Use Smaller Words, Socrates: Using Words and Particles to Predict Conversation Types in the Socratic Dialogues

Benjamin Thompson

University of Georgia – BenThompson@uga.edu

Abstract

How might Socrates' dialogues be characterized by more back and forth and less lecturing? Plato's Dialogues of Socrates is a collection of conversations between Socrates and at least one counterpart. This paper attempts to determine what characteristics of a Socratic speech might elicit a response of similar length from his counterpart. This paper examines the entire Socratic Dialogue corpus in ancient Greek to examine several possible variables that might change the type of conversation: number of words used, number of questions asked, and the presence of certain particles (such as γε and δή). This paper finds that several features are correlated with different types of conversation.

1 Introduction

Plato's Dialogues of Socrates are famous for their depiction of philosophical conversation. It is from these dialogues that we derive the Socratic method, whereby a teacher asks questions to elicit thought in his pupil. There are many lengthy conversations in the dialogues, and one might wonder how strictly Socrates adheres to the Socratic method. How do his questions affect the nature of the dialogue taking place? One might imagine that the more he asks questions, the more his counterpart will respond. We can think of the conversation as falling into one of two categories: either it is more of a lecture, where Socrates does most of the talking, or a dialectic, where both parties speak for roughly equal amounts. A conversation may take on both qualities at different times, but what causes it to take one mode or another at any given point? If Socrates says more words, perhaps his counterpart will have to respond to. Or perhaps there are conversational markers, words such as "like" and "so" in English that serve more of a conversational than a semantic role, that would indicate to his counterpart that they should treat him less

as a teacher and more as an equal. In the Greek language, these small words are called PARTICLES, and they are very similar to what Pragmatics calls DISCOURSE MARKERS, as this paper will discuss.

There are four particles covered in the scope of this paper: the particles γε ⟨gɛ⟩, δή ⟨deɪ⟩, οὖν ⟨uːn⟩ and ἄρα ⟨ˈɑːrə⟩. I hypothesize that a Socratic dialogue becomes more dialectic with questions, the particle γε, and the particle δή, while it becomes more like a lecture with the more words Socrates says, the particle οὖν and the particle ἄρα. As part of this analysis, I hope to also demonstrate that these particles serve more of a pragmatic function than a semantic one, thus making them discourse markers rather than true parts of speech.

The corpus analyzed is all of Plato's works in Greek, narrowed down to Socratic dialogues alone, further narrowed down by Socratic dialogues in which there is more than one speaker present and presented in turn-by-turn form (meaning the *Republic* is excluded, as there was only one tagged speaker). This paper was motivated in part by the fact that Socrates tends to talk a lot and his counterpart does a lot less talking. Figure 1 shows a chart of all the characters across all the dialogues, sorted by what percentage of their dialogues they were speaking. A chart of each dialogue, and the colors represent the portion of words spoken by each of them. In each dialogue, most of the talking is done by one character, and that one character is usually Socrates, which is why he is third on the charts. Across all the dialogues, 61.4% of the words are spoken by Socrates. The number would be much higher were it not for the few dialogues in which Socrates appears as a side character rather than the main character.

