative (Guiraud 1962, 189–198, Lanérès 1994, 81–83). Lanérès (1994, 163–165) notes that the status of nouns or adjectives followed by infinitives (with or without copula) is ambiguous, as the predication can be interpreted both as existential ("there [is] the necessity to do something") or as attributive ("doing something [is] a necessity") (see also Kahn 1973, 449). More fine-grained distinctions, e.g., the one between the attributive predications expressing formal identity, class subsumption, or inclusion in a set, are generally not considered relevant for the linguistic analysis of the constructions (Benveniste 1966, 187–188).

One aspect that is generally agreed upon is that the expression of person, tense and modality other than third-person (singular or plural) present indicative tends to favor the use of the copula.[6] Other aspects that favor/disfavor nominal or copular constructions, however, have been noted only sporadically, and no systematic quantitative assessment on the possible correlation with other features has been produced.

As for the problem of the history of the phenomenon, much of the work has been devoted to the earliest documents. Guiraud (1962) does not analyze texts later than the 5th century BCE. Lanérès (1994, 1), whose scrutiny is limited to

---

6 However, it is not true that the nominal sentence can express *only* this tense, person and modality: marks of other persons, tenses and moods can be realized lexically, e.g., with the use of personal pronouns as subjects (which are indeed quite frequent in nominal constructions), or temporal adverbs; the point is rightly stressed by Lanérès (1994, 667).

the earliest surviving Greek literary text (the *Iliad*), only notes in passing that the nominal construction, though still vital at the end of the classical age, is virtually absent (and mostly explained as an imitation of Homer's style) from the work of an Hellenistic poet like Apollonius Rhodius (3rd century BCE). We will see that this simple evolutionary model, according to which the nominal clause falls into irrelevance after the end of the classical age, does not hold true for Polybius (2nd century BCE).

As we saw, Benveniste's choice of texts (Herodotus, opposed to Homer and Pindar) as testbed for his interpretation is based on the idea that historiography, as a narrative genre, is less suitable to the kind of predication expressed through the nominal construction. The hypothesis is made plausible by the fact that copular constructions are significantly more frequent in prose (and historiography in particular) than in the poetical texts of the archaic and classical age.[7] However, as we see below, a non-negligible number of nominal structures are attested even in the texts of the Greek historians. This makes an investigation of that corpus all the more interesting, as no special account for the opposition in historiography has ever been attempted.

2 Corpus and methodology

2.1 *The Ancient Greek Dependency Treebank*

In what follows, we draw our data from version 2.1 of the *Ancient Greek Dependency Treebank* (AGDT), the first syntactically annotated corpus of Greek literary texts, published since 2009 (Bamman, Mambrini, and Crane 2009). The latest release of AGDT includes several samples of prose texts ranging from the 5th century BCE to the 2nd/3rd CE; all of them have been annotated by V. Gormann. Figure 1 represents a tree-shaped visualization of a copular sentence from our sample, the sentence that is glossed in Example 5. A summary of the prose selection of AGDT is reported in Table 1.[8]

---

7   A previous survey of the AGDT treebank reveals that, in a restricted sample limited to main clauses only, with verb at third person (implied or present), with subjects overtly expressed, 77.88% of the constructions attested in poetry are nominal (817 vs 232 copular); for prose, the amount drops to 45.32% (194 vs 234). I discussed those results in the 2016 annual meeting of the *Societas Linguistica Europaea* in Naples (2016). The slides of the presentation are available at: https://github.com/francescomambrini/research-projects/tree/master/ SLEPub/MambriniSLE2016.pdf.

8   Dates refer to century (B)CE and are at times speculative. The AGDT contains also a quite substantial selection of poetry from the archaic and classical age, including the full text of the Homeric poems, the work of Hesiod, Aeschylus and five tragedies of Sophocles. In total, the
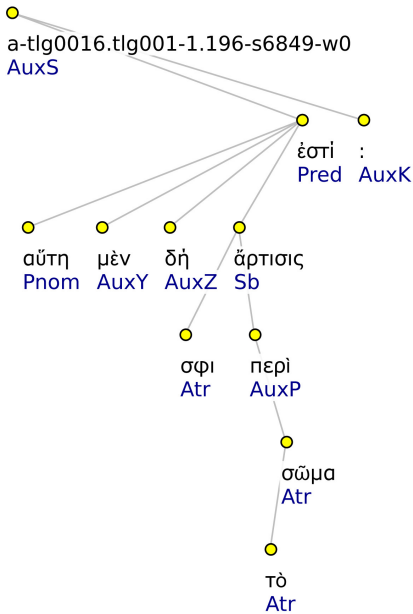
a-tlg0016.tlg001-1.196-s6849-w0
AuxS

ἐστί   :
Pred AuxK

αὕτη   μὲν   δή   ἄρτισις
Pnom   AuxY   AuxZ   Sb

σφι   περὶ
Atr   AuxP

σῶμα
Atr

τὸ
Atr

FIGURE 1
A tree from the AGDT: Herodotus, 1.196.1
(Example 5)

