Thomas Broadwater

July 21, 2022

κὴ γᾶς κὴ ϝυκίας:

A Computer Assisted Epigraphic Study on the Isoglosses of Central Greece

Linguistics is poised to take great advantage of the processing power available to modern computers. This is especially true for the epigraphist, the dialectologist, or any other researcher who encounters large volumes of text. Plentiful tools already exist for the analysis of textual data. First and foremost is the programming language called *R*, which was purpose-built for data management and analysis. The study will be built in this language. Also available is *Tidyverse*, a collection of packages — i.e., precomposed lines of code — which greatly expand upon the capabilities of R, especially with regards to textual data. This combination, at once robust and straightforward, enables the linguist to work large volumes of data at a rate much quicker than they might achieve by hand. Of course, programming may not solve every problem, but to ignore it when it is applicable is unreasonable.

This thesis is inspired by Luo et al.'s article *Deciphering Undersegmented Ancient Scripts using phonetic Prior*. The piece focuses on applying statistical and, most significantly, programmatic methods in the pursuit of deciphering lengths of unreadable ancient texts. More light-hearted influence is found in *The LOTR Project* by Emil Johnson, which demonstrates the applications of text mining, natural language processing (NLP), and expressive data visualization in a philological context.

My approach will rely on the use of text mining as the principal method of data collection and organization (a.k.a. "wrangling"). This will include using a web

scraper, a program that systematically retrieves data from the internet. The resulting inscriptions will then be cleaned of any editorial marks, such as brackets "[ ]" indicating missing text and dots "◌̣" (where ◌ is a placeholder for any character) indicating a reconstructed letter. For example, in Inscriptiones Graecae volume seven, inscription five hundred and four (IG VII 504) contains the name ⟨Ὀν̣[ασι]μ̣ίδας / Ἀμουνίαο⟩ split across the first two lines. After cleaning, the name will read ⟨Ὀνασιμίδας Ἀμουνίαο⟩.

This process will continue through the entirety of the Inscriptiones Graecae as available on the Packard Humanities Institute website. The raw corpus will then be reduced to capture inscriptions from seven cities:

- Megara, Page, and Aegosthena will represent the Dorians.
- Oropos and Tanagra will represent the Boeotians.
- Eleusis and Athens will represent the Athenians.

The ultimate result of this thesis will be a map of isogloss lines across three centuries between the regions of Megaris, Boeotia, and Attica.

This thesis will consist of four chapters: two informative then two analytical. The first chapter will cover the methodology behind the study. This thesis being an initial foray into the subject, the study will be primarily concerned with finding and explaining the evidence for known isoglosses, and so this chapter will cover the sources for those arguments. Given that the methodology is contingent on programming, this chapter will also include a breakdown of what the program does and how.

The second chapter will contain an overview of the dialects in question. The chapter will focus primarily on dialect phonologies, as phonological differences are the basis for the study. However, issues such as phonotactics and morphology will be considered as necessary throughout the article, since the study will include little to no close reading from the inscriptional corpus. The reader, as much as possible, will be given the relevant tools to examine any inscription at any time individually. As far as the "narrative" of the piece is concerned, it may be convenient to consider the program as having run between this chapter and the next.

The Attic-Ionic group, being most typically taught and having the highest immediate recognizability, will function as a "standard" for comparison across all chapters, especially the third and fourth. All analyses will be made from the perspective of comparing the Doric and Boeotian inscriptions to the Attic. A distinction with little real-world difference, this mindset will inform the organization of the thesis as a whole.

The third chapter deals with non-Attic isoglosses. For example, IG VII 504 shows Boeotian <κὴ> corresponding to Attic <καὶ> (Boe κὴ ≙ Att καὶ). Therefore, this correspondence is marked as an isogloss that includes Boeotian and excludes Attic. This chapter will proceed clockwise around the map, starting with the Doric inscriptions, then finishing with the Boeotian inscriptions.

The fourth chapter will examine the isoglosses which include Attic. For instance, IG VII 43 (3rd c. bce) shows Dor καὶ ≙ Boe κὴ ≙ Att καὶ. This correspondence qualifies the isogloss as including Doric and Attic while excluding Boeotian.

This chapter will follow much the same structure as the former. The examination will begin with the Doric evidence, then proceed to the Boeotian.

References

Beekes, R. S. P., and Lucien van Beek. Etymological Dictionary of Greek. Brill, 2016.

Brixhe, C. "A Modern Approach to the Ancient Greek Dialects." A History of Ancient Greek from the Beginnings to Late Antiquity, vol. 1, Cambridge University Press, 2015, pp. 486–99.

Dosuna, J. Méndez. "The Aeolic Dialects." A History of Ancient Greek from the Beginnings to Late Antiquity, vol. 1, Cambridge University Press, 2015, pp. 460–74.

---. "The Doric Dialects." A History of Ancient Greek from the Beginnings to Late Antiquity, vol. 1, Cambridge University Press, 2015, pp. 444–59.

Karali, M. "The Classification of the Ancient Greek Dialects." A History of Ancient Greek from the Beginnings to Late Antiquity, translated by Chris Markham, vol. 1, Cambridge University Press, 2015, pp. 387–94.

Levshina, Natalia. How to Do Linguistics with R: Data Exploration and Statistical Analysis. John Benjamins Publishing Company, 2015.

Luo, Jiaming, Frederik Hartmann, Enrico Santus, Regina Barzilay, and Yuan Cao. 2021. "Deciphering Undersegmented Ancient Scripts Using Phonetic Prior." Transactions of the Association for Computational Linguistics 9 (February): 69–81. https://doi.org/10.1162/tacl_a_00354.

McLean, Bradley H. An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great down to the Reign of Constantine (323 B.C. - A.D. 337). University of Michigan Press, 2011.

Panayotou, A. "Ionic and Attic." A History of Ancient Greek from the Beginnings to Late Antiquity, translated by Chris Markham, vol. 1, Cambridge University Press, 2015, pp. 405–16.

Papanastassiou, G. C., and E. B. Petrounias. "The Morphology of Classical Greek." A History of Ancient Greek from the Beginnings to Late Antiquity, translated by W. J. Lillie, vol. 1, Cambridge University Press, 2015, pp. 571–89.

Papazarkadas, Nikolaos, editor. The Epigraphy and History of Boeotia: New Finds, New Prospects. Brill, 2014.