

A Survey of Phrase Projectivity in *Antigone*

Jonathan Sterling

April 2013

In this paper, I will show how phrase projectivity (which corresponds to lacking hyperbaton) is linked to register and meter in Sophocles’s *Antigone*, by developing a quantitative metric for projectivity and comparing it across lyrics, trimeters and anapaests.

1 Dependency Trees and Their Projectivity

A dependency tree encodes the head-dependent relation for a string of words, where arcs are drawn from heads to their dependents. We consider a phrase *projective* when these arcs do not cross each other, and *discontinuous* to the extent that any of the arcs intersect. Figure 1 illustrates the various kinds of projectivity violations that may occur.

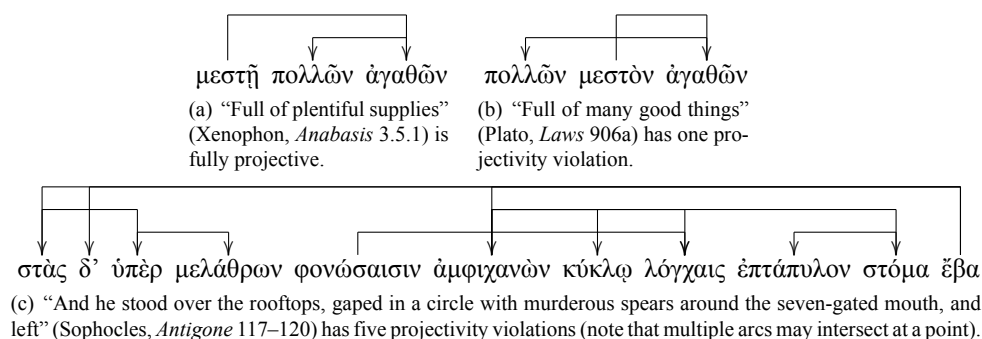


Figure 1: A dependency path wrapping around itself is a projectivity violation, as in (b); interlacing adjacent phrases also violate projectivity, as in (c). Examples (a–b) drawn from Devine & Stephens.

In this paper, we shall use a concrete metric of projectivity, ω , given by the following ratio:

$$\omega = \frac{\text{number of violations}}{\text{number of arcs}}$$

The number of violations is simply the total number of intersections that occur in a tree. Appendix A deals with the development of a model and algorithm in the programming language Haskell to compute this quantity for a particular dependency tree.

2 The Perseus Treebank

The Perseus Ancient Greek Dependency Treebank is a massive trove of annotated texts that encode the all dependency relations in every sentence. The data is given in an XML (Extensible Markup Language) format resembling the following:

```
<sentence id="2900759">
  <word id="1" form="χρῆ" lemma="χρῆ" head="0" />
  <word id="2" form="δὲ" lemma="δὲ" head="1" />
  . . .
</sentence>
```

Every sentence is given a unique, sequential identifier; within each sentence, every word is indexed by its linear position and coreferenced with the linear position of its dominating head. In the case of the data for *Antigone*, the maximal head of each sentence has its own head given as 0. Appendix B deals with parsing these XML representations into dependency trees for which we can compute ω .

3 Projectivity in Antigone

To observe the variation of projectivity within a text, then, one may make a selection of sentences that have something in common (such as meter), compute their trees and thence derive ω , and then average the results. Then that quantity may be compared with that of other selections.

To that end, I have selected passages from *Antigone* and organized them by type. Table 1 enumerates the lyric passages of the play, along with their computed mean ω values, and a final mean of means with the standard deviation of the set. Table 2 does the same for anapaests. Lastly, Table 3 gives the data for dialogue (which is in iambic trimeters), divided between medium-to-long speeches and stichomythia.

