



A computational study on preverbal and postverbal accusative object nouns and pronouns in Ancient Greek

Giuseppe G. A. Celano

Tufts University, USA

Abstract

Many studies try to determine whether Ancient Greek is an OV or VO language. All of them, however, fail to conduct a research whose method is entirely clear. This paper presents the first attempt to quantify the number of verbs governing preverbal or postverbal accusative object nouns or pronouns in single or coordinate independent clauses in Homer's *Iliad* and *Odyssey*, Herodotus' *Histories*, and the New Testament, by providing results which are fully verifiable and reproducible. I prove that as for the parameter OV vs. VO there is great variation in the texts, which suggests a change over time from OV order in Homer to VO order in the New Testament. The figures for Herodotus' Greek prove a quasi-exact match between OV order and VO order.

1. Introduction

Ancient Greek (AG) is an Indo-European language allowing great freedom of word order at both clausal and subclausal level. A great variety of studies were conducted on the position of subject (S), verb (V), and object (O) to establish the "normal" order of such constituents (see, among others, Ebeling (1902); Friederich (1975); Cervin (1990); Kwong (2005)). They however provide discordant results, which are impossible to evaluate (see, for example, Cervin (1990); Taylor (1994)): the sample analyzed is often limited and, what is worse, the method employed to count the instances of a given word order is usually not precisely defined: e.g., Friederich (1975) counted 195 constructions in *Iliad* 5.1–296, but it is not clear what exactly he means by a construction.

Even when the criteria for counting seem to be clearer, the amount of data to manually process is so large that the research is unlikely to be reproducible (see, for example, Kwong (2005)). The present study, therefore, aims to be the first attempt at scientific quantification of AG word order: more precisely, I search the Ancient Greek Dependency Treebank (AGDT)¹ and the Pragmatic-Resources-in-Old-Indo-European-Languages corpus (henceforth the PROIEL Treebank, PROIELT)² for verbs governing preverbal or postverbal accusative object nouns or pronouns in single or coordinate independent clauses in Homer's *Iliad* and *Odyssey*, and the annotated parts of Herodotus' *Histories* and the New Testament (henceforth also referred to as OV and VO order, respectively).

In Section 2, I describe the data on which the research is based, i.e., how the morphosyntactic annotation of the texts has been encoded in the AGDT (Section 2.1) and the PROIELT (Section 2.2). Section 3 shows the method of the research: in Section 3.1 and 3.2 the queries used to search the AGDT and the PROIELT, respectively, are presented and explained. The results of the queries are given in Section 4 and discussed in Section 5. Section 6 summarizes the article.

2. Data

The data of the present research are drawn from the AGDT and the PROIELT, the two open source dependency treebanks currently available for AG. The AGDT provides Homer's *Iliad* and *Odyssey*, while Herodotus' *Histories* and the New Testament are contained in the PROIELT. The three texts belong to different varieties and stages of the language: the Homeric poems are likely to have been created in the 8th century BC; the historian Herodotus lived approximately between 485 and 424 BC; the New Testament was written in the 1st century AD.

All the texts have been automatically divided into sentences on the basis of the punctuation of the source text. Each word of such sentences has been semi-automatically annotated for morphology and manually annotated for syntax. The categories for the morphological annotation are those elaborated by traditional grammar (see more in Section 2.1 and 2.2): e.g., the noun is annotated for number, gender, and case, while the verb is annotated for number, person, tense, mood, and voice.

The syntactic annotation, for which specific guidelines (Bamman and Crane (2008); Haug (2010)) exist, mostly relies on the Prague Dependency Treebank 2.0 (PDT 2.0) Manual for Analytic Annotation (Hajičová et al. (1999)). Although the annotation style of the AG treebanks is different (see Section 2.1 and 2.2), they share the same core structure: according to the theory of Dependency Grammar as developed for the analytical level of the PDT 2.0, each text is a set of sentences, and each sentence is a set of words connected to each other in a head-dependent relationship.