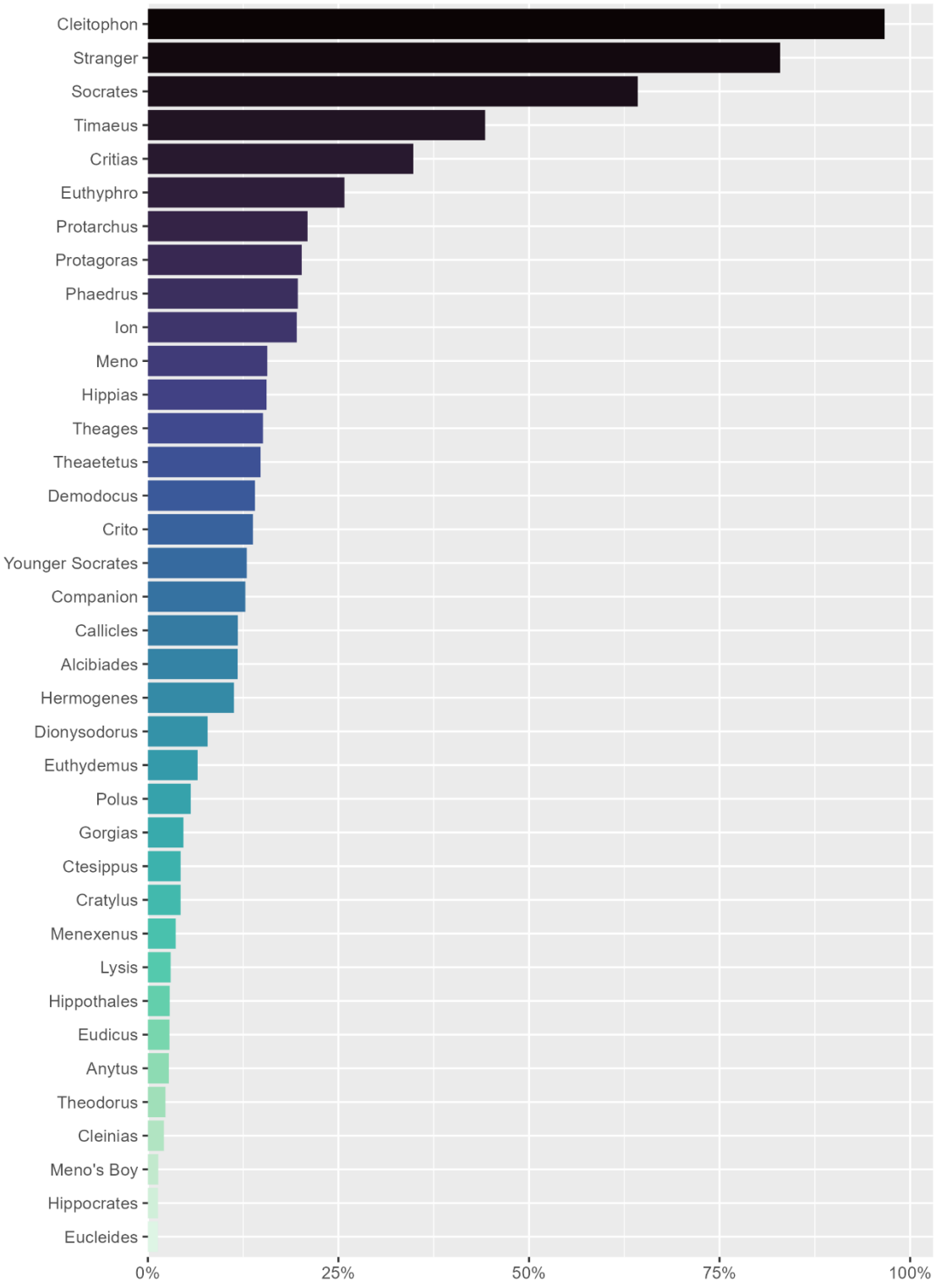
This paper looks at the conversational features that might explain or predict whether a given set of conversations is more lecture-oriented, where one speaker does most of the talking, or whether it is more dialectic, with a back-and-forth. To understand precisely what is meant by that, we must first understand the field of pragmatics.

2 Pragmatics

Pragmagtics is a linguistic field that focuses on the practical parts of human communication; how we use words and sentences in actual situations. It is less concerned with the literal meaning of words than it is with how that meaning is constructed. A subfield of Pragmatics, Conversation Analysis, looks at what speech does, rather than what it means. In Conversation Analysis, the fundamental unit is what is referred to as the *conversational turn*, which we can understand as akin to lines of dialogue. One person talks, which constitutes one *turn,* then another person talks, which constitutes another *turn*. In total, two people speaking one time would be two *turns*.