(5) *haútē   mèn dé sphi   ártisis   perì*
     DEM.NOM.SG PTC PTC 3PL.DAT equipment:NOM.SG around
     tò   sôma   estí
     DET.ACC body:ACC is:PRS.3SG
     'This is the equipment of their persons' (Herodotus, 1.196.1)

Although the AGDT also includes samples from the historical works of Diodorus
and Plutarch, the number of relevant constructions (both copular and nomi-
nal) that met the conditions for our study (see Section 2.2) were too few to be
meaningful for comparison with the other three authors. We decided, there-
fore, to focus our study only on the two historians of the 5th century BCE
(Herodotus and Thucydides) and Polybius (2nd century BCE).

Though the corpus that we obtained is less than ideal, it allows us to conduct
a quantitative study based on several different co-occurring phenomena, and
to extend the chronological limits of the investigation beyond the usual bound-
ary of the classical age. Our work will at least lay the foundation for further and
more systematic diachronic investigations that will become possible as soon as
new annotated data is made available.

poetry section amounts to more than 354,000 annotated tokens. The whole treebank can be
downloaded from https://perseusdl.github.io/treebank_data/.

TABLE 1 AGDT: prose (*=dates CE)

Author	Works	Genre	Date	N. nodes
Herodotus	bk.1	History	5th	33,105
Thucydides	bk.1	History	5th	25,360
Lysias	1,12,14,15,23	Oratory	4th	12,857
Plato	*Euthyphro*	Dialogue	4th	6,349
Polybius	bk.1,9,10	History	2nd	54,950
Diodorus Sic.	bk.11	History	1st	25,692
Plutarch	bk.11	History	1st/2nd*	22,152
Athenaeus	bk.12,13	Dialogue	2nd/3rd*	45,588
Total				226,053

2.2 *Methodology*

As no query software exists that works natively with the AGDT, we decided to convert the treebank into the format of the *Prague Dependency Treebank* in order to use PML Tree Query (Štěpánek and Pajas 2010) to query the corpus.[9]

As for the particular constructions we are concerned with, the annotation guidelines of the AGDT (Bamman et al. 2007) treat nominal clauses as cases of verb ellipsis.[10] One problem that our investigation faces is therefore how we can query an annotated corpus for something that is *not* there. The AGDT guidelines require the annotators to reconstruct a dummy, lexically empty node for the elided tokens, which is then assigned a syntactic label and governs arguments and satellites as any normal lexical node would do (Bamman et al. 2007, 36); thus, for instance, the subject and predicate nominal of example 2 above in the AGDT's formalism are governed by a reconstructed node representing a missing verb with the function of main predicate. The syntactic structure that is obtained is easy to query, but is indistinguishable from proper cases of ellipsis, such as elisions of non-initial conjuncts in coordination. For this reason, we have avoided coordinated verbs (or coordinate dummy nodes) and manually reviewed the matches in order to prune cases of proper ellipsis from the sample.

---

9   The converted files can be downloaded from: https://github.com/francescomambrini/pAGDT.

10   The previous guidelines of Greek used to annotate the AGDT, now updated to v. 2.1, were adapted from the Latin with only minor modifications.

Although, as we saw (Section 1.2), nominal and copular constructions are attested in different types of predications, the label PNOM is reserved only for predicate nominals in identity/attributive sentences and in constructions with infinitive verbs as subjects. For the existential, locative and possessive predications, the AGDT guidelines adopt different notations. In this study, we therefore limit our investigation to identity/attributive predications, which are in any case the most frequently attested and easiest to trace in a treebank.[11]

A few other constraints were added to reduce the number of features that impact the choice of construction to a controlled, limited set, or to make sure that certain aspects (such as the properties of the subject clause) are represented in all sentences of the sample. As we said, expressions of modality, person and tense other than the third-person, present indicative tend to favour the copula over the nominal construction; for this reason, we decided to limit our study to sentences that display (or would require, if a copula were used) either a third-person (singular or plural), indicative, present or a present infinitive. While it is easy to restrict the forms of *eînai* to match the aforementioned constraints, it is obviously impossible to apply a similar filter to the nominal clauses that are governed by a dummy node lacking any morphological feature. Nominal sentences that, if construed with the copula, would require a form of *eînai* other than the ones admitted in our study for the copular constructions were discarded from our sample after manual review, as well as all the ambiguous cases.

To sum up, our sample includes sentences from the works of Herodotus, Thucydides and Polybius from the AGDT that match the following requirements:

– display a node labelled as PNOM (nominal predicate) governed by either a form of *eînai* 'to be', or a reconstructed node;
– the verb *eînai* is (or would be, if the nominal clause were expressed with a copular construction) third-person (sing. or plur.), indicative present or infinitive present;
– the subject of the clause is expressed;
– alternative explanations for the omission of the copula (e.g., ellipsis of a coordinated conjunct, or nominal glosses in the text) are excluded; ambiguous cases were discarded from the sample after manual revision.