¹<http://nlp.perseus.tufts.edu/syntax/treebank/greek.html>

²http://foni.uio.no:3000/users/sign_in

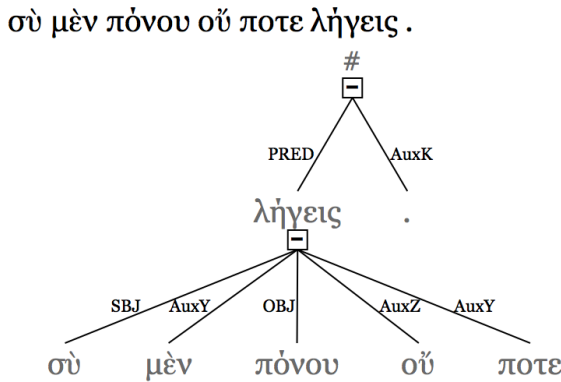


Figure 1. Example of a dependency tree from the AGDT

Each sentence can be represented as a top-down tree where the topmost word (typically the main verb or a coordinate conjunction) directly or indirectly governs the whole structure (excluding the final punctuation mark):

Example 1

σὺ μὲν πόνου οὐ ποτε λήγεις
 you.NOM.SG PTC work.GEN.SG not ever cease.PRS.IND.ACT.2SG
 ‘You never cease from working’

(Il. 10.164)

As Figure 1 shows, each word of Example 1 corresponds to a node, which is associated with its grammatical function: e.g., πόνου is annotated as an object (OBJ) of the predicate λήγεις, which takes the function PRED, i.e., “predicate”, because it is the (main) verb of the sentence. Note that a word can have only one head, but a head can have more than one dependent: e.g., the verb λήγεις has five dependents.

The function OBJ, as well as the function SBJ, marks the arguments of a verb (for a more detailed definition see Sgall et al. (1986); Hajičová et al. (1999)). Depending on the meaning of the verb, the function OBJ can be conveyed not only by different cases in AG, but also by different parts of speech (e.g., nouns and adverbs) and constructions (e.g., infinitive and participial clauses). For the purposes of the present research, I restrict the analysis to only OBJ constituents being accusative nouns or pronouns in single or coordinate independent clauses.

It is important to underline that from the perspective of linguistic theory two different kinds of dependency are captured in the syntactic annotation: grammatical dependencies and “technical” dependencies. The former correspond to (rather) uncon-

troversial linguistic dependencies, such as that between the verb and its subject, while the latter concern dependencies such as coordination, whose annotation is rather governed by theory-dependent rules. In both the AGDT and the PROIELT, a coordinate conjunction is taken to technically govern its conjuncts and depend on the word to which such conjuncts are grammatically related. In the following two sections, I will present the AGDT and the PROIELT: the reader is referred to the online documentation and XML files for a full description.³

2.1. The AGD Treebank

The AGDT includes Homer's Iliad and Odyssey (232.339 words, including punctuation). As an illustration, the following text shows how Example 1 is coded in the XML file:

```
<sentence subdoc="urn:cts:greekLit:tlg0012.tlg001:10.164" id="2277214"
  document_id="Perseus:text:1999.01.0133" span="σὸ0:.7">
  <primary>alexlessie</primary>
  <primary>millermo</primary>
  <secondary>nicanor</secondary>
  <word id="1" cid="34282848" form="σὸ" lemma="σού1" postag="p-s----n-" head="6"
    relation="SBJ" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="2" cid="34282849" form="μὲν" lemma="μέv1" postag="g-----" head="6"
    relation="AuxY" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="3" cid="34282850" form="πόνου" lemma="πόνοç1" postag="n-s---mg-" head="6"
    relation="OBJ" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="4" cid="34282851" form="οὖ" lemma="οὐ1" postag="d-----" head="6"
    relation="AuxZ" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="5" cid="34282852" form="ποτε" lemma="ποτέ1" postag="g-----" head="6"
    relation="AuxY" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="6" cid="34282853" form="λήγειç" lemma="λήγω1" postag="v2spia---" head="0"
    relation="PRED" cite="urn:cts:greekLit:tlg0012.tlg001:10.164"/>
  <word id="7" cid="34282854" form="." lemma="πериод1" postag="u-----" head="0"
    relation="AuxK" cite=""/>
</sentence>
```