As can be seen from the data, lyrics have the highest degree of non-projectivity, followed by speeches, then anapaests, and then stichomythia. However, the standard deviation of the ω for anapaestic passages is so high that it may be difficult to say much of interest about them at all in respect to the questions that we are considering. So, let

Lines		ω
100 ... 154	<i>First choral ode</i>	0.64
332 ... 375	<i>Second choral ode</i>	0.40
583 ... 625	<i>Third choral ode</i>	0.44
781 ... 800	<i>Fourth choral ode</i>	0.43
806 ... 816	<i>Antigone's Kommos</i>	1.24
823 ... 833	<i>Antigone's Kommos (cntd.)</i>	0.61
839 ... 882	<i>Antigone's Kommos (cntd.)</i>	0.47
944 ... 987	<i>Fifth choral ode</i>	0.34
1116 ... 1152	<i>Sixth choral ode</i>	0.60
1261 ... 1269	<i>Kreon's Kommos</i>	0.28
1283 ... 1292	<i>Kreon's Kommos (cntd.)</i>	0.77
1306 ... 1311	<i>Kreon's Kommos (cntd.)</i>	0.29
1317 ... 1325	<i>Kreon's Kommos (cntd.)</i>	0.98
1239 ... 1246	<i>Kreon's Kommos (cntd.)</i>	0.44
mean $\omega = 0.57$, sdev = 0.27		

Table 1: Lyrics, including odes and kommoi.

Lines		ω
155 ... 161	<i>Kreon's Entrance</i>	0.17
376 ... 383	<i>Antigone's Entrance</i>	0.62
526 ... 530	<i>Ismene's Entrance</i>	0.05
626 ... 630	<i>Haimon's Entrance</i>	0.44
801 ... 805	<i>Antigone's Entrance</i>	1.08
817 ... 822	<i>Chorus to Antigone</i>	0.46
834 ... 838	<i>Chorus to Antigone</i>	0.03
929 ... 943	<i>Chorus, Kreon and Antigone</i>	0.25
1257 ... 1260	<i>Chorus before Kreon's Kommos</i>	0.00
1347 ... 1353	<i>Final anapaests of the Chorus</i>	0.36
mean $\omega = 0.35$, sdev = 0.33		

Table 2: Anapaests.

(a) Speeches and Dialogue

Lines		ω
162...210	Kreon: <i>ἄνδρες, τὰ μὲν δὴ...</i>	0.25
249...277	Guard: <i>οὐκ οἶδ'· ἐκεῖ γὰρ οὔτε...</i>	0.34
280...314	Kreon: <i>παῦσαι, πρὶν ὀργῆς...</i>	0.39
407...440	Guard: <i>τοιούτον ἦν τὸ πρᾶγμ'...</i>	0.40
450...470	Antigone: <i>οὐ γὰρ τί μοι Ζεὺς...</i>	0.39
473...495	Kreon: <i>ἀλλ' ἴσθι τοι...</i>	0.49
639...680	Kreon: <i>οὔτω γὰρ, ὦ παῖ...</i>	0.44
683...723	Haimon: <i>πάτερ, θεοὶ φύουσιν...</i>	0.38
891...928	Antigone: <i>ὦ τύμβος, ὦ νυμφεῖον...</i>	0.31
998...1032	Teiresias: <i>γνώση, τέχνης σημεῖα...</i>	0.35
1033...1047	Kreon: <i>ὦ πρέσβυ, πάντες...</i>	0.18
1064...1090	Teiresias: <i>ἀλλ' εὔ γέ τοι...</i>	0.64
1155...1172	Messenger: <i>Κάδμου πάροικοι καὶ...</i>	0.35
1192...1243	Messenger: <i>ἐγώ, φίλη δέσποινα...</i>	0.27
mean ω = 0.37, sdev = 0.11		

(b) Stichomythia

Lines		ω
536...576	<i>Ismene, Antigone and Kreon</i>	0.23
728...757	<i>Haimon and Kreon</i>	0.27
991...997	<i>Kreon and Teiresias</i>	0.53
1047...1063	<i>Kreon and Teiresias</i>	0.09
1172...1179	<i>Chorus and Messenger</i>	0.16
mean ω = 0.26, sdev = 0.17		

Table 3: Dialogue (Trimeters)

us put the anapaests aside for the moment and deal exclusively with lyrics, speeches and stichomythias.

