This Great Flood of Words*:

A Brief Phasal Analysis of the Dialogue between Socrates and Thrasymachus

Thomas Broadwater

GREK 8010: Readings and Research in Greek Literature

April 26, 2022

* Rep. 1: 344 d1

Goals and Methodology

This study aims to establish a high-level phasal structure of the dialogue between Socrates and Thrasymachus in Book 1 of Plato's Republic, as translated by G.M.A. Grube and revised by C.D.C. Reeve. The elenctic style of the argument presented in Plato's writing involves metatextual details as much as the arguments themselves, using the characterization of the interlocutors to position these individuals' stances relative to one another. This sort of metatextual characterization is apparent in Thrasymachus' very name: ϑρασύ-μαχ-ο-ς: bold-in-battle. (Brill, SV ϑράσος, ϑάρσος; cf. Ther. Θhαρρύμαϙhος) A simple reading name-reading prepares the audience for the rough treatment to come.

Deeper characterizations abound. "Give an answer yourself, and tell us what you say the just is," he demands of Socrates. "And don't tell me it's the right, the beneficial, the profitable, the gainful or the advantageous, but tell me clearly and exactly what you mean; for I won't accept such nonsense from you." (336 c4-d3) He cannot resist mockery and derision even in his most reasonable requests. He knows the correct answer — his correct answer — and all else is so much nonsense.

Socrates, for his part, replies with his fateful sarcasm as Thrasymachus resses him for an answer. When Thrasymachus, gloating, accuses Socrates of being deceitful (337 a3, a5) and claims to have foretold such an occasion, Socrates cedes this conflict, but only with a barb meant to draw Thrasymachus into the dialectic. "That's because you're a clever fellow, Thrasymachus." (337 a7) One must, of course, never assume themselves clever around Socrates.

There are means of deepening this sort of analysis. By examining phrases and the specifics of their wording — the chosen vocabulary, word arrangement, and length of sentence, among others — emergent patterns are made to reveal themselves. These patterns recontextualize the piece, demarcating interwoven segments called phases.

This study will follow the dialogue as presented, with all narrative elements removed. The piece is framed as a direct quote from Socrates as he retells a previous conversation and so includes occasional narrative asides meant to guide the reader. These include shorter phrases such as "he said" and "he replied," as well as lengthier elements, such as the description of Thrasymachus blushing. While certainly part of a higher-level discourse, they impede any discourse analysis and so are removed from the data.

Five individuals participate in this dialogue: Cleitophon, Glaucon, Polemarchus, Socrates, and Thrasymachus. The speakers typically take turns speaking with Socrates, but not exclusively. For example, there is a short digression between Cleitophon and Polemarchus as they attempt to pin down the specific meaning of Thrasymachus' stance on the nature of justice (340 a1 - 340 b9), which Socrates quickly stops. Glaucon also enters the conversation at two points, once addressing Thrasymachus in support of Socrates (337 d7 - d8) and later asking Socrates for clarification on a point (347 a7 - 348 b6). These asides have little impact on the dialogue, but they are undeniably a part of it and are retained in the data.

Whenever a sentence is quoted directly, it will be presented in a specific notation from this point on. First will come the Stephanus Pagination. Second will come a token specifying the speaker and the turn. For example, Glaucon's first turn will have the token G1, his second G2. Third will come a letter indicating the sentence in the turn. For example, the first sentence will be labeled a./, the second b./, and so on. Last will come the text. To exemplify, Socrates' seventh turn in the conversation reads as follows.

337 d2-d4, S7 a./ What else than the appropriate penalty for one who doesn't know, namely, to learn from the one who does know? b./ Therefore, that's what I deserve.

This study will use the programming language R to analyze three key markers: the number of words in each turn, the number of sentences in each turn, and the sentiment of each turn. This study uses G.M.A. Grube's and C.D.C. Reeve's translation of the Republic, as mentioned above, to facilitate the use of programmatic methods. To attempt this study on the Greek text would require three significant undertakings.

First, the programmer would need to create a list of "stop words." These are short words that, while grammatically significant, are less semantically relevant. The list would also need to be sensitive to context. For example, the script may safely remove the particle "δέ" from a great many sentences, but removing it from a "ὁ μέν . . . ὁ δέ" construction would remove important semantic material. The task would be of much greater scope than this study.