Each <sentence> element has some attributes identifying it: particularly relevant are the @id attribute, which specifies the number of the sentence, and the @subdoc attribute, which specifies the reference of the sentence. Within the <sentence> element there are two kinds of children. The <primary> and <secondary> elements contain the names of the annotators of the sentence: the primary ones annotated the sentence independently of each other, with the secondary one deciding when the primary annotations were different. The second kind of children are the <word> elements: there are as many as the actual words of the sentence. Each <word> element has 8 attributes:

- The value of the @id attribute indicates the position of the word in the sentence.
- The value of the @cid attribute is the unique number of the word in the database.

³ The relevant documentation and the exact files on which I performed my queries can be downloaded at https://github.com/gcelano/PBML101_2014.

- The value of the @form attribute is the actual form of the word in the sentence.
- The value of the @lemma attribute corresponds to the dictionary form of the word.
- The value of the @postag attribute is the morphological annotation of the word: it consists of a string of 9 ordered positions, each of which can be filled by a character or, if a value is missing, a hyphen. The 9 positions correspond to: part of speech (1), person (2), number (3), tense (4), mood (5), voice (6), gender (7), case (8), and degree (9). For example, the postag value n-s---mg- of the word πόνου means noun (n), singular (s), masculine (m), genitive (g).⁴
- The value of the @head attribute coincides with the value of the @id attribute of the head word: e.g., the word πόνου is syntactically annotated as depending on the word λήγεις because the value of its @head attribute and the value of the @id attribute of λήγεις match (6 in both). If the head value of a word is 0, this means that the word has no head.
- The value of the @relation attribute indicates the kind of relation the word entertains with its head: e.g., the word πόνου has an OBJ relation, which means that it is taken to be an object with respect to the predicate λήγεις, its head (see Bamman and Crane (2008) for the list and meaning of all relations).
- The value of the @cite attribute shows the exact reference of the word.

2.2. The PROIEL treebank

The PROIEL corpus is a parallel corpus containing the morphosyntactic annotation of almost all the New Testament in AG (124.991 tokens) and its translations in Old Church Slavonic, Classical Armenian, Gothic, and Latin. The corpus also includes the annotation of some literary works, such as (part of) Herodotus' Histories (71.719 tokens). They are made available in XML format. The following is an example of how a sentence of the New Testament is encoded in the PROIELT:

```
<sentence id="47594" status="reviewed" presentation-after=" ">
<token id="266750" form="Ἀμιναδάβ" citation-part="MATT 1.4" lemma="Ἀμιναδάβ"
  part-of-speech="Ne" morphology="-----n" head-id="266752" relation="sub"
  antecedent-id="266748" information-status="old" presentation-after=" "/>
<token id="266751" form="δέ" citation-part="MATT 1.4" lemma="δέ"
  part-of-speech="Df" morphology="-----n" head-id="266752" relation="aux"
  presentation-after=" "/>
<token id="266752" form="ἐγέννησεν" citation-part="MATT 1.4" lemma="γεννώω"
  part-of-speech="V-" morphology="3saia---i" relation="pred"
  presentation-after=" "/>
<token id="266753" form="τὸν" citation-part="MATT 1.4" lemma="ὁ"
  part-of-speech="S-" morphology="-s--ma-i" head-id="266754" relation="aux"
  presentation-after=" "/>
<token id="266754" form="Ναασώων" citation-part="MATT 1.4" lemma="Ναασώων"
  part-of-speech="Ne" morphology="-----n" head-id="266752" relation="obj"
```

⁴The full value list can be found at https://github.com/gcelano/PBML101_2014.

```

      information-status="acc_gen" presentation-after=","/>
</sentence>

```

The text has been divided into sentences corresponding to <sentence> elements. The words of each sentence are annotated as <token> children having up to 11 attributes:

