Anna Novokhatko – Matt Munson

Philological Vocabulary before Plato: Semantic Networks and Digital Approaches [data, linked data]

Anna Novokhatko

In my research project I study pre-Platonian philological vocabulary, arguing that it did exist already as technical vocabulary of early scholarship. The form and amount of this vocabulary is what I am discussing in my research project. The material I use for that is for the large part fragmentary. The texts of proper scholars (even phiolologists – I am aware of the problematic use of this term) do not survive. On the one hand there is the corpus of natural philosophic and sophistic fragments which contain significant material on early linguistics, etymology, lexicology, literary criticism etc. I examine this corpus analysing the use of (technical) language. On the other hand, comedy is an important source for understanding the development of scholarship. Sicilian comedy (Epicharmus) and Old Attic comedy reflect and mock contemporary philological studies. Again, I examine the corpus of Sicilian and Old comedy analysing the use of (technical) language this time as reflected and mocked in comedy.

Today we will work with one example from my project.

I argue that the word γλῶσσα ('tongue, language') aquires a new meaning ('obsolete epic word') as a terminus technicus coined around 430s-420s BCE.

The first definition of the term γλῶσσα we find in Aristotle, who calls so rare and obsolete words opposed to standard words, *kyria*, and typical for heroic poetry.

Arist. *Poet.* 22, 1459a9-10: τῶν δ' ὀνομάτων τὰ μὲν διπλᾶ μάλιστα ἁρμόττει τοῖς διθυράμβοις, **αἱ δὲ γλῶτται τοῖς ἡρωικοῖς**, αἱ δὲ μεταφοραὶ τοῖς ἰαμβείοις. ('Of word types, compound words particularly suit dithyramb, *glottai* **suit heroic poetry**, and metaphors suit iambic verse')

Arist. *Rh.* 1406b1-5: διὸ χρησιμωτάτη ἡ διπλῆ λέξις τοῖς διθυραμβοποιοῖς (οὗτοι γὰρ ψοφώδεις**), αἱ δὲ γλῶτται τοῖς ἐποποιοῖς** (σεμνὸν γὰρ καὶ αὔθαδες), ἡ δὲ μεταφορὰ τοῖς ἰαμβείοις (τούτοις γὰρ νῦν χρῶνται, ὥσπερ εἴρηται). ('Thus, diction using compound words

is most useful to dithyrambic poets (for they are sensitive to sound), **but *glottai* to epic poets** (for they are stately and self-assured), metaphor to iambic poets (for they now use these, as has been said')')

Arist. *Rhet.* 3, 1404b23: ὄντων δ' ὀνομάτων καὶ ῥημάτων ἐξ ὧν ὁ λόγος συνέστηκεν, τῶν δὲ ὀνομάτων τοσαῦτ' ἐχόντων εἴδη ὅσα τεθεώρηται ἐν τοῖς περὶ ποιήσεως, τούτων **γλώτταις** μὲν καὶ διπλοῖς ὀνόμασι καὶ πεποιημένοις ὀλιγάκις καὶ ὀλιγαχοῦ χρηστέον (ὅπου δέ, ὕστερον ἐροῦμεν, τό τε διὰ τί εἴρηται· ἐπὶ τὸ μεῖζον γὰρ ἐξαλλάττει τοῦ πρέποντος). ('Since speech is made up of nouns and verbs, and the species of nouns are those examined in the books on poetry, from among these one should use **glottai** and compound words and neologisms rarely and in a limited number of situations (we shall later explain, where; the reason has already been given: the usage departs from the appropriate in the direction os excess))')

Thus it is clear that in the mid 4th century BCE the term **γλῶσσα** clearly belonged to technical vocabulary of scholarship. As the material of the 5th century BCE reveals, Aristotle' use is based on the previous tradition.

In Aristophanes' early comedy *Daitales* (427 BCE) fr. 233 PCG (Galen 19, 66, 1ss. Kühn) we find a dialogue perhaps between a father and a "bad" son, representing a discussion of Homeric glosses, which may have been influenced by the popular Homeric studies that were increasingly common during the 5th century BCE.

Ar. *Daitales* fr. 233, 1-2 PCG

πρὸς ταύτας δ 'αὖ λέξον **Ὁμήρου** **γλώττας**, τί καλοῦσι κόρυμβα;

...τί καλοῦσ' ἀμενηνὰ κάρηνα;

And further, now explain **Homeric *glottai***; what do the call *korymba*?

... what do they call *amenena karena*?

The meaning of the word *glossa* here corresponds entirely to Aristotle's definition of the term.

Apart from Aristophanes, it was Democritus who may have used this word as terminus (before 370 BCE). Democritus showed an intense interest in linguistic and literary matters and apparently compiled (according to Diogenes Laertius) a Homeric dictionary explaining rare and ancient words **Περὶ Ὁμήρου ἢ Ὀρθοεπείης καὶ γλωσσέων**. The title suggests a Homeric-Attic dictionary explaining rare and ancient words, a D-

scholia (scholia minora) form of dictionary. This seems to place him in the mainstream of linguistic and Homeric discussion of his time.