Conversational turns characterize all of the dialogues, so it would be helpful to

Figure 1


Cleitophon
Stranger
Socrates
Timaeus
Critias
Euthyphro
Protarchus
Protagoras
Phaedrus
Ion
Meno
Hippias
Theages
Theaetetus
Demodocus
Crito
Younger Socrates
Companion
Callicles
Alcibiades
Hermogenes
Dionysodorus
Euthydemus
Polus
Gorgias
Ctesippus
Cratylus
Menexenus
Lysis
Hippothales
Eudicus
Anytus
Theodorus
Cleinias
Meno's Boy
Hippocrates
Eucleides
0%
25%
50%
75%
100%

have a way to label turns. Beňuš et al. (2011) do just that, examining turn-taking in spoken exchanges between two interlocutors. Their concern is with what they refer to as "conversational dominance," and they have many metrics to define and understand it, including one they call the ENTRAINMENT INDEX, or latency divided by the rate of speech (Beňuš et al., 2011, p. 3005). The metrics they use are largely constructs of an audio corpus, which of course is not and interruptions, which are absent The Socratic dialogues do not feature interruptions the same way we would expect in verbal communication, but the notion of "conversational dominance" is similar to what this paper is concerned with. With that in mind, I define the LECTURE QUOTIENT as the difference between the number of words, squared, that each speaker utters in turn. The higher the lecture quotient, the more two turns resemble a lecture:

$$\text{LECTURE QUOTIENT} = (\text{turn } i \text{ word count} - \text{turn } i+1 \text{ word count})^2 \quad (1)$$

Stenström (2014) analyzed whether the choice and frequency of DISCOURSE MARKERS might be able to distinguish chatting from more informative speech. Discourse markers are a widely-studied feature of languages, and there are many definitions and nuances that this paper will gloss over. A good summary is found in Bonifazi et al. (2022): "In general, what connects words that are considered discourse markers is not their form, but their guiding function in utterance processing or in interaction. … What all discourse-marker studies share, however, is a focus on functions and meanings that transcend the transfer of referential information" (Bonifazi et al., 2022, I.3.2§15). In other words, discourse markers are words with function rather than meaning.

Stenström looked at four instances of chatting ("phatic talk") and one instance of a lecture given by a teacher and compares shorter back-and-forth exchanges with longer, uninterrupted speech. She found that the shorter exchanges have a higher ratio of discourse markers, that discourse markers are often used to avoid conversational gaps, and that the discourse markers in the shorter exchanges tend to either limit the scope of the speaker's statement or to indicate their counterpart should reiterate theirs (Stenström, 2014, p. 46). The analysis of discourse markers by breaking down the exchanges into short and long is similar to the focus of this paper, which attempts to categorize exchanges as either more like a lecture or more like a dialectic.

It should be noted, too, that the Platonic dialogues are not unaltered conversations. They are mediated through Plato's interpretive writings, and thus much of the honesty a recording would have provided is lost to history.

3 Greek particles

In order to understand how Greek particles relate to the field of Pragmatics and Conversation Analysis, we must first understand Greek particles in their own right. Particles are short words in Greek that can take on a range of meanings depending on context, making them famously difficult to translate.

The most famous treatment of the Greek particle comes from Denniston's work titled the same (Denniston, 1954). He explains that a particle in Greek is "a word expressing a mode of thought, considered either in isolation or in relation to another thought, or a mode of emotion" (Denniston, 1954, p. xxxvii). Another scholar, Smyth, explains in his *Greek Grammar,* "Greek has an extraordinary number of sentence adverbs (or particles in the narrow sense) having a logical or emotional (rhetorical) value ... [which] often resist translation by separate words" (Smyth, 1956, p. 631). The most recent comprehensive treatment of Greek Grammar, *The Cambridge Grammar of Classical Greek* (van Emde Boas et al., 2019, p. 663), explains that "[b]ecause they have a (often rather abstract) functional meaning rather than a referential meaning, and because there is not always an English word with the exact same function, there is often no one-to-one equivalent for a particular Greek particle in English translation." In short, there is difficulty translating these words.

3.1 Choosing four particles

I chose four particles to analyze: γε, δή, οὖν, and ἄρα. The particles γε and δή are similar because they both convey a sense of emphasis. Denniston himself states that "γε and δή, as emphatic particles, share a good deal of common ground" (Denniston, 1954, p. 244). Smyth explains that γε could indicate "assent, concession, banter, scorn, deprecation, irony, etc." Smyth, 1956, p. 642). Similarly, Tomaka (2022) explains that δή "is one of the subtlest and most elusive of particles ... its function is still not fully evident nor it is transparently known what force it brings to the utterance" (Tomaka, 2022, p. 27). In practice, it is not always apparent how a δή or a γε ought to be translated.

