



HAL
open science

Late Greek Poetry: a Computational Approach

Erik Henriksson

► **To cite this version:**

| Erik Henriksson. Late Greek Poetry: a Computational Approach. 2016. hal-01279948v2

HAL Id: hal-01279948

<https://hal.science/hal-01279948v2>

Preprint submitted on 6 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Late Greek Poetry: a Computational Approach

Erik Henriksson¹

¹University of Helsinki, Finland

Abstract

Late antiquity (c. 300-600 CE) saw the rise of classically inspired Greek poetry, leaving behind a large corpus of texts that has hardly received the scholarly attention it merits. This article presents a work in progress that aims to introduce new digital tools in the analysis of Greek poetry. To evaluate the benefits of a computational approach, I have chosen a relevant case study in late Greek literary history: the question whether, and how, explicitly Christian hexameter verses differed from those with non-Christian content in terms of metrics and diction. A major methodological challenge is posed by the fact that few texts from late antiquity are available in open source digital format. This article presents a number of features that may be legally extracted from licensed texts available online, as well as some preliminary remarks on how to make use of the collected datasets in a large-scale statistical analysis.

Keywords

Late antiquity; Greek poetry; hexameter; classicism; distant reading; digital humanities

I INTRODUCTION

Late antiquity (c. 300 - 600 CE) witnessed a resurgence of poetry after a less productive period in the first centuries CE. This comeback has been accounted for by the renewed social stability attained by political reforms in the early 4th century (Dihle [1994]) and more recently, by the dynamic interactions between the emerging Christian literary culture and the classical tradition (Cameron [2006], Shorrock [2013]). Throughout late antiquity, classical poetry – in particular, Homer and Virgil – formed the basis of Roman literary education, imposing strong stylistic and metrical conventions into poetic production (e.g. Agosti [2012]). The normative force of classicism made late antique versification an entirely scholarly affair that was at odds with the phonological developments of spoken language. Despite numerous studies on poetic meters and diction since the nineteenth century (Hermann [1816], Ludwich [1874], Maas [1929], West [1982] etc.), a comprehensive account of late Greek poetry is still nonexistent (Agosti [2012]).

In this work, I present a new digital “distant reading” (Moretti [2013]) approach to studying the meters and language of late Greek poetry. As opposed to the traditional process of manual data collection and analysis, which typically focuses on predetermined phenomena in small selections of texts, distant reading involves employing computer algorithms to uncover previously unseen patterns in large textual corpora. As noted by Fusi [2009], metrical analysis is particularly attractive from a computational perspective. First, collecting metrical data from thousands of poetical units is liable to produce errors when performed manually (e.g. Maas [1929] and Maas [1972], with contradictive figures). Second, studying metrics along with other linguistic phenomena, such as phonological processes, is hardly feasible by hand due to the abundance of variables that have to be taken into account.

In order to assess the advantages of using digital methods to trace the evolution of Greek poetry in late antiquity, I have chosen a case study that has a high relevance for an ongoing discussion about the relationship between classicism and religion in this period (e.g. Agosti [2014]). Classicism in late antiquity has traditionally been seen as inspired by an active “pagan” (itself a

Christian derogatory term) opposition to Christianity. In recent years, however, the pagan/Christian dichotomy is understood to have been completely misleading (e.g. Kahlos [2013]), giving way to a more accurate view of a sophisticated dialogue between the two.

1.1 The case of Christian and “pagan” hexameters

Classical Greek verses were based on a number of poetic meters, the choice depending on purpose and literary genre. Thus, epic poetry was associated with the dactylic hexameter, drama with the iambic trimeter and lyric poetry with a handful of complex metrical types (Golston [2016]). In late antiquity, however, hexameter gradually supplanted nearly every other metre, becoming the “generic meter” thought of as proper for any occasion (Miguélez-Cavero [2008]). Consequently, a relatively large number of (several tens of thousands) lines of late antique hexameters survive to this day, showing a wide variety of functions and genres including monumental dedications and philosophical treatises as well as works with a Christian or “pagan” subject matter.

In the light of the current scientific view that emphasizes the common classical ground of all Greek poetry, one is inclined to ask whether there are any stylistic or metric differences between non-Christian and explicitly Christian verses with regard to classical ideals. As regards metrics, an important preliminary study (Agosti and Gonnelli [1995]) already indicated that no such division exists (see also Agosti [2004]), and that Christian poetry is in itself rather versatile when it comes to metrical practices. In this project, my aim is to augment and evaluate these results by virtue of introducing new digital tools in the metrical and stylistic analysis of Greek poetry. Furthermore, by incorporating a phonological feature in the dataset, I wish to be able to shed new light on the parallel development of metrics and the spoken language in late antiquity.

II MATERIALS AND METHODS

Practically every published Greek text from antiquity (and beyond) can be found online courtesy of the Thesaurus Linguae Graecae (TLG) project at the University of California, Irvine. TLG’s copyright license states, however, that these vast materials can only be used by subscribers through TLG’s own web interface and can under no circumstances be downloaded to a remote database, even for purely scholarly use. These restrictions make it difficult to use these texts for any large-scale analysis purposes. Consequently, most existing digital methods dealing with ancient texts have been based on smaller corpora, such as those provided by the Perseus Digital Library (per).