- The value of the @id attribute identifies the word in the PROIELT.
- The value of the @form attribute corresponds to the actual word form in the text.
- The value of the @citation-part shows the reference of the token.
- The value of the @lemma is the dictionary form of the word.
- The @part-of-speech attribute contains the part of speech of the word.
- The value of the @morphology attribute is the morphological annotation of the word: it consists of a 10-long character string, where each position corresponds to an ordered value: e.g., the @morphology value of the word ἐγέννησεν is 3saia---i, which means that the word contains the following features: third person (3), singular (s), aorist (a), indicative (i), active (a), inflecting (i).
- The value of the @head-id attribute contains the @id value of the governor word.
- The value of the @relation attribute identifies the relation of the word with respect to its governor: e.g., the word Ναασσών is the object of the verb ἐγέννησεν (see Haug (2010) for the list and meaning of all relations).
- The value of the @antecedent-id attribute is the @id of the token with which a given token is coreferential.
- The attribute @information-status describes the cognitive state related to a word.
- The value of the @presentation-after attribute contains the punctuation mark that may follow the word. Note that in the AGDT punctuation marks are annotated not as attributes but as <word> elements.

3. Method

The treebanks have been searched for the number of verbs governing preverbal or postverbal accusative objects, which are nouns or pronouns, in the single or coordinate independent clauses of the following works: Homer's Iliad and Odyssey and the annotated parts of Herodotus' Histories and the New Testament.⁵ To query the corpora I used XQuery 1.0 as implemented in BaseX 7.8.⁶

3.1. The query for the AGDT

The following is the query used to extract the data from the AGDT:

```

xquery version "1.0";
let $r :=

```

⁵The texts used are made available at https://github.com/gcelano/PBML101_2014.

⁶ <http://basex.org/products/download/all-downloads/>.

```

for $t in //word[@relation= ("PRED", "PRED_CO")]
let $o := $t/preceding-sibling:word[@relation = "OBJ"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head = $t/@id]
let $d := $t/parent::sentence/word[@relation= "COORD"]
    [@head = $t/@id]
let $y := $t/preceding-sibling:word[@relation= "OBJ_CO"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head = $d/@id]
let $h := $t/following-sibling:word[@relation = "OBJ"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head = $t/@id]
let $z := $t/following-sibling:word[@relation= "OBJ_CO"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head = $d/@id]
let $x := $t/parent::sentence/word[@relation= "COORD"]
    [@id = $t/@head]
let $a := $t/preceding-sibling:word[@relation = "OBJ"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head= $x/@id]
let $s := $t/following-sibling:word[@relation = "OBJ"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head= $x/@id]
let $j := $t/parent::sentence/word[@relation= "COORD"]
    [@head = $x/@id]
let $b := $t/preceding-sibling:word[@relation = "OBJ_CO"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head= $j/@id]
let $c := $t/following-sibling:word[@relation = "OBJ_CO"]
    [@postag[matches(., "(n|p).....a.")]]
    [@head= $j/@id]
(: the following where clauses can be changed thus:
where $o or $y or $a or $b
where $h or $z or $s or $c
where ($o and $h) or ($o and $z) or ($o and $s) or ($o and $c)
    or ($y and $h) or ($y and $z) or ($y and $s) or ($y and $c)
    or ($a and $h) or ($a and $z) or ($a and $s) or ($a and $c)
    or ($b and $h) or ($b and $z) or ($b and $s) or ($b and $c) :)
where $o or $y or $a or $b
return <r cid= "{$t/@cid}" v= "{$t/@form}" cite= "{$t/@cite}">
    <pre_object>{data($o/@form)}</pre_object>
    <passage>{data($o/parent::sentence/word/@form)}
    </passage>
    <pre_object_co>{data($y/@form)}</pre_object_co>
    <passage>{data($y/parent::sentence/word/@form)}
    </passage>
    <pre_obj_attached_to_coord>{data($a/@form)}</pre_obj_attached_to_coord>
    <passage>{data($a/parent::sentence/word/@form)}
    </passage>
    <pre_obj_co_attached_to_coord>{data($b/@form)}</pre_obj_co_attached_to_coord>
    <passage>{data($b/parent::sentence/word/@form)}
    </passage>