Whereas in prose, hyperbaton corresponds to *strong focus*, which “does not merely fill a gap in the addressee’s knowledge but additionally evokes and excludes alternatives” (Devine & Stephens 303), hyperbaton in verse only entails weak focus, which emphasizes but does not exclude (ibid. 107).

As a result, hyperbaton in verse may be used to evoke a kind of elevated style without incidentally entailing more emphasis and other pragmatic effects than intended. And so it should not be surprising that lyric passages, which reside in the most poetic and elevated register present in tragic diction, should have proved in *Antigone* to have the highest proportion of projectivity violations.

Appendices

A Algorithm & Data Representation

Dependency trees are a recursive data structure with a head node, which may have any number of arcs drawn to further trees (this is called a *rose tree*). We represent them as a Haskell data-type as follows:

```
data Tree  $\alpha$  =  $\alpha \curvearrowright$  [Tree  $\alpha$ ]
```

This can be read as “For all types α , a Tree of α is constructed from a *label* of type α and a *subforest* of Trees of α ,” where brackets are a notation for lists.

Given a tree, we can extract its root label or its subforest by pattern matching on its structure as follows:

```
getLabel :: Tree  $\alpha$   $\rightarrow$   $\alpha$   
getLabel ( $l \curvearrowright \_$ ) =  $l$   
getForest :: Tree  $\alpha$   $\rightarrow$  [Tree  $\alpha$ ]  
getForest ( $\_ \curvearrowright ts$ ) =  $ts$ 
```

A.1 From Edges to Trees

We shall consider each word index to be a *vertex*, and each pair of vertices to be an Edge, which we shall write as follows:

```
data Edge  $\alpha$  =  $\alpha \leftrightarrow \alpha$  deriving Eq
```

An Edge α is given by two vertices of type α ; the **deriving** Eq statement generates the code that is necessary to determine whether or not two Edges are equal using the (\equiv) operator. In order to perform our analysis, we should wish to transform the raw list of edges into a tree structure. The basic procedure is as follows:

First, we try to find the root vertex of the tree. This will be a vertex that is given as the head of one of the words, but does not itself appear in the sentence:

```
rootVertex :: Eq  $\alpha \Rightarrow$  [Edge  $\alpha$ ]  $\rightarrow$  Maybe  $\alpha$   
rootVertex  $es$  = find ( $\notin$  deps) heads where  
  heads = [ ( $\lambda(x \leftrightarrow y) \rightarrow x$ )  $es$  ]  
  deps  = [ ( $\lambda(x \leftrightarrow y) \rightarrow y$ )  $es$  ]
```

If the data that we are working with are not well-formed, there is a chance that we will not find a root vertex; that is why the type is given as `Maybe`.

Then, given a root vertex, we look to find all the edges that it touches, and try to build the subtrees that are connected with those edges.

```
onEdge :: Eq α ⇒ α → Edge α → ℬ
onEdge i (x ↔ y) = x ≡ i ∨ y ≡ i

oppositeVertex :: Eq α ⇒ α → Edge α → α
oppositeVertex i (x ↔ y)
  | x ≡ i      = y
  | otherwise = x
```

This is done recursively until the list of edges is exhausted and we have a complete tree structure:

```
treeFromEdges :: Ord α ⇒ [Edge α] → Maybe (Tree α)
treeFromEdges es = [(buildWithRoot es) (rootVertex es)] where
  buildWithRoot es root = root ∘ sortedChildren where
    roots      = [(oppositeVertex root) localVertices]
    children   = [(buildWithRoot foreignVertices) roots]
    localVertices = filter (onEdge root) es
    foreignVertices = filter (¬ ∘ onEdge root) es
    sortedChildren = sortBy (compare `on` getLabel) children
```