In this project, my aim is to explore the different ways in which TLG’s resources, and poetry in particular, might be leveraged in a text mining project. Though an important goal in itself given the untapped potential of all of TLG’s materials, it was also dictated by the fact that no other database containing the whole corpus of late Greek poetry currently exists. In order to illustrate the components of the metrical analysis, I first briefly describe the principal parts of a hexameter line.

2.1 Greek hexameter in outline

Greek meter, in essence, is based on patterns of light and heavy syllables, where a light syllable (◡) contains one so-called mora (a minimal unit of timing) and a heavy syllable (–) two, all syllables except the ones ending in a short vowel being heavy. A dactylic hexameter line consists of six feet (also known as metra), each of which has four morae in one of two patterns: – ◡ ◡ (dactyl) or – – (spondee) except for the last foot, which is always a spondee. These alternating patterns alone yield 32 different lines; in addition, the hexameter line tends to break into two

distinct chunks (colae), the break (caesura) being usually somewhere around the middle of the line. Moreover, there are places (bridges) where word breaks are rarely found. From the hellenistic period (c. 300- BC) onward an increasingly strict system of metrical laws evolved, regulating the positions of caesuras, bridges and even word-accents (e.g. Whitby [1994]).

2.2 Data collection

Both the TLG and the Perseus Project provide their materials encoded in Beta Code, a standard format representing Greek characters, diacritics and punctuation using only ASCII characters. For example, consider the following hexameter line from the Dionysiaca by Nonnus, a major late antique (early 5th century) poet:

Εἰπέ, θεά, Κρονίδαο διάκτορον αἴθοπος εὐνήης
 *EI)PE/, QEA/, *KRONI/DAO DIA/KTORON AI)/QOPOS EU)NH=S

A detailed account of the Beta Code format is not possible here, but the above example serves to illustrate the way Beta Code represents texts written in Greek.

2.2.1 Metrical data

The first task is to divide the line into a string of syllables and analyze the metrical weight of each syllable. In the case of public domain materials, this process can be partly automated. To this end, I have developed a script that performs the syllable division (according to, i.a., the criteria outlined in Chapter 2.1) and carries out an analysis of every possible metrical reading of each line, yielding the most probable one. For the example shown above, it would output the following syllable divisions and metrical analysis:

[*EI], [PE/,], [QE], [A/,], [*KRO], [NI/], [DA] ,[O], [DI], [A/], [KTO], [RON],
 [AI)/], [QO], [POS], [EU)], [NH=S]
 _ 0 0 | _ 0 0 | _ 0 0 | _ 0 0 | _ 0 0 | _ _

Using these data, the automatic collection of the positions of caesurae, bridges, punctuation and word-accents is a matter of observing syllables and their positions within the metrical feet. In a variable representation, these data can be summarized as follows:

n	Weight	Word-start	Word-end	Foot-start	Foot-end	Caesura	Diaeresis	etc.
n	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1

2.2.2 Phonetic data

In order to extend the scope of the metrical analysis, a phonological feature is included in the syllable based dataset. Two restrictions apply: first, we have only approximate knowledge of the pronunciation of Greek in late antiquity; and second, as noted earlier, data collection from TLG is limited owing to the copyright license. By virtue of studies such as Allen [1987] and Horrocks [2009], however, our approximation of late Greek phonology is now fairly reliable. The copyright issue, on the other hand, effectively means that the phonetic transcription must be performed manually. Just as with the collection of metrical data outlined above, an automatic transformation of the text is only allowed when working with texts other than those provided by the TLG. Using the previous example, a transcription would roughly look as follows:

[i],[pe],[θe],[a],[kro],[ni],[δa],[o],[δi],[a],[kto],[ron],[e],[θo],[pos],[ev],[nes]

This transcription, though superficially similar to the Beta Code version, can be considered an independent scholarly transformation of the raw text, and therefore may be stored in a local database even when the originals are protected by a license.

2.2.3 *Morphological data*

A third type of data to be collected from the texts are the morphological analyses of each word. This feature is made up of two components: dictionary headword (lemma) and the description of the morphological features (part of speech, number, gender, etc.) of each word form. The collection can be performed semi-automatically with the open source Morpheus (mor) tool developed as part of the Perseus Project (per). In most cases, however, either the lemma or the morphological interpretation, or both, are ambiguous, making a completely automated analysis difficult to achieve.

2.2.4 *A technical note*

The TLG license strictly prohibits the use of scripts. Contrarily to what was proposed in an earlier version of this draft, even client-side scripts invisible to the TLG servers (e.g. a JavaScript extension) are forbidden, as TLG's director, professor Maria Pantelia has kindly pointed out to me. As regards public domain texts, however, data collection can be automated in various ways. I am currently in the process of developing several sscripts in JavaScript and Python to speed up the collection of metrical readings, phonetic transcriptions and morphological analyses.