In total, our sample includes 440 sentences, as reported in Table 2.[12]

---

11   Kahn (1973, 87) claims that 80 to 85% of all uses of *eînai* in *Iliad* 1–12 are non-existential, while 65% of the occurrences belong to the "copulative construction in a narrow sense, with noun, adjective, pronoun or participle as predicate".
12   Further methodological problems, concerning in particular the statistical procedures

TABLE 2   Sample of historians from the AGDT:
                    summary

Author	Nominal	Copular	Total
Herodotus	27	148	175
Thucydides	51	40	91
Polybius	54	120	174
Total	132	308	440

3   Predictors of the nominal and copular construction

We discuss below a series of factors that, upon review of the grammars and of the previous studies, may plausibly impact on the choice of one construction over the other. The first section (Section 3.1) considers whether the sheer distribution of copular and nominal clauses changes significantly across authors and whether it varies from the 5th-century historians to Polybius.[13] The other subsections are dedicated to morpho-syntactic phenomena that concern the subject of the predication (the part of speech of the subject) or the clause itself (main vs subordinate clause, the order of predicate nominal and subject, the length of the clause).

3.1   *Author and chronology*
As it can be seen from Table 2, the nominal constructions account for approximately 30% of the total of our sample. While this number is certainly lower than the figures that can be observed in Homer or in other poetic texts (see above, note 4 for an estimate), such clauses are by no means marginal.

---

adopted, are discussed in the following sections. The sample of 440 sentences and the code that we used to obtain our results (written partly in Python, partly in R) is publicly available at: https://github.com/francescomambrini/research-projects/tree/master/SLE Pub.

13   For the tests of statistical significance we follow the methodology discussed by Gries (2013). We also follow Gries (2013, 26–29) in adopting the Null Hypothesis Significance Testing paradigm, which means that, for every hypothesis discussed, we will calculate the probability p that our observations are compatible with a null hypothesis (i.e., an hypothesis that logically contradicts the starting assumption). If p is below a threshold of significance (generally set at 0.05, with 0.01 and 0.001 marking further levels of significance), the null hypothesis can safely be rejected.
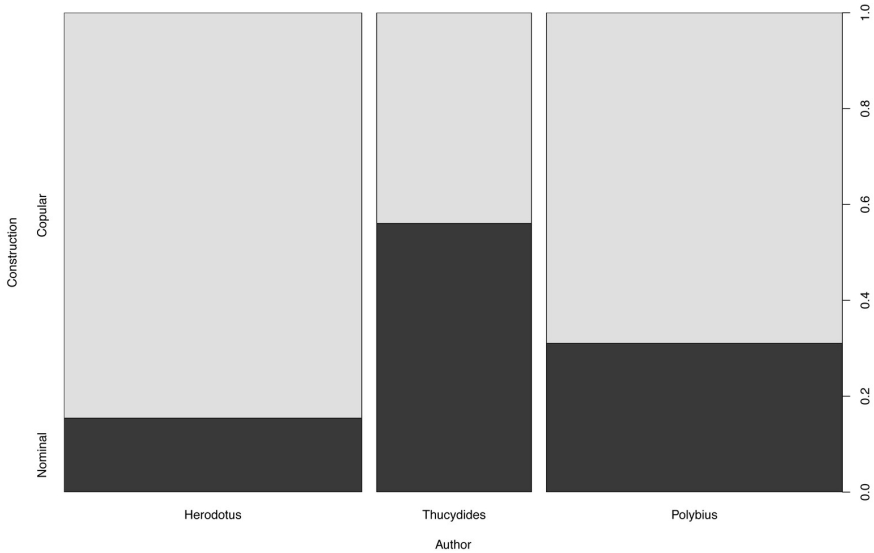
FIGURE 2   Nominal and copular constructions: distribution among authors

More interesting, however, is the distribution among the authors (Figure 2). As it can be seen, the distribution is not quite compatible with the simplistic theory that sees the nominal construction disappear after the end of the classical age. On the contrary, the widest difference that we observe is between the two authors of the 5th century; Herodotus, who seems to avoid the nominal construction (15.4% of nominal) and Thucydides, who is more prone to use it (56%). Though not exactly occupying a middle ground (31% of nominal clauses), Polybius is clearly more prone than Herodotus to resorting to the nominal construction.

The two historians of the 5th century account in total for 78 nominal vs 188 copular clauses of the sample; for Polybius, on the other hand, the sample includes 54 nominal vs 120 copular clauses. A chi-squared test for independence confirms that this variation between the authors grouped by century is not significant (χ^2 = 0.1467, df = 1, p = 0.7017, > 0.05). On the contrary, the distribution of the constructions among the authors differs very significantly (χ^2 = 47.175, df = 2, p = 5.703 e^{-11} < 0.001), with a moderate effect size (Cramer's V) of 0.327.[14]

---

14   Although the correlation indexes used in this study (Cramer's V and the ϕ score) theoretically range from 0 to 1, scores between 0.3 and 0.5 are generally considered indexes of a medium-strength correlation; the other commonly used thresholds are generally > 0.1 (weak correlation) and > 0.5 (strong correlation).