A.2 Counting Violations: Computing ω

Violations are given as an integer tally:

```
type Violations = ℤ
```

The basic procedure for counting projectivity violations is as follows: flatten down the tree into a list of edges cross-referenced by their vertical position in the tree; then traverse the list and see how many times these edges intersect each other.

```
type Level = ℤ
```

The vertical position of a node in a tree is represented as its `Level`, counting backwards from the total depth of the tree. That is, the deepest node in the tree is at level 0, and the highest node in the tree is at level n , where n is the tree's depth.

```

levels :: Tree  $\alpha$   $\rightarrow$   $[[\alpha]]$ 
levels  $t$  = fmap (fmap getLabel) $
  takeWhile ( $\neg \circ$  null) $
    iterate ( $\gg$  getForest) [ $t$ ]

```

```

depth :: Tree  $\alpha$   $\rightarrow$   $\mathbb{Z}$ 
depth = length  $\circ$  levels

```

We can now annotate each node in a tree with what level it is at:

```

annotateLevels :: Tree  $\alpha$   $\rightarrow$  Tree (Level,  $\alpha$ )
annotateLevels  $tree$  = aux (depth  $tree$ )  $tree$  where
  aux  $l$  ( $x \rightsquigarrow ts$ ) = ( $l, x$ )  $\rightsquigarrow$   $[(aux (l - 1)) ts]$ 

```

Then, we fold up the tree into a list of edges and levels:

```

allEdges :: Ord  $\alpha$   $\Rightarrow$  Tree  $\alpha$   $\rightarrow$  [(Level, Edge  $\alpha$ )]
allEdges  $tree$  = aux (annotateLevels  $tree$ ) where
  aux ( $(-, x) \rightsquigarrow ts$ ) =  $ts \gg$  go where
    go  $t@((l, y) \rightsquigarrow -)$  = ( $l, edgeWithRange [x, y]$ ) : aux  $t$ 

```

```

edgeWithRange :: (Ord  $\alpha$ )  $\Rightarrow$  [ $\alpha$ ]  $\rightarrow$  Edge  $\alpha$ 
edgeWithRange  $xs$  = minimum  $xs \leftrightarrow$  maximum  $xs$ 

```

A handy way to think of edges annotated by levels is as a representation of the arc itself, where the vertices of the edge are the endpoints, and the level is the height of the arc.

If one end of an arc is between the ends of another, then there is a single intersection. If one arc is higher than another and the latter is in between the endpoints of the former, there is no violation; but if they are at the same level, or if the latter is higher than the former, there is a double intersection. Otherwise, there is no intersection.

```

checkEdges :: Ord  $\alpha$   $\Rightarrow$  (Level, Edge  $\alpha$ )  $\rightarrow$  (Level, Edge  $\alpha$ )  $\rightarrow$  Violations
checkEdges ( $l, xy@(x \leftrightarrow y)$ ) ( $l', uv@(u \leftrightarrow v)$ )
  |  $y \in_E uv \wedge u \in_E xy$            = 1
  |  $x \in_E uv \wedge v \in_E xy$            = 1
  |  $x \in_E uv \wedge y \in_E uv \wedge l \geq l'$  = 2
  |  $u \in_E xy \wedge v \in_E xy \wedge l \leq l'$  = 2
  | otherwise                         = 0

```


We determine whether a vertex is in the bounds of an edge using $\cdot \in_E \cdot$.

$$\begin{aligned} \cdot \in_E \cdot &:: \text{Ord } \alpha \Rightarrow \alpha \rightarrow \text{Edge } \alpha \rightarrow \mathbb{B} \\ z \in_E x &\leftrightarrow y = z > \text{minimum } [x, y] \\ &\quad \wedge z < \text{maximum } [x, y] \end{aligned}$$