```

```

<fol_object>{data($h/@form)}</fol_object>
<passage>{data($h/parent::sentence/word/@form)}
</passage>
<fol_object_co>{data($z/@form)}</fol_object_co>
<passage>{data($z/parent::sentence/word/@form)}
</passage>
<fol_obj_attached_to_coord>{data($s/@form)}</fol_obj_attached_to_coord>
<passage>{data($s/parent::sentence/word/@form)}
</passage>
<fol_obj_co_attached_to_coord>{data($c/@form)}</fol_obj_co_attached_to_coord>
<passage>{data($c/parent::sentence/word/@form)}
</passage>
</r>
for $k at $n in $r
return <f number= "{$n}">{$k}</f>

```

In the AGDT the verbs of independent clauses are the only verbs to receive the relation PRED: any other kind of verb is annotated differently. In our study we have disregarded the case of empty verbs, i.e., verbs which do not surface in the sentence but the annotator assumes to be implied. Since the objects I am looking for are nouns or pronouns in the accusative case, I have used a regular expression to match the desired values of the @postag attribute.

The AGDT identifies coordinate constituents by adding the _CO suffix to the value of the @relation attribute. This means that in order to capture coordinate verbs, one has to also take into account verbs having relation PRED_CO. Furthermore, since coordinate objects are annotated as direct dependents of a conjunction, with such a conjunction depending on the verb grammatically related to the coordinate objects, it is possible to link such coordinate objects to their predicate only bypassing the coordinate conjunction (see the let \$d, let \$y, and let \$z clauses).

Single OBjs are returned by means of the let \$o and let \$h clauses. When there are two or more PREDs coordinated, there may be an object depending on both of them: these kinds of objects are meant to be captured by the let \$a and let \$s clauses. If there are coordinate objects that depend on coordinate predicates, they are identified by the let \$b and let \$c clauses. As specified in the comment within the query, one has to change the where clause to get the following results:

- If where \$o or \$y or \$a or \$b is added, the query returns verbs with single and coordinate preverbal objects.
- If where \$h or \$z or \$s or \$c is added, the query returns verbs with single and coordinate postverbal objects.
- If where (\$o and \$h) or (\$o and \$z) or (\$o and \$s) or (\$o and \$c) or (\$y and \$h) or (\$y and \$z) or (\$y and \$s) or (\$y and \$c) or (\$a and \$h) or (\$a and \$z) or (\$a and \$s) or (\$a and \$c) or (\$b and \$h) or (\$b and \$z) or (\$b and \$s) or (\$b and \$c) is added, the query returns PRED verbs having at least one preverbal and one postverbal object at the same time.

3.2. The query for the PROIELT

The query used to interrogate the PROIELT is the following:

```
xquery version "1.0";
let $a :=
for $t in //token[@relation = "pred"][@form]
let $i := $t/parent::sentence/token
    [@part-of-speech = "G-"]
    [@id = $t/@head-id]
let $n := $t/parent::sentence/token[@relation = "pred"]
    [@part-of-speech = "C-"]
    [@id = $t/@head-id]
let $z := $t/parent::sentence/token
    [@part-of-speech = "G-"]
    [@id = $n/@head-id]
let $p := $t/parent::sentence/token[@relation = "obj"]
    [@part-of-speech = "C-"]
    [@head-id = $t/@id]
let $e := $t/preceding-sibling::token[@relation = "obj"][@morphology[matches(., ".....a...")]]
    [@part-of-speech = ("Pp", "Ne", "Ps", "Pi", "Pt", "Pk", "Px", "Pc", "Pd", "Nb")]
    [@head-id = $t/@id]
let $o := $t/preceding-sibling::token[@relation = "obj"][@morphology[matches(., ".....a...")]]
    [@part-of-speech = ("Pp", "Ne", "Ps", "Pi", "Pt", "Pk", "Px", "Pc", "Pd", "Nb")]
    [@head-id = $p/@id]
let $w := $t/following-sibling::token[@relation = "obj"][@morphology[matches(., ".....a...")]]
    [@part-of-speech = ("Pp", "Ne", "Ps", "Pi", "Pt", "Pk", "Px", "Pc", "Pd", "Nb")]
    [@head-id = $t/@id]
let $k := $t/following-sibling::token[@relation = "obj"][@morphology[matches(., ".....a...")]]
    [@part-of-speech = ("Pp", "Ne", "Ps", "Pi", "Pt", "Pk", "Px", "Pc", "Pd", "Nb")]
    [@head-id = $p/@id]
```