To sum up, the data from our sample do not support the conclusion that the number of copular vs. nominal constructions is influenced by the chronology of the author. The distributions observed are compatible with the null hypothesis that the two features (chronology and number of constructions) are not related. On the other hand, we can discard the null hypothesis that the distribution of the two constructions is not influenced by the style of single authors.

3.2   *Part of speech of the subject*

The grammars of Ancient Greek list several classes of nouns and adjectives that tend to be constructed nominally. Among them, there are several words that are joined with an infinitive functioning as subject, especially in clauses that express necessity, obligation, likelihood, impossibility or difficulty (see e.g., Example 6 below for a striking example from our corpus).

In this section, we inspect the impact of the part of speech of the subject, starting with a null hypothesis that the choice of construction is not affected by the type of subject used in the clause.

In order to analyze the impact of this factor, we have considered the part of speech encoded in the morphological tags available for each word in the AGDT (which also reflects the other morphological categories, such as mood, tense, case, number, etc). Five POS tags are attested for the subjects in our sample: nouns, adjectives, pronouns, articles and verbs. The latter category includes both infinitives (as in Example 6) and participles. As all adjectives are substantivized and articles are used as third-person pronouns, we reduced the number of categories to three, grouping adjectives with nouns, while articles were classified with the pronouns. Within the group of verbs, we used the supplementary morphological information to class the "infinitive" in a special group (coded as "vn"), while the substantivized participles were classified together with the nouns. The final classification includes: verb infinitive (vn), nouns (n) and pronouns (p).

Table 3 summarizes the distribution of the two constructions with the different classes of subjects. A chi-squared test for independence confirms that the correlation between the subject's part of speech and the choice of nominal vs copular construction is strongly significant ($\chi^2 = 67.198$, $df = 2$, $p = 2.559\ e^{-15}$, < 0.001), with a rather high correlation coefficient for the effect size (Cramer's $V = 0.391$).

As it is to be expected from the numbers reported in Table 3, the combination that affects the significance test the most is the one with infinitive verbs as subject and nominal construction. This fact is clearly visible in the association plot represented in Figure 3. In the plot, black boxes above the lines

TABLE 3   POS of the clause's subject

POS	Nominal	Copular	Total
Noun (n)	42	188	230
Pronouns (p)	21	74	95
Infinitives (vn)	69	46	115
Total	132	308	440

represent the values that exceed the expected frequency: the height of the
box is (to simplify) proportional to the variance with the expected frequency,
while the width is proportional to the square root of the expected frequency
(Gries 2013, 187–188, for a more detailed explanation). As can be seen, the
highest black box that extends itself above the line of expected frequency is
the one representing infinitive subjects with nominal constructions (on the
bottom-left corner); conversely, the grey box for copular constructions with
infinitive verbs (bottom-right corner of Figure 3) is also considerably below
the line of expected frequency. As we will see, this preference for nominal
structures in clauses with infinitive as subjects is strong in Thucydides and Poly-
bius, while the whole pattern (with both constructions) is overall rather rare in
Herodotus (18 cases: 3 nominal, 15 copular). Example 6 reproduces a sentence
from Polybius, where we find two nominal clauses of this type in the same sen-
tence.

(6) *perì   gàr hôn   adýnaton   gnônai   prìn   è*
     about PTC REL.GEN.PL impossible:NOM.SG know:INF before that
     genésthai   perì   toútōn   oudè synthésthai prò
     happen:INF about DEM.GEN.PL nor   agree:INF   before
     toû   dynatón
     DEM.GEN.SG possible:NOM.SG
     'For for those events that it [is] impossible to know before they happen,
     for them it [is] also not possible to make arrangements beforehand' (Poly-
     bius, 10.45.4)

To sum up, the data presented in this section allow us to disprove the initial
null hypothesis that the POS of the subject and the choice of copular vs nomi-
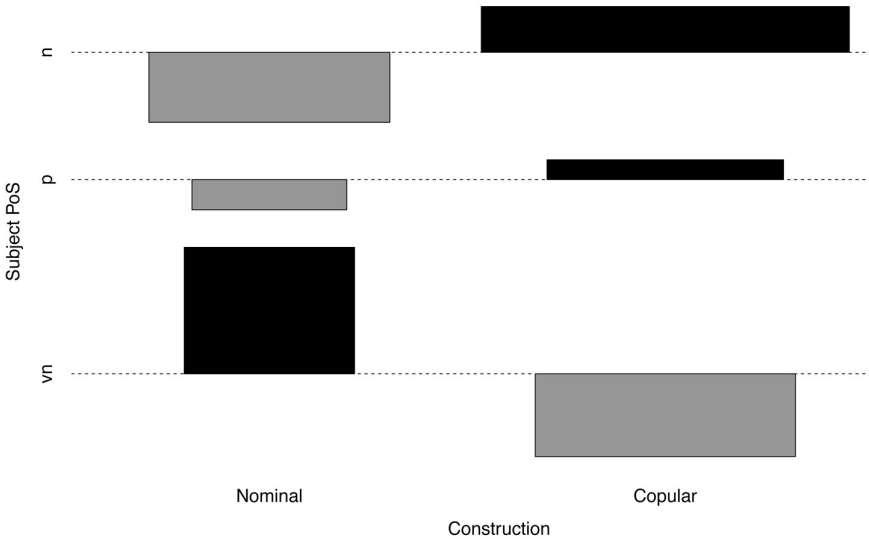nal construction are unrelated.