Γλῶσσαι, the exegetical practice of the rhapsodes, were used in schooling perhaps in the form of lists with rare words. These words were discussed by teachers and students during readings of Homer (cf. Ar. fr. 233 PCG). The first Homeric-Attic dictionaries, which formed the basis of the so-called D-scholia (from their attribution to Didymus) and as a result of later Homeric commentaries, had their origins in these lists. The interpretations of Homeric words sometimes required a knowledge of the religious and historical background as well as of the Homeric language. The contemporary poet and the first editor of Homeric text Antimachus of Colophon was known to use *glossai*: he employed Homeric words in his own poetry.

Pace Major 2013, who discusses the term *glossa* in his "Catalog of terminology related to rhetoric" and notes that "Fr. 233 seems to use *glotta* in its later technical meaning of 'gloss', but this most likely reflects the usage of Galen (our source for the fragment), rather than an early attestation of this meaning" I believe that the word γλῶσσα aquires a new meaning ('obsolete epic word') around 430s-420s BCE.[1]

Matt Munson

Γλῶσσα and Distributional Semantics

Anna makes an excellent case that γλῶσσα (γλῶττα) has become a technical term in certain works of the 4th century. What I will do is to investigate whether this specialization is also detectable in the larger corpus of pre-Aristotelean literature. To do this, I will use what is called distributional semantics. This term was coined by the American linguist Zellig Harris in his 1954 article "Distributional Structure,"[2] in which he asserted that words demonstrate their meanings in texts by means of the words that occur around them.[3] The most developed expression of this hypothesis came in a series of lectures he did in 1986 in which he stated, "The most precise way of determining a word/s meaning is by investigating the meanings of the words that occur along with that

---

[1] Cf. Wilfred E. Major, The court of comedy: Aristophanes, rhetoric and democracy in 5th-century Athens, Ohio (2013), 188, n. 7.
[2] Zellig Harris, "Distributional Structure," Word 10, no. 23 (1954): 146–62.
[3] See especially section 2 of this article, "Distribution and meaning", ibid., 151–8.

word."[4] So if the meaning of a word is conveyed primarily by the words with which it occurs, i.e., its co-occurrents, then if the co-occurrence patterns of a word change, we should expect that its meaning has changed.

To use distributional semantics to determine diachronic semantic change, which is what we are investigating here, we first need two or more corpora to investigate, one (or more) that we expect to demonstrate the accepted or base or original meaning of a word and one (or more) in which we expect to find a change in the meaning or the word. For this study, I have chosen a 1.5 million word corpus of pre-Aristotelean literature (including Aristotle) culled from the Perseus Digital Library.[5] The comparison corpora are, then, made up of those three authors in which Anna has found the first traces of this specialized meaning of γλῶσσα, i.e., Aristophanes, Democritus and Aristotle.

Once we have chosen the corpora to be compared, we move on to the analysis. This consists of counting how many times each word in our corpora occurs within a certain context window with every other word in our corpus. This produces a series of values for each word, called a vector, that represent the number of co-occurrences of the two words in question. But this vector of raw counts can often be misleading. For instance, if we had two words in our corpus that both occurred 100,000 times and they co-occur with each other 5,000 times, this is probably less interesting and less important to us than two words that both occur only 10 times but always co-occur, i.e., that have a co-occurrence count of 10. For this reason, we need to implement some sort of significance measure that can tell us whether the co-occurrence counts that we have for each of our words are unexpected and, therefore, more meaningful, or whether they are precisely what we expected to find. The significance algorithm I am using here is one called log-likelihood, which was first applied to the study of distributional semantics by Ted Dunning.[6] Log-likelihood is a mathematically complex algorithm that has the advantage over more other similar measures that it deals very well with sparse data, which is precisely what we have in natural languages where half of our data points, i.e., the words, occur only once.

The log-likelihood calculations then produce new vectors for each word that are populated now by statistical significance values for the co-occurrence of each word. It is these vectors that will be the focus of the digital part of this study. By investigating which words co-occur more significantly with γλῶσσα in our two test corpora (Aristotle and

---

[4] Zellig Harris, "How Words Carry Meaning" (Lecture, Language and Information: The Bampton Lectures, Columbia University, 1986), http://www.ircs.upenn.edu/zellig/3_2.mp3

[5] https://github.com/PerseusDL/canonical-greekLit and http://www.perseus.tufts.edu/hopper/

[6] Ted Dunning, "Accurate Methods for the Statistics of Surprise and Coincidence," Computational Linguistics 19 (1993): 61–74.

Aristophanes) than in our base corpus, we will be able to determine whether this specialization in meaning of γλῶσσα is also visible in the larger picture drawn by these statistical, distributional methods.