Unlike Greek, typical English prose is less reliant on sentence particles, and so a computer may remove stop words from a text with less trepidation. In addition, the tools for removing English stop words already exist, and their engineers have had years to solidify their lists. Taking advantage of these preexisting tools clears the way for quicker and simpler analyses.

Second, the programmer would need to create a tool to derive the lemma form — i.e., the dictionary form — of each word encountered. Computers cannot simply identify such elements as verbal roots in the same way a person can, so some method of disambiguating forms would be necessary. Without this tool, the computer would have no understanding that different forms may be from the same word. For example, the words “ἔλυσα” and “ἐλύθην” would read as completely distinct elements, ignoring the apparent relationship between the two.

English again differs from Greek in this regard. Given that English lacks the thorough conjugation and declension patterns found in Greek, an adequately written program can derive English lemma forms much more easily. The R community has also produced such tools side-by-side with stop word lists, making them more immediately available and further simplifying the study.

Third, the programmer would need to create a sentiment dictionary for Ancient Greek. A sentiment dictionary is a lexicon of the moods or feelings connoted by the words it contains. The standard R libraries include four sentiment dictionaries. The Bing and Loughran lexicons use a simple binary for their definitions: positive-negative. For example, ”prison” is marked as negative, while “free” is marked as positive. The NRC lexicon attempts to define specific emotions, such as “trust,” “fear,” and “anger;” as well as the more ambiguous “positive” and “negative.” The AFINN lexicon strikes a compromise between the binary approach of Bing and Loughran and the specific definitions of NRC. Each word is placed on a sentiment range from -5 (most negative sentiment) to +5 (most positive sentiment). The effort required to replicate one of these dictionaries in Greek, a process that would rely on preexisting stop word lists and lemmatizers, would be tremendous. Teams of programmers typically undertake projects of this size, and the lack of options should speak to the difficulty of their construction.

This study will use the AFINN lexicon's sentiment range to identify transitions in and out of phases. However, using the AFINN lexicon does require compromise. Finn Årup Nielsen designed it for analyzing the sentiment of microblogs, and the sentiment scores are specifically tailored for Twitter. Any sentiment analysis undertaken will have a modern colloquial underpinning, and the analysis will specifically represent a 21st-century reading of the text.

Sentences per Turn (Turn Density)

The first portion of the code tallies the number of sentences per turn, hoping to identify a correlation between turn "density" and the occurrence of phase transitions. It counts the data in two ways. First, it gathers "global" calculations, i.e., the relevant information from the entire dialogue, and second, it gathers speaker-by-speaker calculations. In order to construct a high-level view of turn density, the script returns a table of calculations. The mean represents the average number of sentences per turn, the median represents the middle value when the data is organized numerically, and the mode represents the middle value when organized. The standard deviation represents the overall variation within the data. The table is printed below.

The table shows notably regular results. Each speaker's average sentence density is close to the global, $P_{\bar{x}g} = 1.68$. Cleitophon and Socrates deviate the most, $P_{\bar{x}c} = 1.33$ and $P_{\bar{x}s} = 1.99$, respectively, but the variations are slight, making it difficult to tell what difference

```
# A tibble: 6 × 5
  speaker      mean median  mode    sd
  <chr>       <dbl>  <dbl> <int> <dbl>
1 global       1.69    1       1 1.79
2 cleitophon   1.33    1       1 0.577
3 glaucon      1.25    1       1 0.463
4 polemarchus  1.75    1.5     1 0.957
5 socrates     1.99    1       1 1.90
6 thrasymacus  1.41    1       1 1.72
```

Table 1. Turn Density Data

is significant. The script performs a t-test comparing the global mean to each speaker's to solve this problem. This test will return a figure called a p-value for each speaker. For the purposes of this study, if a p-value is less than 0.05, then the difference between the speaker's mean and the global mean are significantly different. The results of the test are printed below.

```
[1] "cleitophon: 0.0755533142252797"
[1] "glaucon: 0.0953570075963998"
[1] "polemarchus: 0.0102844982210078"
[1] "socrates: 0.0514802153411887"
[1] "thrasymacus: 0.0581347985214711"
```

Table 2. P-Values: Average Turn Density by Speaker

Interestingly, Only Polemarchus shows a statistically significant deviation from the global mean. In Polemarchus' case, the small sample size may skew the data. Polemarchus only has four turns in the entire dialogue. His third turn (P3) is three sentences long, P2 is two sentences long, and P1 and P4 are both one sentence long. As such, his result here provides little insight.