FIGURE 3   Subject POS and construction: association plot (*vn* = verb infinitve, *p* = pronoun, *v* = noun)

3.3   *Type of clause*

Guiraud (1962, 209–280) devoted a long chapter to the study of the opposition between nominal and copular constructions in subordinate clauses. His analysis starts from the observation that nominal constructions are rarer in subordinates than in main clauses. Lanérès (1994, 671) has also noted that the stronger the syntactic link of subordination between main and subordinate clause, the more the nominal construction seems to be disfavored.

The data of the treebank may help ascertain whether these views are confirmed, or whether a null hypothesis holds true that the two constructions do not vary significantly between subordinate and main clause.

Table 4 reports the distributions in our sample. The data confirm the observations quoted above (see also Figure 5): the nominal construction is avoided more frequently in subordinate clauses than in main clauses; according to a chi-squared test for independence, this correlation is very significant ($\chi^2 = 21.564$, $df = 1$, $p = 3.423e^{-06} < 0.001$), and is observed in all the authors. The effect size, however, is rather weak ($\phi = 0.221$).

We are thus allowed to discard the null hypothesis that type of clause and choice of construction are unrelated.

3.4   *Constituent order*

In some languages that allow the nominal construction, the omission of the copula seems to be linked to the order of constituents. In modern Hebrew,

TABLE 4   Type of clause: main vs subordinate

Clause	Nominal	Copular	Total
Main	91	138	229
Subordinate	41	170	221
Total	132	308	440

TABLE 5   Constituent order: pnom-sb vs sb-pnom

Order	Nominal	Copular	Total
Pnom-Sb	77	125	202
Sb-Pnom	55	183	238
Total	132	308	440

for example, while the nominal structure is favored with subject-pnom order, the personal pronoun functioning as copula is mandatory in the opposite case (Doron 1986).[15] In Greek, Lanérès (1994, 546–547) has noted that both orders are possible with either construction; the fact that the nominal construction seems to be preferred with the more marked pnom-subject order is interpreted as a sign that the nominal clause is used as an enunciative strategy to stress the validity of an assertion (p. 673).

To our knowledge, however, no study on the tendencies associated with the order of constituents is available for Greek authors. A null hypothesis that we can test would state that the order of subject and nominal predicate and the choice of construction are not related. Table 5 reports the distribution of the two constructions per order pattern.

As can be seen, in our sample the nominal construction is used more often when the order is pnom-subject, while the opposite is true when the order is reversed, thus confirming the tendency noted by Lanérès (1994, 673). According to a chi-squared test for independence, this correlation is very significant (χ^2 = 11.722, df = 1, p = 0.0006), but the effect size is small (Cramer's V = 0.163).[16] This

---

15   See also Moro (2010, 217) for a lengthier discussion from a generativist perspective.

16   It must be noted, however, that it is not always easy to distinguish between predicate nominal and subject in an Ancient Greek sentence; in Figure 1 / Example 5, the annotator has chosen to label the demonstrative *haútē* 'this' as the predicate nominal and *ártisis*

tendency is visible in all the authors of our sample (Herod.: 17.9% of nominal constructions with P-Sb order; 16.1% with Sb-Pnom; Thuc.: 66% nominal with Pnom-Sb, 43.9% with Sb-Pnom; Polyb.: 37.6% nominal with Pnom-Sb, 24.7% with Sb-Pnom).

These data do not support the null hypothesis that the choice of construction is unrelated to the respective order of nominal predicate and subject.

3.5 *Length of the clause*

Brevity is one of the most important stylistic features that can be associated with proverbial idioms. Therefore, given the stress that studies on the nominal constructions have placed on traditional sayings and sentential statements of general truths, it is interesting to verify whether copular and nominal clauses differ in their length, and in particular whether nominal clauses are shorter than the copular ones.

Dependency treebanks allow for a handy operationalization of clause length, which is independent from the punctuation marks introduced by the modern editors to mark sentence boundaries. By counting the nodes that depend (directly or indirectly) on the root of the copular or nominal predication, we can get an accurate index of the clause length. Thus, the tree represented in Figure 1 has a length of 8, as there are 8 nodes in its subtree that depend (directly and indirectly) on the main verb *estí* 'is' (3SG.PRES.IND of *eînai*).[17]

Figure 4 displays a boxplot of the subtree lengths grouped by the copular and nominal construction. The different length values are distributed rather irregularly; both constructions show elongated tails on the upper part and are positively skewed, with the presence of several outliers (i.e., data points that are more than 1.5 interquartile ranges above the third quartile), which are especially numerous for copular clauses.