The particles ἄρα and οὖν, while of course not without nuance, do not represent the same difficulty. The particles ἄρα and οὖν are both inferential, and are generally translated "so" or "then." (Smyth, 1956, pp. 635/664). Table 1 summarizes and compares how Denniston, Smyth, and Boas et al. explain these four particles, looking both at the function the particles serve as well as how those sources propose to translate the given particle.

[1] γε: pp. 114-162, δή: pp. 203-259, ἄρα: pp. 32-43, οὖν: pp. 415-448
[2] γε: pp. 642-643, δή: pp. 646-647, ἄρα: pp. 635-637, οὖν: pp. 664-665
[3] γε: p. 692, δή: pp. 686-688, ἄρα: pp. 685-686, οὖν: pp. 681-682

Table 1

Particle		Denniston[1]	Smyth[2]	Boas et al.[3]
γε	Function:	Concentration, limitation, attention on one idea	Intensive and restrictive particle	Concentration, Limitation, Emphasizing words
	Proposed Translations:	'At least,' often best by inflection or italics	'At least, at any rate, even, certainly, indeed'	'At least, to be precise,' often best translated by means of stress
δή	Function:	A thing really and truly is so	Immediately present, clear to the mind, greater precision	Evident, Clear, Precise
	Proposed Translations:	'Actually, indeed, verily'	'Certainly, indeed'	'In fact, actually, indeed, precisely'
ἄρα	Function:	Connection (consequence or succession)	Consequence drawn from the connection of thought, and expresses impression or feeling	Cannot but make the contribution
	Proposed Translations:	'So'	'Then, perhaps'	'Apparently, then, so'
οὖν	Function:	Connective, meaning	Confirmatory or inferential	Transition
	Proposed Translations:	'So, then'	'Therefore, accordingly'	'So, then, therefore'

3.2 Fixed uses of particles

Here too must be explained a crucial part of Greek particles. They are used so commonly that they often pick up other words that, when combined, become their own pseudo-particle. For instance, *πάνυ γε* is a phrase that means 'certainly,' and it is often an entire conversational turn on its own. Two of the most common with ge are *πάνυ γε*, 'certainly,' and *εὖ γε*, 'excellent.' The particle *γε* appears 2,502 times throughout the entirety of the dialogues, of which 293 are either *πάνυ γε* or *εὖ γε*. A strict definition of the term FIXED PARTICLE is beyond the scope of this paper, as is a comprehensive list of such particles, but the concept, and the most common examples, will be relevant later in the analysis.

4 Greek particles as discourse markers

With a fresh understanding of particles in Greek, we can now begin to explore how they might be thought of as discourse markers. Tronci (2017) examined the use of four particles, *ἄρα, οὖν, οὐκοῦν*, and *τοίνυν* in Plato's *Theaetetus,* choosing those four particles because they tend to be translated similarly and are often viewed as interchangeable. She argues that each particle appears in different parts of the speech turns, and as a result, she splits them into three classes (Tronci, 2017, p. 213). The classes are not of relevance to this paper, but what is of interest is the Pragmatic approach: looking at particles by dint of their appearance in conversational turns, looking at other conversational correlates, and the overall data-driven approach to particles.

If Tronci (2017) took a step in the right direction of a pragmatic approach to Greek particles, Bonifazi et al. (2022) made a leap. *Particles in Ancient Greek Discourse: Exploring Particle Use Across Genre* contains lengthy analyses of the Greek particle explicitly as a discourse marker. In their words, "There is a clear similarity, then, between the way Greek particles are defined in modern scholarship and the way discourse markers are defined in discourse-marker studies" (Bonifazi et al., 2022, I.3.5§61). After a lengthy examination of the different ways particles are portrayed in Greek tragedy and comedy, the researchers conclude that "particles reveal how turns relate to each other and to the structure of an ongoing interaction."