Without any useful p-values to rely on, it becomes helpful to visualize the data. Below is a histogram representing each turn in the order of occurrence. The x-axis represents the index — a number identifying each turn in the data set — and the y-axis represents the number of sentences in each turn. The blue line represents the global mean density over time.

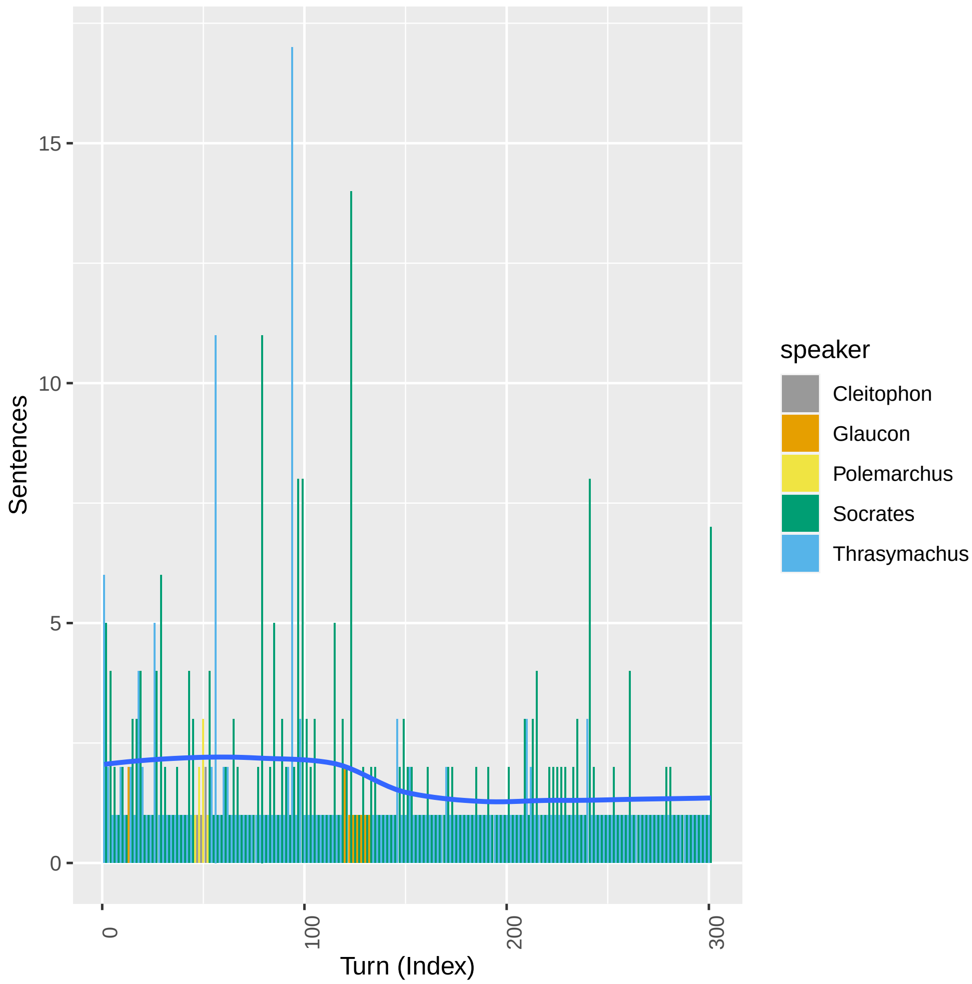

Figure 1. Sentence Density by Speaker

As reported by the initial data, the dialogue shows a consistent regularity in sentence density. However, Socrates and Thrasymachus both show greater variation than the others, especially in the early half of the text. This variation is measured by the standard deviations above. A similar t-test on them reveals that both have a significantly high range of sentence densities: $P_{\sigma s} = 0.02$ and $P_{\sigma t} = 0.01$ (both rounded), respectively. These p-values imply that significantly denser turns than the average should be considered when defining phases.

```
[1] "cleitophon: 0.301818447656808"
[1] "glaucon: 0.339266024338994"
[1] "polemarchus: 0.187958111608137"
[1] "socrates: 0.0175810404193463"
[1] "thrasymacus: 0.0129304749109537"
```

Table 3. P-Values: Standard Deviation of Turn Density by Speaker

Words per Turn (Turn Length)