Copular clauses range from a minimum length of 2 to a maximum of 87 nodes; nominal range from 2 to 51. Though they are characterized by a far larger range of values and a higher number of outliers on the upper part of the box plot, copular clauses are in fact shorter than the nominal ones on average. The

---

'equipment' as subject; others (including the author of this article) would find the opposite interpretation more natural. In similar cases, we have chosen not to change the annotation.

17   Note that, although this is not shown in the figure, punctuation marks like commas or parentheses can be easily excluded from the count, as we did for this study. Also, note that we did not include the root of the copular/nominal structure itself in the count.
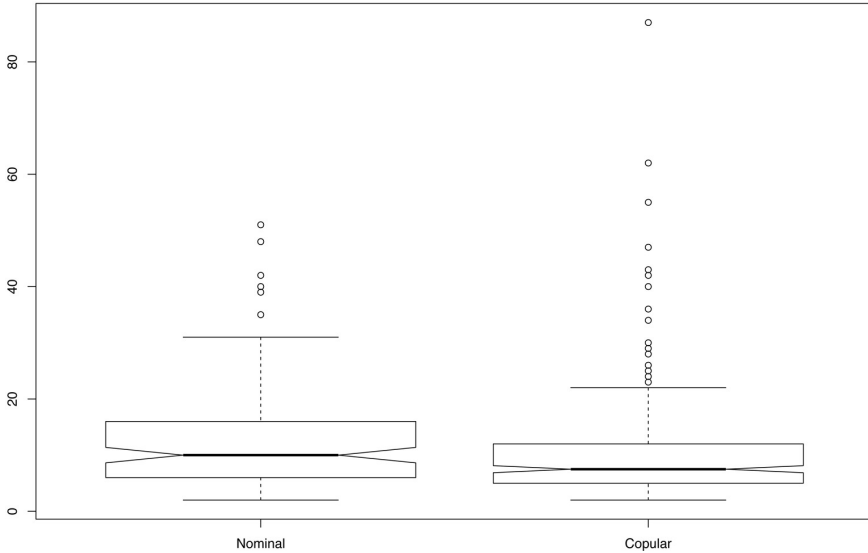
mean and median of copular sentences are 10.25 and 7.50 respectively, while those of nominal clauses are 12.27 and 10.00. This dispersion is not compatible with the hypothesis that nominal constructions are used predominantly with shorter clauses.

4   A multifactorial analysis

In the previous section we discussed a series of factors that impact the distribution of the two constructions. The potential predictors were considered separately, without any indication of the possible interaction of one feature over the other; although it is always possible to break down the distributions into the total per each of the three authors, it is hard to assess how the impact of the single factor might change in time and across the authors.

In what follows we address this crucial question with the help of a multifactorial analysis, following the methodology discussed by Gries (2013, 253–316).

A logistic regression model was fitted using the construction (with "nominal" and "copular" as factors) as the dependent variable; we started from a maximal model that used all the predictors discussed in Section 3, plus all the two-way interactions between them. In order to avoid too much data sparseness, we grouped noun and pronoun subjects in a single class, so that greater stress can be assigned to the opposition between clauses with infinitive sub-

jects and clauses with all other subjects; the predictor is thus a binary category: "infinitive" vs "noun/pronoun". In order to get a more symmetrical distribution, we have also used log-transformed values for subtree length.

In a backward selection process (Gries 2013, 259–261, 285–293), we proceeded to remove the insignificant predictors from the maximal model, until after seven steps a final model with only significant predictors of the higher order was reached. The final model preserves all the main effects, plus the interactions between the following predictors: Subject POS and Order, Subject POS and Author, and Order and Log.-Subtree Length.

As it emerges from the previous discussion, all the interactions between the predictor "Author" and the other effects were discarded as non-significant, except for that of (binary version of) subject POS. This implies that the impact of the other factors (length, clause type, order) does not decrease or increase significantly between the three authors.

The final model is highly significant, with a *G* score of the likelihood ratio test of 139.28 (*df* = 10, *p* < 0.001), and is sufficiently accurate: the predicted variation (C-score) is 82.2%, but the Nagelkerke R^2 index, which quantifies the variability accounted for, is rather low at 38.5%. The model is able to classify 79.5% of the cases correctly, which improves on a baseline that always predicts the most frequent construction (copular: 70%) by 9.5%. In what follows, we discuss the main effects and interactions by means of visualizations of the predicted probabilities.

