Syntactic Complexity in Virgil's Aeneid

Russell Bateman
Syntactic Complexity in Virgil's *Aeneid*

Russell Bateman
Computational Linguist
ALPNET, Inc.
(formerly Automated Language Processing Systems, Inc.)
Provo, Utah

One is struck in the reading of Latin poetry by the sheer syntactic diversity of expression with which the author hopes to control what the reader is feeling. The use of some if not most rhetorical devices and syntactic liberty-taking may have an impact on the reader's immediate comprehension. In the case of Virgil’s *Aeneid*, it has been posited that a passage of emotional content will often be fraught with syntactic difficulty or conversely that a passage containing difficult syntax will betray emotional content. Challenged by this suggestion, I set out to determine whether there might be some substance to it and to see if I could evolve a simple computational method of identifying syntactic complexity.

What Is Syntactic Complexity?

After the previously mentioned suggestion, I become more than vaguely aware as I read that some passages of the *Aeneid* posed more problems than others in comprehension. I began to contemplate different syntactic phenomena and how they might relate to this issue. My initial list included:

- Sentence length
- Ratio of different syntactic categories to the total count of words
- Ratio of modifiers to substantives
- Hypotactic content (relative, subordinate constructions, indirect discourse)
- Parataxis, particularly asyndeton\(^1\)
- Subject, verb and object placement (SVO, SOV, etc.)
- Conservation of the verb *sum*
- Abundance or lack of conditional constructions
- Rhetorical devices\(^2\) which might tend to complicate the text or to render it less readable such as anacoluthon, apostrophe, breviloquenta, chiasmus\(^3\), ellipsis, hyperbaton, synchysis and others which might make it more readable like anaphora and pleonasm.

Given the daunting nature of the task and these virtually limitless parameters, I eliminated a number of items which seemed too great to accomplish. Too, I wanted to succeed in establishing a view of syntactic difficulty which would be based on simple, easily understandable and demonstrable concepts.

---

\(^1\) Or the lack of conjunctions.

\(^2\) An exhaustive list of these, including definitions and examples, may be found in Pharr, Clyde, *Virgil's Aeneid, Books I-VI*, Grammatical Appendix, items 411—447 (1964, D.C. Heath and Company).

\(^3\) And other liberties with word order and collation.
I chose therefore: the storm stirred up by Juno and Aeolus in the beginning of Book I, Sychaeus' death at the hand of Dido's brother Pygmalion, an excerpt from Sinon in Book II as he weaves his story of his relation to Palamedes, the gobbling-up of Laocoön and his two sons by the sea serpents, the episode where Aeneas returns to Troy in search of his wife Creūsa whom he forgot in his haste to flee the city, and in Book IV, one of Dido's interminable and delightfully contorted soliloquies, the one where she berates Aeneas for thinking of leaving her. Together, these passages represent some 120 lines or about 800 words of syntactically complete text.

Additionally, I chose three passages of at least 100 lines each\(^{11}\) to serve as control to this experiment. Except to exclude the previous six passages, I placed no criterion on their choosing other than the same familiarity of the original subject passages. I took the first from Book II:234-335, which describes the opening of the walls of Troy to the horse, the partying, the city going to sleep, Hector's apparition to Aeneas and the beginning of the sack. The second was chosen from Book IV:522-627, which treats the night before Aeneas' departure from Carthage including another example of Dido's spectacular soliloquies where she tries to reason her way out of fate,\(^{12}\) the visit to the sleeping Aeneas by Mercury and yet another voyage through the raging woman's mind. Last, I chose from Book VI:212-312, a passage dealing with Misenus' proper funeral, the nocturnal sacrifices and the plunging of Aeneas and the Sibyl into Hades and on to their encounter with the ferryman Charon.

**Conclusions**

Following are the statistical results which I will attempt to interpret in conclusion. Here, "parataxis" means nothing more than a count of coordinating conjunctions and "hypotaxis" a count of relative pronouns and subordinating conjunctions.

<table>
<thead>
<tr>
<th></th>
<th>Aeneid as a whole</th>
<th>Control passages</th>
<th>Subject passages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>64215</td>
<td>2048</td>
<td>804</td>
</tr>
<tr>
<td>per full stop</td>
<td>19.01</td>
<td>14.84</td>
<td>26.51 (a)</td>
</tr>
<tr>
<td>per half stop</td>
<td>13.27</td>
<td>11.44</td>
<td>15.02 (b)</td>
</tr>
<tr>
<td><strong>Parataxis</strong>, per sentence</td>
<td>2.05</td>
<td>1.71</td>
<td>2.47 (c)</td>
</tr>
<tr>
<td>per 100 words</td>
<td>10.76</td>
<td>11.52</td>
<td>10.10 (d)</td>
</tr>
<tr>
<td><strong>Hypotaxis</strong>, per sentence</td>
<td>0.37</td>
<td>0.27</td>
<td>0.53 (e)</td>
</tr>
<tr>
<td>per 100 words</td>
<td>1.93</td>
<td>1.81</td>
<td>1.99 (f)</td>
</tr>
<tr>
<td><strong>Substantives</strong> per 100 words</td>
<td>—</td>
<td>32.47</td>
<td>33.67 (g)</td>
</tr>
<tr>
<td><strong>Modifiers</strong> per 100 words</td>
<td>—</td>
<td>20.31</td>
<td>23.85 (h)</td>
</tr>
<tr>
<td><strong>Modifiers/Substantives</strong></td>
<td>—</td>
<td>0.63</td>
<td>0.64 (i)</td>
</tr>
</tbody>
</table>