We can now use what we've built to count the intersections that occur in a collection of edges. This is done by adding up the result of `checkEdges` of the combination of each edge with the subset of edges which are at or below its level:

$$\begin{aligned} \text{edgeViolations} &:: \text{Ord } \alpha \Rightarrow [(\text{Level}, \text{Edge } \alpha)] \rightarrow \text{Violations} \\ \text{edgeViolations } xs &= \text{sum } \llbracket \text{violationsWith } xs \rrbracket \textbf{ where} \\ \text{rangesBelow } (l, _) &= \text{filter } (\lambda(l', _) \rightarrow l' \leq l) \ xs \\ \text{violationsWith } x &= \text{sum } \llbracket (\text{checkEdges } x) (\text{rangesBelow } x) \rrbracket \end{aligned}$$

Finally, ω is computed for a tree as follows:

$$\begin{aligned} \omega &:: \text{Ord } \alpha \Rightarrow \text{Tree } \alpha \rightarrow \mathbb{Q} \\ \omega \text{ tree} &= \frac{\text{edgeViolations edges}}{\text{length edges}} \textbf{ where} \\ \text{edges} &= \text{allEdges tree} \end{aligned}$$

B Parsing the Perseus Treebank

We can express the general shape of such a document as follows:

```
type Document = [Sentence]
data Sentence = Sentence { sentenceId :: ℤ, sentenceEdges :: [Edge ℤ] } deriving Show
```

To construct a Document from the contents of an XML file, it suffices to find all of the sentences.

```
documentFromXML :: [Content] → Document
documentFromXML xml = catMaybes [ sentenceFromXML elems ] where
  elems = onlyElems xml >>= findElements (simpleName "sentence")
```

Sentences are got by taking the contents of their `id` attribute, and extracting edges from their children.

```
sentenceFromXML :: Element → Maybe Sentence
sentenceFromXML e = [ Sentence (readAttr "id" e) (pure edges) ] where
  edges = catMaybes [ edgeFromXML children ]
  children = findChildren (simpleName "word") e
```

An edge is got from an element by taking the contents of its `id` attribute with the contents of its `head` attribute.

```
edgeFromXML :: Element → Maybe (Edge ℤ)
edgeFromXML e =
  case findAttr (simpleName "form") e of
    Just x | x ∈ [ ". ", " ", " ", " ", " ", " ", " " ] → Nothing
    otherwise → [ (readAttr "head" e) ↔ (readAttr "id" e) ]
```

Thence, turn a sentence into a tree by its edges using the machinery from Section A.1.

```
treeFromSentence :: Sentence → Maybe (Tree ℤ)
treeFromSentence (Sentence _ ws) = treeFromEdges ws
```

By applying `treeFromSentence` to every sentence within a document, we can generate all the trees in a document.

```
treesFromDocument :: Document → [Tree ℤ]
treesFromDocument ss = catMaybes [ treeFromSentence ss ]
```

By combining the above, we also may derive a document structure from a file on disk.

```
documentFromFile :: FilePath → IO Document
documentFromFile path = [ (documentFromXML ∘ parseXML) (readFile path) ]
```

C Analysis of Data

We compute the mean ω of the trees contained in a document as follows:

```
analyzeDocument :: Document → ℚ
analyzeDocument doc = mean [ ω (treesFromDocument doc) ]
```

We will wish to compare the ω for parts of *Antigone*. A section is given by a two sentence indices (a beginning and an end):

```
data Section = ℤ ··· ℤ
```

Then, the entire document can be cut down into smaller documents by section:

```
restrictDocument :: Section → Document → Document
restrictDocument (start ··· finish) = filter withinSection where
  withinSection (Sentence i _) = i ≥ start ∧ i ≤ finish
```

Auxiliary Functions

```
simpleName :: String → QName
simpleName s = QName s Nothing Nothing

readAttr :: Read α ⇒ String → Element → Maybe α
readAttr n = fmap read ∘ findAttr (simpleName n)
```

```
mean :: Fractional n ⇒ [n] → n
mean = [ sum / length ]
```

```
sdev :: Floating n ⇒ [n] → n
sdev xs = √ (sum [ (λx → x2) [ -(mean xs) + xs ] ] / (length xs - 1))
```