As it can be seen from Figure 5 (top-left corner), the model shows clearly the impact of the type of clause on the choice of construction. In subordinate clauses, the probability of finding a copular construction is increased to 87%, versus a 62% in the case of main clauses. While clauses with pnom-subject order do not seem to be sensitive to the length of the clause (although with very high confidence intervals at both end), in the case of clauses with preposed subjects the probability of a copular construction increases steadily with the length of the subtree (lower-right corner of the figure).

More interesting for the purpose of a diachronic study is the effect of the interaction between the POS of the subject and the author. As we expect (Section 3.2), the probability of a copular construction drops when the subject is an infinitive. However, this tendency is more marked for Polybius and especially Thucydides; while for the former and the latter the predicted probability of a copular expression in case of an infinitive is 29.7% and 11% respectively, for Herodotus it remains very high at 80.6%; it must be noted, however, that for Herodotus the confidence intervals are extremely wide, due to the paucity of examples with infinite verbs as subjects that was already noted. Although on a more superficial observation, like the one we suggested in Section 3.1, Poly-
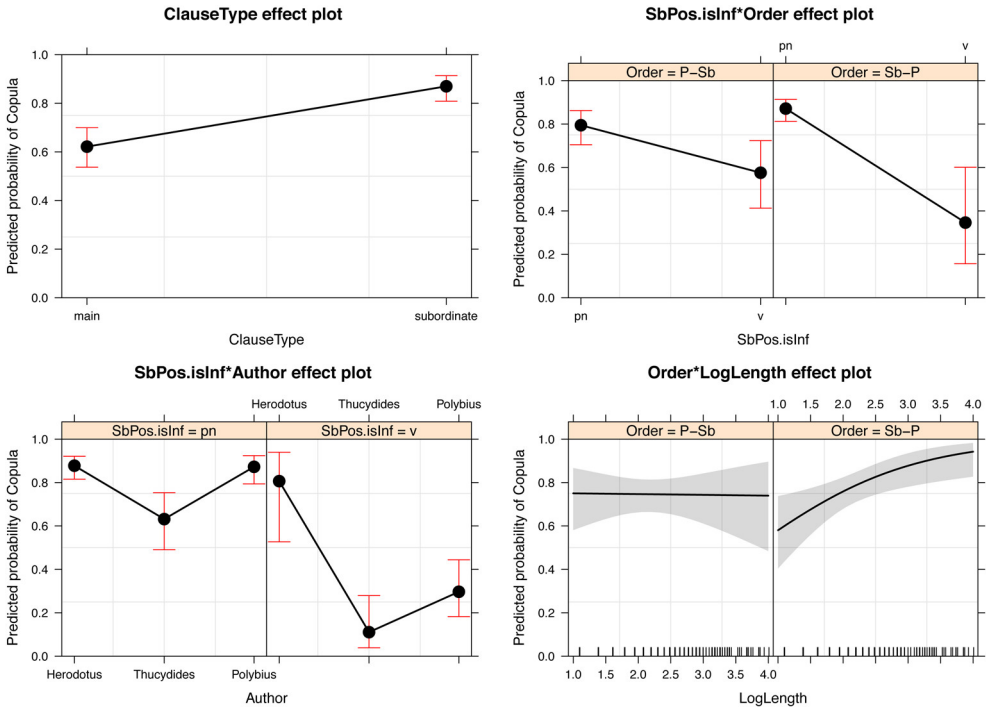
FIGURE 5　　Effects and interactions in the Logistic Regression model

bius seemed to strike a middle path between Herodotus and Thucydides, this predictor enriches the picture with other details.

In general, however, the model confirms that there is no trend towards a diminished role of the nominal construction in Polybius. On the contrary, the major difference appears to be the stylistic one that we already noted between Herodotus (with a marked preference for the copular construction in this author that was already observed by Benveniste) and Thucydides. Polybius, on the one hand, is comparable to Herodotus in his preference for the overt copula when the subject is a noun or a pronoun, but for him too the probability of a nominal construction increases significantly in the clauses where an infinitive is used as subject.

As for the other factors, as we said, the difference between authors does not seem to impact their predictive force.

5   **Conclusion**

In this paper we have seen how data from the AGDT can provide new evidence even for one of the most debated topics in the history of Ancient Greek. We have used the treebank to explore a rather marginal area in the thoroughly studied syntax of the nominal vs copular clauses; namely, we have focused on the genre of historical prose, using the annotated section of an author (Polybius) who lies outside the traditional chronological boundaries of research in that area. The interesting data that one can draw from such a treebank-based investigation clearly highlight the benefits of combining preexisting hypotheses and linguistic theories (that are often already available in abundance for classical languages) with corpus-based, quantitative methods for studying syntactic change.

The evidence clearly leads us to question a simplistic evolutionary model, according to which the nominal clause is reduced to a very marginal role after the end of the classical age in the 4th century BCE.

The data support the view that some morphosyntactic factors mentioned in the literature are arguably correlated with the choice of one construction over the other. But their explanatory force does not seem to be altered much by the difference observed in the three authors. The exception is the interaction between the authors and the part of speech of the subject, but again the difference is not compatible with a simplistic explanation of a reduced role of the nominal construction.