I would be remiss if I did not also include their description of the particle *γε*, the focus of this project: "*γε* is comparable to the prosodic prominence rendered by an exclamation mark: a verbal equivalent, we can say, of banging one's fist on the table, or of stamping one's feet" (Bonifazi et al., 2022, 4.5.2§64).

As discourse markers are characteristic of dialectic rather than lecture, it would follow that *γε* and *δή*, which are looser in meaning and more emphatic, should decrease the lecture quotient. The particles *οὖν* and *ἄρα*, on the other hand, have tighter definitions and are less like discourse markers, and in turn, should increase

the lecture quotient. Particles aside, more words spoken in a turn should increase the lecture quotient, and asking questions should decrease it.

5 The corpus

Putting the database together took by far the longest of all the parts of this paper. Perseus .xml files were downloaded and converted into a corpus database where each row contains information about one conversational turn. More specifically, each row contains the content and number of words spoken, the number of particles γε, οὖν, ἄρα, and δή (as well as many other particles that were not used in the analysis)[4], how many questions were asked, and how many words were spoken in the next turn. That last variable, words spoken in the next turn, was used to create the lecture quotient.

There are several reasons why this paper analyzed Plato in Greek and not in English. First, all Plato's works have been put together into one downloadable corpus by academics at Tufts University. Second, any translation into English has artifacts that would render any fine-tuned analysis impossible (word count variations, vocabulary choices, and so forth). Third, and most importantly, this paper was motivated in part by a desire to better understand Greek particles, which are unrecoverable from any English translation.

Another corpus was created as a subset of these which contained only those lines spoken by Socrates. This corpus was created because, as the main character of most of the dialogues, Socrates might have effects different than his counterparts, and I wanted to be able to capture that difference.

6 The results

With both corpora, I looked at three models to determine how Socrates' turns might affect conversational quality, compared in Table 2 and Table 4. In all of these models, the LECTURE QUOTIENT was used as the dependent variable. The first model, COUNTING, counted the number of particles, the number of questions, and the total number of words spoken as independent variables. The second model, BINARY, looked at the same things, but instead of counting particles, it contained a 1 if any particles were present, and a 0 if none were. The final model, IDIOMS INCLUDED, is identical to COUNTING, but it includes the idiomatic "fixed" uses of γε, which the other models exclude.

To compare these models, I used three standard metrics: Adjusted R^2, AIC, and BIC. Adjusted R^2 is a measure of how much of the variance in the data is explained by the model, and a higher Adjusted R^2 is better. AIC and BIC are

[4]These particle counts excluded two fixed usages of γε: πάνυ γε and εὖ γε

more complicated, but serve the same purpose of comparing models with the same dependent variable. AIC and BIC scores are like golf scores in that a lower score is better.

6.1 Socrates only

Beginning with Socrates, the three models are compared below in Table 2. Of the three models, the IDIOMS INCLUDED performed the best, and its results are in Table 3. Each variable has its own row, and the numbers to the right of each variable are its marginal effect on the LECTURE QUOTIENT (e.g., for every question asked, Socrates decreases the lecture quotient by over 34,000). The asterisks denote statistical significance.

Table 2

Socrates	Adj. R^2	AIC	BIC
Counting	0.7592260	107416.6	107466.9
Binary	0.7404658	107713.0	107763.2
Idioms included	0.7604145	107397.1	107447.3

The first thing to note is that all of the results are statistically significant. More importantly, they are not what I expected. The only two variables that increase the LECTURE QUOTIENT are the number of instances of the particle *δή* and the number of words in Socrates' turn. The other three particles all decrease the LECTURE QUOTIENT, marking a dialectic. This is not quite I had hypothesized, but fits well with the Conversation Analysis idea that discourse markers have function more than meaning. It is also notable that *γε* has the largest negative impact on LECTURE QUOTIENT. Banging his fist on the table gets Socrates a response.

6.2 Entire Platonic corpus

It turns out that the models assessing all of the Socratic Dialogues, rather than just the Socratic sub-corpus, are both more predictive overall and turn out differently. This indicates that there really is a linguistically meaningful difference between Socrates and his counterparts. The comparisons between the different models are contained in Table 4, using the same metrics as before, and the results of the best model of the three, COUNTING, are in Table 5.