(: the following two clauses serve to identify coordinate verbs sharing an object.

Change the direction of the greater than sign in all the following clauses at the same time:

> captures preverbal objects, while < captures postverbal objects :)

```
let $b := let $l := $t/parent::sentence/token[@relation = "pred"]
    [@head-id = $n/@id]/slash[@relation = "obj"][@target-id = $e/@id]
    where $l/parent::token/xs:integer(@id) > $e/xs:integer(@id)
    return $l
let $c := let $v := $t/parent::sentence/token[@relation = "pred"]
    [@head-id = $n/@id]/slash[@relation = "obj"][@target-id = $p/@id]
    where $v/parent::token/xs:integer(@id) > $o/xs:integer(@id)
    return $v
let $m := let $r := $t/parent::sentence/token[@relation = "pred"]
    [@head-id = $n/@id]/slash[@relation = "obj"][@target-id = $w/@id]
    where $r/parent::token/xs:integer(@id) > $w/xs:integer(@id)
    return $r
let $f := let $y := $t/parent::sentence/token[@relation = "pred"]
    [@head-id = $n/@id]/slash[@relation = "obj"][@target-id = $p/@id]
    where $y/parent::token/xs:integer(@id) > $k/xs:integer(@id)
    return $y
```

(: the following where clause can be changed thus:

where (not(\$i) and not(\$z)) and (\$e or \$o)

where (not(\$i) and not(\$z)) and (\$w or \$k)

```

where (not($i) and not($z)) and (($e and $w) or ($e and $k)
      or ($o and $w) or ($o and $k))
where (not($i) and not($z)) and ($b or $c or $m or $f) :
where (not($i) and not($z)) and ($e or $o)
return
<r id= "{$t/@id}" locus= "{$t/@citation-part}" v= "{$t/@form}">
<pre_object>{data($e/@form)}</pre_object>
<passage>{data($e/parent::sentence/token/@form)}</passage>
<pre_object_co>{data($o/@form)}</pre_object_co>
<passage>{data($o/parent::sentence/token/@form)}</passage>
<fol_object>{data($w/@form)}</fol_object>
<passage>{data($w/parent::sentence/token/@form)}</passage>
<fol_object_co>{data($k/@form)}</fol_object_co>
<passage>{data($k/parent::sentence/token/@form)}</passage>
<object_of_co_verb1 v= "{$b/parent::token/@form}">{data($e/@form)}</object_of_co_verb1>
<passage>{data($b/parent::token/parent::sentence/token/@form)}</passage>
<object_of_co_verb2 v= "{$c/parent::token/@form}">{data($o/@form)}</object_of_co_verb2>
<passage>{data($c/parent::token/parent::sentence/token/@form)}</passage>
<object_of_co_verb3 v= "{$m/parent::token/@form}">{data($w/@form)}</object_of_co_verb3>
<passage>{data($m/parent::token/parent::sentence/token/@form)}</passage>
<object_of_co_verb4 v= "{$f/parent::token/@form}">{data($k/@form)}</object_of_co_verb4>
<passage>{data($f/parent::token/parent::sentence/token/@form)}</passage>
</r>
for $m at $x in $a
return
<r number= "{$x}">{$m}</r>

```

In the PROIELT the function of verbs in single or coordinate independent clauses is invariantly labeled as “pred”; the same label, however, is used to annotate the function of the verb of clauses introduced by a subordinate conjunction. This means that, in order to only capture single or coordinate independent clauses, one has to filter the number of predicates by using the where clause not(\$i) and not(\$z), which allows exclusion of predicates (not(\$i)) and coordinate predicates (not(\$z)) depending on a subordinate conjunction.