Our model, however, with its limited explanatory power, leaves plenty of room for improvement. Possibly, other features that lay outside the scope of the current morphosyntactic annotation of the AGDT (e.g., animacy of the subject) can play a role which is not possible to assess with the help of this treebank alone. Further exploration of the literature (also of other Indo-European languages) is needed to identify additional co-occuring features whose influence on the choice of construction can be tested.

Moreover, the choice of authors must be extended to include a fairer distribution of authors among different periods of the Greek language from the classical age to late Antiquity. Other genres of Greek prose that are well represented throughout the history of Greek literature, such as oratory or technical and military treatises, should be taken into account as well.

For studies in the diachrony of Greek syntax as reflected in the different genres of prose texts, the current release of the AGDT is far from ideal. Historiography, however, which is better represented than e.g., oratory or philosophy, is in better shape than other genres; some preliminary scrutiny, like the one attempted here, can already be conducted. The main goal of the present study

was in fact to lay the foundation for a treebank-based approach to the history of Ancient Greek syntax as reflected in the annotation.

References

Bamman, David, Francesco Mambrini, and Gregory Crane. 2009. "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank." In *Proceedings of the eigth international workshop on Treebanks and Linguistic Theories (TLT 8)*, 5–15. Milan: EDUCatt.

Bamman, David, Marco Passarotti, Gregory Crane, and Savina Raynaud. 2007. "Guidelines for the syntactic annotation of Latin treebanks (v. 1.3)." Tufts University, Medford (MA).

Benveniste, Émile. 1950. "La phrase nominale." *Bulletin de la Société de Linguistique* 45: 19–36.

Benveniste, Émile. 1960. "« Être » et « avoir » dans leurs fonctions linguistiques." *Bulletin de la Société de Linguistique* 55: 113–134.

Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.

Delbrück, Berthold. 1900. *Vergleichende Syntax der indogermanischen Sprachen. Dritter Theil*. Strassburg: Trübner.

Doron, Edit. 1986. "The Pronominal 'Copula' as Agreement Clitic in The Syntax of Pronominal Clitics." *Syntax and semantics* 19: 313–332.

Eckhoff, Hanne Martine, Silvia Luraghi, and Marco Passarotti. 2018. "The added value of diachronic treebanks for historical linguistics." *Diachronica* 35 (3): 297–309.

Gries, Stefan Th. 2013. *Statistics for linguistics with R*. 2nd ed. Berlin: De Gruyter Mouton.

Guiraud, Charles. 1962. *La phrase nominale en grec d'Homère à Euripide*. Paris: Klincksieck.

Hjelmslev, Louis. 1948. "Le verbe et la phrase nominale." In *Mélanges de philologie, de littérature et d'histoire anciennes offerts à J. Marouzeau*, 253–281. Paris: Les Belles Lettres.

Kahn, Charles H. 1973. *The verb "Be" in Ancient Greek*. Boston: Reidl.

Kühner, Raphael, and Bernard Gerth. 1898. *Ausführliche Grammatik der griechischen Sprache. Zweiter Teil: Satzlehre. Bd. 1*. 3rd ed. Hannover und Leipzig: Hahn.

Lanérès, Nicole. 1994. *Les formes de la phrase nominale en grec ancien: étude sur la langue de l'Iliade*. Lille: Centre Nationale de la Recherche Scientifique (CNRS).

Lehmann, Winfred P. 1974. *Proto-Indo-European syntax*. Austin: University of Texas Press.

Meillet, Antoine. 1906. "La phrase nominale en indo-européen." *Mémoires de la Société de Linguistique de Paris* 14: 1–26.

Moro, Andrea. 1997. *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge: Cambridge University Press.

Moro, Andrea. 2010. *Breve storia del verbo essere*. Milano: Adelphi.

Rickford, John R., Arnetha Ball, Renee Blake, Raina Jackson, and Nomi Martin. 1991. "Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English." *Language Variation and Change* 3 (1): 103–132.

Ruijgh, Cornelis. 1991. *Scripta minora ad linguam Graecam pertinentia*. Amsterdam: Gieben.

Schiefer, Erhard. 1974. "Zur Abgrenzung von Nominalsatz und Ellipse." *Zeitschrift für Vergleichende Sprachforschung* 88: 199–209.

Schwyzer, Eduard. 1950. *Griechische Grammatik. Zweiter Band: Syntax und syntaktische Stilistik*. München: Beck.

Stassen, Leon. 2013. "Zero Copula for Predicate Nominals." In *The World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Štěpánek, Jan, and Petr Pajas. 2010. "Querying Diverse Treebanks in a Uniform Way." In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 10)*, 1828–1835. La Valletta, Malta: European Language Resources Association (ELRA).

Walker, J.A. 2006. "Copula Variation." In *Encyclopedia of Language and Linguistics*, edited by Keith Brown, 2nd ed., 197–202. Oxford: Elsevier.