Notable again is the fact that all of the independent variables are statistically significant. This time, though, the biggest negative impact on LECTURE QUOTIENT no longer comes from the particle *γε*, but instead from the particle *ἄρα*,: for each additional instance of the particle ara, the LECTURE QUOTIENT is expected to decrease by over 680,000. The biggest positive impact comes from the particle *δή*. Both of these findings are precisely the opposite of what I had hypothesized.

Table 3

	Dependent variable:
	Lecture Quotient
(Intercept)	-50,693.780***
	(3,634.823)
γε present	-73,172.560***
	(4,994.951)
δή present	100,979.400***
	(4,971.612)
οὖν present	-20,270.080***
	(5,476.205)
ἄρα present	-21,327.960***
	(8,270.844)
Number of questions	-34,463.180***
	(2,079.930)
Number of words	2,386.184***
	(56.927)
Observations	3,950
R^2	0.761
Adjusted R^2	0.760
Residual Std. Error	193,781.800 (df = 3943)
F Statistic	2,089.939*** (df = 6; 3943)
Note:	*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table 4

Overall	Adj. R^2	AIC	BIC
Counting	0.8949946	315519.8	315577.6
Binary	0.8713152	317592.2	317650.0
Idioms included	0.8940936	315606.8	315664.7

Table 5

	Dependent variable:
	Lecture Quotient
(Intercept)	-113,693.700***
	(14,328.460)
γε present	-667,370.000***
	(24,884.740)
δή present	984,350.500***
	(23,577.970)
οὖν present	-87,024.470***
	(28,147.110)
ἄρα present	-680,833.400***
	(48,547.750)
Number of questions	-119,401.700***
	(11,745.630)
Number of words	8,354.489***
	(293.366)
Observations	10,191
R^2	0.895
Adjusted R^2	0.895
Residual Std. Error	1,278,163.000 (df = 10184)
F Statistic	14,476.440*** (df = 6; 10184)
Note:	$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

7 Conclusion

It turns out that, from the variables chosen, there are lots of things Socrates can do if he wants to make the conversation more of a dialectic than a leccture. Using the particle γε is his best option. The distinction between emphatic and inferential particles ended up making no difference in this analysis.

There are many limitations to this approach. First, it included only four particles, while Greek has over twenty. It could be that other particles contribute more to the type of conversation. Second, it did not identify how the particles were being used. Sometimes γε has a limiting force; sometimes it has more of an emphatic force. A more thorough analysis might be able to tease that out. Third, it ignored the relationship particles have both with other types of words (like nouns and verbs) and with each other. Particles used with a noun might have a different effect than particles used with a verb, for instance, and particles used in pairs often have a similar function to the excluded πάνυ γε and εὖ γε. Further analysis might take that into account.

Looking at Greek particles through the lenses of Conversation Analysis and Pragmatics, the meaning is less important than the function. By and large, in the Platonic dialogues, γε, οὖν, and ἄρα serve the function of reducing the lecture quotient, while δή serves the opposite function.

References

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001–3027. https://doi.org/10.1016/j.pragma.2011.05.011

Bonifazi, A., Drummen, A., & de Kreij, M. (2022). *Particles in ancient greek discourse: Exploring particle use across genres*. Harvard Univ Press.

Denniston, J. D. (1954). *The greek particles* (2nd). Clarendon Press.

Stenström, A.-B. (2014). Avoid silence! keep talking! *Discourse Linguistics: Theory and Practice*, *21*(1), 30–49. https://doi.org/10.1075/fol.21.1.03ste

Tomaka, K. (2022). λέγε δή: What we know about the ancient greek particle δή – the current state of research. *Roczniki Humanistyczne*, *70*(3), 27–46. https://doi.org/10.18290/rh22703.2

Tronci, L. (2017). Chapter 10. on the distribution of some interactive/conclusive discourse markers in plato's theaetetus. In *Studies in language companion series* (pp. 213–234). https://doi.org/10.1075/slcs.190.10tro

van Emde Boas, E., Rijksbaron, A., Huitink, L., & de Bakker, M. (2019). *The cambridge grammar of classical greek*. Cambridge University Press.