The predicate [@form] in the for \$t clause serves to select only verbal predicates which appear in the texts (i.e., to exclude implied verbs, which are annotated as tokens having no @form attribute). The let \$e and let \$o clauses capture the single and coordinate preverbal objects, respectively: the predicates [@part-of-speech = (“Pp”, “Ne”, “Ps”, “Pi”, “Pt”, “Pk”, “Px”, “Pc”, “Pd”, “Nb”)] and [@morphology[matches(., “.....a...”)]] are meant to select only nouns and pronouns in the accusative case. The let \$w and let \$k clauses are the counterparts of the let \$e and let \$o clauses, in that they concern following objects.

The where clause always contains the logical expression not(\$i) and not(\$z), which excludes the single and coordinate verbs of subordinate clauses. The same where clause has also to contain the expression and (\$e or \$o) if the single and coordinate preverbal objects are searched for, or the expression and (\$w or \$k) if the single and coordinate postverbal objects are searched for, or the expression and ((\$e and \$w)

	preverbal objects	postverbal objects
Homer	4995	2451
Herodotus	804	753
the New Testament	997	2084

Table 1. Figures for preverbal and postverbal accusative object nouns and pronouns

or (\$e and \$k) or (\$o and \$w) or (\$o and \$k)), if the co-occurrences of preverbal and postverbal objects are searched for.

In the PROIELT, constituents relating to more than one finite verb are annotated differently than in the AGDT: one object is attached to a verb and, if it is also related to another coordinate verb, this verb contains a <slash> element pointing to the object (see Haug (2010) for a full description). This phenomenon is searched for through the let \$b, let \$c, let \$m, and let \$f clauses. More precisely, to find the preverbal objects of coordinate verbs, one has to use the greater than sign in all of the four clauses; to find postverbal objects, the less than sign is to be used.

The results of the where (not(\$i) and not(\$z)) and (\$b or \$c or \$m or \$f) clause are added to those obtained by applying the where (not(\$i) and not(\$z)) and (\$e or \$o) clause when the greater than sign is used; when the less than sign is used, the results are added to those returned by limiting the query with the where (not(\$i) and not(\$z)) and (\$w or \$k) clause. There are no coordinate objects with a <slash> element pointing to two objects, so there is no possibility for a coordinate verb to have both a preverbal and a postverbal object in the same clause.

4. Results

The results are presented in three tables. Table 1 contains the figures for verbs governing preverbal or postverbal objects, both single and coordinate; Table 2 contains the figures for verbs having preverbal and postverbal objects in the same clause; Table 3 shows the figures of the first table minus those of the second table. If we count two or more objects being on the same side of a verb as one object and the object relating to two coordinate verbs as two objects, we can say that the tables report the numbers of preverbal and postverbal objects.

5. Discussion

The results prove that in my files single or coordinate accusative object nouns or pronouns in single or coordinate independent clauses in Homeric poems are usually placed before the verb, while in the New Testament the same kind of objects are typically placed after the verb. Interestingly, Herodotus' text shows no real predominant order. This result is essentially in accordance with that found by Dunn (1988) (but he pro-

	co-occurrences
Homer	163
Herodotus	14
the New Testament	18

Table 2. Figures for co-occurrences of preverbal and postverbal accusative object nouns and pronouns

	preverbal objects	postverbal objects
Homer	4832	2288
Herodotus	790	739
the New Testament	979	2066

Table 3. Figures for non-co-occurring preverbal and postverbal accusative object nouns and pronouns

vides figures only for the first Book of the Histories, and it is not clear to me whether, as seems probable, he only analyzed single and coordinate independent clauses).