The simplest and most easy phenomenon to understand is the number of words per sentence (a and b) and indeed the results show that sentence length may be the most identifiable feature of complex syntax. Due to the greater length of the sentences, one immediately suspects that Virgil will be employing all manner of rhetorical devices and

---

\(^{11}\) For a total of 2048 words.

\(^{12}\) This may have been better off in the set of experimental passages or excluded from the control set.
word relationships which will contribute to a naturally longer—and as we’ll see—more syntactically complex sentence.

That a greater number of coordinating conjunctions (e) occurred in the subject passages did not surprise me given that the sentences tend to be longer. I took pleasure in noting that on a per word basis (d), it was actually somewhat lower—confirming my suspicion of asyndeton or other some other reason for the omission of conjunctions. Sometimes these are replaced by editors with semicolons,13 but in spite of this and because of my “half stop” counts,14 the statistics show that even this device is not completely responsible for the reduced number. Clearly, the passages are heavier for lack of coordinating conjunctions.

I suspect that homography with interrogatives is at fault for the questionable figures15 for subordinate construction (e and f). It appears that there tends to be a bit more of this in the emotional passages—a more sophisticated syntactic expression. I regard the control passage numbers as more accurate. Perhaps an adjustment to the software to avoid counting words found in initial position would yield more correct and contrasting results. There was little interrogation though a fair number of relative connectives in the subject passages.

The ever-so-slightly higher ratio of substantives and modifiers (g and h) to other categories was predictable and equates directly with sentence length (a and b). For me however, the ambivalence of the ratio (i) yields credibility to my experiment because initial observations had led me to hope for a clear divergence here.16 That this was not ultimately born out caused me great though temporary concern for I feared that the loss of this “flagship of my hypothesis” would invalidate the entire undertaking and leave me with nothing conclusive to report. Moreover, those counts took the lion’s share of my time since the computer could not perform them. In fact, this caused me to look all the more deeply into the figures I did obtain and I drew conclusions which now seem entirely more reasonable and likely.

Although I have not hit upon an interpretation which will allow me to enumerate unequivocally the elements of Virgil’s syntactic complexity, I see vivid symptoms from this exercise. It now seems clear that the answer lurks in the author’s heavier grammatical relationships forced by a lack of conjunction or rather, that his grammatical device precluded the use of conjunctions in much the same way we in English and French add subordination and otherwise lengthen our sentences in situations where we want to raise our register of speech. In contrast, an interesting question spawned by this conclusion is why we do just the opposite from Virgil when we become emotional; that is, why do we reduce our speech to a rush of short and highly paratactic emissions?

---

13 I do not propose however nor do I presume to question the current punctuation in the framework of this undertaking.

14 Except for words, the “half stop” statistics are not reported in this condensation of the final numbers. Please refer to the appendix for this.

15 C’est-à-dire, that the numbers for the Aeneid as a whole and for the control passages should be much closer. My suspicion is that somewhere there is an especially great number of interrogatives contributing to such a high ratio for the work as a whole.

16 Indeed, I reported in a short essay in December 1988 covering initial work in Books I and II of the Aeneid that one of the symptoms of complexity would be a higher ratio of adjectives to nouns in the subject passages over the control text.
Methodology

A preliminary process was drawn up and applied experimentally to a small sampling of Books I and II. Out of these experiments quickly emerged encouraging results—confirming that there might indeed be something to the complexity hypothesis. Imposing certain limits and refining some of the steps, I arrived at an easily implementable methodology which consisted mostly of maintaining a set of counters relating to the following aspects:

- **Substantive count.** This count was increased each time a noun was encountered. Excluded were those nouns in the genitive case (voir plus bas). When an adjective was found to be used substantively, it figured in the count also. All pronouns were ignored.

- **Modifier count.** Each time any adjective or noun in the genitive case occurred, this count was increased. Included were pronouns when used as determiners, e.g.: *illo Hectore.* A fine distinction was be made between participles which were more adjectival than verbal in sense.

- **Hypotaxis.** This count was simply incremented each time a subordinating conjunction or relative pronoun was found.