2.3 **Data analysis methods**

A thorough investigation concerning the available methods to analyse the collected data is the subject of further work and will be only very briefly discussed here. Focusing on the case study discussed in chapter 1.1, I have chosen a technique known as the Euclidean distance (see Jockers [2014]) as my starting point. In summary, calculating the Euclidean distance means applying Euclidean metric to discover the closeness or distance between books, determined by a number of variables associated with each book. In the case of analysing Greek hexameters, these variables range from syllable weight to frequencies of parts of speech in a line or sentence. The versatile selection of features should be sufficient enough to yield some interesting results with regard to the discussion about Christian and “pagan” poetry. Another data analysis methodology I will be exploring in the future is based on algorithms known as topic models (Blei [2012]), which have proved to be highly useful in finding hidden patterns in text collections.

III **CONCLUSIONS**

Considering the potentially huge benefits of using digital tools and methods in analysing the meters and language of Greek poetry, surprisingly few such tools exist. This may be in part due to the poor availability of freely downloadable digitized texts, and in part, perhaps, to the common suspicion among classicists that the numerous hand-made statistics from the past centuries have the final word in metrical studies. In this work, I have proposed a new approach to collecting metrical and linguistic data from online databases of Greek texts, whether in the public domain or not. The presented methods demonstrate to what extent computer algorithms can help to analyse Greek verses on the one hand, and the untapped potential of electronic resources on the other. The newly collected datasets can be useful for many purposes, such as scrutinizing old statistics in metrics handbooks. In this project, I will analyse the data specifically with a view to contributing to an ongoing discussion about the relationship between late antique classicism and the emerging Christian literary culture. Future work will show whether a “distant reading” approach will provide the field of late antique literary studies with useful results.

References

- Thesaurus linguae graecae. <http://stephanus.tlg.uci.edu>. Accessed: 2016-02-28.
- Morpheus tool. <http://www.perseus.tufts.edu/hopper/morph>. Accessed: 2016-02-28.
- Perseus digital library. <http://www.perseus.tufts.edu>. Accessed: 2016-02-28.
- G. Agosti. Alcuni problemi relativi alla cesura principale nell'esametro greco tardoantico. In F. Spaltenstein and O. Bianchi, editors, *Autour de la césure. Actes du colloque Damon des 3 et 4 novembre 2000*, pages 61–80. Peter Lang SA, 2004.
- G. Agosti. Greek poetry. In Scott Fitzgerald Johnson, editor, *The Oxford Handbook of Late Antiquity*. Oxford University Press, 2012.
- G. Agosti. Classicism, paideia, religion. In R. L. Testa, editor, *The Strange Death of Pagan Rome. Reflections on a Historiographical Controversy*. Brepols, 2014.
- G. Agosti and F. Gonnelli. Materiali per la storia dell'esametro nei poeti cristiani greci. *Studia di metri classica*, 10(1):289–434, 1995.
- W. S. Allen. *Vox Graeca: The Pronunciation of Classical Greek*. Cambridge University Press, 1987.
- D. M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.
- A. Cameron. Poetry and literary culture in late antiquity. In S. Swain and M. Edwards, editors, *Approaching Late Antiquity: The Transformation from Early to Late Empire*. Oxford University Press, 2006.
- A. Dihle. *Greek and Latin literature of the Roman Empire: from Augustus to Justinian*. Routledge, 1994.
- D. Fusi. An expert system for the classical languages: Metrical analysis components. *Lexis*, 2009.
- C. Golston. Metrics. In *Encyclopedia of Ancient Greek Language and Linguistics*. Brill Online, 2016.
- G Hermann. *Elementa doctrinae metricae*. Fleischer, 1816.
- G. Horrocks. *Greek: A History of the Language and its Speakers*. John Wiley & Sons, 2009.
- M. L Jockers. *Text analysis with R for students of literature*. Springer, 2014.
- M. Kahlos. *Debate and dialogue: Christian and pagan cultures c. 360-430*. Ashgate Publishing, Ltd., 2013.
- A Ludwich. Hexametrische untersuchungen ii: Ein betonungsgesetz des nonnos und seiner schule. *NJb*, 109: 441–53, 1874.
- P. Maas. *Griechische Metrik*. Teubner, 1929.
- P. Maas. *Greek Metre*. Clarendon Press, 1972.
- L. Miguélez-Cavero. *Poems in context: Greek poetry in the Egyptian Thebaid 200-600 AD*. Walter de Gruyter, 2008.
- F. Moretti. *Distant reading*. Verso Books, 2013.
- R. Shorrock. *The Myth of Paganism: Nonnus, Dionysus and the World of Late Antiquity*. Bloomsbury, 2013.
- M. L. West. *Greek Metre*. Oxford University Press, 1982.
- M. Whitby. From moschus to nonnus: the evolution of the nonnian style. In N. Hopkinson, editor, *Studies in the Dionysiaca of Nonnus*. Cambridge Philological Society, 1994.