It is possible that further research will show that other kinds of texts of the Classical age had a marked preference for one order or the other, as occurs in Homer and the New Testament. Notwithstanding, my data prove that Classical Greek admitted at least a variety of the language (i.e., the literary one of historical texts) in which OV and VO orders are virtually equally licensed. Although the figures for the three texts clearly suggest a change from OV to VO order (which has always been assumed but never adequately justified), further texts will have to be analyzed to properly understand the conditions of this change, (e.g., how the genre of a text influenced word order).

The figures given do not explain why accusative object nouns and pronouns can be preverbal or postverbal: as it is known, this question can be successfully explored only in the light of studies on information structure: word order in AG is heavily determined by the categories of topic and focus and the phonology associated to them (this is prototypically shown by the distinction between nouns and pronouns; see Celano (2013a,b) and the bibliography therein). Notwithstanding, scientific quantification does prove to be a necessary means to guide research by exactly describing the language: the present analysis suggests that a word order change from OV to VO occurred over time.

6. Conclusions

In this article I have quantified the number of verbs with preverbal or postverbal object nouns or pronouns in single or coordinate independent clauses in Homer's Iliad and Odyssey, and the annotated parts of Herodotus' Histories and the New Testament. The results show that over time there was a shift from OV to VO order in AG. Although this has already been argued, the present research turns out to be the first attempt to quantify AG word order scientifically.

Dedication

Dedico questo articolo al Prof. Gregory Crane, difensore e promotore degli studi classici nel mondo.

Bibliography

- Bamman, David and Gregory Crane. Guidelines for the syntactic annotation of the Ancient Greek Dependency Treebank (1.1), 2008. URL <http://nlp.perseus.tufts.edu/syntax/treebank/agdt/1.7/docs/guidelines.pdf>.
- Celano, Giuseppe Giovanni Antonio. Argument-focus and predicate-focus structure in Ancient Greek: Word order and phonology. *Studies in Language*, 37(2):241–266, 2013a. URL <https://benjamins.com/#catalog/journals/sl.37.2.01cel/details>.
- Celano, Giuseppe Giovanni Antonio. Word order. In *Encyclopedia of Ancient Greek language and linguistics*. Brill Online, 2013b. URL http://brillonline.nl/entries/encyclopedia-of-ancient-greek-language-and-linguistics/word-order-EAGLL_COM_00000378?s.num=0&s.q=word+order.
- Cervin, Richard Stuart. *Word order in Ancient Greek: VSO, SVO, SOV, or all of the above?* PhD thesis, University of Illinois at Urbana-Champaign, 1990.
- Dunn, Graham. Syntactic word order in Herodotean Greek. *Glotta*, 66:63–79, 1988.
- Ebeling, H. L. Some statistics on the order of words in Greek. In *Studies in Honor of Basil L. Gildersleeve*, pages 229–240. Johns Hopkins University Press, Baltimore, 1902.
- Friederich, Paul. Proto-Indo-European syntax: The order of meaningful elements. *Journal of Indo-European Studies*, 1, 1975.
- Hajičová, Eva, Zdeněk Kirschner, and Petr Sgall. A manual for analytic layer annotation of the Prague Dependency Treebank (English translation), 1999. URL <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html>.
- Haug, Dag Trygve Truslew. Proiel guidelines for annotation, 2010. URL http://folk.uio.no/daghaug/syntactic_guidelines.pdf.
- Kwong, Ivan Shing Chung. *The word order of the gospel of Luke: Its foregrounded messages*. T&T Clark, London, 2005.
- Sgall, Petr, Eva Hajičová, and Jarmila Panevová. *The meaning of the sentence in its semantic and pragmatic aspects*. Reidel Publishing Company and Academia, Dordrecht & Prague, 1986.

Taylor, Ann. The change from SOV to SVO in Ancient Greek. *Language Variation and Change*, 6 (1):1–37, 1994.

Address for correspondence:

Giuseppe G. A. Celano
giuseppe.celano@tufts.edu
Tufts University
Perseus Project/Department of Classics
134C Eaton Hall
Medford, MA 02155, USA