- **Parataxis.** Each coordinating conjunction was counted as it was encountered (*-que or -ue* at the end of words were not forgotten).

In accordance with the foregoing methods and as the subscripts indicate, consider the following three excerpts as examples in applying the methodology:

\[ O \text{ terque}_4 \text{ quaterque}_4 \text{ beati}_1 \]
\[ \text{quis}_3 \text{ ante ora}_1 \text{ patrum}_2 \text{ Troiæ}_2 \text{ sub mœnibus}_1 \text{ altis}_2 ... \]
\[ \text{raptatus}_2 \text{ bigis}_1 \text{ ut quondam, ...} \]
\[ \text{uel Danaum}_2 \text{ Phrygios}_2 \text{ iaculatus}_2 \text{ puppibus}_1 \text{ ignis}_1 ! \]

I thought it might be useful for the purposes of the word-to-sentence ratios to break sentences not only in the traditional places marked by the period and the question and exclamation marks (called *full stops* in some terminologies), but also at what I will call, for want of a better term, *half stops* including the colon and semicolon which always fall at complete syntactic pauses. The results of this appeared in the larger set of reported statistics and confirmed what was already born out through the normal breaks.

To speed up the process and enable it to include as much of the *Aeneid’s* text as possible, the implementation of the counting process was divided between that which could be done by the computer and that which required biological intelligence. Of course, it was necessary that the software require only a minimal amount of time spent in design and implementation of its programs. An eye was also to be kept toward moving more and

---

4 A nominalized adjective.

5 This may have been a mistake though it ultimately eliminated some subjective decision making in the counting process.

6 This encourages the thought that a mistake was made previously in leaving out pronouns although few instances of this actually occurred.

7 This was occasionally difficult—and close attention was necessarily paid to this issue.
more of the computational process from intelligent activity to the machine as increased understanding of the problem permitted.

The machine\(^8\) first began counting only words and sentences. All ratios could of course be produced quickly and accurately on the computer. Then, by brute force,\(^9\) coordinating and subordinating conjunctions as well as relative pronouns were counted this way since they all exist in lexically finite sets and since due to time constraints—as mentioned earlier—I gave up much of what I wanted to include in counting instances of hypotaxis, particularly indirect discourse. In fact, hypotaxis became too big a concept considering what was ultimately placed under this heading. It remains however a useful alias for two other simple but very long terms: subordinating conjunction and relative pronoun. Similarly, though more accurately, parataxis ended up only including a count of coordinating conjunctions.

For lack of a dictionary as well as a morphological reduction and look-up capability, there remained the task of counting substantives and modifiers. This seems simple at first glance: simply count nouns and adjectives. But both notions become more complex considering the vast verbal system of Latin. Still, a simple approach needed to be found in order to complete the study and be able to interpret the results.

A number of practical difficulties plagued me as I made the counts. It was often hard in spite of my familiarity with the text,\(^10\) to determine when to count a present or past participle as a modifier and when to leave it out as a verb. By convention, one is encouraged to construe Latin participles as clauses, but since such a construction might yield a clause simply containing a copula (when not deponent of course in the case of the past participle), it still nevertheless passes as a modifier. On the other hand, similar clausal phenomena, among them relative clauses of characteristic, etc., were not counted as modifiers since this violated the criterion of simplicity. In view of this, there was a tendency to leave out many past participles as well. Ultimately, I made the choice based on how “adjectival” I felt the form was at the moment of choice and left it at that.

Other problems surfaced in the computer program I wrote to make the easy counts and later the subordinate construction counts. Some errors were almost certainly produced by over-detection of *quis* and its composites in the case where it has supplanted *aliquis*. Interrogative pronouns when homographic with relatives and subordinating conjunctions caused the program to err a bit in favor of hypotaxis. All errors committed by the software were done so uniformly across the entire *Aeneid* however.

**Choice of Passages**

I selected six passages of greatly differing length in the most subjective manner possible, namely when I felt that I had had a difficult time deciphering the passage or when I felt that it had good potential as an emotional passage. While there is perhaps some flaw in the latter with respect to scientific method, I was impatient to see whether statistical analysis of the passage supported the conclusions I hoped to draw.

---

\(^8\) The main program and all supporting software was implemented in the C programming language under SunOS (a derivative of Berkeley UNIX) on a Sun Microsystems 3/50\(^\text{TM}\) using a corrected version of the IBYCUS text.

\(^9\) In this case, simple *boolean* searching.

\(^10\) I used no passages which I hadn’t read at least two or three separate times nor with which I felt any insecurity.
Complete statistical output for the *Aeneid* as a whole and for each book and the passages examined in this paper, as well as the C source code listings which are too long to reproduce here, are available at no charge from the author.

Russell Bateman  
742 West 400 North  
Provo, Utah 84601  
(801) 375-8733